



Département Electronique Electrotechnique Automatique
Equipe Électronique
Tronc Commun - UE STI
Cours STI tc1 Systèmes Électroniques
2016-2017

**Unité d'Enseignement
Science et Technologie de l'Information
Systèmes Électroniques**

Sébastien Le Beux, David Navarro, Hai Son Nguyen, Ian O'Connor, Pedro Rojo-Romeo
Equipe d'Enseignement d'Electronique
Département EEA, bâtiment F7
e-mail : ian.oconnor@ec-lyon.fr

Version du 08 juillet 2016

1 Pré-requis

1.1 Électronique analogique

1.1.1 Grandeur

Les grandeurs caractéristiques sont les **tensions** ou **potentiels** (notées U , u , V ou v), et les **courants** (I ou i). Nous adoptons les conventions suivantes : majuscule pour les grandeurs continues (statiques), et les minuscules pour les signaux alternatifs (dynamiques).

Les unités sont les volts (tension¹) et les ampères (courant).

1.1.2 Impédances

Les composants passifs usuels (résistance, capacité, inductance) sont caractérisés par leur impédance Z . Nous pouvons définir l'impédance comme la "résistance équivalente" à une fréquence donnée. L'admittance est l'inverse de l'impédance $Y = 1/Z$ (unité: siemens). Tous les composants que nous lisons voir dans cette section sont des **dipôles passifs**.

Résistance

La résistance R a comme unité l'ohm. Son impédance est réelle et constante avec la fréquence :

$$Z_R = R$$

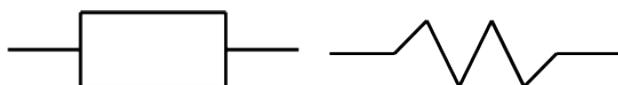


Figure 1: Symboles de résistance

Équation en statique : $U = R \cdot I$

Équation en dynamique : $u(t) = R \cdot i(t)$

Capacité

La capacité (ou condensateur) C a comme unité le farad. Son impédance est complexe (ce composant introduit un déphasage entre tension et courant), et varie avec la fréquence :

$$Z_C = 1/jC\omega$$

ω est la pulsation (unité : rad.s⁻¹) et vaut $\omega = 2\pi f$

Son impédance est donc infinie en continu, élevée pour des signaux de faible fréquence, et faible pour des signaux de haute fréquence. Sa caractéristique est de type $1/x$.

¹ "Voltage" est un mot anglais, qui n'a pas sa place dans le langage scientifique français en électronique ou électricité.

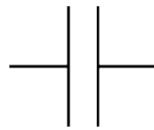


Figure 2: Symbole de la capacité.

Equation temporelle en statique : $I = C.U / t$ soit $Q = C.U = I.t$ (coulombs)

Equation temporelle en dynamique : $i(t) = C.dv / dt$

Equation fréquentielle : $u = i / jC\omega$

Les valeurs usuelles des condensateurs sont faibles (classiquement de 1pF à 1mF).

Inductance

L'inductance L (ou self, appelée aussi bobine) a comme unité le henry. Son impédance est complexe (ce composant introduit un déphasage entre tension et courant), et varie avec la fréquence :

$$Z_L = jL\omega$$

Son impédance est donc nulle en continu, faible pour des signaux de faible fréquence, et élevée pour des signaux de haute fréquence. Sa caractéristique est linéaire.



Figure 3: Symbole de l'inductance.

Equation temporelle en statique : $U = L.I / t$ soit $L.I = U.t$ (weber)

Equation temporelle en dynamique : $u(t) = L.di / dt$

Equation fréquentielle : $u = i.jL\omega$

Les valeurs usuelles des inductances sont faibles (classiquement de 1 μ H à 1H).

Associations de dipôles

Les dipôles peuvent être mis en série ou en parallèle. Dans le cas d'une mise en série de dipôle, le courant traversant chaque dipôle est identique égal au courant qui rentre dans la série de dipôle et la tension aux bornes de l'ensemble est égale à la somme des tensions présentes aux bornes de chaque dipôle. L'**impédance équivalente** est égale à la somme des impédances des dipôles soit

$$Z_{eq} = \sum_{i=1}^n Z_i$$

Dans le cas d'une mise en parallèle de dipôle, le courant traversant l'ensemble est égal à la somme des courants de chacune des dipôles et la tension aux bornes de l'ensemble est égale à la tension présente aux bornes de chaque dipôle. L'**admittance équivalente** est égale à la somme des admittance des dipôles soit

$$Z_{eq} = \frac{1}{\sum_i^n \frac{1}{Z_i}}$$

Conclusion

Quelle que soit l'impédance Z , la relation $u(t) = Z \cdot i(t)$ reste valable.

Les impédances non fixes en fonction de la fréquence ($1/jC\omega$ et $jL\omega$) permettent de créer des circuits ayant une réponse en fréquence caractéristique (non constante).

Ainsi, il est par exemple possible de réaliser des circuits coupant des fréquences faibles, élevées, ou des combinaisons (conception de filtres). Cette notion sera vue dans le paragraphe filtres linéaires du 1^{er} ordre.

1.1.3 Alimentations

Il existe deux types de circuits permettant d'alimenter un système électronique : la source de tension et la source de courant. Leur symbolique est rappelée figure 4. Ce sont des **dipôles actifs**.

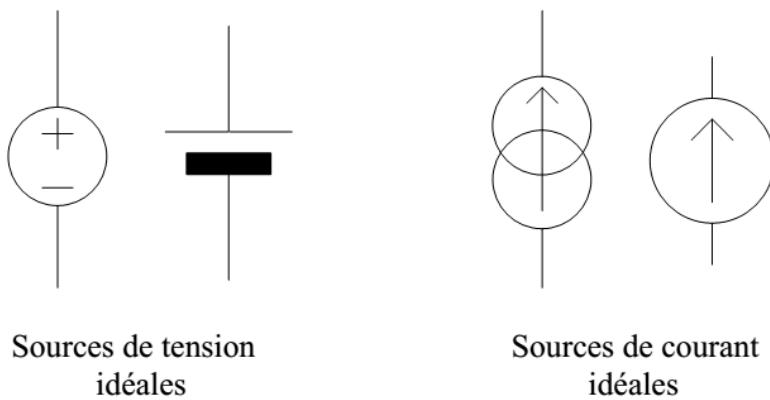


Figure 4: Symbole des sources de tension et de courant.

Source de courant

Nous appelons source de courant idéale, un dispositif qui délivre un courant constant quelque soit la tension à ses bornes (cela implique que la puissance que peut délivrer la source est théoriquement infinie).

Si le courant diminue lorsque la tension aux bornes de la source augmente, nous parlons alors de source de courant non-idéale. Une source de courant non-idéale est représentée par une source de courant idéale en parallèle d'une résistance de source R_{source} qui représente la perte en courant quand la tension augmente : plus celle-ci augmente, plus le courant traversant R_{source} est élevé et donc moins il y a de courant disponible pour alimenter le dispositif aux bornes de la source.

Le calcul de la valeur de la résistance de source est aisément : elle est égale à l'inverse de la pente de variation de $I=f(V)$ comme montré figure 5.

Il est à noter que pour une source de courant idéale $R_{source} = \infty$. Une mise en

parallèle de plusieurs sources de courant revient à additionner leur effet.

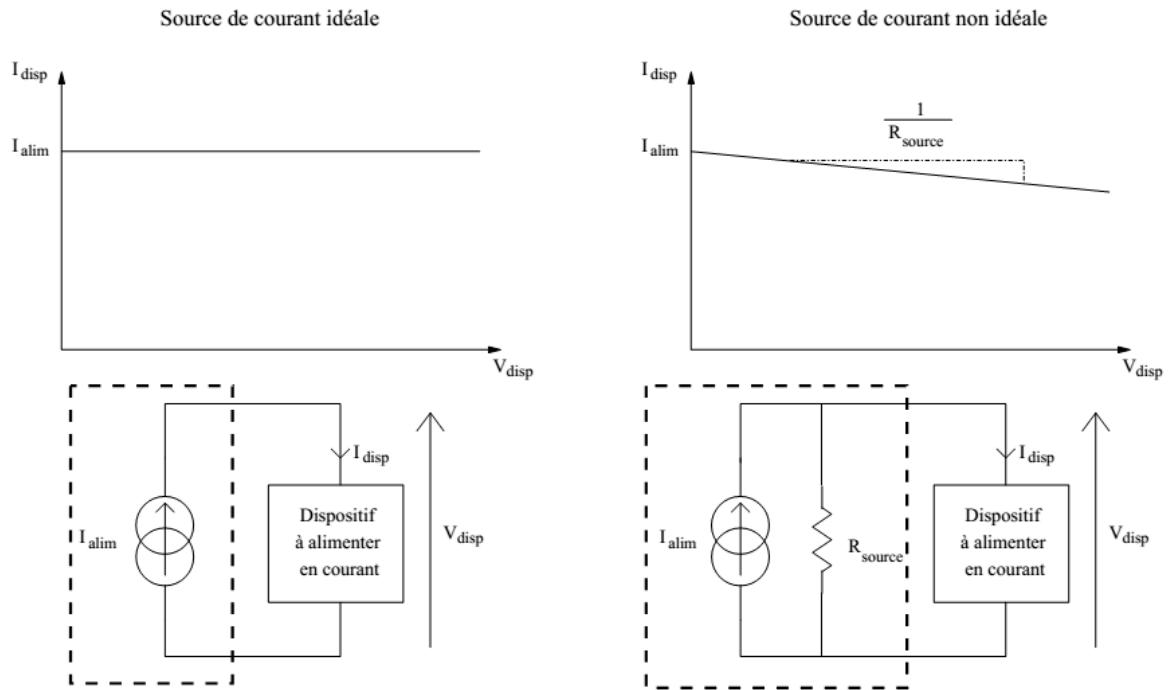


Figure 5: Sources de courant idéales et non idéales.

Calcul de I_{alim} et R_{source} :

pour calculer la valeur nominale du courant d'une source non idéale et sa résistance de sortie sans avoir accès à l'intérieur du composant, la manière de procéder est la suivante²:

- Si nous mettons la source de courant en court-circuit et que nous mesurons le courant traversant le fil de court-circuit, nous obtenons directement I_{alim} puisque la résistance de source est court-circuitee et qu'aucun courant ne la traverse.
- Nous éteignons la source de courant : $I_{alim} = 0$. Il suffit alors d'injecter un courant et de mesurer la tension obtenue – cas bleu – ou l'inverse (appliquer la tension et mesurer le courant – cas vert –) afin d'obtenir la valeur de la résistance (rapport des deux grandeurs).

La démarche est résumée figure 6.

² Il s'agit de la manière de procéder d'un point de vue conceptuel en aucun à reproduire de manière expérimental

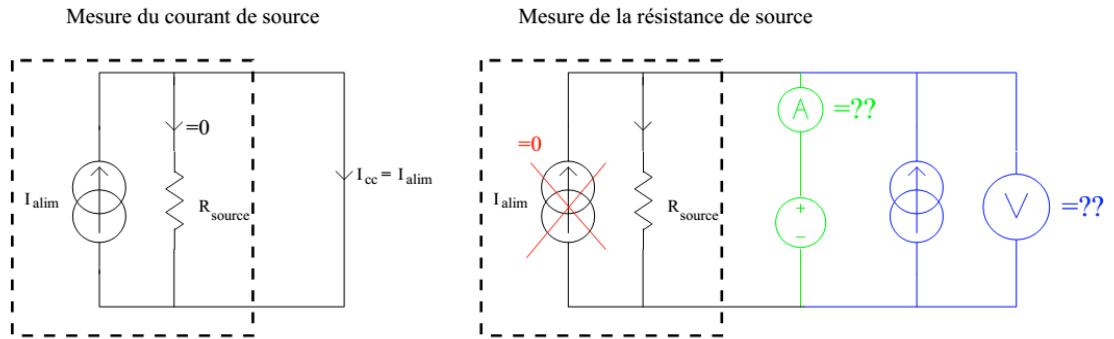


Figure 6: Mesures d'une source de courant.

Source de tension

Nous appelons source de tension idéale, un dispositif qui délivre une tension constante quelque soit le courant demandé à ses bornes (cela implique que la puissance que peut délivrer la source est théoriquement infinie). Si la tension diminue lorsque le courant demandé aux bornes de la source augmente, nous parlons alors de source de tension non-idiéale.

Une source de tension non-idiéale est représentée par une source de tension idéale en série d'une résistance de source R_{source} qui représente la perte en tension quand le courant augmente : plus celui-ci augmente, plus la tension aux bornes de R_{source} est élevée et donc moins il y a de tension disponible aux bornes du dispositif.

Le calcul de la valeur de la résistance de source est aisée : elle est égale à la pente de variation de $V=f(I)$ comme montré figure 7.

Il est à noter que pour une source de tension idéale $R_{source} = 0$. Une mise **en série** de plusieurs sources de tension revient à additionner leur effet.

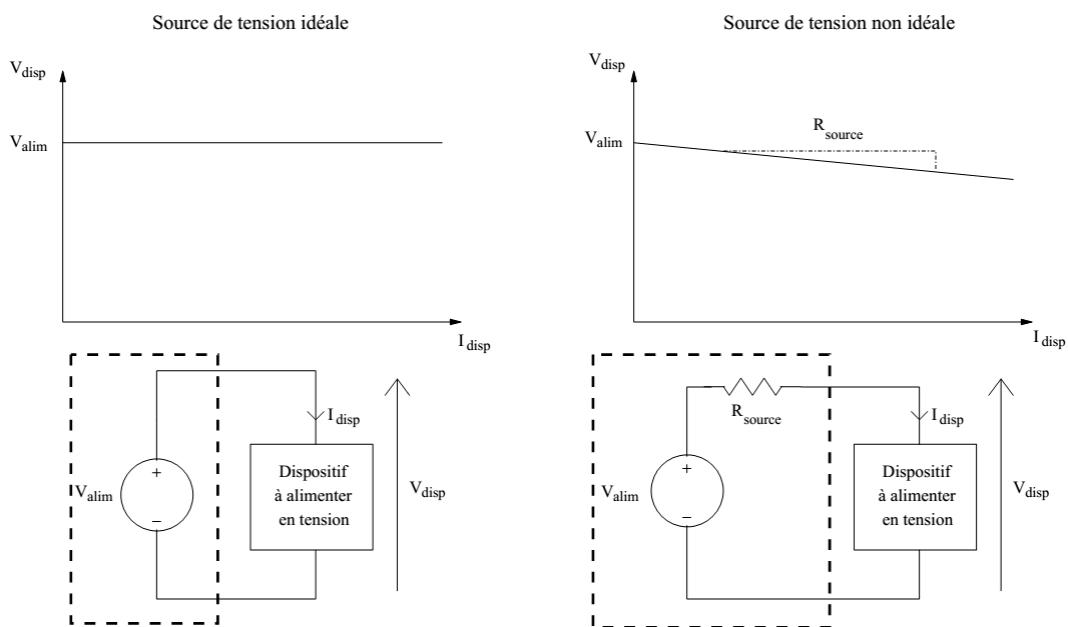


Figure 7: Sources de tension idéales et non idéales.

Calcul de V_{alim} et R_{source} :

pour calculer la valeur nominale de la tension d'une source non idéale et sa résistance de sortie sans avoir accès à l'intérieur du composant, la manière de procéder est la suivante³:

- Il suffit de ne pas tirer de courant de la source de tension : ainsi le courant qui traverse la résistance de source est nul et donc la tension à ses bornes aussi. Il faut donc faire la mesure de la tension aux bornes de la source en boucle ouverte (cela revient à mettre une résistance infinie en sortie de la source et de mesurer la tension).
- Nous éteignons la source de tension : $V_{alim} = 0$. Il suffit alors d'injecter un courant et de mesurer la tension obtenue –cas bleu– ou l'inverse (appliquer tension et mesurer le courant –cas vert–) afin d'obtenir la valeur de la résistance (rapport des deux grandeurs).

La démarche est résumée figure 8.

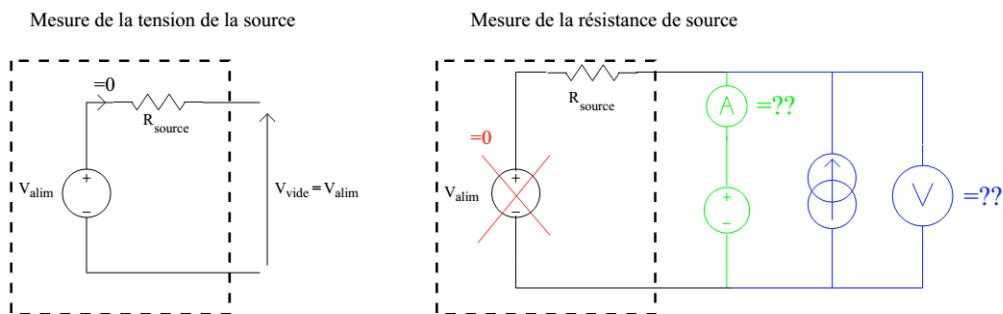


Figure 8: Mesures d'une source de tension.

1.1.4 Mailles

Généralités

Cette section est destinée aux calculs des tensions et courants dans un circuit. Nous considérons le circuit de la figure 9.

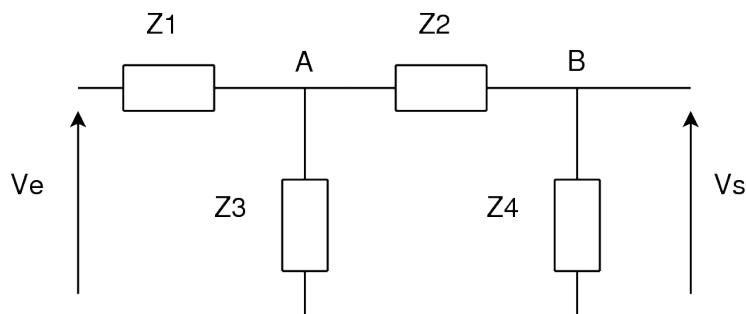


Figure 9: Exemple du calcul de maille.

La question est de calculer la tension de sortie V_s par rapport à la tension d'entrée V_e , en fonction des impédances Z du circuit. Pour cela, il faut s'intéresser aux calculs des

³ Il s'agit de la manière de procéder d'un point de vue conceptuel en aucun à reproduire de manière expérimentale.

tensions aux bornes de chaque impédance, et donc s'intéresser au courant qui traverse chaque impédance. Si rien n'est indiqué, nous considérons que V_s n'est pas connecté, où que la connexion ne modifie pas le potentiel du nœud.

Le courant circulant d'une tension élevée vers une tension faible, les sens des courants sont pris comme représenté⁴ figure 10.

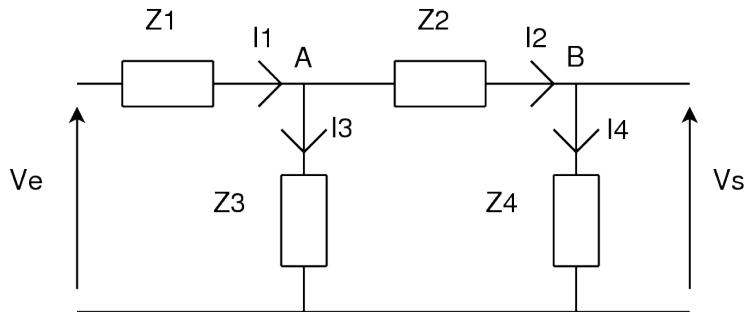


Figure 10: Définition du sens des courants.

Nous considérons V_A et V_B les potentiels en A et B. En appliquant la loi d'ohm $U = ZI$, nous pouvons écrire :

$$V_e - V_A = Z_1 I_1 \text{ et } I_2 = I_1 - I_3$$

$$V_A - 0 = Z_3 I_3 \text{ et } I_4 = I_2$$

$$V_A - V_B = Z_2 I_2$$

$$V_B - 0 = Z_4 I_4 \text{ et } V_B = V_s$$

De ces équations, nous pouvons calculer $V_s = f(V_e)$

De $I_2 = I_1 - I_3$:

$$\frac{V_A - V_s}{Z_2} = \frac{V_e - V_A}{Z_1} - \frac{V_A}{Z_3}$$

or $V_A - V_s = Z_2 I_2$ donc

$$V_A = \frac{Z_2 V_s}{Z_4} + V_s$$

En remplaçant V_A :

$$\frac{V_s}{Z_4} = \frac{V_e}{Z_1} - \frac{Z_2}{Z_1 Z_4 V_s} - \frac{V_s}{Z_1} - \frac{Z_2}{Z_3 Z_4} \cdot V_s - \frac{V_s}{Z_3}$$

De là , il vient :

$$V_s \cdot \left(\frac{1}{Z_4} + \frac{Z_2}{Z_1 Z_4} + \frac{1}{Z_1} - \frac{Z_2}{Z_3 Z_4} - \frac{1}{Z_3} \right) = \frac{V_e}{Z_1}$$

d'où l'expression finale :

$$\frac{V_s}{V_e} = \frac{1}{\frac{Z_1}{Z_4} + \frac{Z_2}{Z_4} - \frac{Z_1 \cdot Z_2}{Z_3 \cdot Z_4} - \frac{Z_1}{Z_3} + 1}$$

⁴ Dans le cadre de la convention récepteur, courant et tension sont notés en sens opposés dans un dipôle passif.

Méthodes de résolution

Pour faciliter les calculs, des formules mathématiques écrivent différemment la loi d'ohm :

Loi de Kirchhoff : en tout nœud la somme des courants est nulle :

$$\sum_n I_n = 0$$

Sur l'exemple au dessus, en A : $I_1 - I_2 - I_3 = 0$

$$\frac{V_e - V_A}{Z_1} + \frac{V_A}{Z_3} + \frac{V_A - V_B}{Z_2} = 0$$

Loi de Millman : en tout nœud la tension vaut la somme des courants divisée par la somme des admittances :

$$V = \sum_n \frac{\sum V_n \cdot Y_n}{\sum_n Y_n}$$

Sur l'exemple au dessus, en A :

$$V_A = \frac{V_e \cdot Y_1 + 0 \cdot Y_3 + V_B \cdot Y_2}{Y_1 + Y_2 + Y_3}$$

Ces équations sont *in fine* identiques, mais le résultat est parfois plus direct, selon la topologie du circuit.

Équivalences Thévenin - Norton

Pour résoudre les calculs, il est parfois plus simple de modifier le montage. Le passage d'un montage Thévenin à Norton ou inversement peut alors être utilisé (cf. figure 11).

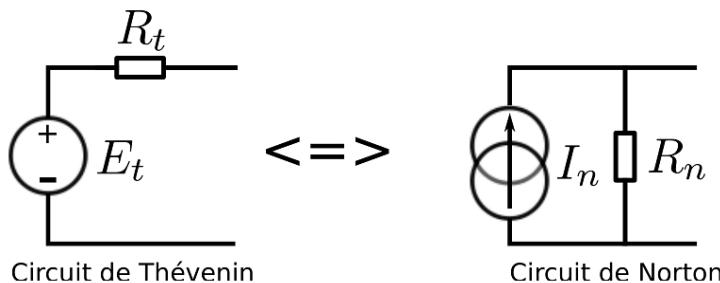


Figure 11: Équivalence Thévenin-Norton.

Circuit de Thévenin Circuit de Norton

Dans ces circuits, pour passer de Thévenin à Norton: $R_t = R_n$ et $I_n = E_t / R_t$.

Et inversement: $R_n = R_t$ et $E_t = I_n \cdot R_n$

1.1.5 Approche quadripolaire

Un quadripôle (cf. figure 12) est une boîte noire dont nous ne représentons que les entrées et les sorties en terme de courant et de tension (dans un formalisme de Kirchhoff). C'est une méthode couramment employée en électronique pour représenter et formaliser n'importe quel système électrique. Si de plus le système est linéaire, nous pouvons

représenter les relations de transfert entrées-sorties en termes matriciels et utiliser tout le formalisme matriciel pour la résolution des calculs.

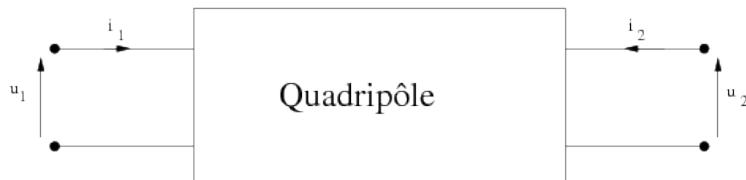


Figure 12: Quadripôle.

Il existe trois manières de représenter un quadripôle : formalisme en z (nous travaillons sur les impédances), formalisme en y (nous travaillons sur les admittances) et formalisme en h (formalisme hybride mélangeant impédances et admittances).

Paramètres z

Nous appelons (Z) la matrice d'impédance du quadripôle. Les équations associées à ce quadripôle sont :

$$u_1 = z_{11} \cdot i_1 + z_{12} \cdot i_2 \quad (1.1)$$

$$u_2 = z_{21} \cdot i_1 + z_{22} \cdot i_2 \quad (1.2)$$

Ce qui donne la représentation matricielle suivante :

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} \cdot \begin{pmatrix} i_1 \\ i_2 \end{pmatrix}$$

De cette formalisation, nous pouvons peut en tirer les équations des éléments de la matrice, c'est à dire représenter un dispositif sous forme quadripolaire ce qui revient à déterminer les éléments de la matrice à partir des entrées courant/tension du dispositif. Nous obtenons :

- $z_{11} = \frac{u_1}{i_1} \Big|_{i_2=0}$: impédance d'entrée à sortie ouverte (Ω),
- $z_{12} = \frac{u_2}{i_1} \Big|_{i_2=0}$: impédance de transfert directe à sortie ouverte (Ω),
- $z_{21} = \frac{u_1}{i_2} \Big|_{i_1=0}$: impédance de transfert inverse à entrée ouverte (Ω),
- $z_{22} = \frac{u_2}{i_2} \Big|_{i_1=0}$: impédance de sortie à entrée ouverte (Ω).

Cela donne le schéma quadripolaire équivalent de la figure 13.

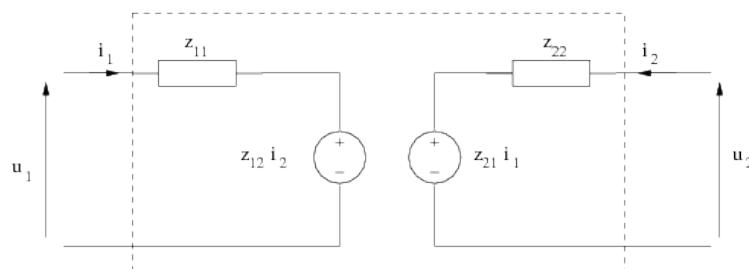


Figure 13: Quadripôle en paramètre z.

Paramètres y

Nous appelons (Y) la matrice d'admittance du quadripôle. Les équations associées à ce quadripôle sont :

$$i_1 = y_{11} \cdot u_1 + y_{12} \cdot u_2 \quad (1.3)$$

$$i_2 = y_{21} \cdot u_1 + y_{22} \cdot u_2 \quad (1.4)$$

Ce qui donne la représentation matricielle suivante :

$$\begin{pmatrix} i_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

De cette formalisation, nous pouvons en tirer les équations des éléments de la matrice. Nous obtenons :

- $y_{11} = \frac{i_1}{u_1} \Big|_{u_2=0}$: admittance d'entrée à sortie court-circuitée (Ω^{-1}),
- $y_{12} = \frac{i_2}{u_1} \Big|_{u_2=0}$: admittance de transfert directe à sortie court-circuitée (Ω^{-1}),
- $y_{21} = \frac{i_1}{u_2} \Big|_{u_1=0}$: admittance de transfert inverse à entrée court-circuitée (Ω^{-1}),
- $y_{22} = \frac{i_2}{u_2} \Big|_{u_1=0}$: admittance de sortie à entrée court-circuitée (Ω^{-1}).

Cela donne le schéma quadripolaire équivalent de la figure 14.

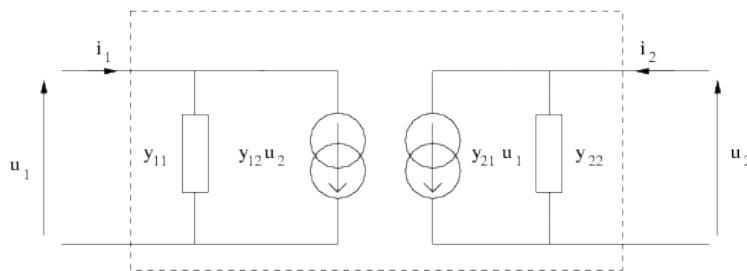


Figure 14: Quadripôle en paramètre y.

Paramètres h

Nous appelons (H) la matrice hybride du quadripôle. Les équations associées à ce quadripôle sont :

$$u_1 = h_{11} \cdot i_1 + h_{12} \cdot u_2 \quad (1.5)$$

$$i_2 = h_{21} \cdot i_1 + h_{22} \cdot u_2 \quad (1.6)$$

Ce qui donne la représentation matricielle suivante :

$$\begin{pmatrix} u_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \cdot \begin{pmatrix} i_1 \\ u_2 \end{pmatrix}$$

De cette formalisation, nous pouvons en tirer les équations des éléments de la matrice. Nous obtenons :

- $h_{11} = \frac{u_1}{i_1} \Big|_{u_2=0}$: impédance d'entrée à sortie court-circuitée (Ω),
- $h_{12} = \frac{u_1}{u_2} \Big|_{i_1=0}$: gain de tension à entrée ouverte (-),
- $h_{21} = \frac{i_2}{i_1} \Big|_{u_2=0}$: gain de courant à sortie court-circuitée (-),
- $h_{22} = \frac{i_2}{u_2} \Big|_{i_1=0}$: admittance de sortie à entrée ouverte (Ω^{-1}).

Cela donne le schéma quadripolaire équivalent de la figure 15.

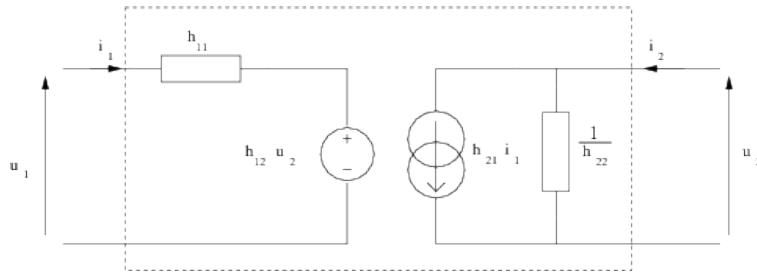


Figure 15: Quadripôle en paramètre h .

Les paramètres hybrides sont excessivement utilisés en électronique. L'interprétation électronique du schéma 15 est la suivante :

- l'entrée du quadripôle présente une **impédance d'entrée** limitant le courant de charge de la source en amont,
- il existe une **tension de réaction** (sortie \leftarrow entrée) proportionnelle à la tension de sortie,
- la sortie est une **source de courant** non idéale contrôlée par le courant d'entrée présentant une **admittance de sortie**.

Passage entre formalismes

Il existe bien évidemment des règles basées sur la théorie des réseaux de Kirchhoff qui permettent de passer d'un formalisme à un autre. Elles ne seront pas développées ici mais se trouvent facilement dans la littérature.

Utilisation d'un quadripôle

Lorsque un quadripôle, représenté par les paramètres hybrides h , est chargé par une impédance en sortie Z_L et alimenté par une source de tension idéale (ou au moins d'impédance d'entrée négligeable, nous sommes dans le cas de la figure 16).

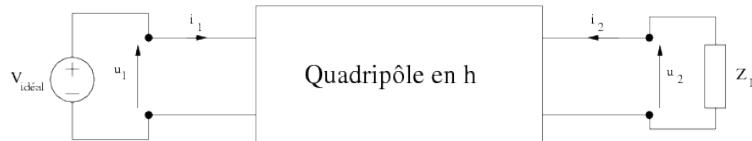


Figure 16: Quadripôle en paramètre h avec résistance de charge en sortie.

Nous obtenons dès lors les relations suivantes :

- **Gain en courant :** $A_i = \frac{i_2}{i_1} = \frac{h_{21}}{1 + Z_L \cdot h_{22}}$
- **Gain en tension :** $A_v = \frac{u_2}{u_1} = \frac{h_{21} \cdot Z_L}{h_{11} + h_{12}h_{21}Z_L + \frac{h_{11}}{h_{22}} \cdot Z_L}$
- **Impédance d'entrée :** $Z_e = \frac{u_1}{i_1} = \frac{h_{11} + h_{12}h_{21}Z_L}{1 + \frac{Z_L}{h_{22}}}$
- **Impédance de sortie :** $Z_s = \frac{u_2}{i_2} = h_{22}$

Ces formules, qu'il n'est pas nécessaire de retenir par cœur, permettent ainsi le calcul aisément de tout montage représenté par un quadripôle et deviennent encore plus intéressante lorsque nous mettons plusieurs quadripôles en cascade.

Remarque importante :

En électronique, les systèmes analogiques que nous cherchons à créer sont des systèmes pour lesquels la fonctionnalité est indépendante de la charge en sortie du système⁵. Ainsi, la plupart des systèmes qui sont rencontrés en électronique auront une tension de retour nulle c'est à dire un paramètre h_{12} nul ou alors que nous chercherons par tous les moyens à minimiser.

Ainsi le modèle quadripolaire le plus couramment utilisé en électronique est celui de la figure 17.

⁵ Cette approche est vitale dans le développement de systèmes réutilisables dans de multiples contextes et pour lesquels il est hors de question que la fonctionnalité soit mise en défaut par une impédance de charge. On cherche ainsi à créer et utiliser des systèmes présentant une quasi-séparation physique entre l'entrée et la sortie

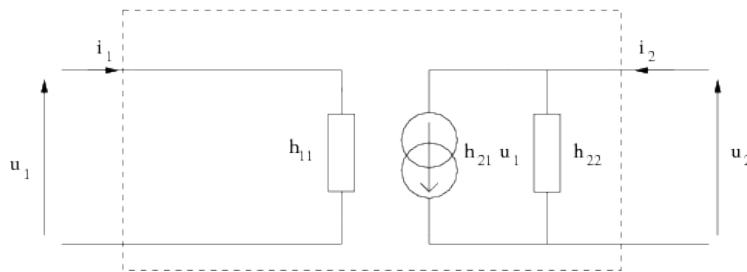


Figure 17: Quadripôle en paramètre h simplifié.

Des différents types d'amplificateurs

Dans le prolongement du paragraphe précédent, nous pouvons découper les quadripôles en deux catégories différentes elle-mêmes découpées en deux catégories. Il existe deux types d'amplificateurs⁶ selon s'ils délivrent un signal de tension ou un signal de courant en sortie. La seconde catégorie concerne le signal qui commande la source de sortie : tension ou courant. Nous obtenons ainsi quatre types amplificateur :

- **Amplificateur de tension** : il délivre en sortie une tension dont la valeur est commandée par la tension en entrée du quadripôle (VCS en anglais pour Voltage Controlled Voltage Source),
- **Amplificateur de courant** : il délivre en sortie un courant dont la valeur est contrôlée par le courant rentrant dans le quadripôle (CCCS en anglais pour Current Controlled Current Source),
- **Amplificateur de transimpédance** : il convertit un courant d'entrée en tension de sortie (CCVS en anglais pour Current Controlled Voltage Source) : c'est la base de toutes les commandes de systèmes à base de détection de lumière type photodiode et de remise en forme de signaux issus de capteurs à base de courant,
- **Amplificateur de transconductance** : il convertit une tension en courant de sortie (VCCS en anglais pour Voltage Controlled Current Source) : c'est le système à la base de toutes les commandes de systèmes de laser et de nombreux systèmes commandant des dispositifs en courant. C'est également le formalisme couramment utilisé pour les transistors MOS.

Les deux derniers cas peuvent être vus comme des combinaisons hybrides des deux premiers amplificateurs.

Amplificateur de tension

Le schéma se trouve figure 18.

⁶ Il est évident d'après ce qui précède qu'adopter une modélisation quadripolaire revient à formaliser le dispositif sous la forme d'un amplificateur.

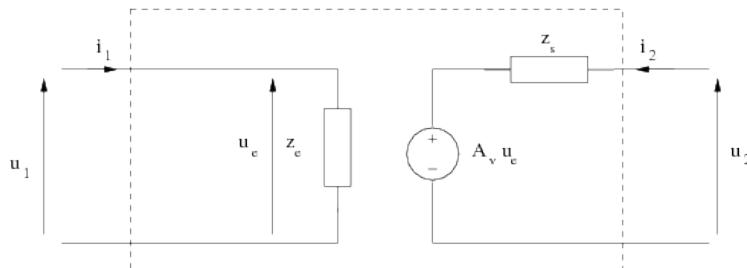


Figure 18: Amplificateur de tension.

Le cas de l'amplificateur idéal de tension est :

- Impédance d'entrée infinie qui garantit qu'indépendamment de la qualité de la source de tension en entrée, la totalité du signal utile se trouve aux bornes de Z_e sans perte dans l'impédance de la source de tension d'entrée soit $Z_e = \infty$,
- Impédance de sortie nulle (afin que tout le signal amplifié soit transmis en sortie sans perte dans l'amplificateur) soit $Z_s = 0$,
- Gain en tension A_v précis et stable et dans le cas d'un AOP infini et de bande passante infinie.

Amplificateur de courant

Le schéma se trouve figure 19.

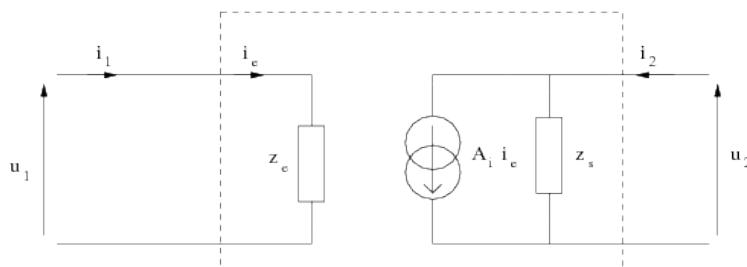


Figure 19: Amplificateur de courant.

Le cas de l'amplificateur idéal de courant est :

- Impédance d'entrée nulle qui garantit qu'indépendamment de la qualité de la source de courant en entrée, la totalité du signal utile entre dans l'amplificateur sans perte dans l'impédance parallèle de la source de courant d'entrée soit $Z_e = 0$,
- Impédance de sortie infinie (afin que tout le signal amplifié soit transmis en sortie sans perte dans l'amplificateur) soit $Z_s = \infty$,
- Gain en courant A_i précis et stable et de bande passante infinie.

Amplificateur de transimpédance

Le schéma se trouve figure 20.

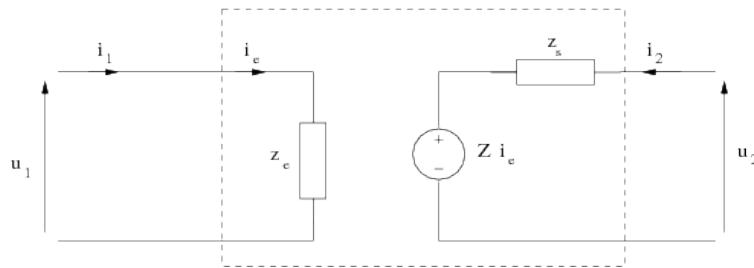


Figure 20: Amplificateur de transimpédance.

Le cas de l'amplificateur idéal de transimpédance (pour lequel il peut être judicieux de se souvenir que cet amplificateur est un hybride des deux cas vus précédemment) est:

- Impédance d'entrée nulle qui garantit qu'indépendamment de la qualité de la source de courant en entrée, la totalité du signal utile entre dans l'amplificateur sans perte dans l'impédance parallèle de la source de courant d'entrée soit $Z_e = 0$,
- Impédance de sortie nulle (afin que tout le signal amplifié soit transmis en sortie sans perte dans l'amplificateur) soit $Z_s = 0$,
- Gain de transimpédance Z précis et stable et de bande passante infinie.

Amplificateur de transconductance

Le schéma se trouve figure 21.

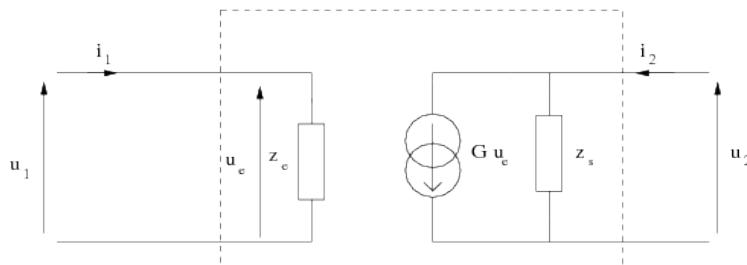


Figure 21: Amplificateur de transconductance.

Le cas de l'amplificateur idéal de transconductance est :

- Impédance d'entrée infinie qui garantit qu'indépendamment de la qualité de la source de tension en entrée, la totalité du signal utile se trouve aux bornes de Z_e sans perte dans l'impédance de la source de tension d'entrée soit $Z_e = \infty$,
- Impédance de sortie infinie (afin que tout le signal amplifié soit transmis en sortie sans perte dans l'amplificateur) soit $Z_s = \infty$,
- Gain en transconductance G précis et stable et de bande passante infinie.

1.1.6 Filtres linéaires du 1^{er} ordre

En combinant résistances, capacités et inductances, il est possible de réaliser des filtres simples. Ces circuits ont comme particularité de laisser passer le signal sur une gamme de fréquence et de le couper (filtrer) sur une autre gamme, d'où le nom de filtre.

Il existe donc différents montages selon les composants utilisés (R,C; R,L; L,C et R,L,C) comme le montre la table 1.

R,C	R,L	L,C	R,L,C
RC	RL	LC	RLC
CR	LR	CL	RCL
			CRL
			CLR
			LCR
			LRC

Table 1: Les différents types de montage RC, RL et RLC.

Soient les deux filtres de la figure 22.

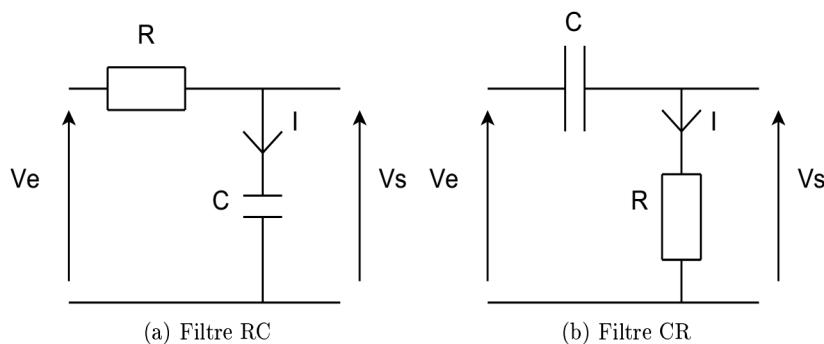


Figure 22: Filtres du premier ordre.

Les calculs peuvent être identiques si nous considérons les impédances : le circuit compose un diviseur de tension. Nous avons :

$$v_s = Z_2 \cdot i \text{ et } v_e = (Z_1 + Z_2) \cdot i \text{ donc } v_s / v_e = \frac{Z_2}{Z_1 + Z_2}.$$

Circuit RC

Dans le circuit RC, $Z_1 = R$ et $Z_2 = 1 / jC\omega$ donc $v_s / v_e = \frac{1}{1 + jRC\omega}$

Pour des fréquences (et donc pulsations) faibles, $v_s / v_e \approx 1$: filtre passe bas

Pour des fréquences (et donc pulsations) élevées, $v_s / v_e \approx 0$

Etude temporelle du circuit RC

Nous considérons que le condensateur est déchargé à $t=0$, et que nous appliquons un échelon de tension E (créneau).

$$E = R \cdot i + v_s \text{ or } i = C \cdot dv_s / dt$$

ce qui donne l'équation finale :

$$RC \frac{dv_s}{dt} + v_s = E$$

En résolvant cette équation différentielle, nous trouvons :

$$v_s = E \cdot \left(1 - e^{-t/RC}\right)$$

De la même manière, pour une décharge:

$$v_s = E \cdot e^{-t/RC}$$

Circuit CR

Dans le circuit CR, $Z_1 = 1 / jC\omega$ et $Z_2 = R$ donc

$$\frac{v_s}{v_e} = \frac{jRC\omega}{1 + jRC\omega}$$

Pour des fréquences (et donc pulsations) faibles, $v_s / v_e \approx 0$

Pour des fréquences (et donc pulsations) élevées, $v_s / v_e \approx 1$: filtre passe haut

N.B: nous écrivons souvent $v_s / v_e = \frac{1}{1 + j\omega/\omega_0}$ ou $v_s / v_e = \frac{j\omega/\omega_0}{1 + j\omega/\omega_0}$ avec $\omega_0 =$

$1/RC$

Nous mettons ainsi en évidence la fréquence particulière où $f = f_0$ ($\omega = \omega_0$), la fréquence (f_0) pulsation (ω_0) de coupure.

Tracé des asymptotes : Pour calculer le gain, nous prenons le module de ces fonctions de transfert:

$$G_{RC} = \frac{1}{\sqrt{1 + (RC\omega)^2}} \text{ et } G_{CR} = \frac{RC\omega}{\sqrt{1 + (RC\omega)^2}}$$

Nous calculons usuellement les fonctions de transfert V_s / V_e en décibels.

Pour les tensions: $\left(\frac{V_s}{V_e}\right)_{dB} = 20 \cdot \log \frac{V_s}{V_e}$. Ainsi nous obtenons les diagrammes de Bode

en gain de la figure 23.

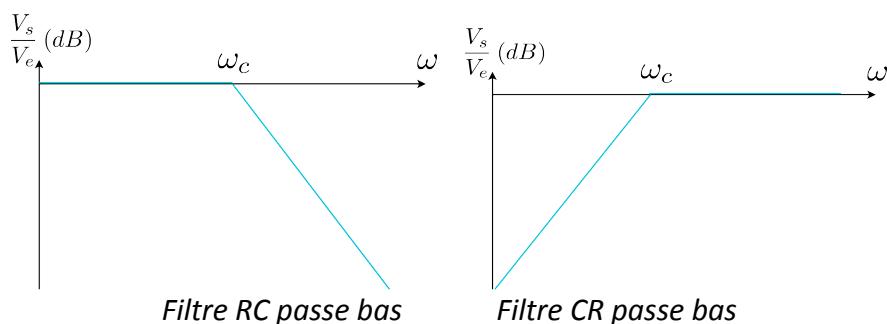


Figure 23: Diagramme de Bode en gain des filtres.

La coupure du filtre commence (RC) ou finit (CR) à $\omega = \omega_0$, pulsation pour laquelle⁷ :

$$V_s / V_e = 20 \cdot \log \left(\log 1 / \sqrt{2} \right) = -3 \text{ dB}$$

⁷ On peut rappeler que la fréquence de coupure est la fréquence à laquelle la puissance du signal de sortie est divisée par deux par rapport à sa puissance maximale. Une puissance étant homogène à une tension au carré, on retrouve le $\sqrt{2}$ de la formule donnée.

1.1.7 Amplificateurs opérationnels (AOP)

Les amplificateurs opérationnels sont des éléments largement utilisés grâce à leurs caractéristiques très intéressantes pour les applications classiques. En effet, les caractéristiques (détaillées plus bas) sont proches de l'idéal pour des applications électroniques classiques à faible fréquence ($< 10 \text{ MHz}$) et à courant fort ($> 10 \mu \text{ A}$).

Son schéma bloc est représenté figure 24.

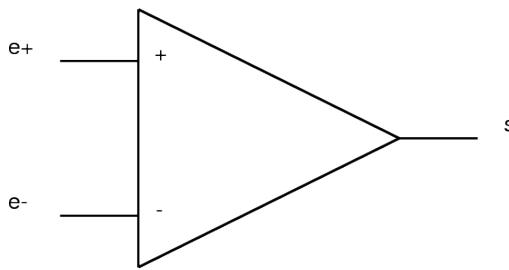


Figure 24: Schéma d'un AOP.

Il est constitué de 2 entrées e^+ et e^- , et d'une sortie s . L'intérieur de cet amplificateur comprend plusieurs étages: un premier étage de différence, un deuxième étage d'amplification. En conséquence, cet amplificateur est un amplificateur de différence à très grand gain :

$$S = (e^+ - e^-) \cdot A_0$$

Les caractéristiques principales sont l'impédance d'entrée Z_e , l'impédance de sortie Z_s , et le gain A_0 (cf. tableau 2) .

	Z_e	Z_s	A_0
Idéal	infinie	nulle	infini
Réel	$10^6 \Omega$	100Ω	10^6

Table 2: Caractéristiques d'un AOP.

En connectant ce circuit directement sur 2 signaux différents et en observant la sortie, le montage fonctionne en comparateur. En effet, la sortie ne peut pas générer une tension qui tend vers l'infini (ou qui vaut 10^6 V si $e^+ - e^- = 1\text{V}$!). Il y a une tension maximale (dite de saturation) qui est proche des tensions d'alimentation $V_{max} = + V_{sat} \approx V_{cc}$ (alimentation haute) et $V_{min} = - V_{sat} \approx V_{ss}$ (alimentation basse). La sortie d'un comparateur est donc bistable ($\pm V_{sat}$).

Nous pouvons ajouter des éléments R,L,C autour de cet amplificateur opérationnel pour en faire un amplificateur de tension, un filtre etc...

1.2 Électronique numérique et Logique

1.2.1 Codage binaire

En numérique classique, nous utilisons des grandeurs logiques à la place de

grandeur physiques. Ces grandeurs logiques sont au nombre de 2: "**0**" logique et "**1**" logique. Ainsi, nous pouvons trouver les correspondances dans le tableau 3.

Grandeur logique	"0"	"1"
Grandeur physique	0v	1v
	0v	3,3v
	0v	5v
	-5v	+5v

Table 3: Grandeurs logiques et grandeurs physiques : quelques exemples de correspondance.

Pour coder des entiers ou des flottants, le codage se fait donc en base 2⁸). Chaque bit (BInary digiT) est donc associé à un poids en base 2 :

$$\dots 2^7 \ 2^6 \ 2^5 \ 2^4 \ 2^3 \ 2^2 \ 2^1 \ 2^0$$

Le vecteur de bits (ou mot binaire) 11001101 vaut donc en décimal :

$$2^7 + 2^6 + 0 + 0 + 2^3 + 2^2 + 0 + 2^0 = 205$$

Il existe d'autre codage binaire comme le code Gray⁹ appelé également binaire réfléchi ou encore le code BCD (Binaire Codé Décimal) où chaque puissance de 10 est codée séparément en base 2.

1.2.2 Codage hexadécimal

Il peut être fastidieux d'écrire un mot binaire (0110011100101101 illisible !) ou en décimal (213 à convertir !). Le codage hexadécimal est donc un codage largement utilisé car il associe chaque quartet (4 bits) d'un mot binaire à un caractère. Comme la base 16 comporte 16 caractères, il serait inopportun d'écrire 213 (= 2,1,3 ou 2,13 ?). Nous utilisons donc les nombres 0 à 9 puis les lettres A à F (voir le tableau 4).

Décimal	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Hexadécimal	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F

Table 4: Codage hexadécimal.

Ainsi, le mot binaire 0110011100101101 pourra s'écrire plus simplement et sans ambiguïté : 672D

1.2.3 Algèbre de Boole

L'algèbre de Boole permet de travailler sur les équations logiques. Si nous considérons un booléen A, il peut prendre les valeurs 0 ou 1, nous notons ainsi A (si A = 1) ou \overline{A} (si A = 0).

Les opérateurs utilisés couramment sont des opérateurs logiques, car les portes

⁸ de manière similaire à la base 10 que nous utilisons quotidiennement d'où l'appellation de binaire naturel.

⁹ Il est à noter que ce code dont le brevet fut déposé en 1953 par Franck Gray remonte au XIX^e siècle où on peut le trouver dans un opuscule intitulé *Théorie du Baguenaudier, par un clerc de notaire lyonnais* par Monsieur Louis Gros et publié à Lyon, en 1872. Il était présenté en lien avec un casse-tête.

logiques de base effectuent des opérations logiques (et non pas arithmétiques comme nous avons l'habitude). Ainsi, le "•" est l'opérateur ET, le "+" est l'opérateur OU. Ci-dessous sont détaillées les tables de vérités (table 5) de ces 2 opérateurs. A et B sont 2 entrées, S est la sortie.

A	B	S = A.B
0	0	0
0	1	0
1	0	0
1	1	1

A	B	S = A+B
0	0	0
0	1	1
1	0	1
1	1	1

Table 5: Table de vérité des fonctions ET et OU.

Le ET logique est vrai si A ET B valent 1; le OU logique est vrai si A OU B OU les deux valent 1.

Exemple de logique de Boole :

L'équation $S = \overline{A} \cdot B$ est donc vraie ($S = 1$) ou fausse ($S = 0$) selon les valeurs de A et B.

Il faut A = 0 (donc $\overline{A} = 1$) et B = 1 pour que S = 1.

Théorèmes : l'ensemble des théorèmes applicables à l'algèbre de Boole sont récapitulés dans le tableau 6.

<i>Idempotence</i>	$A \cdot A = A$
<i>Involution</i>	$A + A = A$ $\overline{\overline{A}} = A$
<i>Éléments neutres</i>	$A + 0 = A$ $A \cdot 1 = A$
<i>Éléments absorbants</i>	$A + 1 = 1$ $A \cdot 0 = 0$
<i>Complémentation (unicité)</i>	$A + \overline{A} = 1$ (loi du tiers exclus) $A \cdot \overline{A} = 0$ (principe de contradiction)
<i>Absorption</i>	$A + A \cdot B = A$ $A \cdot (A + B) = A$
<i>Redondance</i>	$A \cdot B + A \cdot C = A \cdot B + A \cdot C + B \cdot C$
<i>Théorème de De Morgan</i>	$\overline{A \cdot B} = \overline{A} + \overline{B}$ $\overline{A + B} = \overline{A} \cdot \overline{B}$
<i>Commutativité</i>	$A + B = B + A$ $A \cdot B = B \cdot A$
<i>Associativité</i>	$(A + B) + C = A + (B + C)$ $(A \cdot B) \cdot C = A \cdot (B \cdot C)$
<i>Distributivité</i>	$(A + B) \cdot C = (A \cdot C) + (B \cdot C)$ $(A \cdot B) + C = (A + C) \cdot (B + C)$

Table 6: Ensemble des théorèmes liés à l'algèbre de Boole.

Ces théorèmes permettent de modifier ou simplifier des équations logiques.

Ex: $S = \overline{A \cdot B} + \overline{C \cdot D} = \overline{\overline{S}} = \overline{\overline{A \cdot B} \cdot \overline{C \cdot D}}$ (un seul type de portes logiques utilisées)

2 Introduction à la physique des semiconducteurs

Au sens strict, semiconducteur est le nom donné à une famille de matériau caractérisé par des propriétés physiques bien déterminées. Cependant, le mot semiconducteur peut prendre plusieurs significations selon le contexte dans lequel il apparaît et suivant qui l'exprime. Dans la vie quotidienne, le terme semiconducteur évoque l'ensemble des composants, dispositifs et même des systèmes microélectroniques construits avec des éléments semiconducteurs.

L'objectif de ce chapitre est d'introduire les notions de base de la physique des semiconducteurs, en considérant en particulier comme exemple le silicium – le semiconducteur qui a permis la création de la deuxième génération de transistors, puis les circuits intégrés (les « puces »). La deuxième partie de ce chapitre sera consacrée pour illustrer un exemple de composants à semiconducteurs : la jonction p-n.

2.1 Théorie simplifiée de la physique des semiconducteurs

Cette partie a pour but de fournir les bases nécessaires à la compréhension du fonctionnement physique des composants électroniques. Pour avoir une explication détaillée et exhaustive des phénomènes physiques, on se reportera aux cours de Physique Chimie de la Matière ou aux ouvrages cités en bibliographie.

2.1.1 De l'atome au cristal

Atome de Bohr

L'atome (*insécable*, en grec) est l'élément de base qui constitue la matière. Chaque atome est constitué d'un noyau (composé des protons de charge positive de $+e=1.6\times10^{-19}$ Coulombs et des neutrons de charge neutre) autour duquel gravitent des électrons (de charge négative $-e$). L'atome étant neutre, la charge des protons équilibre celle des électrons. Le modèle de Bohr communément employé (remplaçait le modèle planétaire de Rutherford) pose les préceptes suivants :

- Orbite circulaire stable des électrons autour du noyau
- Quantification de l'énergie
- « Sauts » entre différentes orbites stables avec l'émission/absorption d'un photon

Chaque niveau d'énergie discret E_1, E_2, \dots, E_n ne peut être occupé que par un nombre limité d'électrons, et définit une couche électronique (nommée K, L, M...) (cf. figure 2.a). Plus le niveau est élevé, plus l'électron est éloigné du noyau. On choisit comme origine le point énergétique nul ($E=0$) qui correspond à l'énergie de l'électron qui s'est soustrait à l'influence du noyau (distance infinie du noyau). Ainsi, tous les niveaux d'énergie sont négatifs. Ceci traduit le fait qu'il faille fournir un travail pour éloigner l'électron de son noyau. En générale, en physique des semiconducteurs, on exprime les énergie en unité électron-volt (eV) : $1\text{eV}=1.6\times10^{-19}$ Joules.

Parmi ces niveaux d'énergie, on différencie trois types d'électrons :

- Les **électrons internes** qui occupent les premières couches : ils sont fortement liés au noyau,
- Les **électrons de valence** ou périphériques qui occupent la couche la plus externe dans le modèle planétaire donc celle la plus éloignée du noyau. Les électrons sont faiblement liés au noyau.
- Enfin les **électrons libres** dont l'énergie est positive et qui ne sont pas liés au noyau.

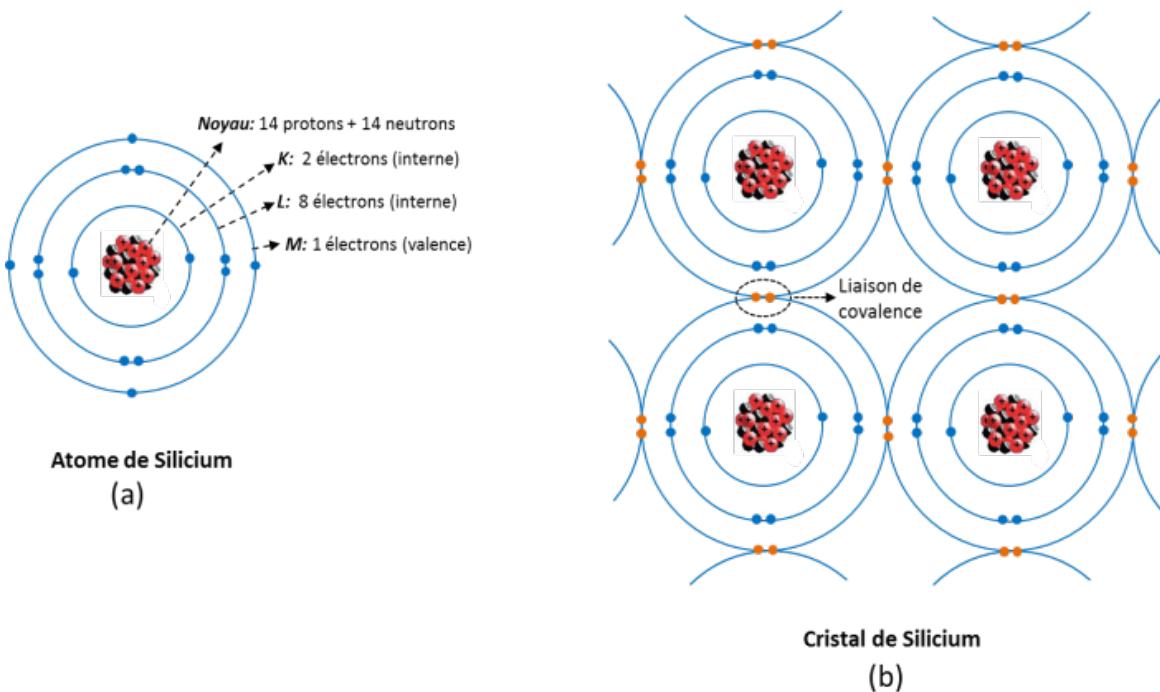


Figure 1: (a) Modèle de Bohr de l'atome de silicium. (b) Cristal de silicium à 0K

La figure 1.a illustre la vision du modèle de Bohr avec pour exemple l'atome de silicium composé d'un noyau (14 protons et 14 neutrons) et de 14 électrons périphériques (2 électrons sur la couche K, 8 électrons sur la couche L et 4 électrons sur la couche M). La couche M (couche de valence) est incomplète avec 4 électrons manquants. L'atome de silicium cherchera donc à mettre les 4 électrons de sa bande de valence en partage afin de compléter cette dernière à 8 électrons. L'agencement des plusieurs atomes entre eux avec mise en commun des électrons de valence forme ce que l'on appelle un cristal (cf. figure 1.b).

Structure cristalline et la notion de bandes d'énergie

Un cristal est un solide dont les noyaux atomiques sont disposés aux nœuds d'un réseau géométrique régulier pour former une **structure cristalline**. La cohésion de cet édifice est assurée par les liens de valence qui résultent de la mise en commun de deux électrons appartenant chacun à deux atomes voisins de la maille cristalline.

Les états énergétiques des électrons au sein du cristal sont quantifiés et se représentent de manière similaire à celui de l'atome ; mais du fait du principe d'exclusion de Pauli¹ les niveaux d'énergie sont devenus des bandes d'énergie permises séparées par des bandes d'énergie interdites (cf. figure 2.b) : on parle de quasi-continuum. Comme pour l'atome, les bandes d'énergie accueillent un nombre limité d'atomes et les niveaux d'énergie les plus faibles sont comblés prioritairement.

¹ deux électrons ne peuvent pas occuper le même état quantique.

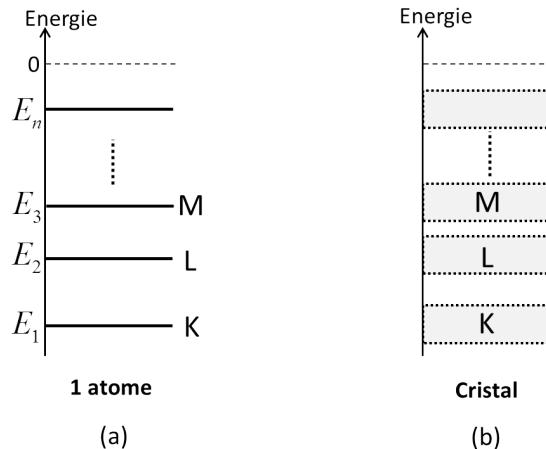


Figure 2: (a) Les niveaux d'énergie électronique discrets d'un atome. (b) Les bandes d'énergie électronique d'un cristal.

2.1.2 Niveau de Fermi

Les électrons sont des fermions² qui obéissent à la statistique de Fermi-Dirac. Ainsi, la probabilité à une température T pour qu'un état d'énergie E soit occupé est donnée par la fonction de distribution³ :

$$f(E) = \frac{1}{1 + e^{\left(\frac{(E-E_F)}{k_B T}\right)}}$$

Avec E_F qui représente le niveau de Fermi des électrons dans le dispositif, T la température et k_B la constante de Boltzmann. La figure 3 présente la fonction $f(E)$ à température nulle ($T=0K$) et à haute température ($T\approx 300K$). On note que :

- **T=0K** : $f(E)$ est une fonction en escalier avec $f(E < E_F) = 1$ et $f(E > E_F) = 0$. Toutes les bandes d'énergie en-dessous du niveau de Fermi sont totalement remplies et toutes ceux qui sont au-dessus du niveau de Fermi sont vides.
- **T>0K** : La fonction d'escalier est « bavée » et on peut avoir $f(E > E_F) > 0$ avec une valeur de E « pas trop loin » de E_F . Ce comportement est d'autant plus important en augmentant la température. Les bandes d'énergie au-dessus du niveau de Fermi peuvent être donc peuplées avec une agitation thermique.

² Des particules de spin $\frac{1}{2}$ dans la mécanique quantique.

³ À haute température, lorsque les effets quantiques ne se font plus sentir, la statistique de Fermi-Dirac tend vers la statistique de Maxwell-Boltzmann ; il en est de même pour la statistique de Bose-Einstein qui régit les bosons.

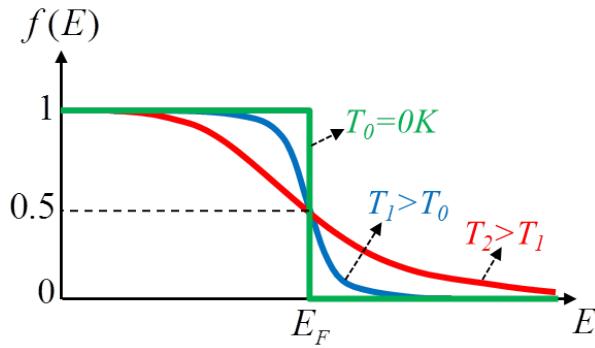


Figure 3: La distribution Fermi à différentes températures : $T_0=0K$, $T_1>T_0$ et $T_2>T_1$

Il est à noter que la population d'électrons dans une bande d'énergie dépend du produit de la **fondation de Fermi** $f(E)$ et la fonction $\rho(E)$ qui est la **densité d'états électroniques**⁴ dans cette bande :

$$n = \int f(E) \cdot \rho(E) \cdot dE$$

2.1.3 Classification des solides

Les matériaux de cristaux solides sont divisés en trois classes : métaux, semiconducteurs et isolants. Dans un premier temps, on peut définir d'une manière simple :

- **Métaux** : les matériaux qui ont une conductivité électrique non nulle même à température nulle (i.e. $T=0K$).
- **Semiconducteurs** : les matériaux qui ont une conductivité nulle à $T=0K$ mais deviennent conductrice à haute température (i.e. $T\sim 300K$).
- **Isolants** : les matériaux qui ont une conductivité nulle à toutes les températures.

La classification des solides peut être expliquée par la théorie des bandes d'énergie. En effet, la conductivité est directement liée au remplissage des bandes : Les électrons d'une bande d'énergie ne peuvent participer à la conduction électrique que si et seulement si cette bande est partiellement remplie :

- Si la bande est complètement remplie, la présence d'un champ électrique extérieur n'apporte aucune énergie supplémentaire aux électrons car tous les états d'énergie sont occupés. Ces électrons sont donc dans des états localisés et ils ne peuvent pas participer aux phénomènes de conduction électrique.
- En revanche, une bande partiellement remplie contient à la fois des électrons et des états non-occupés. Ces derniers permettent des porteurs de gagner en énergie à la présence d'un champ électrique extérieur, et donc participent à la conductivité du cristal. Les électrons d'une bande partiellement peuplée sont alors des **électrons de conductions**. Ils sont délocalisés dans le cristal et « quasiment » libres (ils subissent toujours l'influence du réseau cristallin)⁵.

Dans le cas des métaux, le niveau de Fermi se trouve au milieu d'une bande d'énergie (Figure 4.a). Comme cette bande est partiellement remplie même à température nulle, les électrons de

⁴ Cette fonction est obtenue analytiquement en physique statistique à l'aide du modèle de gaz d'électrons.

⁵ Dans la suite de ce chapitre, on les appelle « électrons libres » par abuse de langage.

conduction sont donc toujours présents dans les métaux⁶.

Dans le cas des semiconducteurs et des isolants, le niveau de Fermi se trouve entre deux bandes d'énergie (Figure 4.b et 4.c). On définit:

- **Band de Valence:** la dernière bande électronique (i.e. celle de plus haute énergie) totalement remplie à $T=0K$ ⁷. Les électrons de la bande de valence sont communs à plusieurs atomes en étant des électrons de valences, et assurent la cohésion du cristal (entre atomes voisins). Au contraire, chaque électron des bandes situées en-dessous de la bande de valence sont liés à un atome donné, en étant un électron interne.
- **Band de Conduction:** la première bande électronique qui est au-dessus de la bande de valence. A $T=0K$, elle est complètement vide.
- **Gap :** le nom donné à la largeur de la bande interdite qui sépare la bande de valence et la bande de conduction. Le niveau de Fermi se trouve dans le gap, et la valeur E_g du gap (donnée en eV) va définir les caractéristiques électriques du matériau puisqu'elle fixe l'énergie qu'il est nécessaire d'acquérir aux électrons de la bande de valence pour passer dans la bande de conduction.

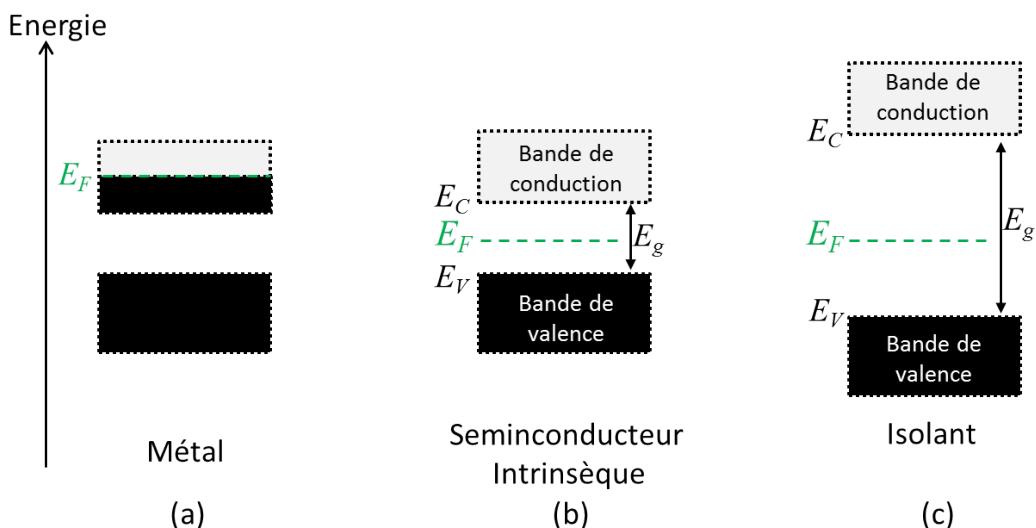


Figure 4: La structure de bande d'un métal (a), semiconducteur intrinsèque (b) et isolant (c). La couleur « sombre » représente le remplissage d'électrons à température nulle.

C'est la valeur du gap qui distingue les semiconducteurs et les isolants :

- **Isolant :** le gap est de plusieurs eV (généralement supérieur à $4eV$)⁸. Il y a donc une très forte séparation entre le niveau de Fermi et la bande de conduction. Cette dernière reste donc vide même à forte température.

⁶ Les comportements métalliques peuvent être aussi obtenus lorsqu'une bande vide et une bande remplie se chevauchent. Ce cas n'est pas présenté sur la figure 3.

⁷ Du fait de l'agitation thermique, les électrons à une température supérieure à celle du zéro absolu sont susceptibles d'acquérir suffisamment d'énergie pour s'échapper de leur orbite pour aller vers des orbites supérieures.

⁸ La limite entre les semiconducteurs et les isolants a été fixée de manière arbitraire à $4eV$.

- Semiconducteur :** le gap est proche de $1eV$, donc une agitation thermique peut rompre les liaisons de covalence pour « arracher » des électrons de valences et les peupler dans la bande de conduction.

Catégorie	Bandgap	Résistivité	Exemples de matériau (E_g à 300K ou résistivité)
Conducteur	$0 eV$	$<10^{-6} \Omega.cm$	Cuivre (16.8 nΩ.m), Or(24.4 nΩ.m)
Isolant	plusieurs eV	$>10^8 \Omega. cm$	Diamant (5.51 eV), SiO ₂ (9 eV)
Semiconducteur	: $1 eV$	$[10^{-3}, 10^3]$	Si(1.11 eV), Ge (0.67 eV), GaAs (1.424 eV), InSb (0.17 eV)

Tableau 1: Semiconducteur, isolant et conducteur : critères et exemples.

Le tableau 1 récapitule les critères de classification isolant/conducteur/semiconducteur⁹. Le grand atout des semiconducteurs est la possibilité qu'ils offrent de pouvoir maîtriser la quantité de porteurs libres (électrons ou trous) par des techniques aussi variées que le dopage, l'injection de courant, les radiations ... Désormais dans ce chapitre, on ne considérera et étudiera que des semiconducteurs.

2.1.4 Formalisme de trou

2.1.4.a Génération d'une paire électron-trou :

Lorsqu'on élève la température ($T>0K$) ou lors d'une absorption de lumière, les électrons de la bande de valence sont susceptibles d'acquérir suffisamment d'énergie pour briser sa liaison covalente, quitter leur orbite initiale et rejoindre la bande de conduction. Chaque électron promu à la bande de conduction laisse une « absence » dans la bande de valence – une « case » vide correspondant à une liaison non assurée. Cette case peut être de son tour occupée par un autre électron de la bande de valence (participant auparavant à une autre liaison covalente). La bande de valence est toujours quasiment pleine (manquant 1 électron) et le mouvement des électrons dans cette bande fait déplacer la « case » vide d'une liaison à une autre. Afin de contourner le traitement d'un problème complexe à plusieurs particules, on étudie le mouvement de la « case » vide à la place de $N-1$ électrons de valence. La bande de valence avec une absence d'un électron est représenté par une particule fictive : **un trou**. En conséquence, on a simultanément d'un électron libre excité dans la bande de conduction et d'un trou créé dans la bande de valence et on parle de **la génération d'une paire électron-trou** (voir Figure 5).

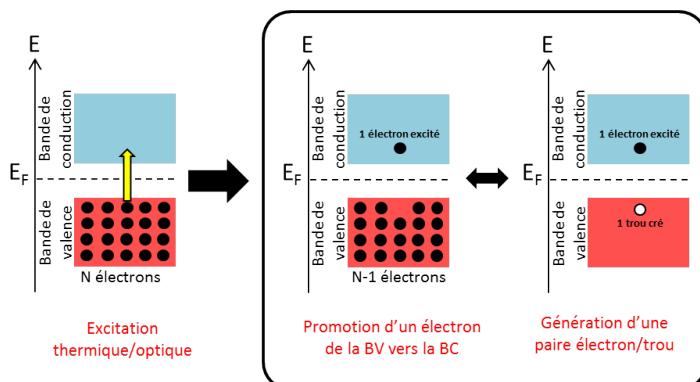
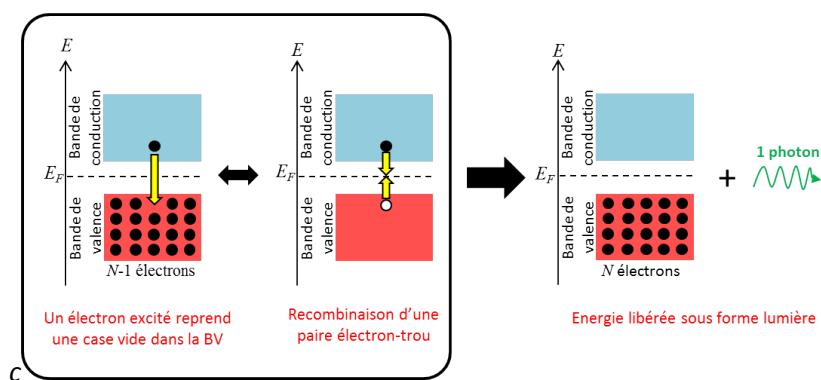


Figure 5: La promotion d'un électron de la bande de conduction à la bande de valence correspond à la génération d'une paire électron-trou.

⁹ Pour rappel, l'énergie thermique $k_B T = 26 \text{ meV}$ à 300K.

2.1.4.b Recombinaison

Le phénomène de recombinaison s'oppose à celui d'ionisation thermique et d'absorption optique. Lorsqu'un électron libre arrive, lors de son déplacement dans le cristal, à proximité d'un ion positif, il peut être « capturé » par ce dernier afin de satisfaire sa liaison de covalence. La liaison de covalence est alors rétablie. Dans le modèle des bandes, un électron de la bande de conduction libère sa place et vient occuper une place libre dans la bande de valence, comblant alors un trou. Le passage de la bande de conduction à la bande de valence s'accompagne d'une restitution d'énergie sous forme de chaleur ou d'émission de lumière¹⁰. Le photon émis a pour énergie $E_g = \frac{h \cdot c}{\lambda}$ avec h constante de Planck, λ longueur d'onde émise et c vitesse de la lumière (voir Figure 6).



2.1.5 Semiconducteur intrinsèque

Il est actuellement possible de produire des cristaux purs avec moins d'un atome étranger pour 10^{11} atomes du semiconducteur pur : on parle alors de **semiconducteur intrinsèque**. Par extension, cela désigne tout **semiconducteur non-dopé**.

Le silicium qui appartient à la colonne IV de la classification périodique des éléments reprise dans le tableau 2 en fait partie. Le silicium intrinsèque a, comme on a pu le voir précédemment, 14 électrons dont 4 de valence, un gap de 1.12 eV, une densité de $5 \cdot 10^{22}$ atomes.cm⁻³ et une densité massique de 2.33 g.cm⁻³. La structure du cristal a déjà été vue (cf. figure 1.b et figure 7) et la structure de bande est celle de la figure 7.

II	III	IV	V
	Bore B (Z=5)	Carbone C (Z =6)	Azote N (Z = 7)
	Aluminium Al (Z = 13)	Silicium Si (Z = 14)	Phosphore P (Z = 15)
Zinc Zn (Z= 30)	Gallium Ga (Z = 31)	Germanium Ge (Z = 32)	Arsenic As (Z = 33)
Cadmium Ca (Z= 48)	Indium In (Z = 49)	Étain Sn (Z = 50)	Antimoine Sb (Z = 51)

Tableau 2: Classification périodique des éléments : extrait le plus couramment utilisé en électronique

¹⁰ Cet effet est utilisé de manière extensive dans les lasers.

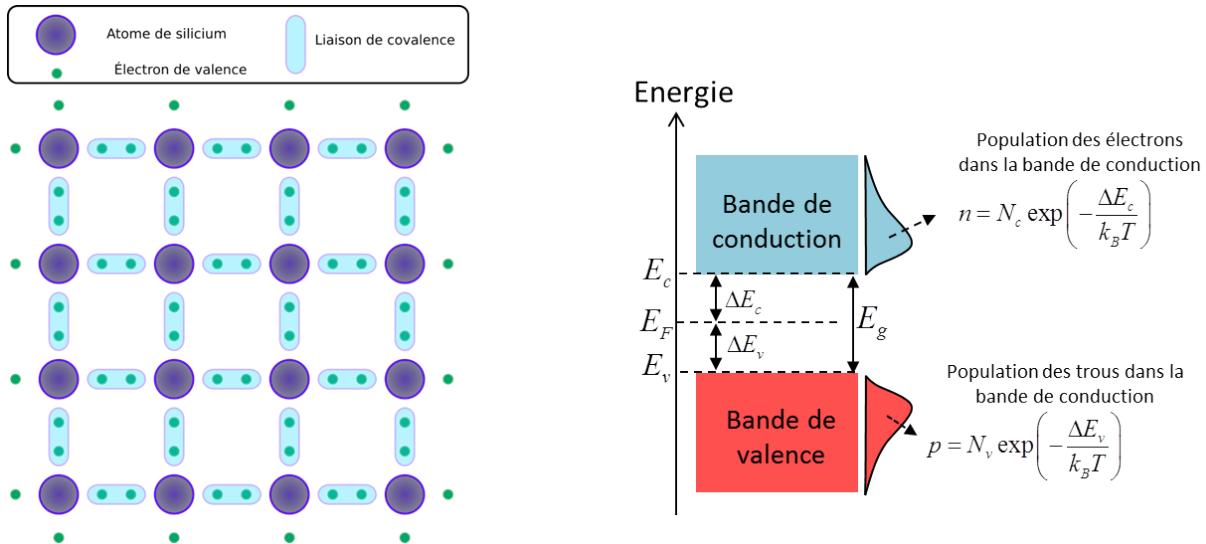


Figure 31: Structure cristalline du silicium intrinsèque et diagramme de bande correspondant.

2.1.5.a La loi d'action de masse

La mécanique statistique montre que la population d'électrons dans la bande de conduction n (cm^{-3}) et la population de trous dans la bande de valence p (cm^{-3}) peut s'écrire¹¹ :

$$\left\{ n = N_c \exp\left(-\frac{E_c - E_F}{k_B T}\right) = N_c \exp\left(-\frac{\Delta E_c}{k_B T}\right) \quad (eq.2.3) \right.$$

$$\left. p = N_v \exp\left(-\frac{E_F - E_v}{k_B T}\right) = N_v \exp\left(-\frac{\Delta E_v}{k_B T}\right) \quad (eq.2.4) \right.$$

avec N_c et N_v les densités effectives d'état respectivement des électrons dans la bande de conduction (égale à $2.82 \cdot 10^{19} \cdot cm^{-3}$ à $300K$ pour Si) et des trous dans la bande de valence (égale à $1.83 \cdot 10^{19} \cdot cm^{-3}$ à $300K$ pour Si)¹². ΔE_c et ΔE_v représentent les différences d'énergie entre la bande de conduction (respectivement la bande de valence) et le niveau de Fermi. (cf . figure 7).

Comme $\Delta E_v + \Delta E_c = E_g$, nous obtenons :

$$p \cdot n = N_v N_c \exp\left(-\frac{E_g}{k_B T}\right) \quad (eq.2.5)$$

A une température donnée, ce produit dépend donc seulement des densités d'états, et du gap – les quantités invariantes lors du dopage d'un semiconducteur (i.e. passage intrinsèque à extrinsèque, c.f. 2.5). Le produit $p \cdot n$ reste donc le même pour un semiconducteur intrinsèque et

¹¹ Ces expressions sont obtenues avec le calcul d'intégral mentionné à la fin de 2.1.2 (i.e. eq.2.2). La mécanique statistique permet d'avoir l'expression de la densité d'état $\rho(E)$.

¹² N_c et N_v varient en $T^{3/2}$.

extrinsèque.

2.1.5.b Calcul de la concentration intrinsèque

Comme le cristal d'un semiconducteur intrinsèque est pur, tous ces électrons et des trous proviennent de l'ionisation thermique (i.e. génération de paire d'électron-trou). La population d'électrons et de trous libres sont donc en quantités égales, appelée **la population intrinsèque** n_i :

$$n_i = n = p = \sqrt{p \cdot n}$$

$$n_i = \sqrt{N_v \cdot N_c} \exp\left(-\frac{E_g}{2k_B T}\right) \quad (\text{eq.2.6})$$

$$\approx 1.4 \times 10^{10} \text{ cm}^{-3} \text{ pour le Silicium à 300K}$$

Finalement, on notera que pour un semiconducteur intrinsèque, le niveau de Fermi se trouve un peu près à la moitié du gap : $\Delta E_c \approx \Delta E_v \approx E_g / 2$.

2.1.6 Semiconducteur extrinsèque

2.1.6.a Dopage d'un semiconducteur

Le silicium est un semiconducteur intrinsèque rarement utilisé en tant que tel¹³ (1 génération de paire électron-trou pour 3.44×10^{12} atomes à 300K). Par contre l'adjonction d'éléments étrangers par le biais de dopage (à raison de 10^{16} à 10^{18} atomes étrangers par cm^3) permet de changer significativement ses propriétés de conduction tout en conservant les propriétés physiques et chimiques du silicium : on parle alors de **silicium extrinsèque dopé N** (si on favorise les électrons) **ou P** (si on favorise les trous). Il y a deux moyens d'introduire des impuretés dans un cristal de semiconducteur (c'est à dire le doper) :

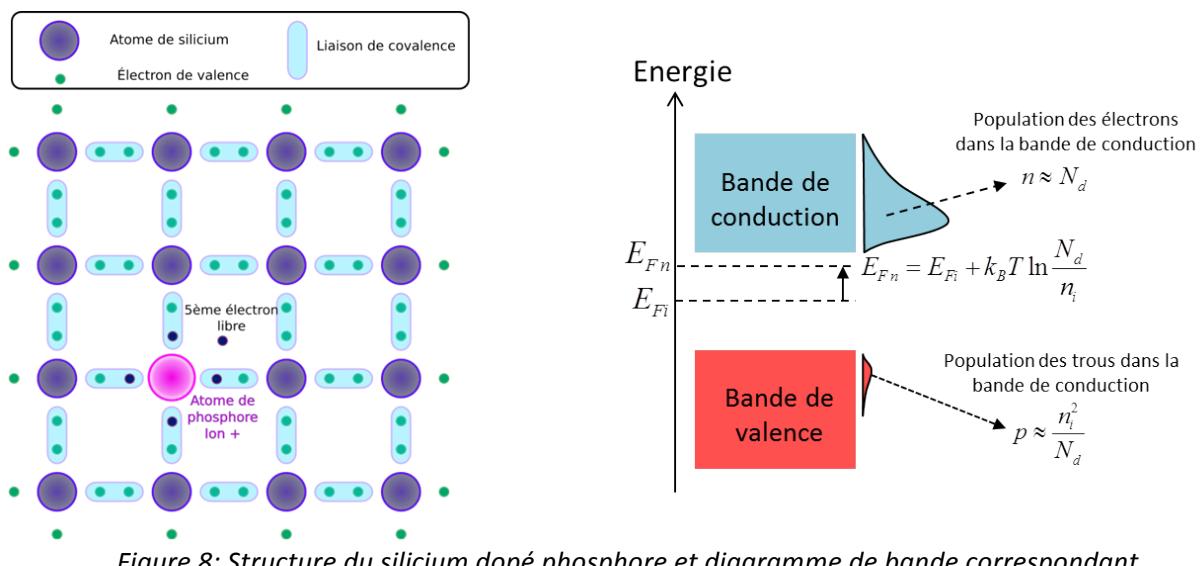
- **Dopage par diffusion** : il s'agit d'une méthode douce qui consiste à mettre le cristal au contact du matériau dopant à forte température (typiquement dans un four de diffusion entre 850°C et 1150°C) et d'attendre la diffusion du dopant dans le matériau. Le dopant peut être en phase solide (auquel cas il est transporté jusqu'au matériau par un gaz vecteur inerte à partir d'un composé solide qui se sublime), en phase liquide (le gaz vecteur baigne dans le liquide ou l'effleure) ou enfin en phase gazeuse (on introduit le gaz contenant l'espèce dopante au contact du matériau).
- **Dopage par implantation ionique** : il s'agit d'une méthode forte qui consiste à accélérer fortement des impuretés ionisées pour leur donner suffisamment d'énergie pour qu'elles puissent pénétrer le matériau à doper. Plus l'accélération est importante, plus la profondeur d'implantation est grande. L'inconvénient de la méthode est le fort désordre cristallin qu'elle engendre, réduisant entre autres la mobilité des porteurs dans le matériau.

¹³ photorésistances et thermistance principalement.

Comme conséquences du dopage, le niveau de Fermi et la population des porteurs libres (électrons dans la bande de conduction et trous dans la bande de valence) sont fortement modifiés par rapport au semi-conducteur intrinsèque. Cependant, d'après la loi d'action de masse, le produit $p \cdot n$ reste invariante quel que soit le dopage ($p \cdot n = n_i^2$).

2.1.6.b Semi-conducteur extrinsèque dopé N

Nous obtenons un semi-conducteur de type N en insérant dans le cristal de silicium des atomes possédant 5 électrons sur leur couche de valence, c'est à dire des éléments de la colonne V tels que le phosphore ou l'arsenic. Prenons l'exemple d'un dopage par le phosphore : le phosphore étant pentavalent, il va mettre en commun (figure 8) quatre de ses cinq électrons pour satisfaire les liaisons de covalence. De fait, le cinquième électron est très faiblement lié à l'atome de phosphore et se retrouve très rapidement "libre" dans la bande de conduction. L'atome de phosphore (appelé atome **donneur**) qui a perdu son électron devient dès lors un ion positif **fixe** du cristal¹⁴.



- **Population des porteurs libres :** A température ambiante la quasi-totalité des atomes donneurs sont ionisés : on considère que la densité de dopage est supérieure à la densité de paire électron-trou que le silicium peut générer à température ambiante, si bien que la densité d'électrons de la bande de conduction est égale à la densité d'atomes donneur. En effet, soit N_d la concentration d'atomes donneurs : la population d'électrons dans la bande de conduction est également égale à N_d (élément qu'on peut retrouver mathématiquement en appliquant la neutralité de la charge qui dit que $n = p + N_d \approx N_d$). On se sert de loi d'action de masse pour calculer la densité de trous dans la bande de valence : $p \cdot n = n_i^2 \Rightarrow p \approx n_i^2 / N_d$. Comme $n >> p$, les électrons sont les porteurs majoritaires et les trous sont les porteurs minoritaires.

¹⁴ La neutralité électrique globale du cristal est toujours conservée puisque l'électron libre reste dans le cristal.

- **Le niveau de Fermi :** Le niveau de Fermi se trouve décalé vers la bande de conduction¹⁵ (c.f figure 8) puisqu'il y a beaucoup plus d'électrons dans la bande de conduction que de trous dans la bande de valence¹⁶. Quantitativement, on peut facilement calculer le nouveau niveau de Fermi E_{Fn} en utilisant l'équation 2.3 :

$$\begin{cases} n_i = N_c \exp\left(-\frac{E_c - E_{Fi}}{k_B T}\right) \\ N_d = N_c \exp\left(-\frac{E_c - E_{Fn}}{k_B T}\right) \end{cases} \Rightarrow E_{Fn} = E_{Fi} + k_B T \ln\left(\frac{N_d}{n_i}\right) \quad (\text{eq.2.7})$$

avec E_{Fi} le niveau de Fermi du silicium intrinsèque pur.

- **Application numérique :** pour un dopage de 10^{18} atomes de phosphore par cm^3 : à 300K, $n = 10^{18} \text{ cm}^{-3}$ et $p = 225 \text{ cm}^{-3}$.

2.1.6.c Semiconducteur extrinsèque dopé P

On obtient un semiconducteur de type P en insérant dans le cristal de silicium des atomes possédant 3 électrons sur leur couche de valence, c'est à dire des éléments de la colonne III tels que le bore ou l'indium. Prenons l'exemple d'un dopage par le bore : le phosphore étant trivalent, il va mettre en commun (figure 9) ses trois électrons pour satisfaire les liaisons de covalence qu'il peut satisfaire. Il va rester une liaison de covalence non satisfaite. Les électrons participant aux liaisons n'étant pas discernables les uns des autres, on peut considérer qu'un atome de silicium voisin a cédé son électron au bore ce qui en fait un atome accepteur dont la neutralité électrique disparait : il devient un ion négatif fixe (puisque'il a plus d'électrons en orbite que de protons dans son noyau).

On se retrouve ainsi avec un trou dans le cristal de silicium. Ce trou aura tendance à vouloir se combler : il cherchera donc à "happer" les électrons libres passant à proximité.

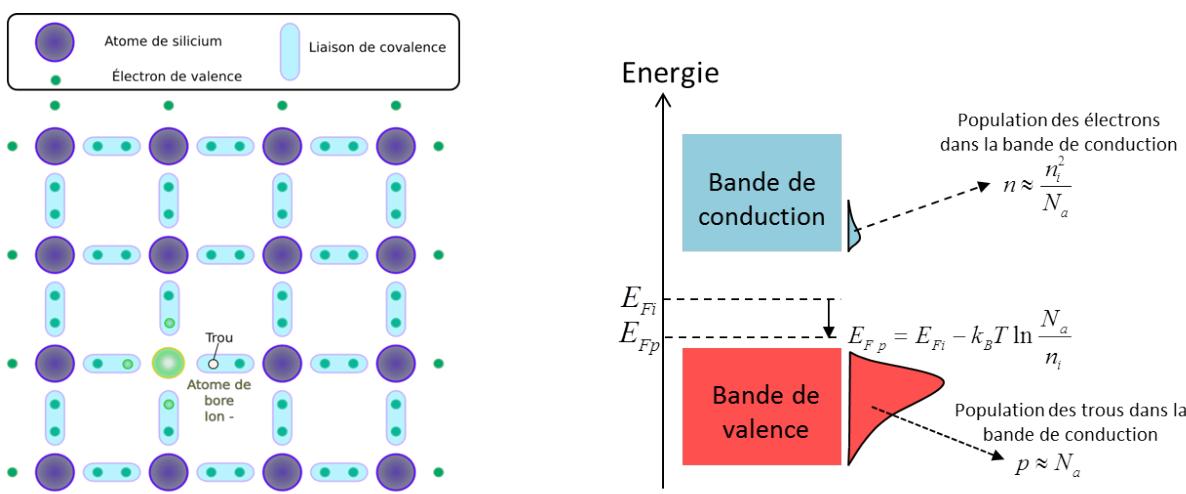


Figure 9: Structure du silicium dopé bore et diagramme de bande correspondant.

¹⁵ Il n'est pas rare dans la littérature de trouver représenté sous le niveau de Fermi des trous ou des ions positifs (en plein milieu de la bande interdite) qui représentent l'origine des électrons qui n'ont pas leur équivalent dans la bande de valence.

¹⁶ On peut voir le niveau de Fermi comme une "jauge à électrons": plus il est élevé, plus le matériau comporte d'électrons libres.

Le raisonnement à température ambiante est similaire à celui tenu pour le semiconducteur extrinsèque dopé N. La quasi-totalité des atomes accepteurs sont ionisées avec la densité de dopage bien supérieure à la densité de paire électron-trou que le silicium peut générer à température ambiante.

- **Population des porteurs libres :** La densité de trous de la bande de valence est égale à la densité d'atomes accepteur $p = n + N_a \approx N_a$ (application de l'équation de neutralité électrique) avec N_a la concentration d'atomes accepteurs. La loi d'action de masse permet de retrouver la population d'électrons dans la bande de conduction (figure 9) : $p \cdot n = n_i^2 \Rightarrow n \approx n_i^2 / N_a$. Comme $p \gg n$, les trous sont les porteurs majoritaires et les électrons sont les porteurs minoritaires.
- **Le niveau de Fermi :** Le niveau de Fermi se trouve cette fois-ci décalé vers la bande de valence¹⁷ puisqu'il y a beaucoup plus de trous dans la bande de valence que d'électrons dans la bande de conduction. Le niveau de Fermi E_{Fp} (voir figure 9) résultant du dopage peut être calculé à l'aide de l'équation 2.4 :

$$\begin{cases} n_i = N_v \exp\left(-\frac{E_{Fi} - E_v}{k_B T}\right) \\ N_a = N_v \exp\left(-\frac{E_{Fp} - E_v}{k_B T}\right) \end{cases} \Rightarrow E_{Fp} = E_{Fi} - k_B T \ln(N_a/n_i) \quad (\text{eq.2.8})$$

avec E_{Fi} le niveau de Fermi du silicium intrinsèque pur.

Application numérique : considérons un dopage de 10^{16} atomes de phosphore par cm^{-3} : à 300K, $p = 10^{16} \text{ cm}^{-3}$ et $n = 2.25 \times 10^4 \text{ cm}^{-3}$.

2.1.6.d Cas général de dopage successif de semiconducteurs

Il est fréquent, au cours de la réalisation de circuits intégrés d'un semiconducteur soit soumis, de manière directe ou pas, à plusieurs dopages successifs. Pour calculer les populations d'électrons dans la bande de conduction et de trous dans la bande valence, il faut recourir aux deux lois suivantes :

- **Loi de la neutralité électrique :** $n + N_a = p + N_d$
- **Loi d'action de masse:** $p \cdot n = n_i^2$

L'application de ces deux lois permet de définir les concentrations en porteurs libres :

¹⁷ Il n'est pas rare dans la littérature de trouver représenté au-dessus du niveau de Fermi des électrons ou des ions négatifs (en plein milieu de la bande interdite) qui représentent l'origine des trous qui n'ont pas leur équivalent dans la bande de conduction.

$$n = \frac{(N_d - N_a) + \sqrt{(N_d - N_a)^2 + 4n_i^2}}{2}$$

$$p = \frac{-(N_d - N_a) + \sqrt{(N_d - N_a)^2 + 4n_i^2}}{2}$$

Ces formules génériques se simplifient selon le dopage prédominant :

- $N_a > N_d$: le matériau est de type P (cas traité précédemment)
- $N_d > N_a$: le matériau est de type N (cas traité précédemment)
- $N_a = N_d$: le matériau est de type intrinsèque par compensation ou neutralisation mutuelle.

2.1.7 Conduction dans les semiconducteurs : courant de conduction et courant de diffusion

Il existe deux phénomènes pouvant mettre des charges en mouvement dans un semiconducteur : le plus connu est sans conteste le phénomène de conduction qui veut que sous l'influence d'un champ électrique, les charges électriques se mettent en mouvement. Un second phénomène, moins connu, peut mettre des charges en mouvement : c'est le phénomène de diffusion selon lequel lorsqu'il y a concentration de porteurs en un point d'un semiconducteur, ces derniers auront tendance à occuper la totalité du volume qui leur est offert. Nous allons, dans cette partie, étudier ces deux phénomènes.

2.1.7.a Un courant de trou ..!

Il est temps d'éclaircir un point qui peut être obscur à savoir la notion de courant de trous : on a vu qu'un trou est par définition une absence d'électron dans la bande de valence que l'on considère comme étant une particule. On considère également que le trou peut être mis en mouvement. Le courant de trou est un phénomène indirect : ce n'est pas le trou lui-même qui bouge mais disons plutôt qu'il se transfère. La figure 10 illustre cela.

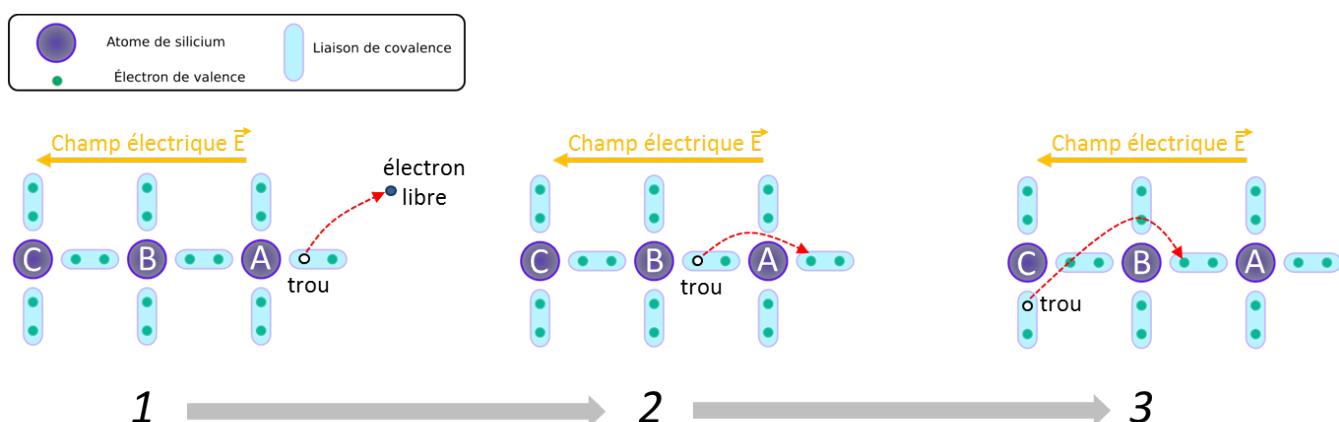


Figure 10: Principe du courant de trou.

Admettons qu'à l'instant t_0 , un électron quitte l'atome A pour partir dans la bande de conduction laissant un trou dans la bande de valence. Celui-ci devient un ion positif et présente une liaison de covalence non-satisfait (étape 1 de la figure 10). Sous l'action du champ électrique présent, un électron ayant acquis suffisamment d'énergie peut quitter l'atome B pour se retrouver happé par l'atome A qui comble ainsi son trou mais fait apparaître un trou en B : l'atome B devient un ion positif (étape 2 de la figure 10). Enfin la même chose se reproduit entre B et C : l'atome C perd un électron au profit de l'atome B (étape 3 de la figure 10) . Si on regarde la situation d'un point de vue global, le trou s'est déplacé de A en C **Le déplacement des trous se faisant par propagation, le mouvement des trous est plus lent que celui des électrons.**

2.1.7.b Courant de conduction

Considérons le cas d'un semi-conducteur isolé. En l'absence de phénomènes extérieurs, les porteurs de charges mobiles (électrons et trous) se déplacent de manière chaotique et aléatoire sans direction privilégiée. Il n'existe pas de circulation de charge à l'échelle macroscopique.

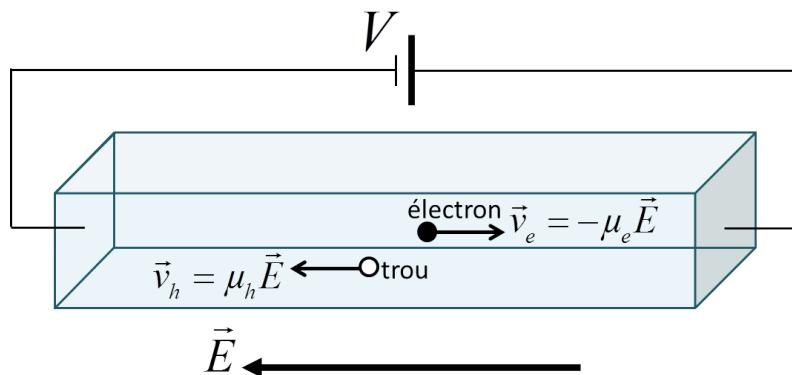


Figure 11: Densité de courant dans un semi-conducteur.

Appliquons maintenant un champ électrique à ce semi-conducteur (par exemple via le biais d'une différence de tension V), le champ électrique va mettre les charges positives (les trous) en mouvement dans son sens et les charges négatives (les électrons) en sens inverse. D'un point de vue macroscopique, on va observer un mouvement des charges dont la vitesse de déplacement peut s'écrire :

$$\begin{cases} \vec{v}_e = -\mu_e \vec{E} & \text{pour les électrons} \\ \vec{v}_h = +\mu_h \vec{E} & \text{pour les trous} \end{cases}$$

Dans ces expressions, μ_e et μ_h représentent respectivement la mobilité des électrons et trous au sein du matériau. Elles dépendent :

- de la **température** : plus la température est élevée, plus le nombre de collisions entre particules est élevé et plus la mobilité est réduite,
- du **dopage** : le dopage va jouer sur les trous et les électrons déjà présents dans le semi-conducteur et donc influencer la mobilité des charges,
- du **champ électrique** qui va jouer sur le temps moyen entre collisions : d'après le modèle simplifié de Drude, la mobilité peut s'écrire $\mu = q\tau / m^*$ avec q la charge, τ le temps

moyen entre deux collisions et m^* la masse effective de la particule¹⁸.

On peut retrouver quelques valeurs de mobilités de semiconducteurs dans le tableau 3 :

Mobilité à $T = 300K$	Électrons ($cm^2 \cdot V^{-1} \cdot s^{-1}$)	Trous ($cm^2 \cdot V^{-1} \cdot s^{-1}$)
Ge	3900	1900
Si	1500	475
GaAs	8500	400

Tableau 3: Quelques mobilités de semiconducteur

Ces déplacements de porteurs correspondent à deux courants de conduction dont l'expression est donné par¹⁹ :

$$\begin{cases} \mathbf{j}_e^{conduction} = ne\mu_e E & \text{pour les électrons} \\ \mathbf{j}_h^{conduction} = pe\mu_h E & \text{pour les trous} \end{cases}$$

Avec n et p la population des électrons et trous libres, et $-e = -1.6 \times 10^{-19} C$ (la charge d'un trou est $+e$).

Cela donne au final la relation suivante²⁰ :

$$\boxed{\mathbf{j}_{conduction} = e(n\mu_e + p\mu_h)E = \sigma E} \quad (eq.2.9)$$

Nous retrouvons donc une résultat très familier : la densité de courant de conduction est proportionnelle au champ électrique E et à la conductivité du matériau σ ($\Omega^{-1} cm^{-1}$).

2.1.7.c Courant de diffusion

La diffusion est un phénomène de transport des particules sous l'effet de l'agitation thermique et qui a pour objectif d'uniformiser la distribution spatiale de ces particules. Elle est donc un processus non réversible dont le "moteur" est le gradient de la densité des particules, et se décrit très simplement à l'aide de la première loi de Fick²¹:

$$\mathbf{j} = -D \nabla n$$

Le signe négatif de l'équation traduit le fait que les particules avec une distribution de densité $n(\mathbf{r})$ diffusent vers les régions à plus faible densité. D est appelé coefficient de diffusion et est généralement exprimé en $cm^2 \cdot s^{-1}$: il traduit la capacité du matériau à diffuser²².

¹⁸ Il s'agit d'une valeur de masse effective des électrons et des trous libres dans le cristal car même s'ils ne sont pas liés à un noyau, leur mouvement est toujours sous influence du cristal (i.e. ils sont « quasi » libres).

¹⁹ Ces expressions sont similaires à celle du courant de déplacement des électrons dans un métal, donnée par le modèle de Drude.

²⁰ Cette relation n'est jamais que l'expression de la loi d'Ohm avec $R=L/(\sigma S)$.

²¹ On pourra noter l'analogie de cette loi avec la loi d'Ohm $\mathbf{j} = \sigma \mathbf{E} = -\sigma \nabla V$.

²² Une analogie qui pourrait être faite pour illustrer D serait la suivante : prenez deux gouttes d'encre et faites en tomber une dans de l'eau et l'autre dans du gel pour cheveu par exemple : l'encre va se diffuser de manière quasi instantanée dans l'eau et va mettre beaucoup plus de temps pour se répartir dans le gel. Ces deux matériaux n'ont pas le même coefficient de diffusion. On peut faire la même expérience avec un matériau poreux et l'autre non pour rester dans le domaine solide.

Dans le cas de semiconducteurs avec une densité d'électrons libres n et une densité de trous libres p , nous aurons deux courants électriques issus de la diffusion des électrons et des trous²³:

$$\begin{cases} \mathbf{j}_e^{\text{diffusion}} = eD_e \nabla n & \text{pour les électrons} \\ \mathbf{j}_h^{\text{diffusion}} = -eD_h \nabla p & \text{pour les trous} \end{cases}$$

Enfin, le courant de diffusion totale est donné par :

$$\mathbf{j}^{\text{diffusion}} = e(D_e \nabla n - D_h \nabla p) \quad (\text{eq.2.10})$$

2.1.7.d La relation d'Einstein

Albert Einstein propose en 1905 une relation très simple et très surprenante qui relie la constante caractéristique du phénomène de diffusion avec celle de la conduction. Il s'agit de la **relation d'Einstein**²⁴ qui s'écrit:

$$\frac{D}{\mu} = \frac{k_B T}{q}$$

Cette équation de toute beauté relie le coefficient de diffusion D qui représente la facilité des porteurs à se déplacer sous l'effet d'une force engendrée par un gradient de la densité avec la mobilité μ qui représente la facilité des porteurs de charge q à se déplacer sous l'effet d'une force engendrée par un champ électrique. Appliquons cette relation au cas des électrons et de trous libres dans un semiconducteur, nous obtenons:

$$\frac{D_e}{\mu_e} = \frac{D_h}{\mu_h} = \frac{k_B T}{e} \quad (\text{eq.2.11})$$

2.1.7.e Densités de courant dans un semiconducteur

En résumé, dans un semiconducteur soumis aux deux phénomènes de conduction (présence d'un champ électrique) et de diffusion (matériau non homogène), la densité de courant totale peut s'écrire :

$$\mathbf{j}_{\text{total}} = q \left(\frac{\eta \mu_4}{4} \hat{z} \frac{\mu_2}{4} \hat{z} \frac{\mu_3}{4} \right) \mathbf{E} + e \left(D_n \frac{\nabla n}{4} \hat{z} D_p \frac{\nabla p}{4} \hat{z} \right)$$

ce qui peut s'écrire aussi :

$$\mathbf{j}_{\text{total}} = \sigma \mathbf{E} + e(D_e \nabla n - D_h \nabla p) \quad (\text{eq.2.12})$$

²³ Les courants de particules sont multipliés par la charge des porteurs pour devenir les courants électriques (- e pour électrons et + e pour trous).

²⁴ Il s'agit d'une des 3 articles célèbre d'Einstein en 1905. Les deux autres sont : la quantification de la lumière, et la relativité restreinte.

Conclusion

Nous avons, au cours de cette partie, rappelé les bases de physique de semiconducteur nécessaire à la compréhension des composants électroniques. De la structure de l'atome à celle du cristal, nous avons abordé les phénomènes permettant de générer des charges au sein d'un semiconducteur et avons établi les équations des mouvements (courant de conduction et de diffusion) de ces charges dans le semiconducteur. Nous allons pouvoir passer maintenant à l'étude de la structure de base de tout composant électronique à savoir la jonction PN.

2.2 La jonction PN

2.2.1 Résumé du fonctionnement

Une jonction PN est un matériau semi-conducteur qui n'est pas dopé de manière uniforme, constitué d'une partie dopée P et d'une partie dopée N. Du fait de la disparité de concentrations de porteurs dans les deux parties de la jonction PN, une migration de trous et une migration d'électrons vont avoir lieu : ce phénomène de diffusion qui a lieu à la jonction des deux parties dopées P et N va laisser les ions donneurs et accepteurs de chaque partie seul, créant ainsi deux zones de neutralité opposées (positive et négative), ce qui va créer une zone de charge d'espace et donc générer un champ électrique qui va s'opposer à la diffusion des porteurs de charges.

Nous allons étudier l'établissement de ce régime stationnaire puis nous étudierons le comportement de la jonction PN soumise à une polarisation extérieure.

2.2.2 La jonction PN en boucle ouverte : établissement du régime stationnaire

Prenons deux semi-conducteurs dopé P et N (figure 12.a) et mettons en contact. Après un bref régime transitoire (figure 12.b), un régime permanent va s'établir (figure 12.c).

2.2.2.a Régime transitoire :

Les porteurs majoritaires de chaque côté vont diffuser vers le côté adjacent : les électrons diffusent vers le silicium dopé P et les trous diffusent vers le silicium dopé N (cf. figure 12.b). Les électrons qui pénètrent dans le silicium dopé P vont se recombiner avec les trous présents en grand nombre du côté P de la jonction et de même pour les trous (recombinaison avec les électrons) qui diffusent dans le silicium dopé N.

On se retrouve alors avec une zone de déplétion (sans trou ni électron) autour de la jonction appelée zone de charge d'espace (cf. figure 12.c). Cette zone n'étant pas neutre, le côté N de la zone est positif puisque les électrons sont partis et le côté P est négatif puisque les trous ont été comblés. Cette zone de charge d'espace (ZCE) polarisée crée un champ électrique (création d'un dipôle électrique) qui va s'opposer au mouvement des porteurs. La zone de charge d'espace va grandir jusqu'à atteindre un équilibre qui est celui pour lequel le champ électrique appelé **champ de rétention de diffusion** est assez fort pour que les pertes par diffusion soient compenser par les courants de conduction induit par ce champ électrique.

2.2.2.b La jonction PN à l'équilibre

Elle présente les caractéristiques électriques de la figure 13. Il existe, entre la région P et la région N, une barrière de potentiel énergétique pour les charges mobiles. L'existence de cette barrière se traduit par une différence de potentiel électrique liée au champ de rétention de la diffusion. L'existence de la barrière de potentiel peut être mise en évidence par le travail W_p qu'il faut fournir pour faire passer un trou de la région neutre P à la région neutre N ou pour faire passer un électron en sens contraire (voir figure 13).

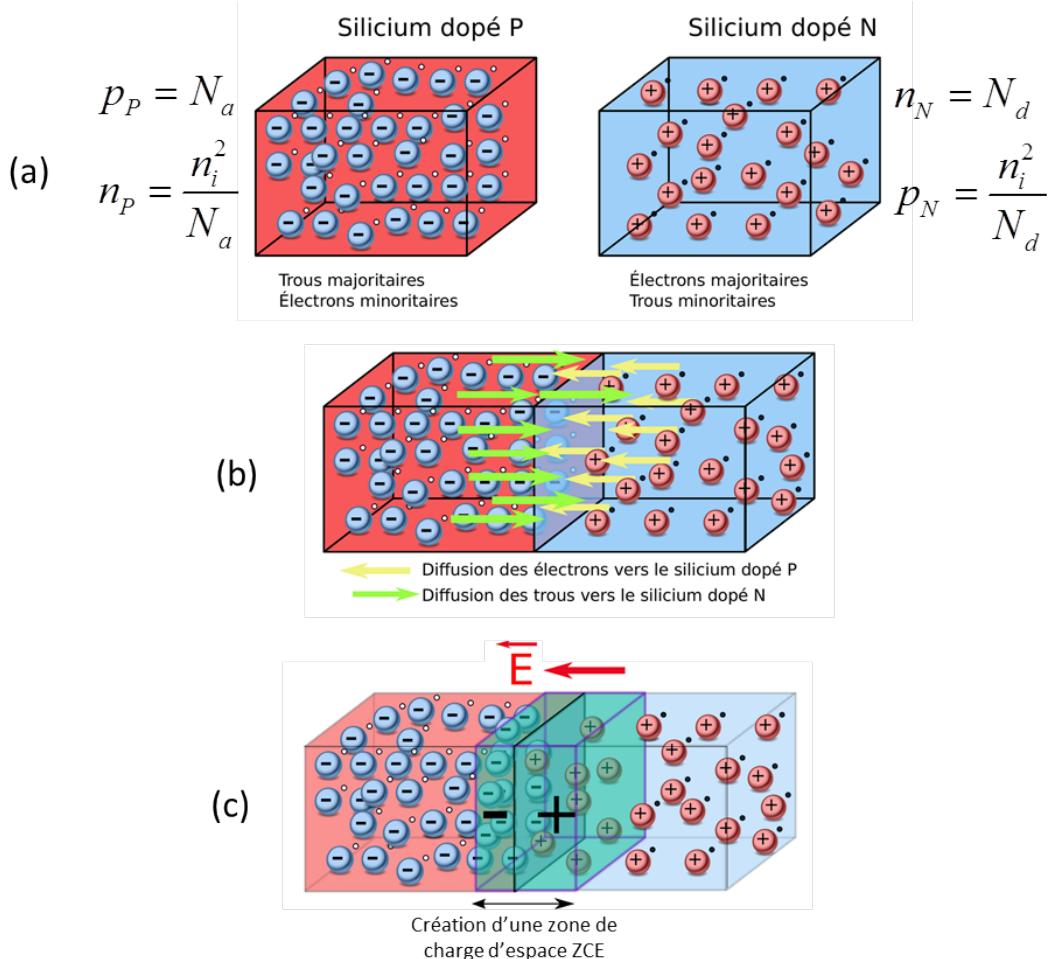


Figure 12: (a) Les deux semiconducteurs P et N avant contact. (b) La diffusion des porteurs majoritaires. (c) La création d'une zone de charge d'espace et l'établissement du régime permanent.

La valeur de la **barrière de potentiel** se déduit de l'équilibre sur les densités de courants qui sont nulles (à la fois pour les électrons et les trous) puisque le courant de conduction compense le courant de diffusion. Cette compensation est traduit par :

$$\begin{cases} J_e = e.n(x).\mu_e E(x) + e.D_e \frac{dn(x)}{dx} = 0 & \text{pour les électrons} \\ J_h = e.p(x).\mu_h E(x) - e.D_h \frac{dp(x)}{dx} = 0 & \text{pour les trous} \end{cases}$$

$$\Rightarrow E(x) = - \frac{D_e}{\mu_e k_B T/e} \frac{1}{n(x)} \frac{dn(x)}{dx} = \frac{D_h}{\mu_h k_B T/e} \frac{1}{p(x)} \frac{dp(x)}{dx}$$

$$\Rightarrow dV = \frac{k_B T}{e} \frac{dn}{n} = - \frac{k_B T}{e} \frac{dp}{p}$$

Le potentiel à calculer étant l'intégrale du champ électrique sur la ZCE. D'une manière cohérente, on obtient:

$$\left. \begin{aligned} V_\Phi &= \frac{k_B T}{e} \int_P^N \frac{dn}{n} \\ &= \frac{k_B T}{e} \ln\left(\frac{n_N}{n_P}\right) = \frac{k_B T}{e} \ln\left(\frac{N_a N_d}{n_i^2}\right) \end{aligned} \right| \quad \left. \begin{aligned} V_\Phi &= -\frac{k_B T}{e} \int_P^N \frac{dp}{p} \\ &= \frac{k_B T}{e} \ln\left(\frac{p_P}{p_N}\right) = \frac{k_B T}{e} \ln\left(\frac{N_a N_d}{n_i^2}\right) \end{aligned} \right.$$

soit la formule finale :

$$V_\Phi = \frac{k_B T}{e} \ln\left(\frac{N_a N_d}{n_i^2}\right) \quad (\text{eq.2.13})$$

Cette tension est la fameuse tension de seuil de la diode V_T souvent aux alentours de 0.7V

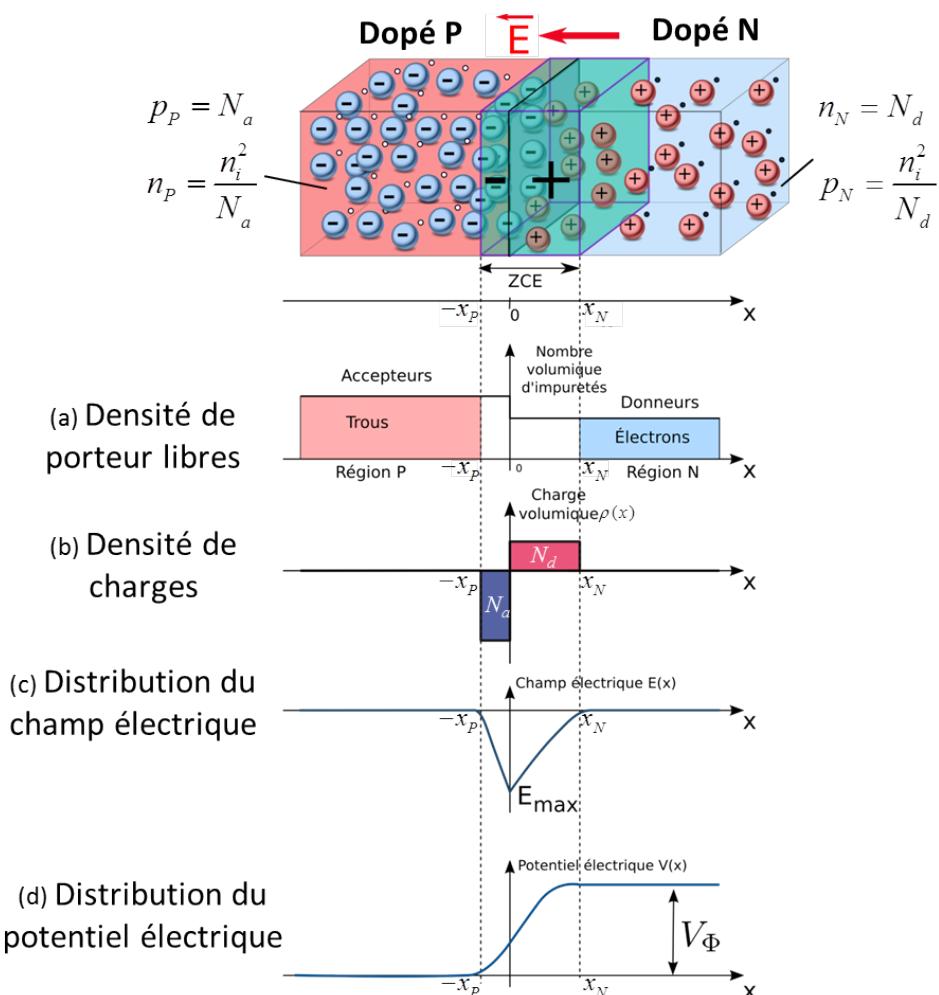


Figure 13: La jonction PN à l'équilibre : porteurs (a) charges (b), champ électrique (c) et potentiel (d).

On peut également calculer la largeur de la ZCE (figure 13):

- Neutralité globale de la ZCE : $x_N N_d = x_P N_a$. On s'aperçoit que la ZCE s'étend du côté le moins dopé donc du côté du silicium dopé N dans notre cas.

- Détermination du champ électrique à l'aide de l'équation de Poisson :

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0 \epsilon_{Si}} \Rightarrow \frac{dE(x)}{dx} = \frac{\rho(x)}{\epsilon_0 \epsilon_{Si}}$$

avec $\rho(x)$ la densité de charge (figure 13.b). Comme est une fonction constante par morceau, on va calculer la répartition du champ sur chaque côté séparément.

- Répartition du champ électrique de rétention du côté P (sachant que $E(-x_p) = 0$) :

$$E(x) = -\frac{eN_a}{\epsilon_0 \epsilon_{Si}}(x + x_p)$$

- Répartition du champ électrique de rétention du côté N (sachant que $E(x_N) = 0$) :

$$E(x) = \frac{eN_d}{\epsilon_0 \epsilon_{Si}}(x - x_N)$$

- Valeur du champ maximal E_{max} (en $x=0$) :

$$E_{max} |_{x=0} = -\frac{eN_a}{\epsilon_0 \epsilon_{Si}} x_p = -\frac{eN_d}{\epsilon_0 \epsilon_{Si}} x_N$$

- On déduit de la formule précédente la largeur de la zone d'espace :

$$W_{ZCE} = x_p + x_N = \frac{\epsilon_0 \epsilon_{Si}}{e} E_{max} \left(\frac{1}{N_a} + \frac{1}{N_d} \right)$$

- On remplace la dernière inconnue (à savoir E_{max}) par le potentiel de diffusion V_Φ qui correspond à l'air du triangle (cf. figure 13 formé par la répartition du champ électrique – notre fameuse barrière de potentiel–) soit:

$$V_\Phi = \int_{-x_p}^{x_N} E(x) dx$$

qui s'exprime également de manière géométrique par:

$$V_\Phi = \frac{1}{2} E_{max} \left(\frac{x_1 + x_2}{2} \right) = W_{ZCE}$$

On obtient donc :

$$W_{ZCE} = \sqrt{\frac{2\epsilon_0 \epsilon_{Si}}{q} \left(\frac{1}{N_a} + \frac{1}{N_d} \right) V_\Phi} \quad (eq.2.14)$$

- Avec la barrière de potentiel déduite dans la partie précédente $V_\Phi = \frac{k_B T}{e} \ln \left(\frac{N_a N_d}{n_i^2} \right)$, on arrive au résultat final :

$$W_{ZCE} = \sqrt{\frac{2k_B T \cdot \epsilon_0 \epsilon_{Si}}{e^2} \left(\frac{1}{N_a} + \frac{1}{N_d} \right) \ln \left(\frac{N_a N_d}{n_i^2} \right)}$$

2.2.3 Jonction PN en court-circuit et en polarisation

Polariser la jonction, c'est lui appliquer une tension V (c.f figure 14). On peut considérer que cette tension V se reporte de manière exclusive sur la zone de déplétion (la ZCE) du fait de sa très grande résistivité en l'absence de charges mobiles.

2.2.3.a Diagramme de bandes

Considérons d'abord le cas de la jonction en court-circuit (i.e. $V=0$). En tant que potentiel chimique des électrons²⁵, il est évident que le niveau de Fermi est le même partout dans un cristal semi-conducteur non soumis à une différence de potentiel et à l'équilibre thermodynamique,. Dans le cas de la jonction PN, les niveaux de Fermi associés, E_{Fn} et E_{Fp} restent alignés dans le schéma de bandes comme le montre la figure 14.b.

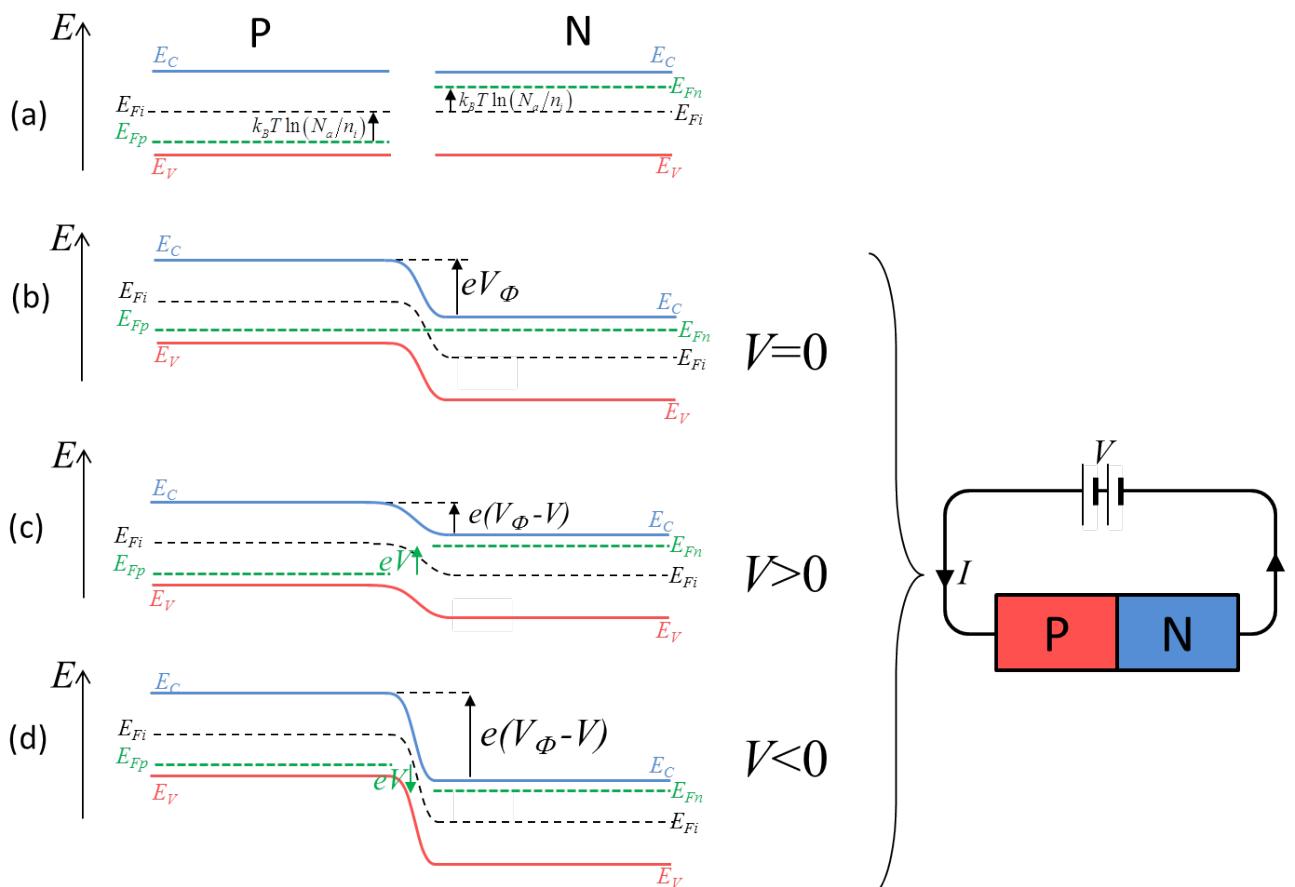


Figure 14: Schéma des bandes de semiconducteur dopé P et N (a), la jonction PN en court-circuit (b), la jonction PN polarisée en direct (c), et en inverse (d).

Sachant que les niveaux de Fermi E_{Fp} et E_{Fn} respectivement associés aux côtés P et N sont alignés, la bande de conduction du silicium dopé P se situe à une énergie plus élevée que celle du silicium N dopé N. Il en est de même pour les bandes de valence. Les bandes de valence et de conduction sont continues dans le cristal : il y a donc une courbure de celles-ci sur la zone de

²⁵ Cela n'est exact qu'à $T=0K$. Il y a une très légère différence (négligeable) entre le niveau de Fermi et le potentiel chimique.

charge d'espace. On peut calculer la différence d'énergie entre les bandes assez facilement :

$$\begin{aligned} E_C|_P - E_C|_N &= E_V|_N - E_V|_P = \left(E_{F_P} - E_{F_N} \right) + \left(E_{F'_P} - E_{F'_N} \right) \\ &\quad k_B T \ln \frac{N_a}{n_i} \text{ (eq 2.7)} \quad k_B T \ln \frac{N_d}{n_i} \text{ (eq 2.8)} \\ &= k_B T \ln \left(\frac{N_a N_d}{n_i^2} \right) \end{aligned}$$

Sans surprise, on retrouve la hauteur de barrière eV_Φ précédemment calculée (c.f. équation 2.9).

La tension V de polarisation appliquée entre P et N correspond à un champ électrique extérieur de $P \rightarrow N$ (algébriquement), qui est de sens opposé de celui au champ électrique de rétention. Par conséquence, la barrière de potentiel que les porteurs majoritaires doivent « surmonter » pour diffuser à travers la jonction devient : $e(V_\Phi - V)$. Quant aux niveaux de Fermi, la tension appliquée impose une différence de potentiel chimique eV entre deux la partie P et N de la jonction. Par conséquent, les niveaux de Fermi E_{F_P} et E_{F_N} ne sont plus alignées mais se diffère par eV (c.f. figure 14.c et d). La largeur de la zone de charge d'espace est également modifiée par rapport au cas en court-circuit. Le résultat dans l'équation 2.14 est modifié en :

$$W_{ZCE} = \sqrt{\frac{2\varepsilon_0\epsilon_{Si}}{q} \left(\frac{1}{N_a} + \frac{1}{N_d} \right) (V_\Phi - V)} \quad (\text{eq.2.15})$$

Alors la ZCE diminue dans le cas de polarisation directe (c.f. figure 14.c) et augmente dans le cas de polarisation inverse (c.f. figure 14.d).

2.2.3.b Courant dans une jonction PN

Il est constitué du courant de diffusion et du courant de conduction :

- **Le courant de diffusion (>0, de P → N)** a pour origine les **porteurs majoritaires** des régions P et N proches de la zone de charge d'espace qui ont assez d'énergie pour franchir la barrière de potentiel $e(V_\Phi - V)$. Il peut s'écrire

$$I_{\text{diffusion}} = I_0 \exp \left[-\frac{e(V_\Phi - V)}{k_b T} \right] \quad (\text{eq.2.16})$$

avec I_0 le courant de diffusion qui existerait en l'absence de la barrière de potentiel.

- **Le courant de conduction (<0, de N → P)** est constitué des **porteurs minoritaires** qui sont en bordure de la zone de charge d'espace et qui se font entraîner par le champ électrique de la ZCE. Ce courant est issu du phénomène de ionisation thermique du silicium (extrêmement sensible à la température), et est appelée **le courant de saturation I_s** :

$$I_{\text{conduction}} = -I_s = -A \cdot T^3 \exp \left(-\frac{E_g}{k_b T} \right) \quad (\text{eq.2.17})$$

avec A une constante du matériau.

Le courant total est donc donné par : $I = I_0 \exp\left[-\frac{e(V_\Phi - V)}{k_b T}\right] - I_s$. Comme la nullité $I_{V=0} = 0$ pour le cas en court-circuit entraîne que $I_s = -I_0 \exp\left(-\frac{eV_\Phi}{k_b T}\right)$, nous avons finalement (c.f. figure 15) :

$$I = I_s \left[\exp\left(\frac{eV}{k_b T}\right) - 1 \right] \quad (\text{eq.2.18})$$

Remarque : Dans le cas d'une tension inverse trop forte, la zone de charge d'espace occupe de plus en plus de volume. Les porteurs peuvent dès lors tirer suffisamment d'énergie du champ électrique pour entrer en collision avec les ions du cristal et leur arracher des électrons qui à leur tour vont reproduire le même phénomène. Cet effet cumulatif est appelé avalanche par multiplication et produit un courant inverse très important conduisant à la destruction de la jonction²⁶

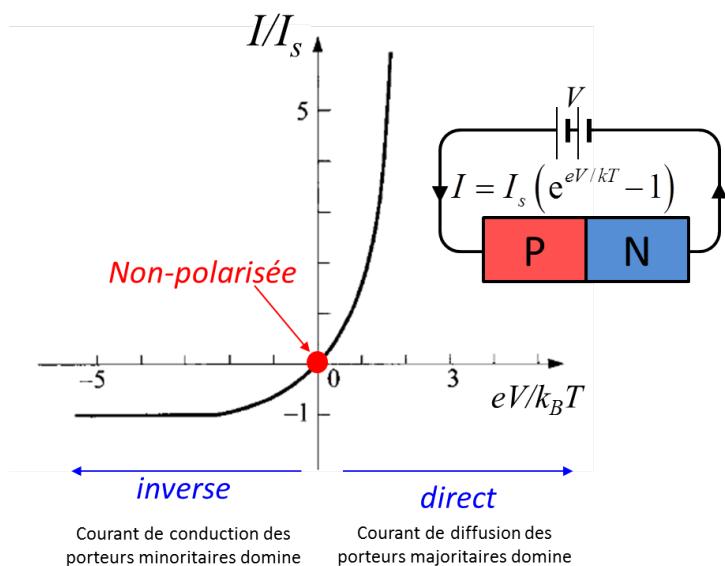


Figure 15: Caractéristique I-V d'une jonction PN.

2.2.3.c Capacité de transition C_T et capacité de diffusion C_d de la jonction PN

La jonction PN présentant une zone de charge d'espace composée de deux charges opposées : elle se comporte donc comme un condensateur appelé capacité de transition C_T (les régions neutres P et N font office d'électrodes et la ZCE en est le diélectrique) de valeur:

$$C_T = \epsilon_0 \epsilon_{Si} \frac{S}{W_{ZCE}}$$

La capacité de transition dépend de la tension de polarisation inverse puisque c'est dans cet état qu'elle est significative. Si on appelle C_{T0} la valeur de la capacité à polarisation nulle, on peut

²⁶ Un autre phénomène lié à un champ électrique intense est l'effet Zener où le champ électrique exerce une force suffisante pour extraire les électrons de leurs liaisons de covalence, créant ainsi des paires électrons-trous et augmentant par là-même le courant inverse. Ce phénomène est réversible (contrairement à celui d'avalanche) et est exploité dans les diodes Zener.

écrire la capacité de la manière suivante :

$$C_T = \frac{C_{T0}}{\sqrt{1 - \frac{V_{inv}}{V_\Phi}}}$$

La valeur de cette capacité varie du pF à la centaine de pF. Elle est parfois notée C_j comme capacité de jonction.

Un second effet capacitif existe dans la photodiode : il ne s'agit pas, contrairement à la capacité de transition, d'une capacité **physiquement** existante dans la structure mais plutôt d'un phénomène dont la représentation physique correspond à celui d'une capacité. Le phénomène de recombinaison locale des trous et des électrons de part et d'autre de la ZCE n'est pas instantané : il dépend d'un temps moyen noté τ_n qui est la durée de vie moyenne des porteurs dans la région neutre concernée (de l'ordre de la nanoseconde). Cela signifie que, temporairement, il existe une accumulation de charge positive et de charge négative de part et d'autre de la ZCE du fait des porteurs de charge non recombinés. Cela peut être assimilé à une capacité dite de diffusion dont la valeur dépend directement du courant I traversant la jonction (c'est donc un phénomène présent principalement dans la jonction polarisée en direct) et dont l'expression littérale est la suivante :

$$C_d = \frac{e\tau_n}{k_B T} I$$

2.3 Approche grand-signal et petit-signal

L'utilisation simple de la jonction PN est la diode. Il en existe diverses catégories et leur champ d'application est multiple :

- **démodulation en amplitude** : première utilisation historique pour les ondes radios,
- **conversion en puissance** : utilisation fréquente pour convertir une tension alternative en tension continue par rectification par pont ou autres procédés,
- **protection des circuits** : les diodes sont très souvent utilisées montées en inverse en parallèle pour la protection d'étages d'entrées de circuits électroniques intégrés : dès que la tension dépasse la valeur nominale autorisée, la diode devient passante, protégeant ainsi le circuit. On les retrouve aussi sur les circuits de contrôle moteur ou de relais où elles absorbent les pics de courant,
- **détection de radiations** : outre l'application de détection de lumière, les diodes sont très utilisées pour détecter tout type de radiations et de particules à haute énergie,
- **détection de température** : du fait de sa forte dépendance en température, la diode peut être utilisée en détection,
- **aiguillage du courant** : utilisée en inverse, la diode permet d'empêcher le fonctionnement des circuits en courant inverse.

La diode a ainsi un nombre d'utilisation multiple. Nous allons dans cette partie faire l'étude d'un circuit classique utilisant la diode, ce qui va nous permettre de présenter la méthodologie classique d'étude d'un circuit et en particulier les notions de régime statique, régime dynamique, analyse grand signal et analyse petit signal.

2.3.1 Le circuit

Étudions le circuit de la figure 16. Il s'agit d'une diode D_1 en série avec une résistance R_1 (de valeur $1\text{k}\Omega$) et alimentée par une source de tension idéale V_s . Ce montage est classiquement utilisé pour la commande en tension de diode électroluminescente avec la résistance qui permet de limiter le courant dans la diode.

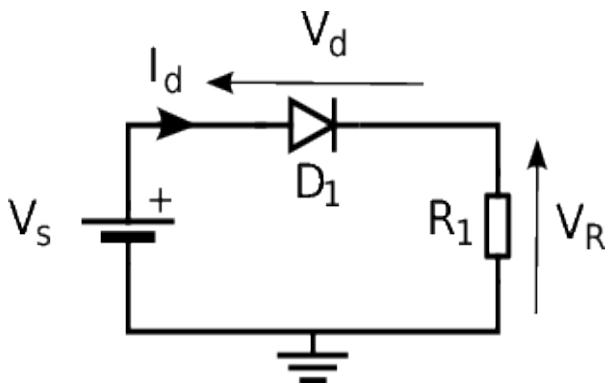


Figure 16: Circuit simple d'une diode montée en série d'une résistance.

2.3.2 Régime statique

Commençons tout d'abord l'étude du circuit par la détermination du point de fonctionnement de circuit en régime établi c'est à dire lorsque V_s est à une valeur fixe appelée V_Q (Q pour quiescent) tension de polarisation ou encore tension de repos. On sait, d'après l'étude physique faite précédemment, que la diode présente la caractéristique²⁷ $I_d = f(V_d)$ de la figure 17.

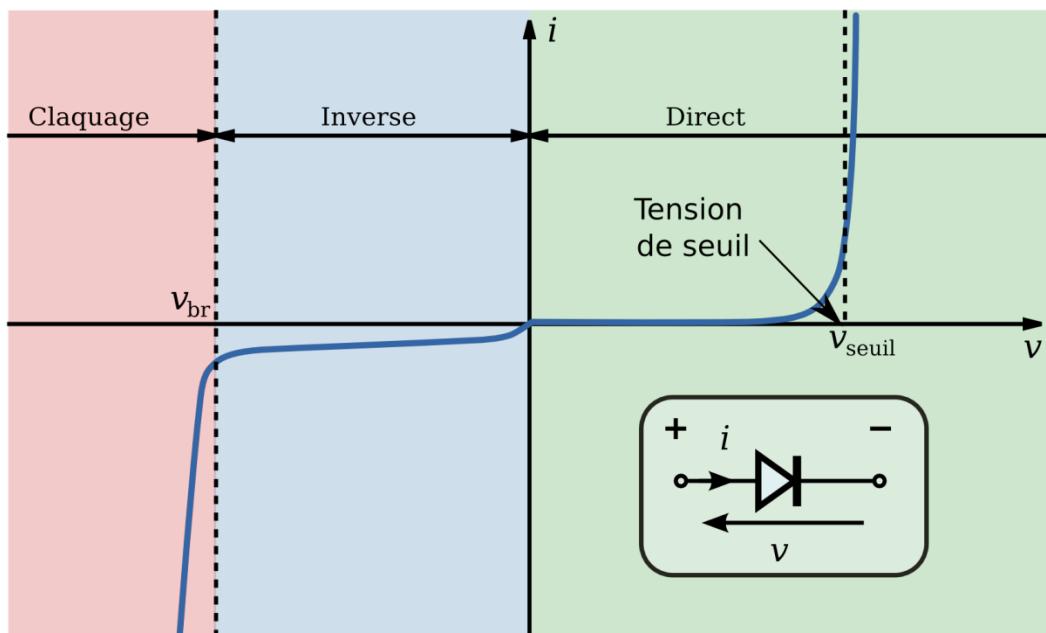


Figure 17: Caractéristique I - V de la diode.

On sait par ailleurs que l'équation qui lie courant et tension dans la résistance est la loi d'ohm soit $V_R = R_1 \cdot I_d$. A ceci on rajoute la loi des mailles qui dit que $V_s = V_d + V_R$ et nous avons la totalité des équations du circuit. En supposant V_s positif et de surcroit supérieur à V_T , on sait

²⁷ La diode idéale est une diode dont la tension de seuil est nulle et dont la tension à ses bornes reste nulle quel que soit le courant qui la traverse (pente infinie).

que la diode est polarisée en direct soit $I_d = I_s \cdot \left(\exp \frac{V_d}{\eta U_T} \right)$ avec η facteur de non-idealité de la diode²⁸ et $U_T = \frac{k_B T}{e}$.

2.3.2.a Calcul explicite du point de fonctionnement

Pour obtenir la totalité des courants et des potentiels du circuit de la figure 17, il suffit d'écrire l'égalité des courants dans la diode et dans la résistance couplé à la loi des mailles :

$$I_s \cdot \left(\exp \frac{V_d}{\eta U_T} \right) = \frac{V_s - V_d}{R_1}$$

Cette équation est non-linéaire. Pour obtenir une expression explicite de cette fonction, il faut utiliser la fonction W de Lambert aussi appelée **Fonction Oméga** qui est la fonction inverse de $f(w) = we^w$ ce qui donne $w = W(f)$.

Nous n'allons pas détailler ici le calcul de la solution mais donner seulement les étapes clés :

- $w = \frac{I_s R_1}{\eta U_T} \left(\frac{I}{I_s} + 1 \right)$ avec $\frac{I}{I_s} = \exp \left(\frac{V_d}{\eta U_T} \right)$ et $V_d = V_s - R_1 \cdot I$
- en utilisant la fonction Oméga $w = W \left(\frac{R_1 I_s}{\eta U_T} \cdot e^{\frac{V_s + R_1 I_s}{\eta U_T}} \right)$
- on pose les hypothèses suivantes : $R_1 I_s = V_s$ et $I / I_s \approx 1$ pour obtenir la solution finale :

$$I ; \quad \frac{\eta U_T}{R_1} W \left(\frac{I_s}{\eta U_T} \cdot \exp \left(\frac{V_s}{\eta U_T} \right) \right)$$

Cette résolution explicite étant assez complexe, on peut recourir à une résolution implicite par calcul itératif.

2.3.2.b Calcul par itération de la solution

En utilisant les outils informatiques, on peut très rapidement calculer la solution par calcul itératif. Les étapes de résolution sont les suivantes :

- on réarrange l'équation de la diode : $\frac{I}{I_s} + 1 = \exp \left(\frac{V_d}{\eta U_T} \right)$
- on utilise les logarithmes népériens de chaque côté de la formule précédente et on remplace la valeur du courant I par l'expression des potentiels (tirés de la loi de mailles) et de la résistance R_1 (loi d'Ohm) :

²⁸ Une diode réelle par rapport à la jonction PN idéale étudiée auparavant présente une résistance série due aux zones P et N neutres de la diode et une génération-recombinaison des porteurs dans la zone de charge d'espace. Ces deux phénomènes qui n'ont pas été considérés dans l'étude précédente sont généralement représenté par un facteur appelé facteur d'idéalité de la diode dont la valeur est comprise entre 1 et 2.

$$V_d = \eta U_T \ln \left(\frac{V_s - V_d}{R_1 I_s} + 1 \right)$$

Il suffit ensuite de partir d'une valeur initiale pour V_d et de calculer la nouvelle valeur obtenue et ainsi de suite jusqu'à stabilisation de cette valeur. On remarquera la formulation en logarithme de la formule préférée à la formulation exponentielle de manière à favoriser la convergence de la solution.

2.3.2.c Résolution graphique

La résolution la plus simple est sans contestation possible la méthode graphique. On trace sur la même courbe la caractéristique I-V de la diode et la caractéristique I-V (loi d'ohm) appelée droite de charge de la résistance. On peut voir ces deux courbes comme les lois d'existence de la diode et de la résistance dans le circuit. Pour que ces deux éléments puissent exister dans le même circuit, il faut qu'ils aient des points communs (ce qui revient à dire que le courant qui traverse la diode est le courant qui traverse la résistance). La solution est donc l'intersection de ces deux courbes : on obtient un couple (I_Q, V_Q) comme le montre la figure 18.

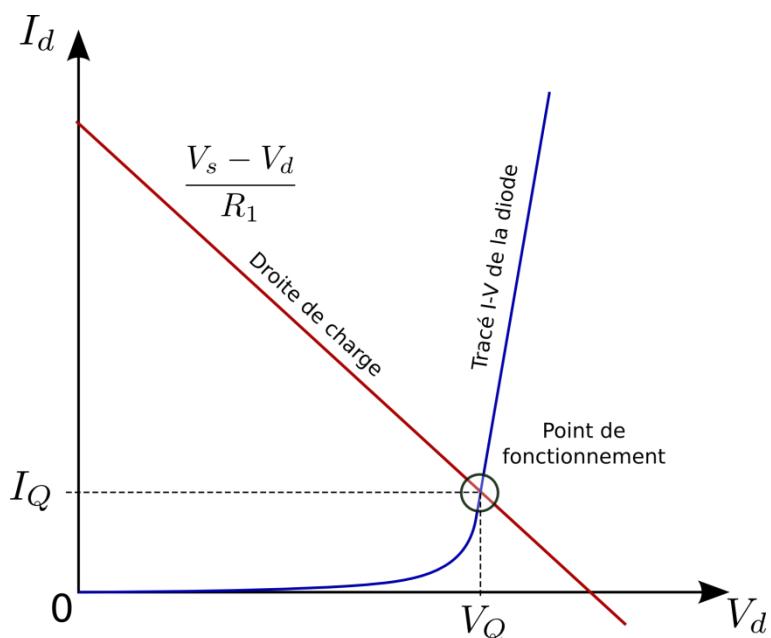


Figure 18: Résolution graphique du point de fonctionnement.

Nous avons maintenant le comportement statique de ce circuit, c'est à dire que nous sommes capables d'en établir les potentiels et les courants fixes. Intéressons-nous maintenant à l'aspect dynamique.

2.3.3 Étude dynamique

Un circuit est généralement soumis à des stimuli externes variables au cours du temps pour lesquels ils donnent une réponse en sortie. Il est fréquent et commode en électronique de séparer le régime statique du régime dynamique. Le régime statique donne le point de fonctionnement du circuit, c'est à dire son état au repos auquel se rajoute (de manière plus ou

moins couplée) le régime dynamique qui caractérise le fonctionnement du circuit et celui qui souvent crée la fonctionnalité.

L'approche ci-dessus développée est d'autant plus facile à appliquer que le circuit est linéaire, le fonctionnement général du circuit étant alors la superposition du régime statique et du régime dynamique. La plupart des composants et circuits électroniques étant non-linéaires, une méthode a été développée pour l'étude des circuits électronique : c'est l'approche petit signal.

2.3.3.a Approche petit-signal

La modélisation petit signal est une technique communément employée en électronique²⁹ pour établir une approximation du comportement non-linéaires des circuits et composants par des équations linéaires. Cette linéarisation se fait autour du point de fonctionnement et reste précise pour de petites variations³⁰. Le modèle petit signal est souvent extrait, en grande partie, des caractéristiques I-V des circuits ou des fonctions de transfert établies en régime statique.

La méthode utilisée est donc la linéarisation autour du point de fonctionnement en recourant aux dérivées partielles par rapport à l'**ensemble** des variables gouvernant le circuit³¹. Ces dérivées partielles peuvent être traduites en terme d'inductances, de capacités et de résistances mais également en terme de sources (tension ou courant) : un circuit électrique équivalent peut être ainsi établi qui donne les signaux de sortie délivrés par le circuit en réponse à des entrées variationnelles de petite amplitude. Le système étant linéarisé, le comportement petit signal est superposé au régime statique pour donner le comportement complet du circuit. Il existe des modèles petits signaux³².

Remarque : l'étude grand signal fait référence à l'étude de signaux dont la variation modifie le point de polarisation : elle prend en compte les phénomènes non linéaires, les limitations dues aux alimentations, ... L'analyse statique est considérée comme faisant partie de l'analyse grand signal (cas extrême de variations extrêmement lente dont la fréquence tend vers 0).

2.3.3.b Conventions de notation

Afin de dissocier de manière visuelle et formelle l'étude d'un circuit, on dénotera les variables de la manière suivante :

- un signal considéré dans une approche grand signal sera noté **intégralement** en majuscule (y compris les indices). Ainsi une tension d'entrée sera notée $V_{IN}(t)$,
- un signal considéré dans une approche petit signal sera noté **intégralement** en minuscule (y compris les indices). Ainsi une variation autour d'un point de polarisation en entrée sera notée $v_{in}(t)$,
- un signal global, composé des deux aspects grand signal et petit signal, sera noté en minuscule pour la variable et en majuscule pour l'indice, ce qui donne, en utilisant les

²⁹ Mais également dans d'autres domaines tels que la mécanique, ..., où on parle plutôt de linéarisation et de résolution au premier ordre.

³⁰ Qui se trouve être le cas pour de nombreux circuits électroniques – communications, traitement du signal, ... – où le signal utile est souvent un signal constant porteur de petits variations contenant l'information.

³¹ Il sera fréquent de négliger l'influence de certaines variables, tout l'art de la modélisation étant dans le choix de ces variables et le domaine de validité de des hypothèses posées.

³² Les modèles petit signal sont souvent présentés sous forme de quadripôle.

deux exemples précédents : $v_{IN}(t) = V_{IN}(t) + v_{in}(t)$.

2.3.3.c Comportement petit signal de la diode

Servons nous de la caractéristique I-V de la diode pour établir son schéma petit signal. Nous appliquons autour du point de fonctionnement V_Q une petite variation et observons ce qui se passe autour du point I_Q (cf. figure 19). Toute se passe comme si la diode agissait comme une admittance de valeur $g_d = \frac{\partial i_d}{\partial v_d} \Big|_{Q \text{ constant}, V_D = V_Q}$ soit

$$g_d = \frac{\partial}{\partial v_d} \left(I_s \cdot \left(\exp \frac{V_d}{\eta U_T} \right) \right) \Big|_{Q \text{ constant}, V_D = V_Q} = \frac{1}{\eta U_T} \cdot I_s \cdot \left(\exp \frac{V_Q}{\eta U_T} \right) = \frac{I_Q}{\eta U_T}$$

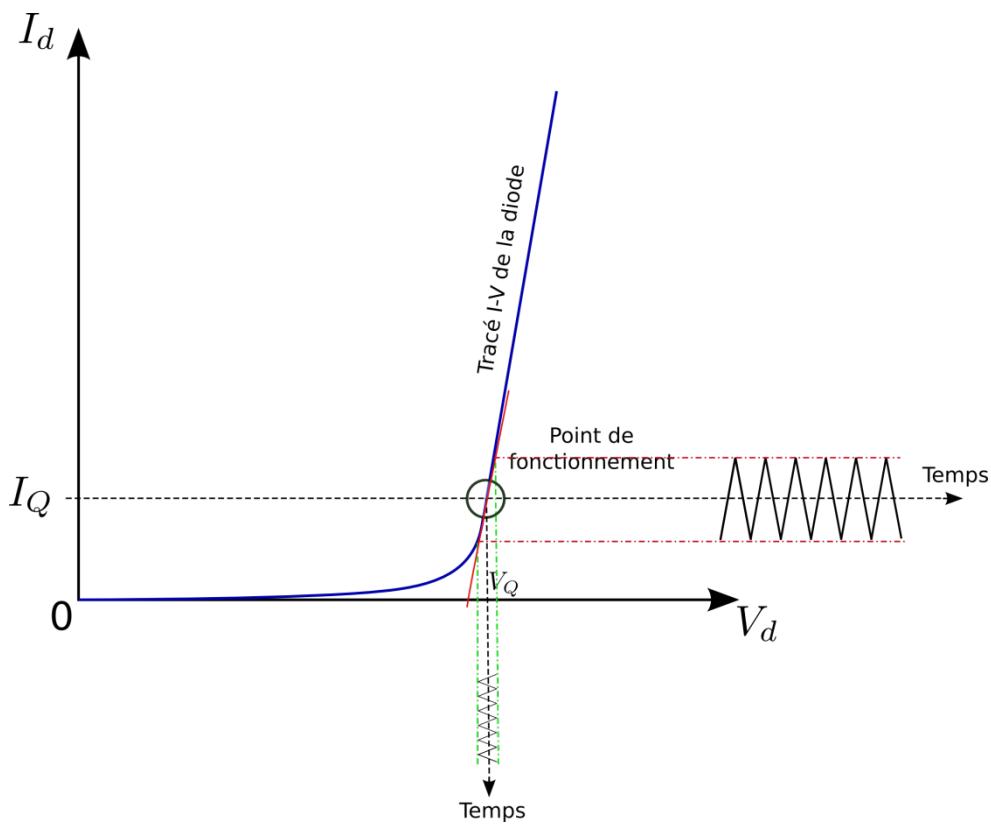


Figure 19: Analyse petit signal de la diode.

On trouve une grandeur qui est bien homogène à une admittance et dont la valeur dépend de celle du courant de polarisation. On peut établir la valeur de la résistance associée à l'admittance calculée r_d qui se trouve être la résistance dynamique de la diode égale à $\frac{\eta U_T}{I_Q}$.

Au comportement de type résistif de la diode, il faut ajouter l'effet capacitif de cette dernière (qui n'apparaît bien évidemment pas la caractéristique I-V du régime statique), le modèle petit signal étant la linéarisation **par rapport à toutes les variables**. On obtient ainsi le schéma équivalent petit signal de la figure 20. La capacité de transition étant de l'ordre du pF en régime

direct et la capacité de diffusion de l'ordre du nF, elle est donc négligeable.

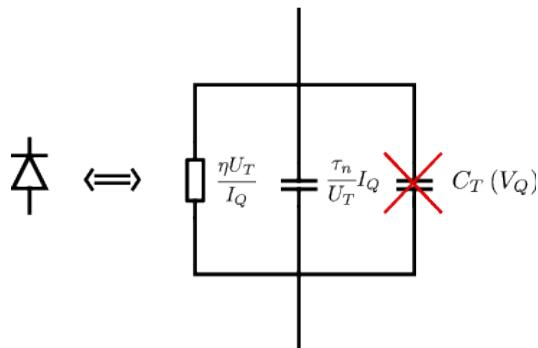


Figure 20: Schéma petit signal équivalent de la diode.

On peut ainsi déterminer le fonctionnement global du circuit.

2.3.4 Modèle global du circuit

Les deux études précédentes ont permis d'établir le modèle global du circuit de la figure 21. De manière plus générale, l'étude de tout circuit électronique commencera par l'étude statique pour établir le point de fonctionnement de ce dernier. Puis on procèdera à l'analyse petit signal du circuit afin d'en extraire les performances dynamiques. On retrouvera cette démarche dans l'établissement du modèle petit signal du transistor MOS du chapitre suivant puis dans l'étude de l'inverseur CMOS en analogique.

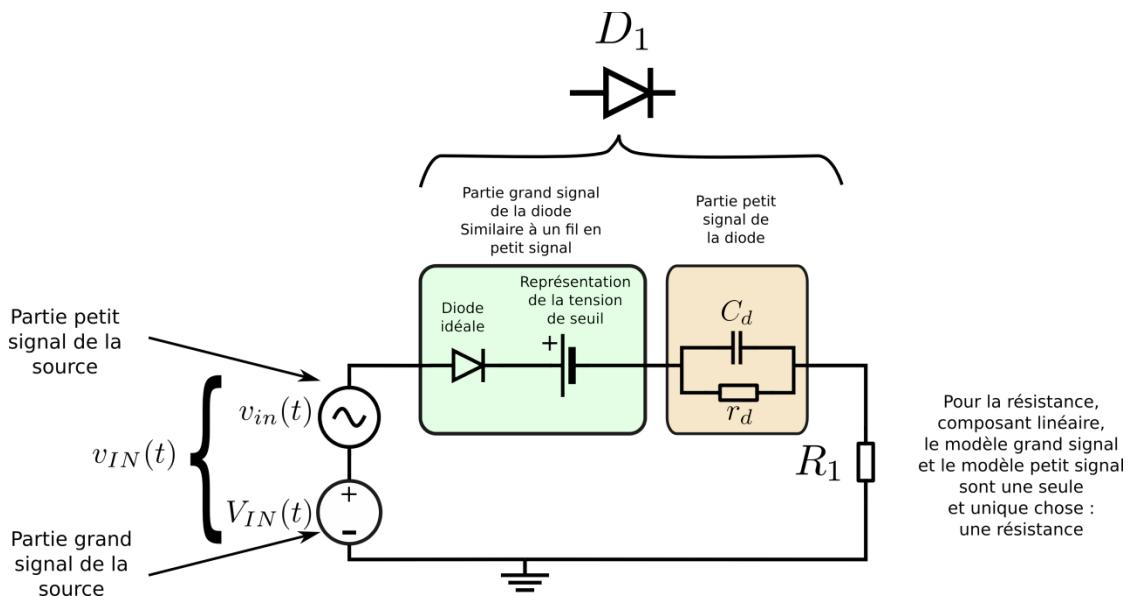


Figure 21: Schéma équivalent global du circuit.

References

- [1] Mathieu Henry Mathieu. *Physique des semiconducteurs et des composants électroniques*, Dunod, 5ème édition, 621.381 52 MAT

[2] Bonnaud Olivier Bonnaud. *Composants à semiconducteurs : de la physique du solide aux transistors*, Bonnaud Olivier, Ellipses, 2006, 621.381 52 BON

[3] Kittel Charles, *Introduction to Solid State Physics*, 8th edition, ISBN 978-0-471-41526-8, November 2004, Wiley edition, 530.41 KIT

[4] Ashcroft Neil W. *Physique des solides*, EDP sciences, Publication Les Ulis, 2002 ISBN-2-86883-577-5, 530.41 ASH

3 Le transistor MOS.

Ce chapitre traite du transistor MOSFET. Il explique simplement le principe de fonctionnement physique du transistor MOSFET, explicite ses différents régimes de fonctionnement et établit les équations fondamentales nécessaires à la formalisation du transistor MOSFET. Enfin, les modèles équivalents grand signal et petit signal basse fréquence seront établis.

3.1 Un bref historique

Le premier transistor date officiellement du 23 décembre 1947 et est à porter au crédit de John Bardeen, Walter Brattain, et William Shockley de Bell Labs. Il s'agit du transistor bipolaire dont la production commerciale débute en 1954 et pour lequel ses trois inventeurs reçurent le prix Nobel en 1956.

Si le principe de base du transistor MOS, fut proposé pour la première fois par Julius Edgar Lilienfeld en 1925, il fallut attendre 1955 pour le premier transistor de Ross et 1959 pour en voir la première réalisation viable par Dawon Kahng et Martin M. (John) Atalla de Bell Labs. Du fait de sa structure différente du transistor bipolaire et de son fonctionnement différent, le transistor MOS ne souffrait pas des phénomènes de dispersions des premiers transistors. Par contre, en raison de son fonctionnement surfacique, sa réalisation a été long à démontrer du fait de la méconnaissance des phénomènes de piégeage surfacique et de passivation (d'ailleurs les travaux de Bardeen, Brattain et Shockley ont contribué à l'avènement du transistor MOS). La nécessité d'obtenir des surfaces très pures a longtemps été un frein pour le transistor MOS.

Le premier circuit à base de transistor MOS date de 1962 (un inverseur) et il fut commercialisé dans des circuits grand public (radios, amplificateurs,...) pour la première fois en 1964. Le procédé planaire inventé par Robert Noyce (Fairchild) en septembre 1959 a permis un développement rapide du transistor MOS du fait du faible nombre d'étapes technologiques et du coût de revient très faible des transistors MOS intégrés.

La technologie CMOS suivra rapidement puisqu'en 1963 Frank Wallas de Fairchild Semiconductor en publia le principe et en démontra l'implémentation physique de manière discrète après avoir échoué à en faire une intégration monolithique (le brevet date lui par contre de décembre 1967).

Ensuite l'histoire du transistor MOS s'accélère avec la loi de Moore en 1965, la fondation d'Intel par Gordon Moore en 1968 et le premier processeur en 1971 : le 4004 (108kHz, 2300 transistors et 60000 opérations par seconde).

3.2 Fonctionnement physique

3.2.1 La Capacité MOS

Structure

Une capacité MOS (Metal - Oxyde - Silicium) est une capacité composée de trois couches (cf. figure 1) :

- Une couche "métallique" : il s'agit typiquement d'aluminium. Dans les dispositifs intégrés, ce "métal" est en fait du polysilicium fortement dopé, qui n'est pas chimiquement

un métal mais en a la structure électronique (absence de gap, forte densité de porteurs), seule caractéristique importante.

- Une couche "isolante" : généralement composée d'oxyde ou de nitrure de silicium. Il s'agit d'une couche ne possédant que très peu de porteurs libres.
- Une couche de semiconducteur constituant la deuxième armature de la capacité : on utilise du silicium (dopé au besoin). Pour l'exemple à suivre, on supposera cette couche comme étant dopé P.

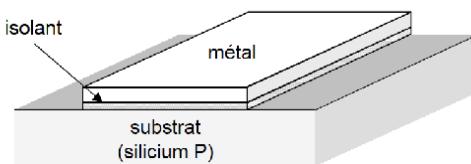


Figure 1: Structure d'une capacité MOS.

Mise en équation

Les dimensions d'un système typique permettent d'écrire le problème sous forme unidimensionnelle, ce qui simplifie l'équation de Maxwell-Gauss. On place le point zéro du repère x à l'interface silicium-isolant, le substrat se situant dans les x négatifs. Le substrat (côté silicium) est placé à la masse, on note V_g le potentiel appliqué à l'électrode métallique.

Une telle structure peut s'étudier en diagramme des bandes de la même manière que la jonction PN. Pour comprendre les différents états possibles sur un tel dispositif on se ramène à l'étude du diagramme de bande d'un système MIS (Métal - Isolant - Semiconducteur). Une telle étude sortant du cadre de ce cours¹, nous n'en donnerons que les conclusions.

Une capacité MOS présente trois régime de fonctionnement qui sont accumulation, déplétion et inversion.

Régime d'accumulation

Du fait de la courbure des bandes au voisinage de l'isolant, toute tension V_g inférieure à une tension appelée V_{FB} (tension pour laquelle les bandes redeviennent plate d'où le nom -FB pour flatband²) attire des électrons en surface de la jonction Isolant - Semiconducteur. La capacité électrique mesurée est constante égale à celle de l'épaisseur d'oxyde.

Régime de déplétion

Une fois la tension V_{FB} dépassée, les charges positives accumulées sur l'électrode métallique ont pour effet d'attirer des électrons sous la surface de l'isolant. Ce dernier étant dopé P, les électrons attirés sont piégés par les ions présents sous la surface Isolant - Semiconducteur. On se retrouve donc avec une charge négative statique (puisque rattachée aux ions négatifs qui sont figés dans la structure cristalline). La capacité électrique mesurée est variable constitué de la capacité d'oxyde augmentée de la capacité de déplétion (dont l'épaisseur dépend de la tension de grille).

¹ Elle fait l'objet d'un cours en troisième année dans l'option MicroNanoBiologie

² On trouvera en annexe 7.5 une explication plus détaillée de la notion.

Régime d'inversion

Une fois que la totalité des ions présents à l'interface Isolant - Semiconducteur ont happé leurs électrons (ce qui correspond à une tension appliquée sur l'électrode métallique dite de tension de seuil V_T), toute augmentation de la tension V_g va attirer des électrons à l'interface Semiconducteur - Isolant. Ces électrons sont mobiles et peuvent participer à un courant électrique. La capacité électrique mesurée est constante égale à la capacité de l'oxyde augmentée de celle de la région déplétée.

L'ensemble des trois régimes de fonctionnement est résumé sur la figure 2 .

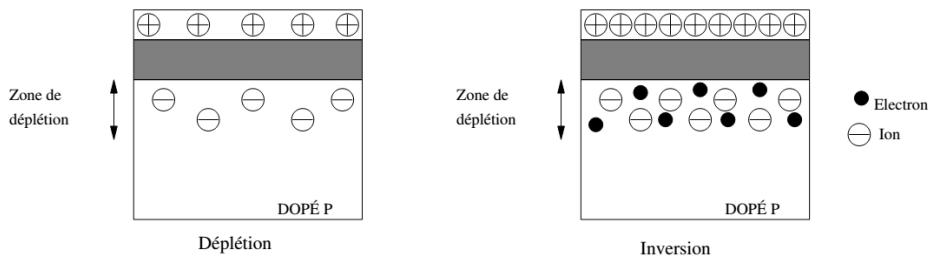


Figure 2: Régimes de fonctionnement de la capacité MOS.

Capacité MOS réelle

Le modèle étudié est basé sur un certain nombre d'hypothèses. En particulier, il a été supposé que le diélectrique était parfait. Un diélectrique réel contient une faible densité de charges. Ces charges sont de plusieurs types. Certaines sont incorporées lors de la production : ce sont des ions (calcium entre autres) piégés dans l'oxyde. Ils sont donc fixes ou très peu mobiles. D'autres charges apparaissent sous l'effet du champ électrique imposé. Elles peuvent prendre plusieurs formes : réorientation de dipôles présents dans le volume, injection d'électrons et de trous aux extrémités, et électrodisassociation de molécules. D'autre part, si le champ électrique devient trop important, le diélectrique claque, il s'agit d'une perte soudaine et irréversible de son caractère isolant.

3.2.2 Le transistor MOS

Principe de fonctionnement

Le transistor MOS est basé sur le principe de la capacité MOS à laquelle on rajoute deux électrodes latérales de manière à générer un champ électrique (via l'application d'une tension entre les deux électrodes) susceptible de mettre en mouvement les électrons présents sous la surface métallique de la capacité (cf. figure 3).

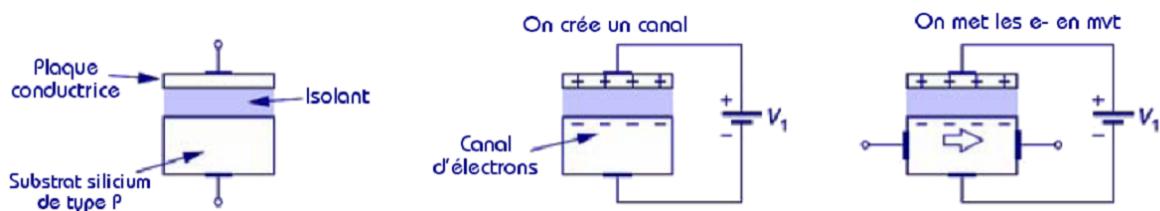


Figure 3: Principe de fonctionnement du transistor MOS.

Une source de courant commandée en tension

Le transistor MOS est donc une source de courant (il fournit des électrons) dont l'intensité est réglée par la tension de la capacité MOS (qui définit la quantité d'électrons susceptibles d'être mis en mouvement) et la tension latérale (qui fixe le champ électrique et donc la vitesse des électrons). On appelle canal, la partie de semiconducteur qui contient des électrons : le canal s'étend entre les deux électrodes de la tension latérale et sa profondeur dépend de la tension appliquée sur la première armature de la capacité.

On appelle **Grille** l'armature métallique de la capacité MOS, **Drain** l'électrode qui reçoit les électrons (elle les draine) et enfin **Source** l'électrode d'où partent les électrons. En effet, sous l'action du champ électrique les électrons sont mis en mouvement dans le canal mais tout électron qui quitte la capacité MOS est remplacé (principe de la neutralité électrique de la capacité) par un autre électron provenant de la source de tension qui génère la tension latérale nécessaire au champ électrique.

On appelle **L** la longueur de canal qui représente la distance que les électrons ont à parcourir et **W** la largeur de canal : elle représente la quantité d'électrons qui transitent de la source vers le drain. Le schéma de la figure 4 résume les notations employées. On ne traitera pour l'instant que du transistor NMOS, c'est à dire le transistor dont le semiconducteur (substrat) est dopé P et où les charges qui sont mises en mouvement sont des électrons³.

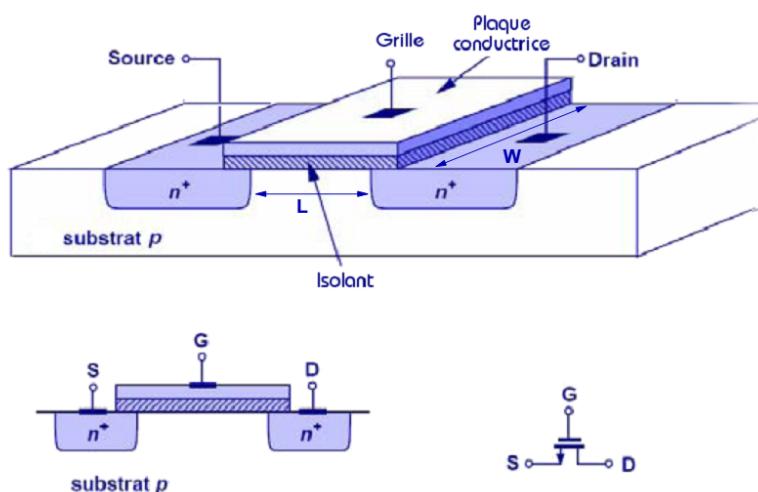


Figure 4: Le transistor MOS : notations.

Régimes de fonctionnement du transistor MOS

Le transistor MOS présente les mêmes états que la capacité MOS : selon la tension de grille appliquée, il se trouve en état d'accumulation, de déplétion ou d'inversion comme récapitulé figure 5.

³ A la différence du transistor PMOS où les charges mises en mouvement sont des trous. Le transistor PMOS fera l'objet d'un paragraphe particulier (cf. 3.6).

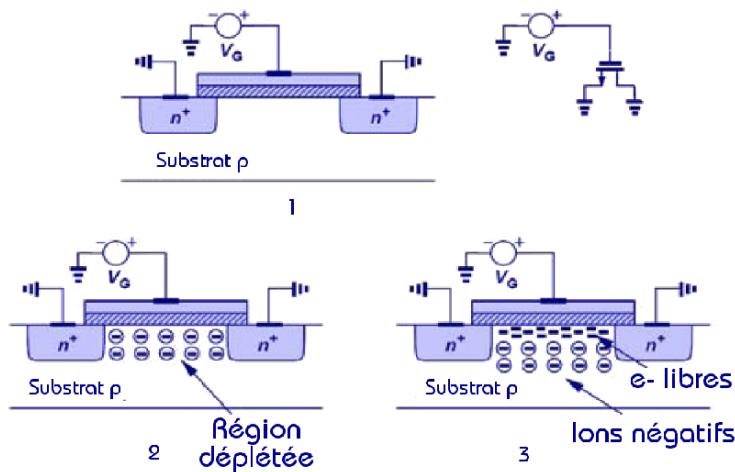


Figure 5: Le transistor MOS : régimes de fonctionnement.

Le fait d'ajouter une tension drain-source a pour effet de modifier légèrement le fonctionnement du canal par rapport à la capacité MOS du fait de la non-uniformité du canal : en effet en l'absence d'une tension V_{DS} , le canal est uniforme. Par contre, si on applique une tension V_{DS} non nulle, la tension de part et d'autre de la grille n'est pas la même : on a V_G du côté de la source et $V_{GD} = V_G - V_D$ du côté du drain ce qui donne un canal non uniforme comme représenté figure 10. Le canal agit comme une résistance soumis à deux potentiels V_G et $V_G - V_D$.

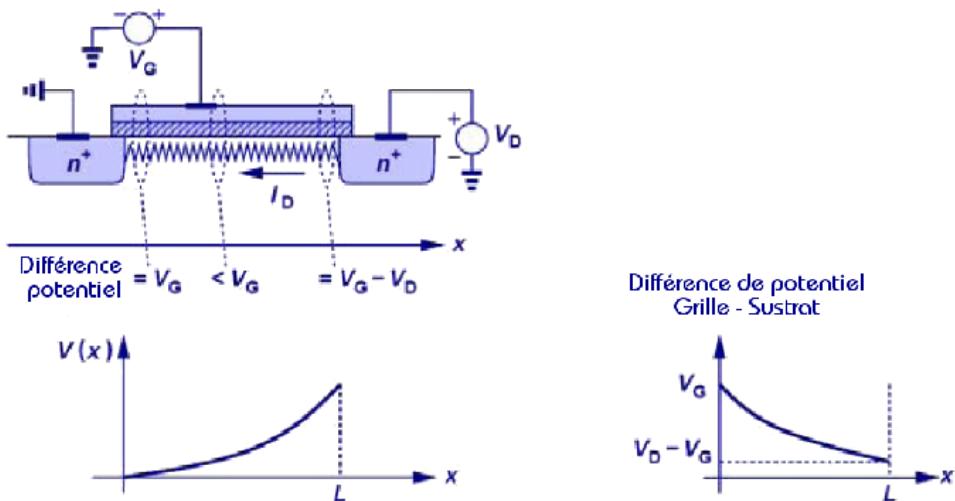


Figure 6: Un canal non uniforme.

Une résistance variable

Le transistor est donc un dispositif qui présente une résistance variable au courant : c'est d'ailleurs l'origine de son nom, transistor étant la contraction de transconductance varistor⁴. On retrouve dans l'appellation transconductance la notion de conversion tension (de grille) – courant (drain–source) et la notion de commande dans le variable de varistor.

⁴ On peut également trouver –plus rarement– la définition transistor = contraction de transfer resistor.

3.2.3 Étude empirique du transistor MOS

Il faut séparer les paramètres structurels des paramètres environnementaux. Les premiers sont inhérents à la structure (dimensions géométriques et caractéristiques physiques) et ne peuvent être modifiés une fois celle-ci fabriquée et les derniers sont des paramètres permettant d'ajuster de manière continue le fonctionnement (tensions appliquées sur les connexions électriques du dispositif).

Influence des polarisations de grille et de drain.

Supposons la source reliée à la tension nulle.

Augmenter la tension de grille signifie augmenter le nombre d'électrons dans le canal donc le courant. Une fois la tension de seuil V_T dépassée (c'est à dire que l'état d'inversion est atteint), il y a une relation linéaire entre la tension de grille et le courant de drain : la caractéristique $I_{DS} = f(V_{GS})$ est semblable à une caractéristique de diode.

Augmenter la tension de drain signifie un champ électrique plus fort et donc un courant plus élevé : la relation entre la tension de drain et le courant est donc a priori linéaire. Il est à noter qu'au-delà d'un certain champ électrique, la vitesse des électrons atteint un maximum v_{sat} et n'augmente plus. Plus la longueur de canal est petite, plus la tension nécessaire à l'obtention de cette vitesse de saturation est faible.

La figure 7 résume ces tendances.

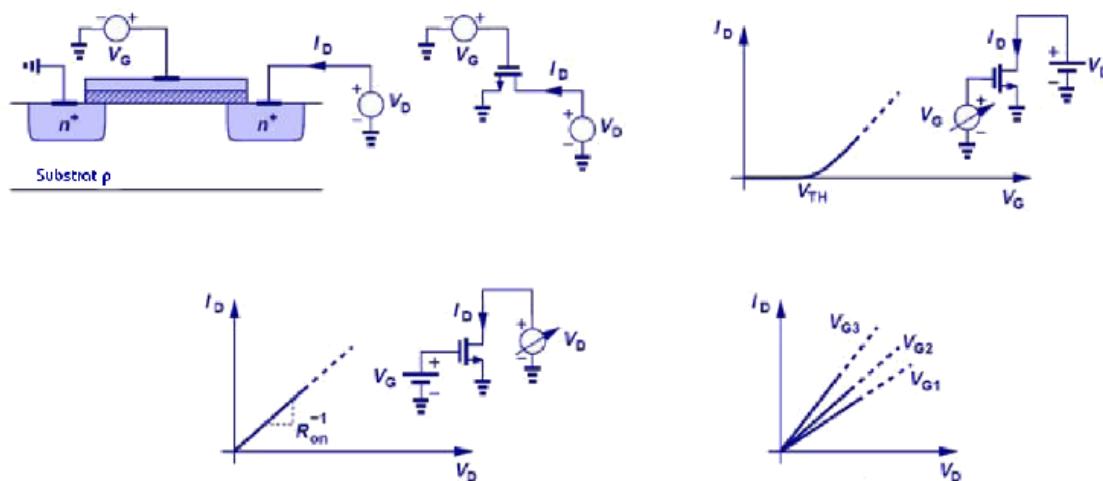


Figure 7: Influence des tensions de polarisation.

Paramètres structurels

On peut étudier l'influence de trois paramètres structurels sur le fonctionnement du transistor MOS : la longueur de canal L , la largeur de canal W et l'épaisseur d'oxyde T_{ox} . Leur influence est résumée sur la figure 8.

Il est à noter que deux phénomènes différents sont en action : jouer sur la longueur de canal et l'épaisseur d'oxyde module le nombre d'électrons présent dans le canal et donc la **densité** d'électrons alors que modifier la largeur de canal module la **quantité** d'électron

qui arrive sur le drain. Pour rappel, un courant est défini par une densité de porteur multipliée par une vitesse le tout multiplié par une surface (voir chapitre sur les semiconducteurs). Dans le premier cas on modifie la densité de courant, dans le second cas on modifie la surface. Bien évidemment pour une technologie donnée, l'épaisseur d'oxyde est fixe, les seuls paramètres physiques que le concepteur de circuit peut utiliser sont la largeur et la longueur de canal.

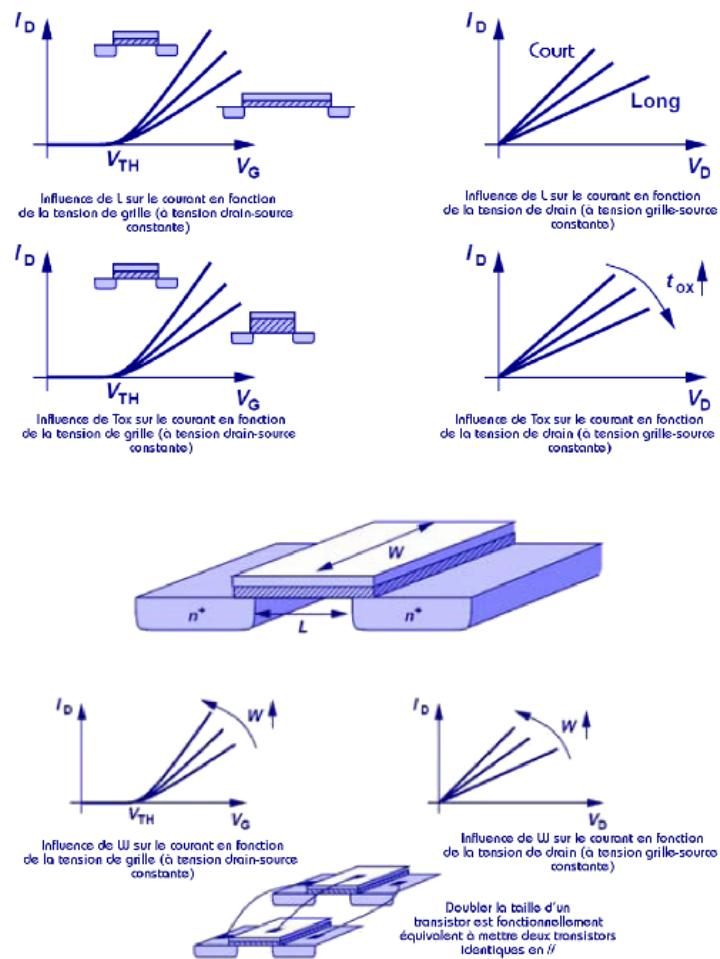


Figure 8: Influence des paramètres structurels sur le courant de drain.

3.2.4 Régime statique : mise en équation

Le transistor MOS pouvant être considéré comme une source de courant commandée en tension, il suffit de calculer le courant généré au travers du canal. Ce courant peut être vu comme un transfert de charge comme montré figure 9.

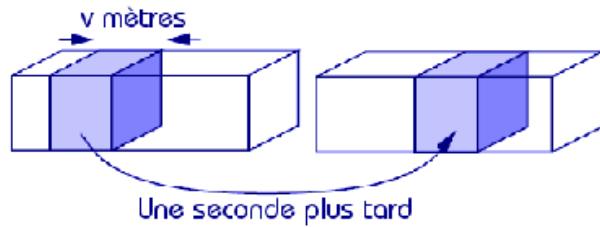


Figure 9: Le courant de drain : un transfert de charge.

La méthode de calcul du courant est simple : il suffit de calculer la charge unitaire du canal, d'appliquer la loi d'ohm puis d'intégrer sur la longueur du canal (cf. figure 10).

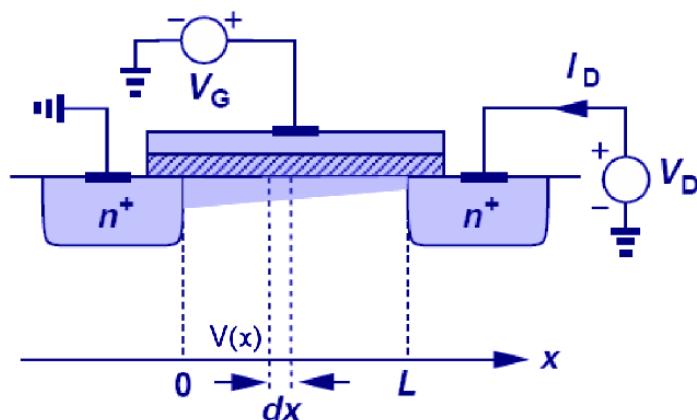


Figure 10: Le courant : une charge unitaire à intégrer le long du canal.

Calcul de la charge unitaire

Considérons un élément en un point du canal. Cet élément de charge dQ est soumis à une tension $v(x)$ tension comprise entre V_S et V_{DS} (cf. figure 11).

$$Q_{I(x)} = -C_{ox} \left(V_{GS} - V_S - v(x) - V_T \right) (\text{coulombs / cm}^3) \quad (3.1)$$

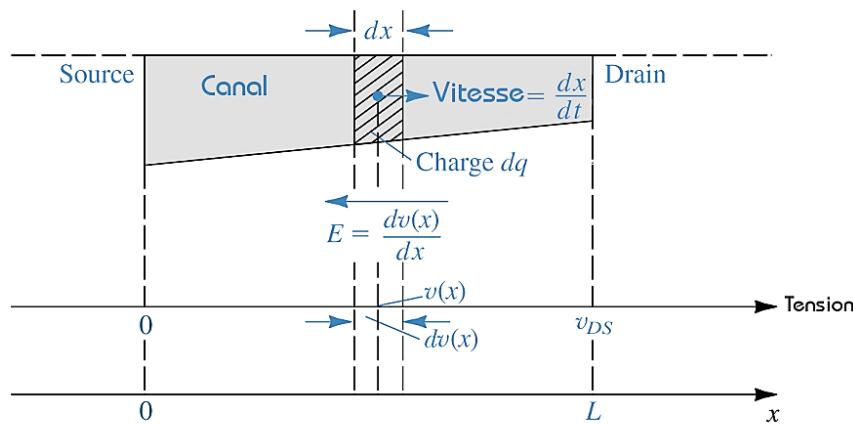


Figure 11: Charge élémentaire du canal.

On en tire la conductivité du canal :

$$\sigma_s = \mu \cdot Q_{I(x)} \left(\frac{cm^2}{v \cdot s} \right) \cdot \left(\frac{coulombs}{cm^3} \right) = \frac{A}{V} = \frac{1}{\Omega \cdot m} = S \cdot m^{-1} \quad (3.2)$$

Loi d'Ohm

Appliquons la loi d'Ohm afin d'obtenir l'expression de la densité de courant dans le canal et le courant associé à la charge unitaire :

$$J_s = \frac{i_D}{W} = -\sigma_s \cdot E_x = -\sigma_s \cdot \frac{dv}{dx} \rightarrow dv = \frac{-i_D}{\sigma_s W} dx = \frac{-i_D dx}{\mu Q_{I(x)} W} \rightarrow i_D dx = -W \mu Q_{I(x)} dv \quad (3.3)$$

Intégration de la charge unitaire le long du canal

Intégrer le long du canal peut être ramené à intégrer sur la tension du canal :

$$\int_0^L i_D \cdot dx = - \int_0^{V_{DS}} W \mu Q_{I(x)} dv = - \int_0^{V_{DS}} W \mu_0 C_{ox} (V_{GS} - v(x) - V_T) dv \quad (3.4)$$

Aux limites :

$$i_D = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) v(x) - \frac{v^2(x)}{2} \right]_0^{V_{DS}} \rightarrow i_D = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (3.5)$$

On obtient ainsi l'équation de base du transistor MOS à savoir :

$$I_{DS} = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \quad (3.6)$$

Tracé de la caractéristique

Le tracé (cf. figure 12) de l'équation 3.6 en fonction de V_{DS} donne une parabole inversée dont le maximum est atteint pour une tension $V_{DS} = V_{GS} - V_T$ et donc la valeur $I_{DSMAX} = \mu C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$. Cette valeur maximale varie avec V_{GS} (comme montré sur la partie droite de la figure 12).

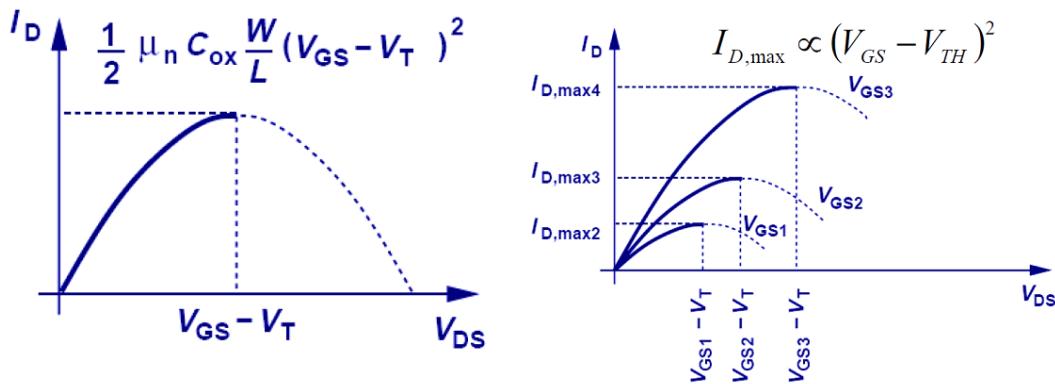


Figure 12: Tracé de l'équation caractéristique $I_{DS} = f(V_{DS})$

Le transistor MOS se comporte comme une résistance (relation quasi linéaire entre V_{DS} et I_{DS} pour des tensions V_{DS} peu élevée) d'où le nom de régime ohmique.

Phénomène de saturation

Il est peu probable que pour une raison physique, l'augmentation de la tension V_{DS} aboutisse à une diminution du courant I_{DS} comme l'indique le tracé de la figure 12. I_{DSMAX} représente la limite de validité des hypothèses posées pour le calcul de l'équation 3.6 : au-delà de $V_{DS} = V_{GS} - V_T$, l'équation 3.6 n'est plus valide.

Étudions de plus près ce point particulier obtenu pour la tension $V_{DS} = V_{GS} - V_T$. En ce point précis, la différence de tension au niveau du drain entre la grille et le drain devient égale à V_T . Cela signifie qu'au niveau du drain, le canal n'est plus inversé : on dit qu'il est pincé (cf. figure 13). Plus la tension V_{DS} augmente et plus le canal "recule". On considère alors que le transistor est dans un état de saturation.

Le canal étant pincé, les électrons sont obligés de "sauter"⁵ au travers de la région pincée pour parvenir au drain, ce qui les ralentit. On obtient donc une valeur fixe de courant pour toute tension V_{DS} supérieure à $V_{GS} - V_T$ (il y a un équilibre qui se crée entre l'augmentation de la vitesse des électrons par la tension drain source et l'augmentation du pincement du canal).

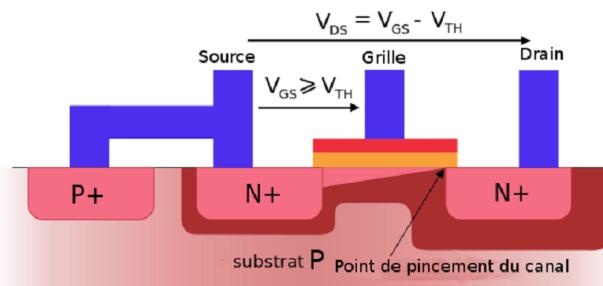


Figure 13: Saturation du transistor MOS

⁵ En fait la région déplétée entre le canal et le drain se comporte comme une zone de charge d'espace au sein de laquelle existe un champ électrique qui permet aux électrons de passer du canal au drain.

Pour toute tension supérieure à $V_{GS} - V_T$, le courant vaut donc la valeur fixe de

$$I_{DS} = \frac{1}{2} \mu C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 \quad (3.7)$$

Le transistor se comporte, dans ce régime, comme une source de courant idéale, c'est à dire comme un dispositif qui délivre un courant constant quelle que soit la tension à ses bornes. La figure 14 représente le réseau de caractéristique résultant pour le transistor NMOS.

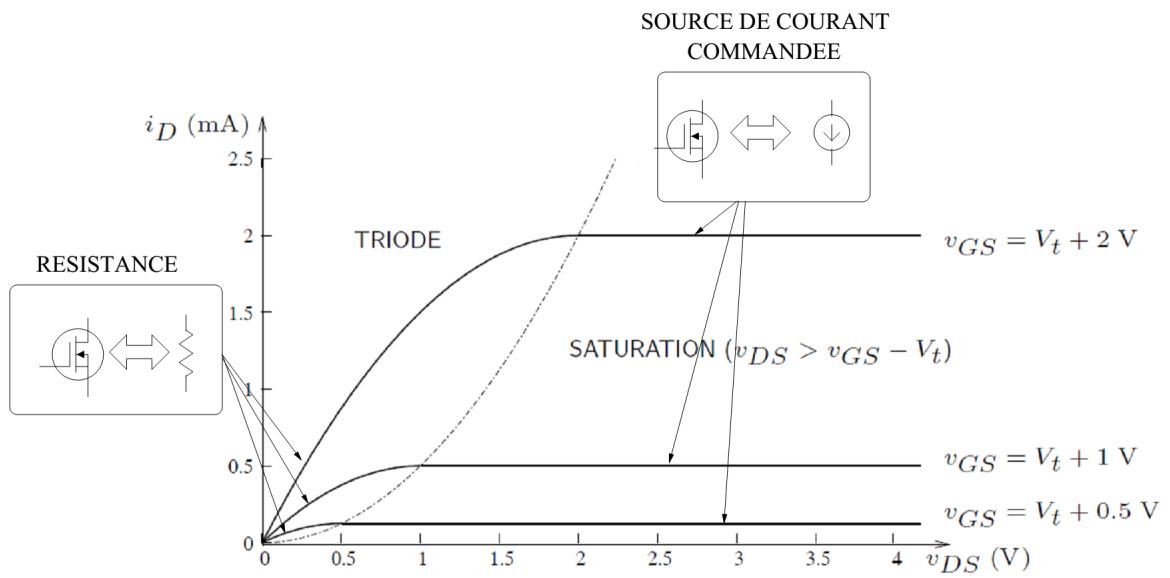


Figure 14: Réseau de caractéristique du transistor MOS

Remarque : en réalité, le courant I_{DS} augmente légèrement du fait du rétrécissement du canal : le canal étant pincé, la longueur à parcourir par les électrons est plus faible donc la longueur à considérer dans les formules n'est plus la longueur physique L mais la longueur effective L_{eff} (se reporter à la figure 15). Le courant augmente donc légèrement avec la tension V_{DS} : la formule 3.7 devient

$$I_{DS} = \frac{1}{2} \mu C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 (1 + \lambda V_{DS}) \quad (3.8)$$

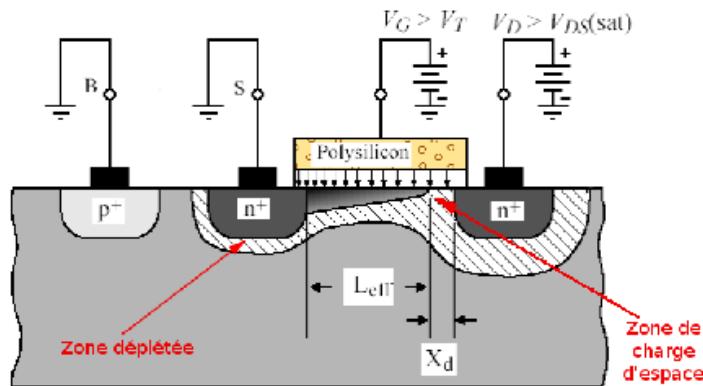


Figure 15: Transistor MOS : longueur effective au niveau du canal

Cela revient à dire qu'en régime saturé, le transistor se comporte comme une source de courant non idéale, la pente du courant représentant la résistance r_0 de la source de courant (cf. figure 16).

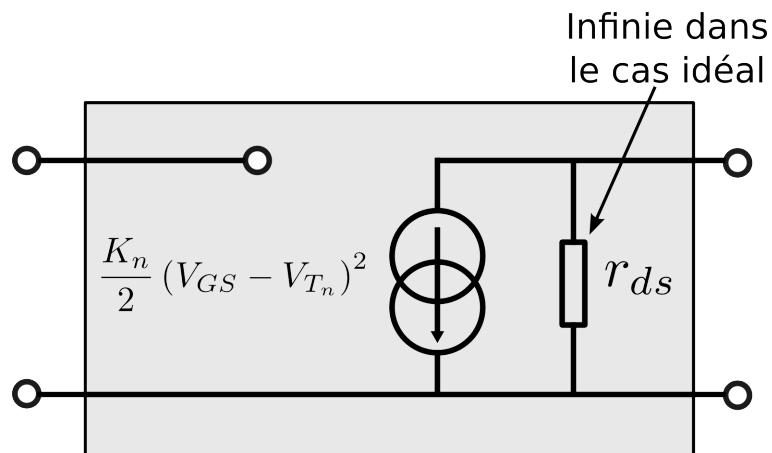


Figure 16: Modèle grand signal du transistor NMOS en saturé.

Remarque : si on prolonge toutes les droites des pentes dues à la longueur effective vers $-\infty$, on peut remarquer qu'elles se croisent toutes au même point, de manière similaire à la tension d'Early pour les transistors bipolaires (cf. figure 17).

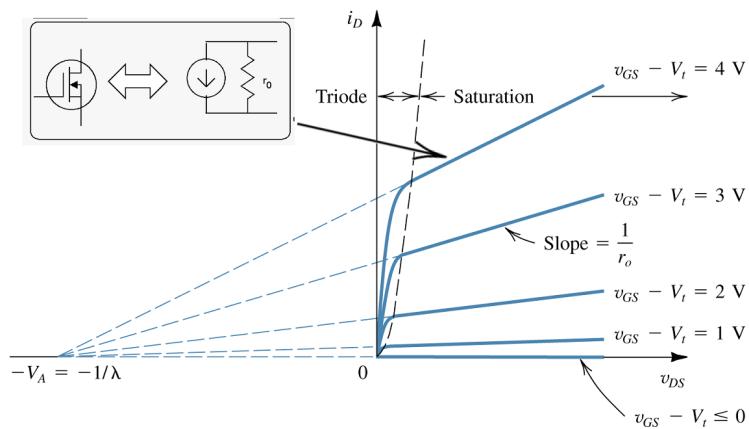


Figure 17: Transistor MOS : influence de la longueur effective sur le courant.

En résumé

Note préliminaire : dans un souci de simplification de notation, il est usuel de regrouper les termes μC_{ox} sous la notation k' et le terme $k' \frac{W}{L}$ sous la notation K . Nous emploierons désormais ces notations⁶.

Le transistor MOS a trois régime : le régime **bloqué** pour lequel il se comporte comme un interrupteur ouvert, le régime **ohmique** pour lequel il se comporte comme une résistance et le régime **saturé** pour lequel il se comporte comme une source de courant plus ou moins parfaite (selon si on considère la longueur de canal effectif).

Conditions	Régime	Équivalence	$I_{DS} =$
$V_{GS} < V_T$	Bloqué		0
$V_{GS} > V_T$	$V_{DS} < V_{GS} - V_T$	Ohmique	
	$V_{DS} > V_{GS} - V_T$	Saturé	<p>Infinie dans le cas idéal</p> <p>$\underbrace{\frac{K_n}{2}(V_{GS} - V_{Tn})^2}_{\text{Transistor idéal}} \underbrace{(1 + \lambda \cdot V_{DS})}_{\text{Transistor réel}}$</p>

Le substrat

Le substrat (bulk en anglais) est le socle dans lequel est fabriqué le transistor MOS :

⁶ Selon si on utilise un transistor NMOS ou un transistor PMOS, les mobilités étant différentes (μ_n et μ_p), on reporterai les indices dans les notations soit $k'_{n/p}$ et $K_{n/p}$.

c'est l'élément qui fournit les électrons qui sont attirés sous la grille. Ce substrat doit être bien évidemment polarisé afin que le transistor MOS fonctionne. En principe, sauf impossibilité due à l'architecture du circuit analogique, le potentiel de substrat est relié à celui de la source : en effet, une différence de potentiel entre la source et le substrat augmenterait d'autant la tension de seuil à fournir pour mettre le canal en état d'inversion (cf. Annexe 7.5). Le potentiel au substrat est appliqué via un dopage P+. Lorsque le substrat est représenté, on utilise le symbole à quatre pattes comme représenté figure 18.

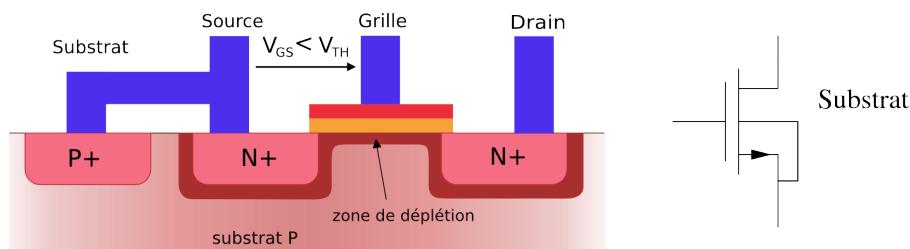


Figure 18: Le substrat du transistor MOS.

Phénomènes de canal court

Le modèle développé est valable pour les transistors à "canal long". Ce modèle présente des limites dès que le canal devient fortement inférieur au micromètre : on parle alors de canal court. Le modèle de canal long reste cependant souvent utilisé pour calculer "à la main" les dimensions des circuits car si elles ne sont pas exactes, elles donnent un bon point de départ pour les calculs numériques⁷ et permettent de dégager les paramètres à faire varier selon le résultat voulu.

Il y a deux phénomènes majeurs à considérer : la **saturation de la vitesse des électrons** et l'existence d'un **courant sous le seuil**.

Vitesse de saturation :

au-delà d'une certaine valeur de champ électrique, la vitesse des particules n'augmente plus de manière linéaire avec la valeur de ce dernier mais tend vers une valeur de saturation comme montrée figure 19.

⁷ On n'oubliera pas que les circuits actuels sont conçus à l'aide d'outils informatiques qui permettent de modéliser et simuler les circuits avec un grand niveau de précision et d'utiliser des algorithmes d'optimisation pour l'obtention des meilleures performances : cependant, pour être efficaces, ces outils ont besoin d'un point de départ pas trop éloigné de la solution final et d'indications sur les paramètres à faire varier, d'où l'intérêt des équations étudiées.

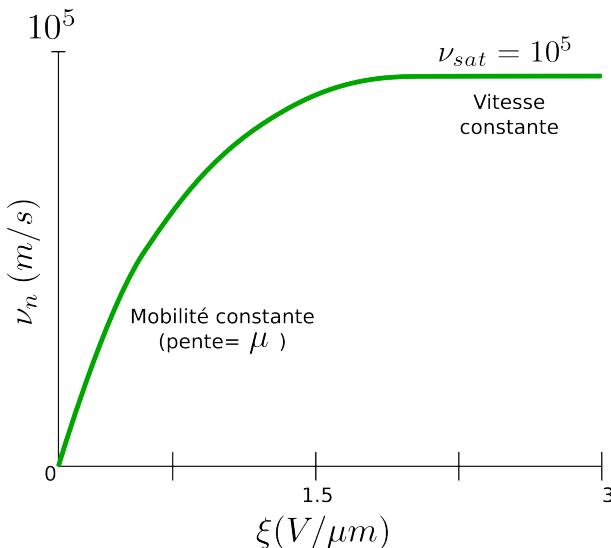


Figure 19: Évolution de la vitesse des électrons en fonction du champ électrique.

La vitesse de saturation dans le silicium est de $8 \cdot 10^6 \text{ cm.s}^{-1}$ pour les électrons et de $6 \cdot 10^6 \text{ cm.s}^{-1}$ pour les trous.

Pour les transistors à canal long, ce phénomène apparaît bien après le pincement du canal. Pour les transistors à canal court, ce phénomène apparaît bien avant la saturation de transistor MOS et devient donc prédominant comme le montre la figure 20.

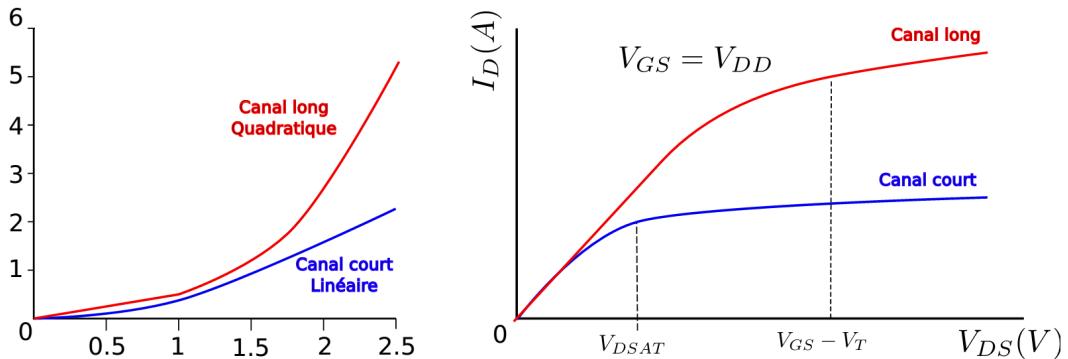


Figure 20: Comparaison entre le modèle de canal long et celui qui prend en compte la saturation de la vitesse des électrons.

La formule à considérer dès lors est une simplification de la formule 3.6. Dans le calcul de l'intégrale, on prend en compte, pour le calcul de la vitesse, le champ électrique sous la forme d'une dérivée de la tension V_{DS} . Ici, on considère directement la vitesse de saturation soit v_{sat} . La formule devient

$$I_{DS} = W \cdot C_{ox} \cdot v_{sat} (V_{GS} - V_T)$$

Par rapport à l'approximation du canal long, cela signifie un dépendance linéaire du courant de saturation avec V_{GS} (au lieu de quadratique), une indépendance vis à vis de la longueur de canal et enfin une indépendance de l'apparition de l'état de saturation par rapport à la tension de grille : la tension drain – source à laquelle apparaît la saturation est la même

quelle que soit la tension de grille!

Courant sous le seuil :

la transition entre l'état bloqué du transistor et l'état passant est moins brutale que ce qui a été affirmé jusqu'à maintenant : pendant l'état de déplétion, un courant existe⁸ qui peut être décrit comme étant de l'ordre de

$$I_{DS} = I_S \cdot e^{qV_{GS}/nk_B T} \text{ avec } n \geq 1$$

Ce courant, négligeable en canal long ne l'est plus dans l'hypothèse d'un canal court comme le montre la figure 21.

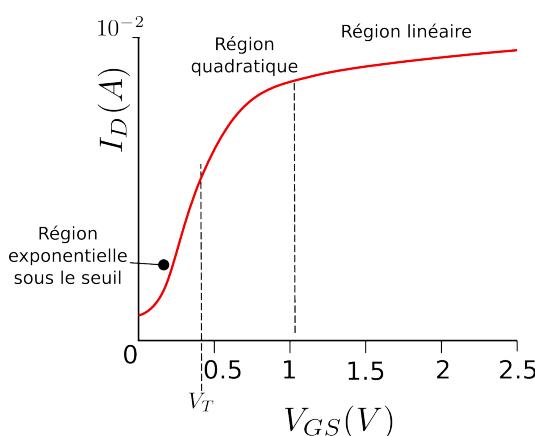


Figure 21: Courant sous le seuil dans un transistor MOS.

Ordres de grandeur et état de l'art

Les transistors utilisés dans le cadre de l'autonomie sont ceux de la technologie CMOS $0.25\mu m$ (datant des années 2000) dont les paramètres technologiques sont rappelés dans le tableau suivant :

Type de MOS	V_{T0} (V)	K (A/V^2)	λ (V^{-1})	V_{DD} (V)
NMOS	0.4	131×10^{-6}	0.1	2.5
PMOS	0.55	45×10^{-6}	0.2	2.5

Tout au long de ce cours, ce sont ces grandeurs qui seront utilisées pour les applications numériques.

Actuellement, l'état de l'art des transistors MOS est la technologie 32nm ce qui correspond à une épaisseur d'oxyde équivalent SiO_2 de 0.9nm.

3.3 Utilisation du modèle grand signal : un exemple applicatif

3.3.1 Montage de source commune

Soit le montage de source commune représenté figure 22. Ce montage est constitué

⁸ caractéristique d'un courant de jonction PN.

d'une résistance et d'un transistor MOS. L'entrée du montage V_E se situe sur la grille du transistor MOS et la sortie V_S sur le drain du transistor MOS. La source est reliée à 0, un point à potentiel fixe du montage d'où le nom de montage à source commune.

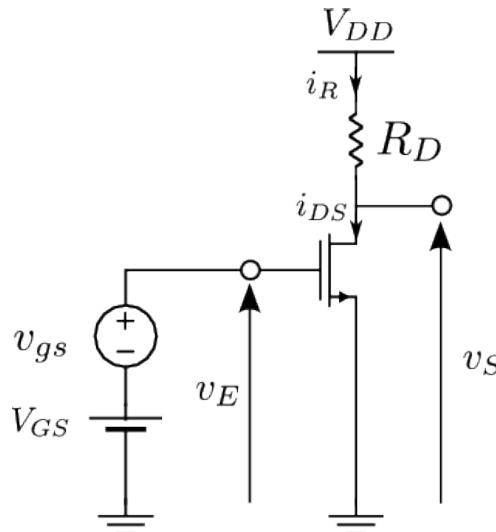


Figure 22: Montage en source commune.

Ce montage est alimenté par une source de tension fixe de valeur V_{DD} . Ce montage est connecté en entrée à une source de tension fixe qui polarise le montage en série avec un signal variable qui est le signal à amplifier. Puisque nous étudions le comportement statique du montage, nous ne prenons pas en compte la source dynamique.

3.3.2 Tracé de la caractéristique de transfert entrée – sortie du montage

Pour tracer cette caractéristique de transfert, il faut étudier l'évolution de la tension de sortie lorsque l'entrée varie entre les deux tensions d'alimentation à savoir de 0 à V_{DD} . Pour connaître la tension de sortie V_S , il suffit d'écrire l'égalité du courant I_{DS} qui rentre dans le drain du transistor MOS et celui qui sort de la résistance I_R . Selon la valeur de la tension V_S égale à $(V_{DD} - R_D \cdot I_R)$, on utilisera l'équation adéquate pour I_{DS} suivant le régime du transistor MOS.

En résumé la démarche est la suivante : selon la valeur de V_{GS} et la configuration du circuit, on fait une supposition sur l'état du transistor MOS, on utilise la formulation de courant adaptée, on calcule V_S puis on vérifie que le résultat obtenu est cohérent avec l'état supposé du transistor (on connaît V_E soit V_{GS} et on connaît V_S soit V_{DS}). Dans le cas où le résultat n'est pas cohérent, on prend une autre hypothèse sur le transistor MOS et on recommence et pour toute valeur de V_E comprise entre 0 et V_{DD} .

Bien évidemment cette méthode itérative est quelque peu fastidieuse. Une méthode plus rapide est de recourir à une résolution graphique illustrée par la figure 23. Les étapes à respecter sont :

1. tracer la loi qui relie le courant I_R qui traverse la résistance à la tension V_S (loi donnée par la loi de mailles $V_{DD} - R_D \cdot I_R = V_S$) relation qu'on appelle **droite de charge** :

$$I_R = \frac{V_{DD} - V_S}{R_D}$$

2. tracer la loi d'existence du courant I_{DS} au sein du transistor MOS en fonction de la tension V_S ce qui revient à tracer le réseau caractéristique $I_{DS} = f(V_{DS})$.

3. prendre l'intersection de ces deux lois d'existence (pour rappel $I_R = I_{DS}$) qui donne les points pour lesquels les deux lois sont simultanément respectées au sein du circuit, soit la caractéristique de transfert entrée – sortie.

On peut dégager trois zones de fonctionnement du transistor MOS sur la courbe de transfert :

- Entre le point X et le point A : le transistor NMOS est bloqué ($V_E < V_T$), le courant traversant la résistance est nulle et donc $V_S = V_{DD}$.
- Entre le point A et le point B, la tension $V_E (= V_{GS})$ est supérieure à V_T et la tension $V_S (= V_{DS})$ est supérieure à $V_E - V_T$; donc le transistor NMOS est en régime saturé.
- À partir du point B, la tension $V_E (= V_{GS})$ est toujours supérieure à V_T et la tension $V_S (= V_{DS})$ devient inférieure à $V_E - V_T$: le transistor NMOS est alors en régime ohmique (ou triode).

Remarque : Si on devait utiliser ce montage pour réaliser l'amplification d'un signal oscillatoire en entrée, il faudrait se placer dans la région où le transistor NMOS est en régime saturé (entre le point A et le point B) afin de profiter d'une amplification maximale (par exemple au point Q).

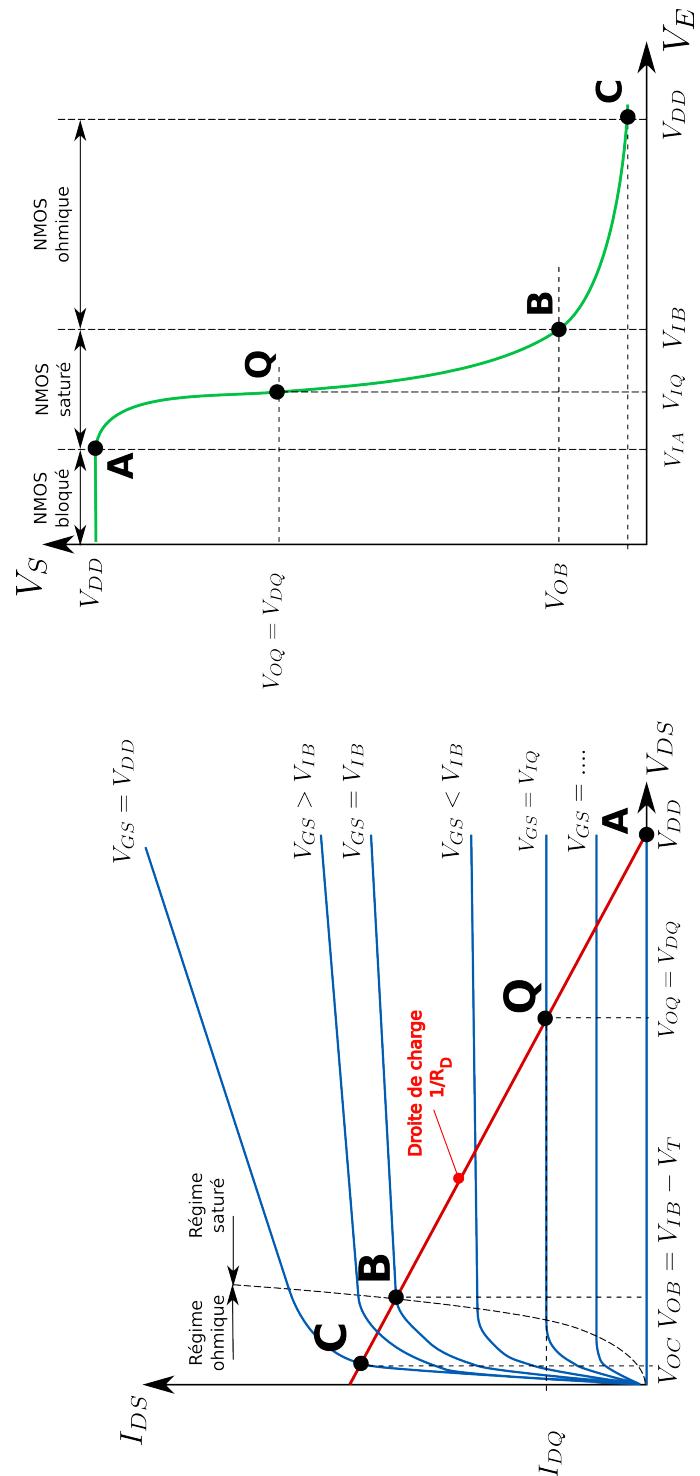


Figure 23: Résolution graphique.

3.3.3 Déterminer un point de polarisation

La méthode précédente est celle qu'il faut adopter lors de l'étude complète d'un montage. Souvent, on se trouve confronté à un montage déterminé dont on nous demande de trouver le point de fonctionnement. Prenons pour exemple le montage de la figure 24.

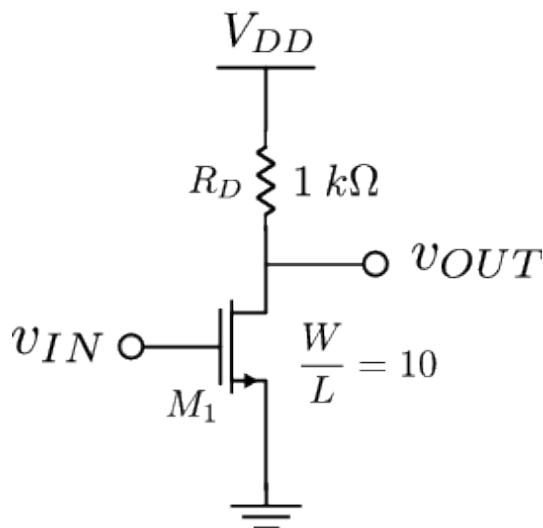


Figure 24: Montage en source commune en technologie $0.25\text{ }\mu\text{m}$.

Le montage est dimensionné et on nous donne une valeur pour $V_{IN} = 1V$. Il faut calculer V_{out} . Soit on procède à l'étude complète du montage (solution un peu longue), soit on procède par hypothèse :

1. on vérifie d'abord si le transistor est bloqué ou passant (sens large pour dire régime ohmique ou régime saturé). Ici $V_{GS} = V_{IN} = 1V > V_{Th}$. Le transistor est donc passant.
2. on ne sait pas si le transistor est passant ou bloqué puisqu'on ne connaît pas V_{DS} (égal à V_{OUT} la valeur recherché). On fait donc une hypothèse sur l'état du transistor MOS qu'on va supposer saturé.
3. on calcule $V_{OUT} = V_{DD} - R_D \cdot I_{DS}$ avec I_{DS} le courant du transistor MOS en saturé.
4. on vérifie avec la valeur V_{ds} obtenue que l'hypothèse de départ est bonne à savoir $V_{DS} = V_{OUT} > V_{GS} - V_{Th} = V_{IN} - V_{Th}$. Au cas où l'hypothèse n'est pas respectée, on recommence les calculs avec l'hypothèse du transistor en régime ohmique.

3.4 Régime dynamique : modèle petit signal

Régime de fonctionnement

Le transistor MOS en électronique analogique est très utilisé pour réaliser des amplificateurs. Pour ce travail en dynamique, il est nécessaire, la plupart du temps, qu'il fonctionne en régime saturé pour obtenir les meilleures performances (cf. paragraphe 3.3.2): on utilise sa fonction de source commandée en tension pour réaliser une amplification⁹.

Le modèle petit signal du transistor MOS est donc celui d'une source de courant dynamique (cf. figure 25) qui s'ajoute à la source de courant statique du transistor MOS en régime saturé. Les caractéristiques petit signal du transistor MOS sont donc celles de la source de courant petit signal associée à savoir le gain de la source (transconductance g_m) et

⁹ Il est évident que dans son fonctionnement ohmique, le transistor MOS ne peut guère réaliser d'amplification – il se comporte comme une résistance variable–.

son impédance de sortie souvent notée $r_d s$ ou r_0^{10} .

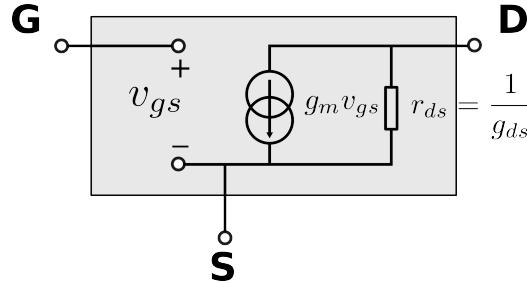


Figure 25: Schéma petit signal du transistor MOS.

Le transistor MOS ayant un comportement qui varie par rapport à deux grandeurs électriques externes que sont v_{GS} et v_{DS} , il faut donc étudier la dérivée du comportement grand signal de la grandeur utile du transistor MOS, à savoir i_{DS} par rapport à ces deux variables externes.

Nous choisissons de nous placer à un point de polarisation correspondant à un courant I_{DQ} fixé par une tension fixe d'entrée V_{IQ} et résultant en une tension fixe de sortie V_{OQ} (cf. figure 26).

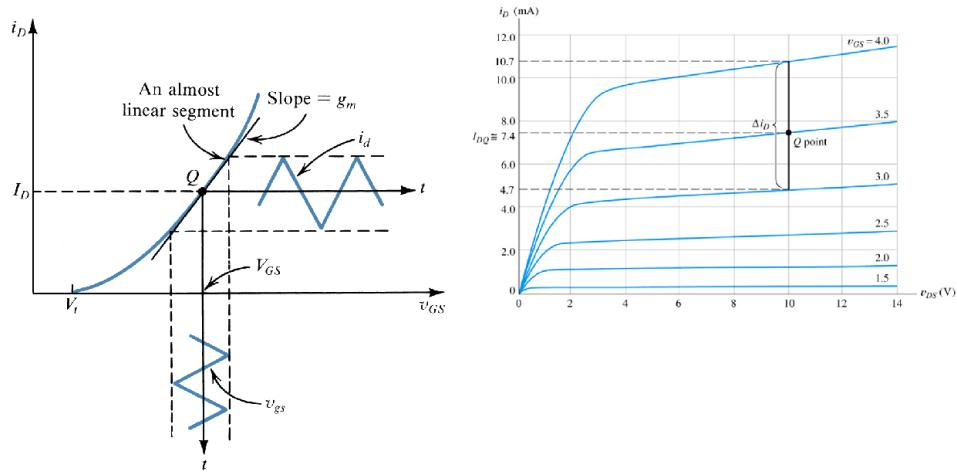


Figure 26: Petites variations autour d'un point de polarisation.

3.4.1 Calcul de la transconductance

Étudions la grandeur g_m (cf. figure 27) définie par

$$g_m = \frac{\partial i_{DS}}{\partial v_{GS}} \Big|_{Q \text{ constant}, V_{DS} \text{ constant}}$$

¹⁰ Pour le transistor MOS, il est habituel d'utiliser plutôt des notations d'admittance à savoir $g_d s$ ou g_0 .

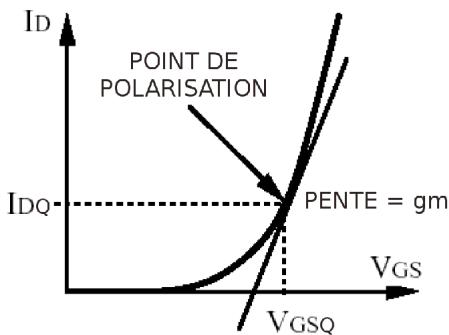


Figure 27: Transconductance.

Connaissant l'expression de $I_{DS} = f(V_{GS}, V_{DS})$, on peut établir l'expression de g_m soit

$$g_m = \frac{\partial}{\partial V_{GS}} \left(\frac{K_n}{2} (V_{GS} - V_T)^2 \right) \Big|_{Q \text{ constant}, V_{GS} = V_{IQ}, V_{DS} = V_{OQ}}$$

$$g_m = K_n (V_{IQ} - V_T)$$

De manière générale, on cherche à exprimer les grandeurs petit signal en fonction d'une grandeur signal de référence, I_{DQ} dans le cas du transistor MOS. Le transistor MOS étant une source de courant commandée lorsqu'il fonctionne en saturé, la grandeur I_{DQ} est la grandeur utilisée pour définir le cahier des charges d'un circuit et le dimensionnement de ce dernier. Sachant que $(V_{GS} - V_T) = \sqrt{\frac{2 \cdot I_{DS}}{K_n}}$, la transconductance g_m s'écrit

$$g_m = \sqrt{2 \cdot K_n \cdot I_{DQ}} \quad (3.9)$$

Le gain de transconductance est donc proportionnel à la racine carrée du courant statique qui traverse le transistor MOS.

3.4.2 Calcul de l'admittance de sortie

Étudions la grandeur g_{ds} (cf. figure 28) définie par

$$g_{ds} = \frac{\partial i_{DS}}{\partial v_{DS}} \Big|_{Q \text{ constant}, V_{GS} \text{ constant}}$$

On remarque immédiatement que si on prend l'hypothèse de transistor idéal, l'admittance de sortie est nulle, ce qui correspond bien à une impédance de sortie infinie caractéristique d'une source de courant idéal (ce qu'est le transistor MOS en régime saturé dans le cas idéal).

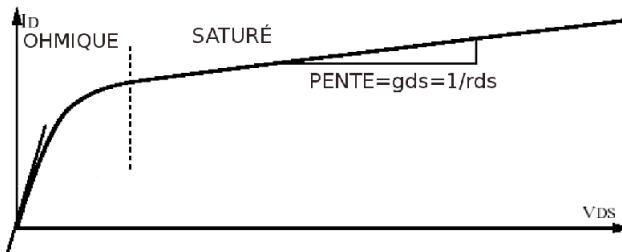


Figure 28: Admittance.

Connaissant l'expression de $I_{DS} = f(V_{GS}, V_{DS})$, on peut établir l'expression de g_{ds} soit

$$\begin{aligned} g_{ds} &= \frac{\partial}{\partial V_{DS}} \left(\frac{K_n}{2} (V_{GS} - V_T)^2 \right) \cdot (1 + \lambda \cdot V_{DS}) \Big|_{Q \text{ constant}, V_{GS} = V_{IQ}, V_{DS} = V_{OQ}} \\ g_{ds} &= \frac{K_n}{2} (V_{IQ} - V_T)^2 \cdot \lambda \\ g_{ds} &\approx \lambda I_{DQ} \end{aligned}$$

L'impédance de sortie est donc inversement proportionnelle au courant statique qui traverse le transistor MOS.

Remarque : Il existe un facteur 100 entre la valeur de g_m et celle de g_{ds} , ordre de grandeur à garder en mémoire lors des simplifications éventuelles.

3.4.3 Le modèle fréquentiel du transistor MOS

Le signal petit signal développé jusqu'ici est le schéma petit signal basse fréquence. En effet, puisqu'il est dérivé du comportement statique du transistor MOS, il ne prend pas en compte les différents aspects dynamiques du transistor MOS. De manière générale, ces aspects dynamiques sont représentés sous la forme de capacités¹¹. Dans le transistor MOS, deux capacités sont à prendre en compte, à savoir la capacité grille-source C_{GS} et la capacité grille-drain C_{GD} . Ces deux capacités ont, de plus, des valeurs qui dépendent du régime du transistor NMOS comme le montre la figure 29.

¹¹ Les capacités peuvent représenter des capacités physiques existant géométriquement dans le dispositif (charges électriques stockées dans des plans parallèles) mais également des capacités "virtuelles" qui permettent de prendre en compte des phénomènes dynamiques comme par exemple les capacités de déplétion qui permettent de prendre en compte la réactivité du transistor lors des changements de sa zone de déplétion.

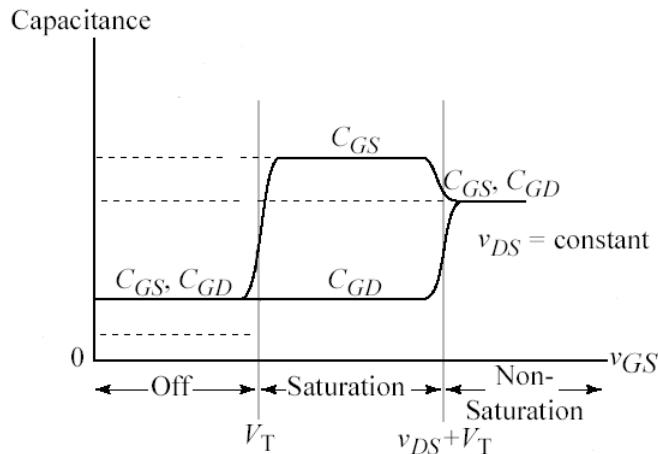


Figure 29: Évolution des capacités selon l'état du transistor MOS.

Le modèle fréquentiel du transistor MOS est celui représenté sur la figure 30.

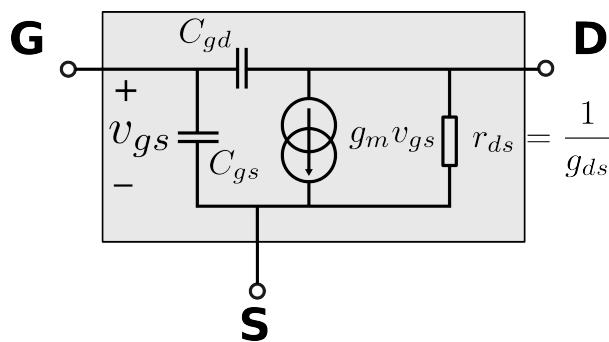


Figure 30: Modèle fréquentiel du transistor MOS.

Le problème majeur de cette représentation provient de la capacité grille-drain qui crée une relation physique entre l'entrée et la sortie (en complète opposition par rapport au modèle quadripolaire classique basé sur une représentation séparée circuit d'entrée et circuit de sortie). Cette limitation peut être contournée en utilisant l'effet Miller.

Effet Miller

Il peut être résumé par la figure 31.

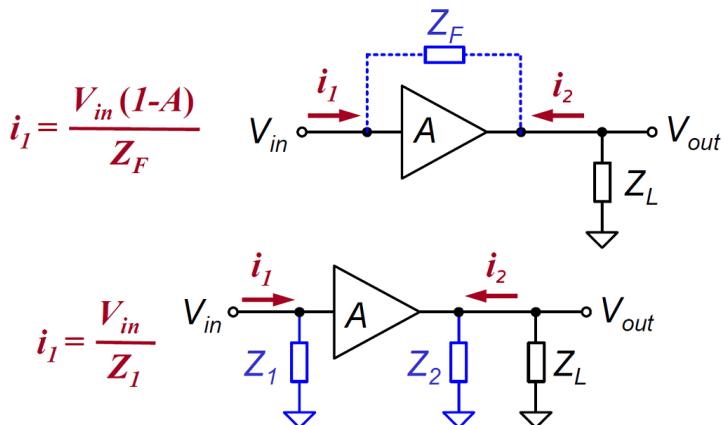


Figure 31: Effet Miller.

Pour un montage dont l'entrée et la sortie sont initialement séparée par un amplificateur dont l'impédance d'entrée est infinie (courant rentrant nul) et de fort gain A , le fait de rajouter une impédance reliant l'entrée et la sortie est équivalent à considérer le montage initial avec l'impédance ramenée en entrée divisée par le gain de l'amplificateur et en sortie multipliée par un gain quasi-unitaire.

La démonstration se fait sur l'égalité des courants.

$$i_1 = \frac{V_{in} - V_{out}}{Z_F} = \frac{V_{in} - A \cdot V_{in}}{Z_F} \approx \frac{V_{in}}{Z_F}$$

$$i_2 = \frac{V_{out} - V_{in}}{Z_F} = \frac{V_{out} - \frac{V_{out}}{A}}{Z_F} = \frac{V_{out}}{Z_F \frac{A}{A-1}} \approx \frac{V_{out}}{Z_F}$$

Fréquence de transition

On définit la fréquence de transition comme étant la fréquence pour laquelle le gain en courant devient unitaire. On démontrera par la suite que cette fréquence correspond en fait au temps de parcours des charges dans le canal du transistor.

Pour calculer la fréquence de transition, il est nécessaire de calculer le gain en courant en se servant du modèle de la figure 30. On se place en condition de court circuit afin de calculer le gain en courant (courant de court-circuit). On obtient dès lors les expressions suivantes :

$$i_{in} = v_{gs} \cdot j\omega(C_{gs} + C_{gd}) \quad (3.10)$$

$$i_{cc} = v_{gs} (g_m - j\omega C_{gd}) \quad (3.11)$$

$$A_i = \frac{i_{cc}}{i_{in}} = \frac{g_m - j\omega \cdot C_{gd}}{j\omega(C_{gs} + C_{gd})} \quad (3.12)$$

Calculons la fréquence pour laquelle ce gain devient unitaire soit

$$|A_i| = \frac{\sqrt{g_m^2 + \omega^2 C_{gd}^2}}{\omega(C_{gs} + C_{gd})} = 1 \quad (3.13)$$

$$f_T = \frac{g_m}{2\pi(C_{gs} + C_{gd})} \quad (3.14)$$

Signification physique de f_T :

Étudions f_T pour le transistor en régime saturé (le cas le plus probable quant à l'utilisation du transistor où f_T pourrait avoir une influence). D'après la figure 29, C_{gd} peut être négligé devant C_{gs} . Par ailleurs, en régime saturé, on peut démontrer assez facilement (étude géométrique qui sort du cadre de ce court) que $C_{gs} \approx (2/3)WLC_{ox}$. L'expression de f_T devient

$$\frac{1}{2\pi f_T} \approx \frac{C_{gs}}{gm} = \frac{(2/3)WLC_{ox}}{K_n \left(V_{DSAT} - V_T \right)} = \frac{(2/3)L}{\mu_n V_{DSAT} / L} \approx \frac{1}{\tau_{transit}}$$

V_{DSAT} μ_n L $\tau_{transit}$
 V_T E V_{DSAT}
 V_{DSAT} V_{DSAT} V_{DSAT}
 Vitesse d'électron

On démontre ainsi que la définition prise pour la fréquence de transition est celle qui correspond au temps moyen de parcours du canal, facteur représentatif des limites fréquentielles des performances du transistor NMOS.

3.5 Utilisation du modèle petit signal : un exemple pratique

3.5.1 Montage à source commune

Reprendons l'exemple précédent de la source commune (cf. figure 22). Nous nous plaçons au point Q, dans la zone où le montage présente la plus forte pente et donc le plus grand gain. Le transistor NMOS est en régime saturé, traversé par un courant fixe I_{DQ} . Appliquons des variations linéaires autour de ce point et calculons les variations résultantes en sortie comme représenté figure 32.

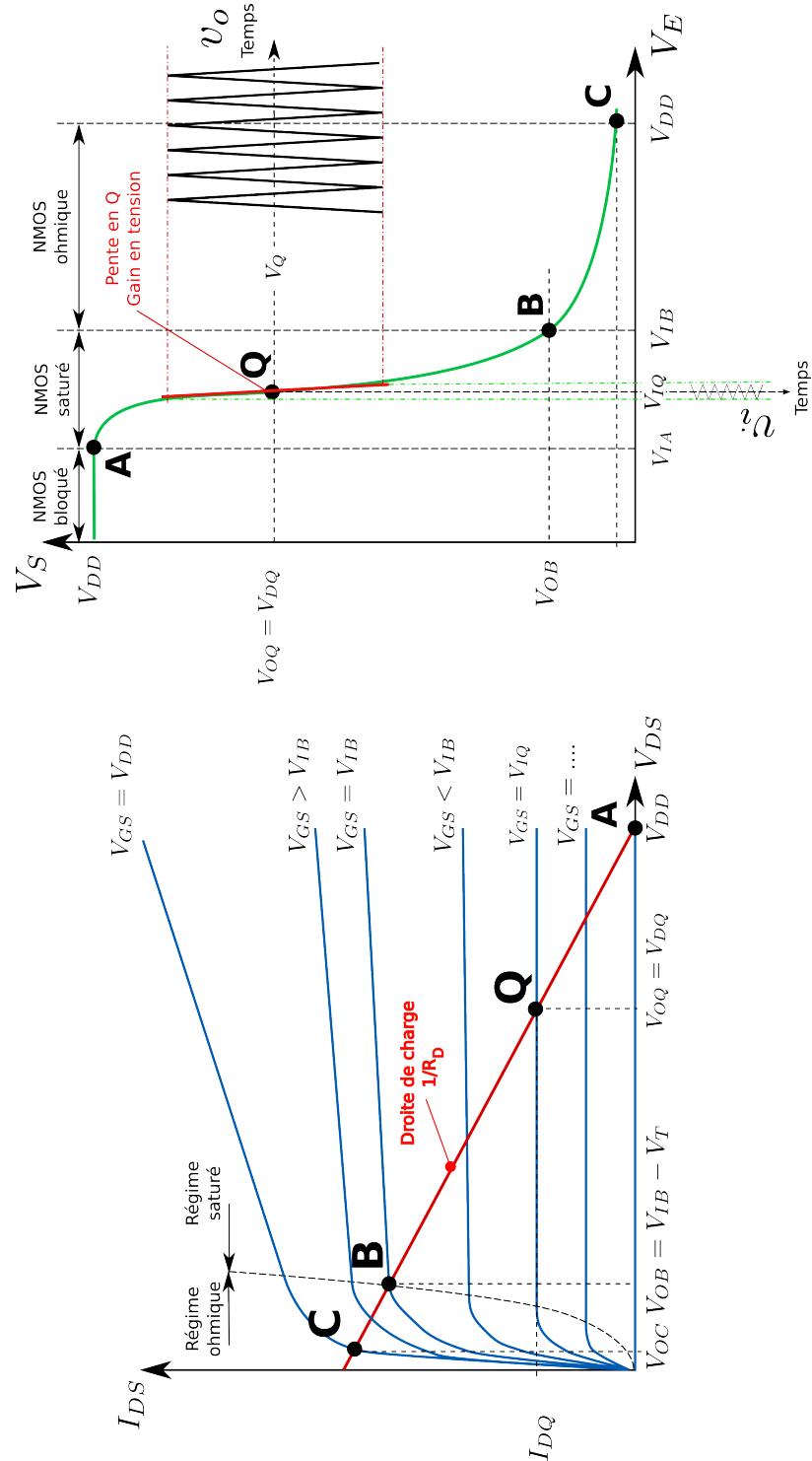


Figure 32: Point de polarisation et variations petit signal.

Autour du point Q, les variations considérées sont d'amplitude suffisamment faible pour que nous puissions linéariser autour du point de fonctionnement, c'est à dire utiliser le modèle petit signal du transistor MOS pour établir le modèle petit signal du montage à source commune.

3.5.2 Schéma petit signal du montage à source commune

Par défaut, la notion d'étude petit signal est basse fréquence, c'est à dire qu'on ne prend pas en compte les effets capacitifs internes du transistor MOS, effets qui se font sentir à haute fréquence. Dans le montage électrique du montage à source commune, pour établir le schéma petit signal équivalent, nous commençons à éliminer toutes les sources de tension ou courant continues puisqu'elles ne présentent pas de comportement variationnel¹². Ensuite, on remplace tous les éléments du montage par leur équivalent petit signal au point de polarisation donné : les résistances, inductances et capacités sont leurs propres équivalents petit signal et le transistor MOS est remplacé par son modèle petit signal. On obtient ainsi le montage de la figure 33.

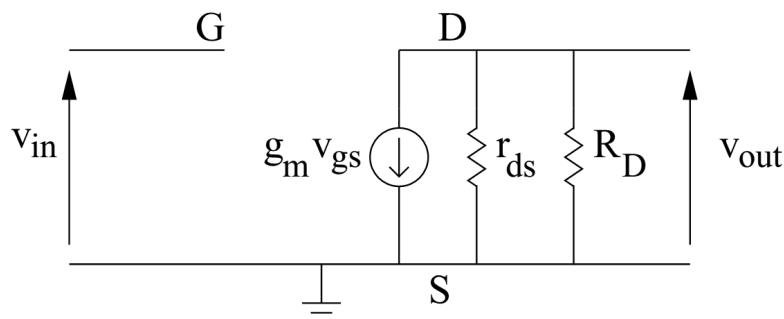


Figure 33: Schéma petit signal du montage à source commune.

Ce montage étant utilisé pour amplifier un signal en tension (cf. figure 32), nous allons établir le schéma quadripolaire équivalent de ce montage, à savoir celui d'un amplificateur de tension commandé en tension, d'impédance d'entrée Z_e , d'amplification A_v et d'impédance de sortie Z_s (cf. figure 34).

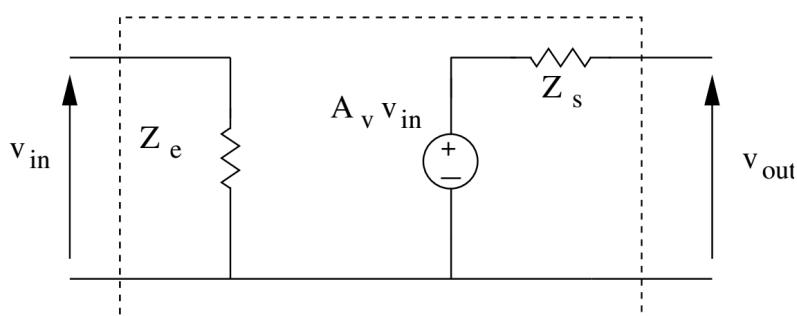


Figure 34: Schéma quadripolaire équivalent du montage à source commune.

Il nous faut donc calculer les valeurs littérales de ces trois grandeurs en fonction des paramètres du montage de la source commune.

¹² On peut également dire qu'on les remplace par leur équivalent petit signal à savoir un fil ($V = 0$) pour une source tension idéale et un interrupteur ouvert ($i = 0$) pour une source de courant idéale

3.5.3 Calcul de l'impédance d'entrée

Il suffit d'observer du point de vue de l'entrée le comportement électrique du montage. Il est évident, par simple observation, que l'impédance d'entrée est infinie. La démonstration formelle de $Z_e = \infty$ se fait en adoptant la démarche classique : pas de charge en sortie du montage, branchement d'une source de tension et observation du courant débité et enfin calcul du rapport de ces deux grandeurs¹³ comme représenté figure 35.

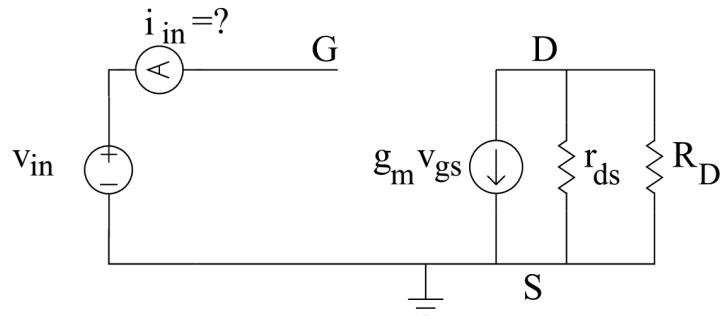


Figure 35: Calcul de l'impédance d'entrée du montage à source commune.

3.5.4 Calcul de l'impédance de sortie

Il suffit d'observer du point de vue de la sortie le comportement électrique du montage. On observe une résistance équivalente constituée de la mise en parallèle de R_D et

r_{ds} . La démonstration formelle de $Z_s = \frac{R_D \cdot r_{ds}}{R_D + r_{ds}}$ se fait en adoptant la démarche classique :

tension nulle en entrée du montage, branchement d'une source de courant et mesure de la tension aux bornes de la source et enfin calcul du rapport de ces deux grandeurs¹⁴ comme représenté figure 36.

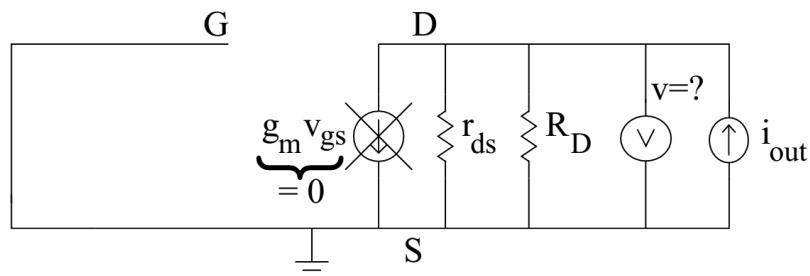


Figure 36: Calcul de l'impédance de sortie du montage à source commune.

¹³ On peut tout aussi bien brancher une source de courant et mesurer la tension en entrée du montage.

¹⁴ On peut tout aussi bien brancher une source de tension et mesurer le courant débité.

3.5.5 Calcul du gain en tension du montage

Pour calculer le gain en tension du montage, il suffit de placer le circuit en boucle ouverte (pas de charge en sortie du circuit), d'appliquer une tension v_{in} en entrée et de mesurer la tension à vide obtenue en sortie (cf. figure 37). Le rapport des deux tensions donne le gain.

$$1. \text{ Calcul de la tension de sortie : } v_{out} = (R_D P r_{ds}) \cdot g_m \cdot \underset{=v_{in}}{\cancel{v_{gs}}}$$

$$2. \text{ ce qui donne } A_v = \frac{v_{out}}{v_{in}} = g_m \cdot Z_s = g_m \cdot \frac{R_D \cdot r_{ds}}{R_D + r_{ds}}$$

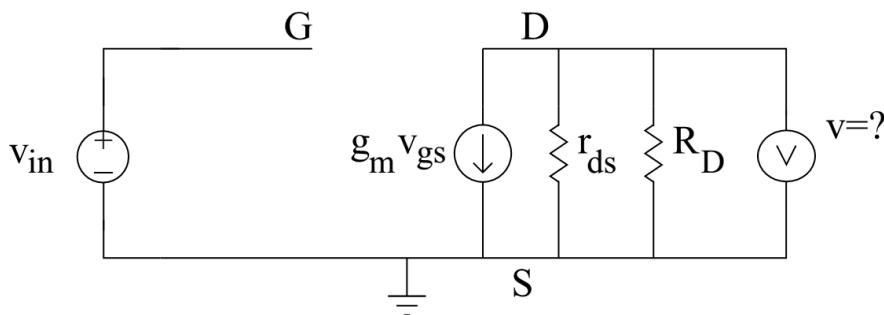


Figure 37: Calcul du gain en tension du montage à source commune.

En résumé, le montage à source commune apparaîtra comme un amplificateur de tension pour tout signal d'entrée oscillant autour du point de polarisation, amplificateur de tension dont le gain en tension égal à $g_m \cdot (R_D P r_{ds})$ est principalement défini par la valeur du courant de polarisation traversant le transistor MOS.

3.5.6 Étude fréquentielle du montage

Reprendons le schéma de la figure 22. Nous allons maintenant considérer les aspects fréquentiels, ce qui revient à inclure dans l'étude petit signal, les capacités C_{gs} et C_{gd} . Le nouveau schéma petit signal à considérer devient celui de la figure 38.

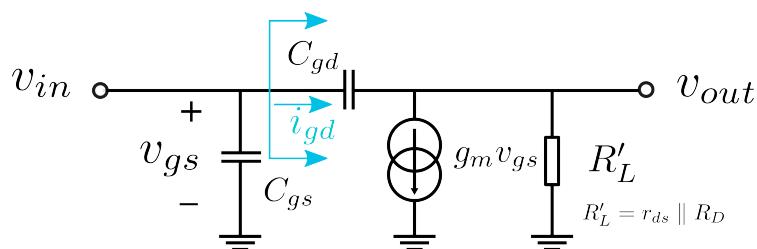


Figure 38: Schéma petit signal : étude fréquentielle.

Établir le schéma petit signal équivalent de ce montage (une impédance d'entrée,

une impédance de sortie et un amplificateur de tension) n'est pas chose aisée du fait de la capacité C_{gd} . C'est le moment de se rappeler du théorème de Miller puisque le montage de la figure 38 n'est jamais que celui de la source commune étudiée précédemment auquel on rajoute une capacité en entrée et une capacité entre l'entrée et la sortie comme le montre le montage de la figure 39.

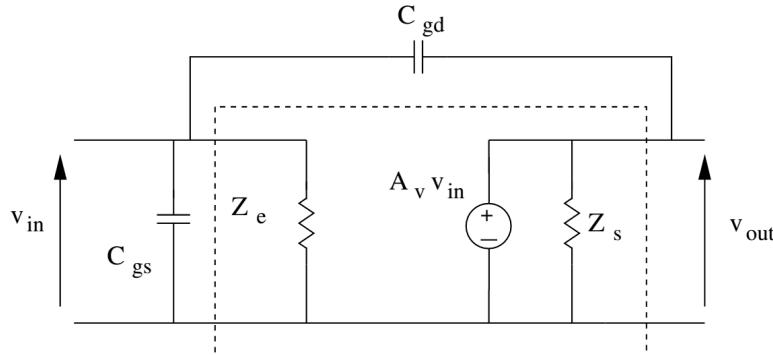


Figure 39: Simplification du montage : Miller à la rescousse.

On obtient, dès lors le montage de la figure 40 ($A_v = g_m R'_L$) avec la capacité ramenée en sortie qu'on néglige du fait de sa faible influence sur le montage.

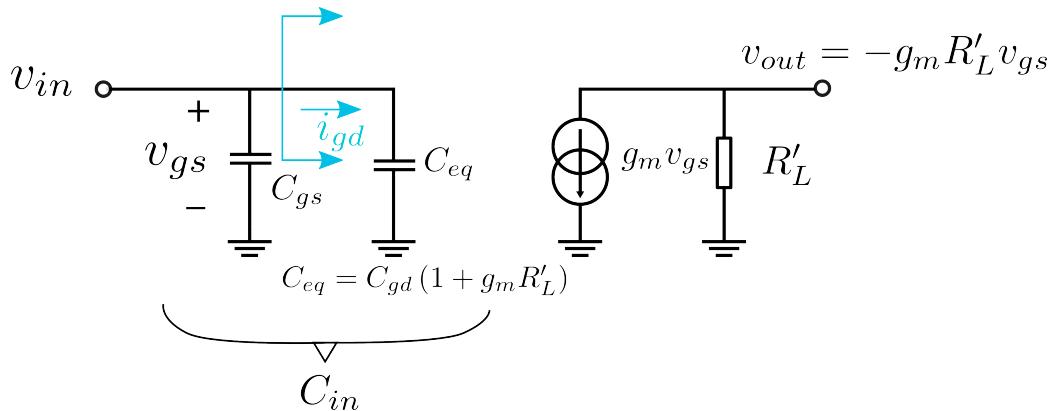


Figure 40: Le schéma petit signal final.

Dès lors, le calcul du modèle quadripolaire est trivial : il s'agit du même que celui pour les basses fréquences avec comme seul changement une impédance d'entrée qui n'est plus infinie mais qui vaut $Z_e = \frac{1}{C_{in}}$, faisant apparaître ainsi une bande passante limitée, dépendant de l'impédance de sortie du montage (source ou autre étage) qui alimente le montage à source commune.

3.6 Le transistor PMOS

3.6.1 Structure et symbole

La structure d'un transistor PMOS est identique à celle d'un transistor NMOS avec une inversion entre les zones dopées P et celles dopées N : source et drain sont dopés P et le

substrat est dopé N¹⁵, comme le montre la figure 41.

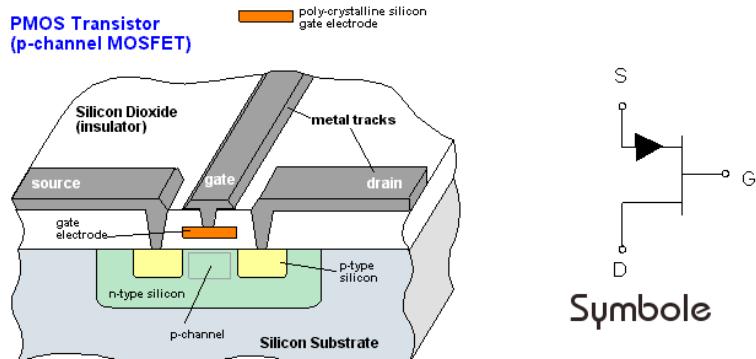


Figure 41: Le transistor PMOS.

Le symbole est le même que celui du transistor NMOS avec comme différence la flèche qui rentre dans la source au lieu d'en sortir pour le NMOS.

Fonctionnement

La différence majeur entre un transistor NMOS et un transistor PMOS réside dans la nature des charges mises en mouvement : des électrons pour le transistor NMOS et des trous pour le transistor PMOS. La mobilité des trous étant plus faible que celle des électrons, on retrouvera, à taille égale, des performances moindres pour le transistor PMOS que pour le transistor NMOS. Ceci étant dit, le principe de fonctionnement reste le même : une tension appliquée sur la grille (au-delà d'une certaine valeur) attire des trous qui sont mis en mouvement par une différence de potentiel (qui crée un champ électrique) entre le drain et la source. Les porteurs étant des trous, toutes les tensions à considérer sont des tensions négatives.

On peut donc reprendre la totalité des équations vues jusqu'ici en gardant à l'esprit que les tensions considérées sont négatives, que les inégalités doivent être inversées par rapport à celles du transistor NMOS.

Une autre méthode beaucoup plus simple d'un point de vue mnémotechnique pour l'utilisation du transistor PMOS en connaissant les formules pour le transistor NMOS est la suivante : on utilise les équations du transistor NMOS en

1. inversant D et S dans toutes les équations : V_{DS} devient V_{SD} , I_{DS} devient I_{SD} ...
2. inversant G et S dans toutes les équations : V_{GS} devient V_{SG} , ...
3. en remplaçant V_{Tn} par $|V_{Tp}|$ ¹⁶
4. en remplaçant la mobilité des électrons μ_n par celles des trous μ_p et donc k'_n devient k'_p et K'_n devient K'_p .

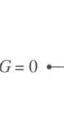
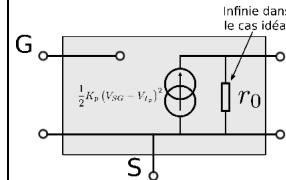
D'un point de vue utilisation, la source du transistor PMOS (donc la zone du transistor

¹⁵ En fait de substrat, le transistor PMOS n'étant pas le plus employé, il est fréquent d'utiliser du substrat dopé P dans lequel on crée des caissons dopé N pour accueillir les transistors PMOS. Ces caissons font office de substrat pour le transistor PMOS.

¹⁶ Pour rappel V_{Tp} est négatif car il faut appliquer une certaine tension négative de seuil pour que les trous attirés sous la grille ne soit pas complété par les électrons libres du matériau dopé N.

PMOS qui "génère" les trous) sera toujours préférentiellement reliée au potentiel le plus haut¹⁷.

3.6.2 Le PMOS en équation : régime statique

Conditions	Régime	Équivalence	$I_{SD} =$	
$V_{SG} < V_{Tp} $	Bloqué		0	
	$V_{SD} < V_{SG} - V_{Tp} $	Ohmique		$K_p \left[(V_{SG} - V_{Tp}) V_{SD} - \frac{V_{SD}^2}{2} \right]$
$V_{SG} > V_{Tp} $ $V_{SD} > V_{SG} - V_{Tp} $	Saturé		$\underbrace{\frac{K_p}{2} (V_{SG} - V_{Tp})^2}_{\text{Transistor idéal}} \underbrace{(1 + \lambda \cdot V_{SD})}_{\text{Transistor réel}}$	

3.6.3 Le PMOS en équation : régime dynamique

En régime dynamique, le transistor PMOS se comporte comme une source de courant¹⁸ commandée en tension. Le schéma petit signal qui en découle est celui de la figure 42.

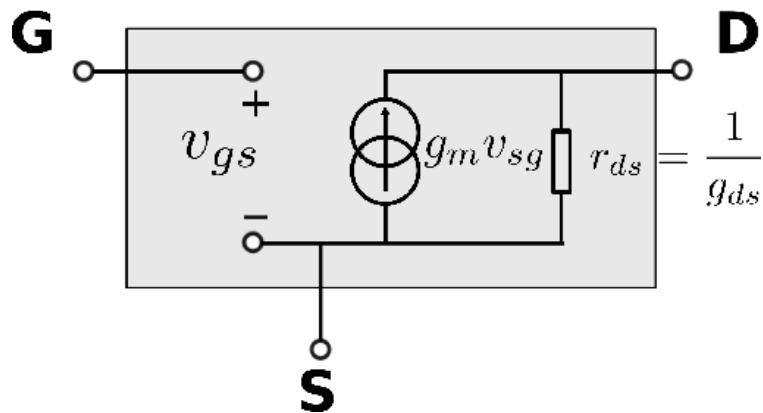


Figure 42: Schéma petit signal du transistor PMOS.

Pour les expressions de la transconductance et de l'admittance de sortie on applique les règles énoncées précédemment (inversion G-S et D-S).

¹⁷ A l'inverse du transistor NMOS pour lequel la source est préférentiellement reliée au potentiel le plus bas du circuit.

¹⁸ Courant qui circule entre la source et le drain – les trous étant assimilés à des charges positives – à l'inverse du transistor NMOS où le courant circule entre le drain et la source – le courant est inverse au flux des électrons – : dans un transistor MOS les porteurs circulent **toujours** de la source vers le drain, mais selon la nature des porteurs, le sens du courant change en conséquence.

References

- [9] Rabaey Jan M., *Digital Integrated Circuits, 2nd Edition*, Prentice Hall.
- [10] Tsividis Yannis, *Operation and modeling of the MOS transistor*, WCB/McGraw-Hill, 621.381 528 TSI
- [11] Razavi Behzad, *Design of analog CMOS integrated circuits*, McGraw-Hill, Publication Boston, MA, 2001, 621.39 RAZ
- [12] Sedra Adel S. et al., *Microelectronic circuits*, Oxford University Press, New York Oxford, 2004, 621.381 5 SED

4 L'inverseur CMOS

Ce chapitre traite de l'inverseur CMOS. L'inverseur CMOS a deux régimes de fonctionnement qui font de lui un amplificateur élémentaire mais aussi le circuit numérique élémentaire, utilisé dans tout système numérique et dont la compréhension du fonctionnement permet :

- d'une part, de mettre en application toute la théorie des circuits analogiques vus jusqu'ici
- et d'autre part, de dégager les bases de l'étude d'un circuit numérique par l'étude des métriques telles que la robustesse, la performance, la consommation et l'efficacité énergétique.

La compréhension de son fonctionnement statique et dynamique permettra d'aborder des circuits analogiques plus complexes telles que la paire différentielle (brique élémentaire de tout amplificateur) et l'étude des métriques précédemment citées permettra de comprendre le fonctionnement et les contraintes temporelles et énergétiques de portes numériques plus complexes telles que des PORTES NAND, NOR, XOR, eux-mêmes blocs élémentaires des multiplicateurs et processeurs.

En tant que tel, l'étude de l'inverseur CMOS permet de faire la charnière entre le monde analogique et le monde numérique.

4.1 Architecture et fonction de transfert

4.1.1 Architecture

L'inverseur CMOS¹ est constitué de deux transistors MOS : un NMOS et un PMOS comme le montre le schéma de la figure 1.

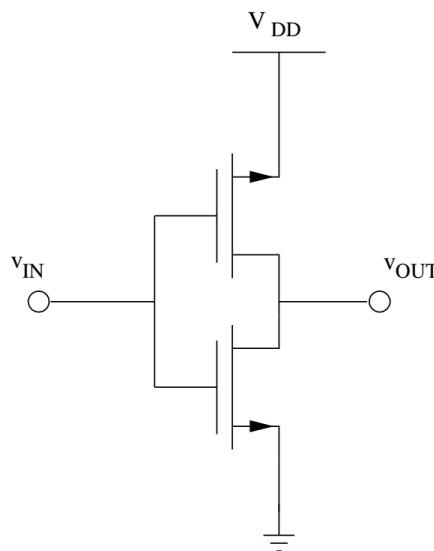


Figure 1: Schéma d'un inverseur CMOS.

Technologiquement, les deux transistors sont réalisés sur un substrat de type P dans

¹ CMOS pour Complementary MOS, acronyme signifiant que l'on utilise des transistors MOS complémentaires dans leur fonctionnement, à savoir des transistors PMOS et des transistors NMOS.

lequel on vient créer un caisson de type N afin d'y implanter le transistor PMOS comme montré figure 2.

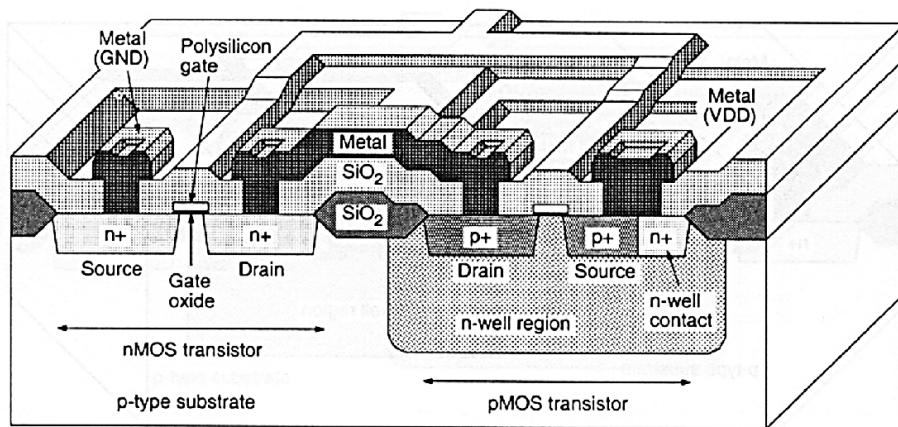
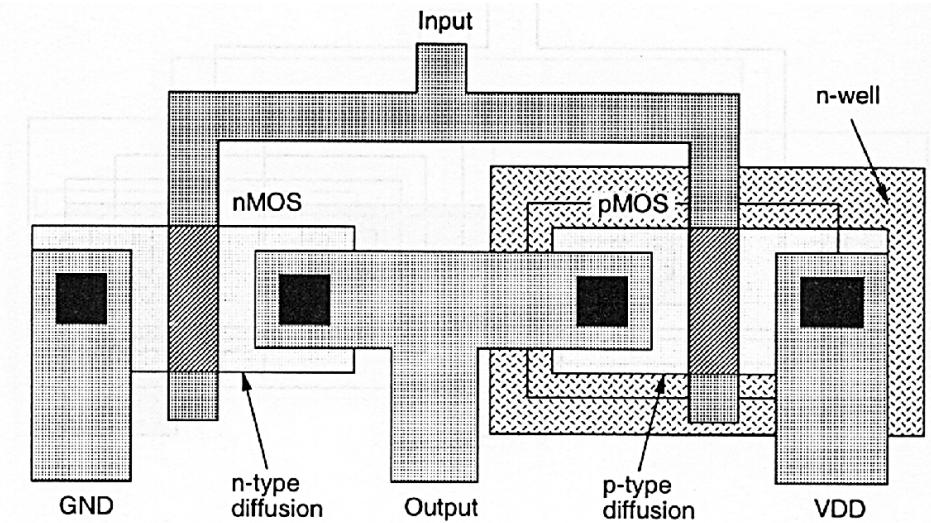


Figure 2: Layout et vue 3D d'un inverseur CMOS d'après *Atlas of IC Technology* de W. Maly.

4.1.2 Fonctionnement statique

Tracé de la courbe de transfert entrée-sortie

Pour obtenir l'évolution de la tension de sortie V_{OUT} en fonction de la tension d'entrée V_{IN} , il suffit d'écrire l'égalité des courants dans les deux transistors. Chaque transistor ayant trois états possibles, cela laisse neuf configurations à étudier pour obtenir le résultat final (sachant qu'un petit nombre de configurations peuvent être impossibles). Le plus simple, une fois de plus, est de tracer les lois d'existence du courant dans chacun des deux dispositifs et de les superposer, les intersections donnant le résultat escompté comme le montre la figure 3.

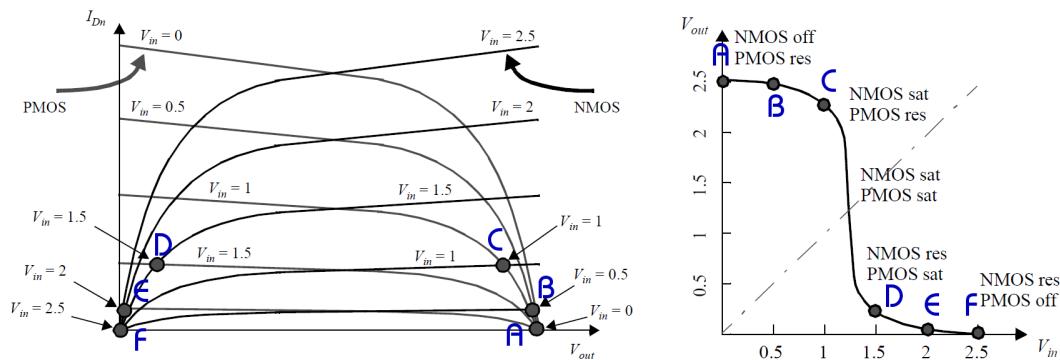


Figure 3: Réseaux de caractéristiques NMOS et PMOS et tracé de la fonction de transfert.

Chaque réseau de caractéristique comprenant 6 courbes (dans l'exemple de la figure 3), on obtient 6 points d'intersection. A chaque point d'intersection correspond un régime de fonctionnement pour le transistor NMOS et pour le transistor PMOS comme récapitulé sur le tableau 1 avec pour rappel $V_{GSn}=V_{IN}$, $V_{DSn}=V_{OUT}$, $V_{SGp}=V_{DD}-V_{IN}$, $V_{SDp}=V_{DD}-V_{OUT}$ et pour une technologie $0.25\mu\text{m}$ des tensions de seuils $V_{Tn}=0.4\text{V}$ et $V_{Tp}=0.55\text{V}$.

Point	Transistor NMOS			Transistor PMOS		
	V_{GSn}	$V_{GSn}=V_{Tn}$	Régime	V_{SGp}	$V_{SGp}- V_{Tp} $	Régime
A	$<V_{Tn}$	-	Bloqué			
B					$> V_{Tp} $	Ohmique
C		$<V_{DSn}$	Saturé			
C vers D						
D					$<V_{DSp}$	Saturé
E			Ohmique			
F				$< V_{Tp} $	-	Bloqué

Tableau 1: Régimes de fonctionnement des transistors dans l'inverseur NMOS

Étude fine des régimes de fonctionnement des transistors MOS

Une fois cette première caractéristique de transfert tracée, on peut procéder à une étude plus fine des différents domaines de fonctionnement des transistors MOS. Pour ce faire, sur la courbe précédente, il suffit de tracer les courbes suivantes :

- Droite $V_{IN}=V_{Tn}$: tant que V_{IN} est inférieure à V_{Tn} , le transistor NMOS est bloqué
- Droite $V_{IN}=V_{DD}-V_{Tp}$: pour toute valeur de V_{IN} supérieure à $V_{DD}-V_{Tp}$, le transistor PMOS est bloqué car $V_{SGp}=(V_{DD}-V_{IN}) < V_{Tp}$
- Droite $V_{OUT}=V_{IN}$ qui va servir de référence pour les deux droites suivantes
- Droite $V_{OUT}=V_{IN}-V_{Tn}$: pour toute valeur de V_{OUT} inférieure à cette droite, le transistor NMOS est en régime ohmique car dès lors $V_{OUT} < V_{IN}-V_{Tn}$ signifie que $V_{DSn}=V_{OUT}$ est inférieure à $V_{GS}-V_{Tn}=V_{IN}-V_{Tn}$

- Droite $V_{OUT}=V_{IN}-V_{TP}$, ce qui revient² à tracer $V_{OUT}=V_{IN}+|V_{TP}|$: pour toute valeur de V_{OUT} supérieure à cette droite, le transistor PMOS est en régime ohmique car dès lors $V_{OUT}>V_{IN}+|V_{TP}|$ signifie que $V_{SDP}=V_{DD}-V_{OUT}$ est inférieure à $V_{SGP}-|V_{TP}|=V_{DD}-V_{IN}-|V_{TP}|$

On obtient ainsi les différents domaines de fonctionnement des transistors MOS comme représentés sur la courbe 4³.

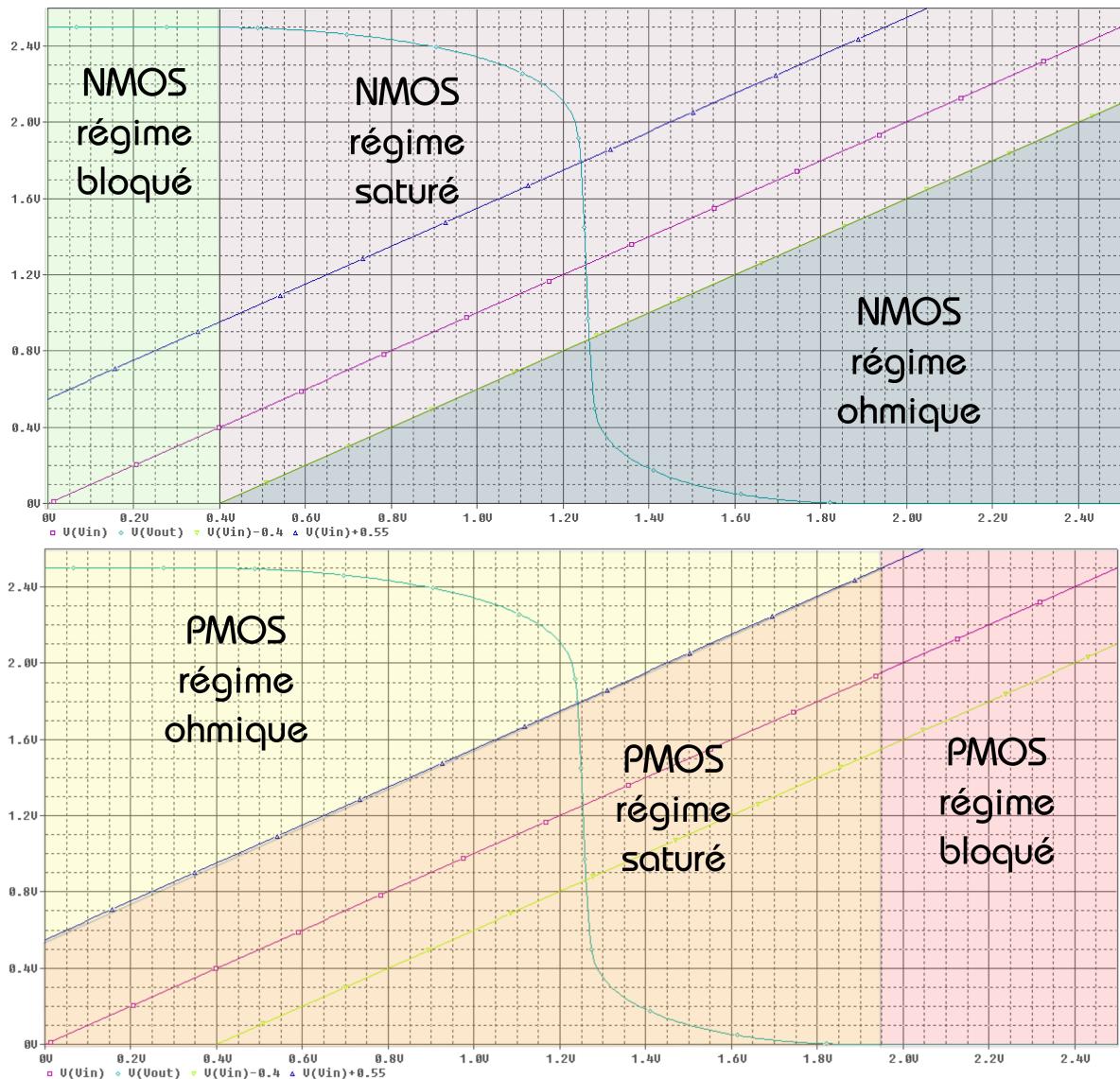


Figure 4: Régimes de fonctionnement des transistors MOS en fonction de la tension d'entrée.

Le point milieu

Définition : On appelle V_M le point tel que $V_{IN}=V_{OUT}$, c'est à dire le point pour lequel la tension de sortie est égale à la tension d'entrée. D'un point de vue mathématique, ce point est un point d'inflexion donc un point pour lequel la dérivée seconde de la fonction $v_{out}=f(v_{in})$

² Pour rappel, V_{TP} est négatif.

³ Ce résultat issu d'une simulation PSpice utilise des transistors NMOS et PMOS de rapport W/L respectifs de $1\mu\text{m}/0.25\mu\text{m}$ et $2.47\mu\text{m}/0.25\mu\text{m}$.

s'annule. Ce point est également le point auquel les deux transistors sont en régime saturé et donc présente un maximum d'amplification. Il peut être intéressant, selon les applications de faire varier la valeur de ce point : elle est fixée par le rapport des dimensions géométriques des transistors NMOS et PMOS.

Expression analytique : Soit V_M le point tel que $V_{IN}=V_{OUT}$. En ce point, les deux transistors sont en régime saturé et il y a égalité entre le courant sortant du transistor NMOS et celui rentrant dans le transistor PMOS soit :

$$\begin{aligned} I_{DSn} &= \frac{K_n}{2} (V_{GSn} - V_{Tn})^2 \cdot (1 + \lambda \cdot V_{DSn}) = I_{DSP} = \frac{K_p}{2} (V_{SGp} - |V_{Tp}|)^2 \cdot (1 + \lambda \cdot V_{SDp}) \\ \Rightarrow K_n (V_{IN} - V_{Tn})^2 \cdot (1 + \lambda \cdot V_{OUT}) &= K_p (V_{DD} - V_{IN} - |V_{Tp}|)^2 \cdot (1 + \lambda \cdot (V_{DD} - V_{OUT})) \\ \Rightarrow K_n (V_M - V_{Tn})^2 \cdot (1 + \lambda \cdot V_M) &= K_p (V_{DD} - V_M - |V_{Tp}|)^2 \cdot (1 + \lambda \cdot (V_{DD} - V_M)) \end{aligned}$$

On obtient une équation du troisième ordre en V_M . Il est donc difficile d'en tirer une solution analytique.

On peut trouver une solution particulière du système, solution couramment utilisée par ailleurs en électronique numérique. L'idée est de chercher la valeur de la tension de sortie pour une tension d'entrée $V_{IN}=V_{DD}/2$.

Il suffit de reprendre les équations précédentes avant la substitution par V_M et de remplacer V_{IN} par $V_{DD}/2$. Cela donne le résultat suivant :

$$K_n \left(\frac{V_{DD}}{2} - V_{Tn} \right)^2 \cdot (1 + \lambda \cdot V_{OUT}) = K_p \left(V_{DD} - \frac{V_{DD}}{2} - |V_{Tp}| \right)^2 \cdot (1 + \lambda \cdot (V_{DD} - V_{OUT})) \quad (4.1)$$

Maintenant si on souhaite obtenir⁴ une tension $V_{OUT}=V_{DD}/2$, on remplace V_{OUT} dans la formule précédente et on obtient la relation suivante (approximation supposant une égalité sur les tensions de seuil $V_{Tn} \approx |V_{Tp}|$) :

$$V_M = \frac{V_{DD}}{2} \Rightarrow \frac{K_n}{K_p} = 1$$

Cela traduit simplement que pour $V_{OUT}=V_{IN}=V_{DD}/2$, il faut que les deux transistors aient un comportement identique voire symétrique, ce qui veut dire que le PMOS doit compenser la faiblesse relative de ses porteurs par une largeur de canal accrue, soit

$$\frac{W_n/L_n}{W_p/L_p} = \frac{\mu_p}{\mu_n}$$

4.1.3 Utilisation d'un inverseur CMOS

L'inverseur CMOS peut être utilisé dans deux zones de fonctionnement distinctes représentées figure 5 :

- Une zone où les deux transistors sont saturés et permettent d'obtenir une amplification en tension : on étudiera principalement cette partie dans les paragraphes suivants.
- Une zone séparée en deux parties où l'inverseur fonctionne en "Tout ou Rien" ; à savoir

⁴ Ce qui revient *in fine* à chercher à obtenir $V_M=V_{DD}/2$.

que soit le transistor PMOS est bloqué et se comporte comme un interrupteur ouvert tandis que le transistor NMOS est passant et agit comme une résistance, soit le contraire (NMOS bloqué et PMOS résistif). Ce mode de fonctionnement qui permet de relier la sortie de l'inverseur CMOS à 0V ou à V_{DD} selon si la tension d'entrée vaut V_{DD} ou 0V est utilisé en électronique numérique pour réaliser la fonction très utilisée d'inversion binaire. On étudiera cette zone un peu plus tard car elle permet de mettre en évidence les principes de consommation, de fiabilité et de robustesse inhérents à tout système numérique, aussi complexe soit-il.

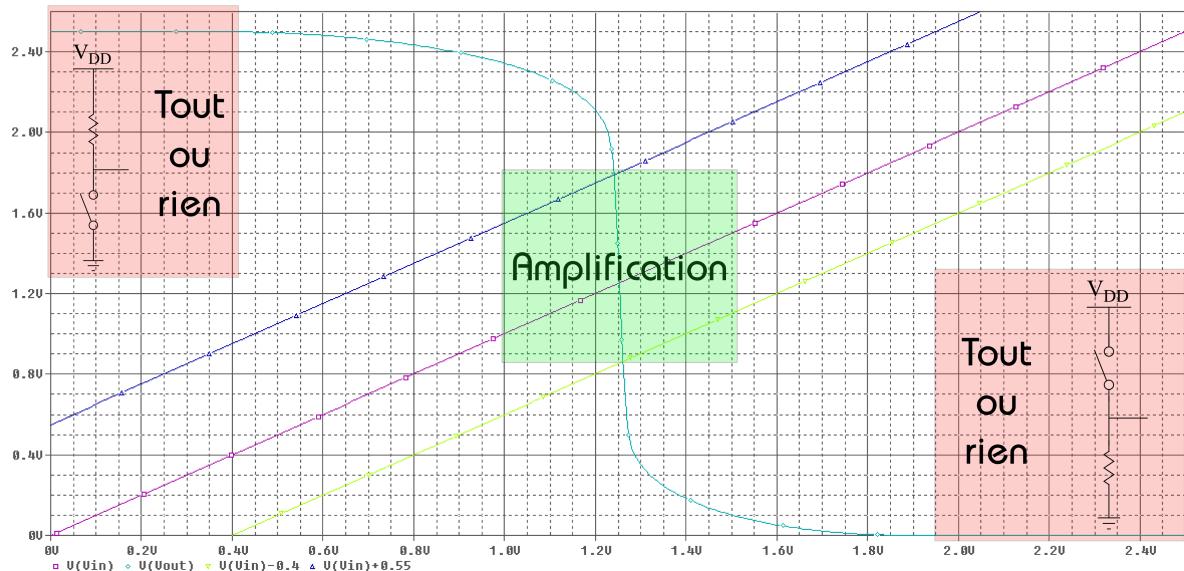


Figure 5: L'inverseur CMOS : deux zones de travail.

4.2 L'inverseur CMOS : étude dynamique petit-signal

La présente partie est consacrée à l'étude des performances de l'inverseur CMOS en amplification. Elle va permettre de mettre en pratique les notions de petit signal abordées dans les chapitres 2 et 3 (jonction PN et transistor MOS) et de procéder à l'étude d'un système à deux transistors complémentaires. A l'issue de cette partie, vous serez à même d'étudier tout système constitué d'un ou plusieurs transistors.

4.2.1 Modèle petit-signal de l'inverseur CMOS

Il a été établi précédemment que l'inverseur CMOS devait être polarisé dans la région où les deux transistors NMOS et PMOS sont saturés. L'étude dynamique sera donc faite dans cette région.

Première étape : remplacement de chaque élément par son équivalent petit-signal

Pour établir le schéma petit-signal de l'inverseur CMOS, on commence par remplacer chacun des transistors par son modèle petit-signal équivalent comme représenté figure 6. Toutes les sources de tensions fixes du montage sont remplacées par leurs équivalents petit signal à savoir un potentiel nul (ce sont des sources de tension constante, donc elles ne présentent pas de variations en temps) et les sources de courant fixes sont remplacées par un interrupteur ouvert, leur équivalent petit-signal.

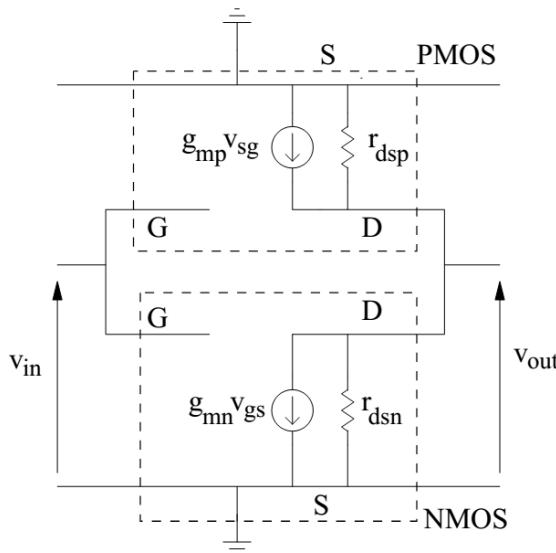


Figure 6: L'inverseur CMOS : première étape du schéma petit signal.

Seconde étape : réorganisation optimale du schéma

Une fois le schéma petit signal établi, il est nécessaire de le réorganiser de manière à le rendre plus lisible. La première chose à faire est de "replier" le circuit sur lui-même de manière à superposer les potentiels nuls. Ensuite, on identifie les éléments que l'on peut superposer ou cascader et donc remplacer par des équivalents (impédances séries ou parallèles, sources de courant ou de tension équivalentes). Ainsi pour notre inverseur CMOS, on obtient le schéma équivalent de la figure 7.

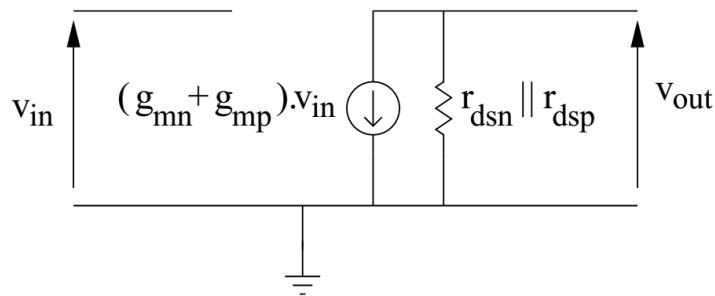


Figure 7: L'inverseur CMOS : schéma petit signal final.

Dernière étape : calcul du modèle quadripolaire équivalent

Établir le modèle quadripolaire de l'inverseur CMOS consiste tout d'abord à identifier le genre d'amplificateur auquel nous avons affaire. Ici, le signal à amplifier est un signal de tension et le signal délivré en sortie est également un signal de tension.

Nous avons donc à calculer le modèle quadripolaire d'un amplificateur de tension commandé par une tension. Il faut donc en calculer l'impédance d'entrée, l'impédance de sortie et le gain en tension de l'inverseur CMOS.

Les calculs sont assez rapides⁵ :

- L'impédance d'entrée du montage est infinie,
- L'impédance de sortie est obtenue en mettant l'entrée à 0V et en se plaçant en sortie pour mesurer l'impédance de sortie. Il est évident que l'impédance de sortie est égale à $r_{dsn} \parallel r_{dsp}$.
- Enfin, le gain en tension A_v consiste à calculer la tension de sortie pour une tension petit-signal v_{in} appliquée en entrée. La tension de sortie vaut, de manière immédiate

$$v_{out} = v_{in} \cdot (g_{mn} + g_{mp}) \cdot (r_{dsn} \parallel r_{dsp})$$

$$\text{soit } A_v = \frac{v_{out}}{v_{in}} = \frac{g_{mn} + g_{mp}}{g_{dsn} + g_{dsp}}.$$

Le modèle quadripolaire du montage de l'inverseur CMOS est donc celui représenté figure 8.

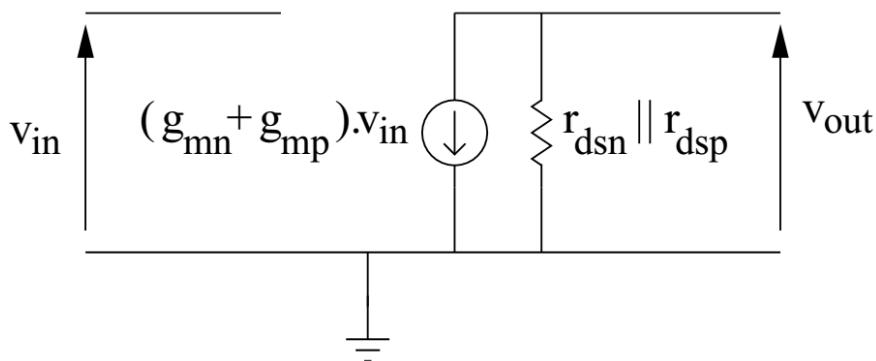


Figure 8: L'inverseur CMOS : modèle quadripolaire.

Remarque sur le point de fonctionnement et son optimisation

Le schéma petit-signal a permis d'établir que le gain est directement lié aux transconductances des transistors MOS. Si l'on veut avoir le plus grand gain possible (en supposant que les admittances de sortie des transistors MOS restent constantes sur la plage de tension considérée), il suffit de maximiser les valeurs des transconductances. Cela revient à étudier le maximum atteint par la courbe $\delta v_{out}/\delta v_{in}$ soit tout simplement dériver la fonction de transfert entrée-sortie de l'inverseur et de regarder pour quelle tension d'entrée le maximum est atteint. Le tracé de la courbe figure 9 représente cette dérivée : on retrouve que le meilleur point de polarisation est atteint pour $V_{IN}=V_{DD}/2$, valeur logique du fait des dimensionnements des transistors choisis tels que $\frac{W_n/L_n}{W_p/L_p} = \frac{\mu_p}{\mu_n}$.

⁵ Un lecteur observateur aura remarqué que le schéma petit signal auquel nous nous sommes ramenés ressemble étrangement à celui de la source commune et donc que l'on peut appliquer directement les résultats de ce dernier.

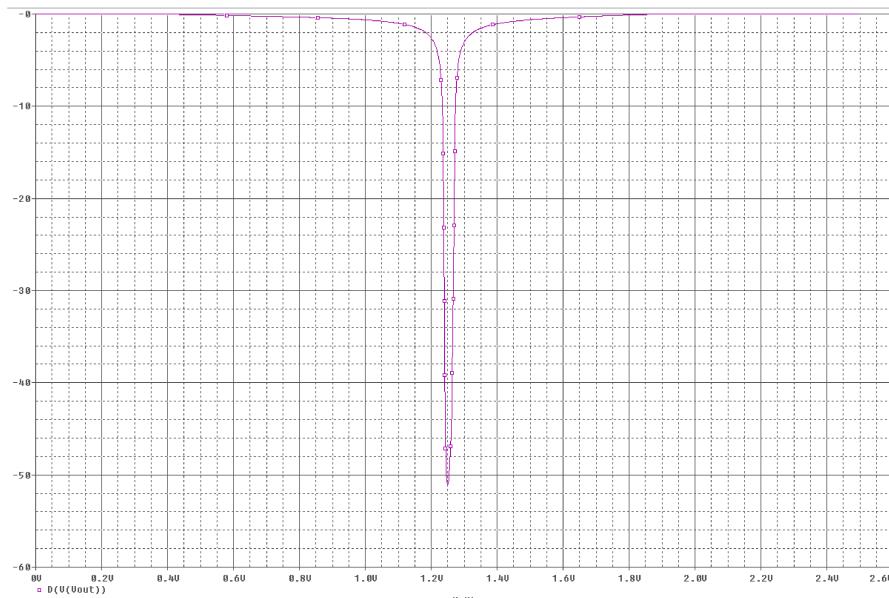


Figure 9: Optimisation du gain de l'inverseur CMOS : dérivée de la fonction de transfert entrée - sortie.

4.3 L'inverseur CMOS numérique

4.3.1 Introduction

Remarque préliminaire :

Pour aborder ce chapitre, il est nécessaire de maîtriser les bases de l'électronique numérique et donc d'avoir lu les rappels d'électronique numérique élémentaire.

Comme nous l'avons vu précédemment, l'inverseur CMOS peut être utilisé en tout ou rien, auquel cas il réalise la fonction numérique d'inverseur. L'inverseur CMOS représente **LE** circuit élémentaire par excellence. Une fois son fonctionnement et ses propriétés comprises, la conception de circuits plus complexes (de la porte NON-ET au microprocesseur) s'en trouve grandement simplifiée. En effet, le comportement électrique de ces circuits numériques complexes peut être complètement extrapolé à partir des résultats obtenus pour la cellule de l'inverseur. L'analyse de la porte inverseuse CMOS se fera au regard des métriques suivantes :

- le coût exprimé sous la forme de la complexité du circuit et de la surface occupée
- l'intégrité et la robustesse qui découlent du comportement statique de l'inverseur
- la performance déterminée par le comportement dynamique de l'inverseur CMOS
- l'efficacité énergétique déterminée par l'énergie et la puissance consommées.

De cette analyse, on extraira les paramètres clés sur lesquels travailler afin d'améliorer les facteurs ci-dessus répertoriés.

4.3.2 Notions préliminaires : robustesse et performances

Dans les paragraphes suivants, nous allons étudier les critères qui permettent de qualifier le comportement d'un système numérique et d'établir son aptitude à fonctionner en présence de perturbations ainsi que ses performances au travers des critères que nous

établirons.

4.3.3 Fonctionnalité

La fonctionnalité d'un circuit est l'aptitude du circuit à réaliser sa fonction logique. Elle peut être mise en défaut par deux catégories de facteurs : d'une part, des facteurs technologiques qui surviennent lors de la fabrication et introduisent des défauts permanents dans le système ; et d'autre part, des facteurs environnementaux (perturbations extérieures, bruit dans les alimentations, parasitage des signaux d'entrées) qui induisent des défauts temporaires dans le système. La capacité qu'aura le circuit à continuer à assurer sa fonctionnalité en présence de ces défauts est ce que l'on appelle la robustesse.

Défauts technologiques

Survenant lors de la fabrication (incertitude sur les dimensions des transistors, sur les dopages, ...), ils ont un impact direct sur les tensions des seuil des transistors ainsi que sur les valeurs nominales des courants. Les modifications de ces valeurs peuvent amener une dérive du comportement du circuit par rapport au comportement théorique. Il y a alors deux cas de figure :

- ces dérives sont mineures et ne mettent pas en défaut la fonctionnalité du circuit mais ont une influence (positive ou négative) sur les performances de ce dernier
- ou alors ces dérives sont telles que le circuit n'est plus fonctionnel et est dès lors rejeté au rebut lors des tests de fonctionnalité en sortie d'usine.

Le parfait exemple de cette dégradation des performances du fait des défauts technologiques est celui des microprocesseurs fabriqués par Intel ou AMD. A la sortie de l'usine, une batterie de tests est effectuée sur les microprocesseurs, et selon leurs réussites au test, ils sont vendus avec l'affichage d'une fréquence plus ou moins haute⁶. Du fait des dispersions technologiques, certains circuits sont moins stables à la fréquence nominale que d'autres et sont donc vendus à une fréquence inférieure : cela ne veut pas dire qu'ils ne peuvent pas marcher à la fréquence nominale, juste qu'ils n'ont pas réussi la totalité des tests assurant la fonctionnalité de manière robuste. C'est pour cela qu'il est plus facile d'"overclocker"⁷ certaines familles de processeur.

Bruit intrinsèque et extrinsèque

On appelle bruit toute variation parasite (en tension ou en courant) qui vient se superposer au signal utile (alimentation, données, ...). Le bruit peut être ramené au niveau du circuit par interférence avec les signaux extérieurs au circuit (couplage capacitif ou inductif, ou alimentation non stabilisée comme montrée figure 10) ou être généré de manière interne par le circuit lui-même.

⁶ Les microprocesseurs sont ainsi fabriqués dans une certaine technologie MOS et leur fréquence n'est déterminée qu'à l'issue des tests en sortie de chaîne de fabrication.

⁷ Sur-cadencement en français : manipulation qui consiste à augmenter la fréquence de fonctionnement d'un processeur au-delà de la valeur nominale affichée.

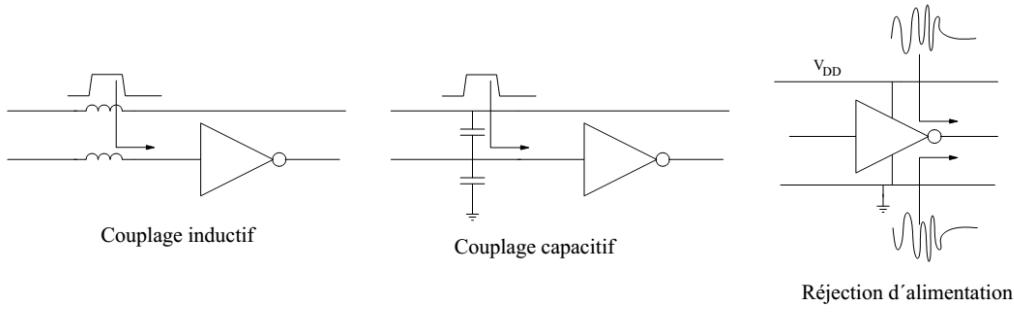


Figure 10: Des exemples de bruit ramenés par couplage.

La figure 11 donne une représentation du phénomène de couplage entre lignes dans les circuits intégrés selon la finesse de la technologie. On considère cinq lignes parallèles et on envoie un échelon dans la ligne centrale ; on mesure le niveau de tension dans les autres lignes et on peut percevoir une impulsion d'autant plus grande que la technologie est fine. C'est donc un phénomène qui prend d'autant plus d'importance que les noeuds technologiques décroissent.

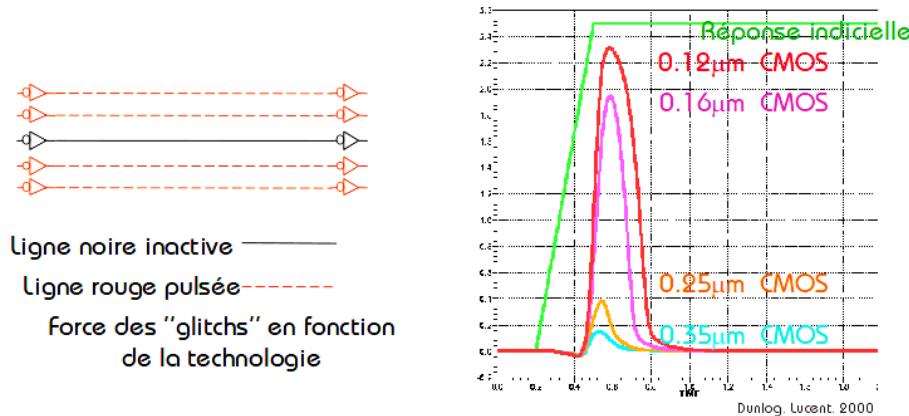


Figure 11: Influence des noeuds technologiques sur le couplage.

Cependant la majeure partie du bruit dans un circuit est celui qu'il génère lui-même. Il y a plusieurs types de bruit : bruit thermique, bruit de grenaille, bruit en $1/f$, ... La notion de bruit étant assez complexe, il suffira de retenir, pour la notion de bruit interne, l'exemple du bruit thermique dans une résistance. En effet, lorsqu'un courant traverse une résistance, cette dernière va dissiper une puissance et donc de la chaleur. La température de la résistance augmentant, un certain nombre d'électrons vont être arrachés à la bande de valence du fait de l'agitation thermique : ces électrons vont être générateurs d'un signal parasite. On a là un parfait exemple de bruit interne généré par le circuit lui-même (d'où l'un des intérêts de réduire la puissance dissipée dans un circuit électronique).

4.3.4 Robustesse

La robustesse est caractérisée par le comportement du circuit vis à vis des dispersions technologiques et des perturbations dues au bruit. Si, comme on l'a vu précédemment, les dispersions technologiques ont un impact définitif sur la fonctionnalité et les performances d'un circuit, tout concepteur peut lors de la conception d'un circuit

mettre en œuvre un certain nombre de méthodes de manière à diminuer l'impact de ces dispersions sur le circuit⁸.

Définition des tensions d'opérations

Niveaux logiques de sortie

La première tâche est d'étudier la fonction de transfert entrée-sortie du montage afin de définir les valeurs de tension que le circuit délivre correspondant au **zéro** logique et au **un** logique. Dans le cas idéal, ces tensions correspondent aux valeurs des tensions d'alimentation hautes et basses du circuit. Ces deux tensions sont nommées V_{OL} et V_{OH} , dénominations représentant respectivement la valeur de tension de sortie associée au **zéro** logique (V_{OL} Voltage Output Low) et la valeur de tension de sortie associée au **un** logique (V_{OH} Voltage Output High).

Niveaux logiques d'entrée

Une fois les deux tensions V_{OL} et V_{OH} déterminées, il faut définir les tensions admissibles en entrée, à savoir les niveaux de tension analogiques qui seront considérés comme étant des **zéros** ou des **un** logiques. La manière mathématique de procéder est la suivante :

1. soit la fonction de transfert *toto* telle que $V_{sortie} = \text{toto}(V_{entrée})$
2. identification sur le tracé de $V_{sortie} = \text{toto}(V_{entrée})$ des tensions V_{OL} et V_{OH}
3. on sait que les valeurs $\text{toto}^{-1}(V_{OL})$ et $\text{toto}^{-1}(V_{OH})$ sont des tensions que le circuit considère comme des **zéros** ou des **un** logiques.
4. on cherche les valeurs pour lesquelles la dérivée de la fonction *toto* devient unitaire : ces valeurs correspondent aux limites des tensions en entrée au-delà desquelles le signal entrant n'est plus considéré comme un signal numérique. Ces deux tensions sont notées V_{IL} et V_{IH} pour Voltage Input Low et Voltage Input High.

Pour éclaircir cette démarche très théorique, prenons l'exemple d'un inverseur dont la fonction de transfert est représentée figure 12. Après avoir tracé les dérivées unitaires, on obtient deux zones qui sont celles pour lequel le signal en entrée est identifié de manière numérique (c'est à dire comme étant un **un** ou **zéro** logique). La zone entre les deux est appelée *zone d'indétermination* : tout signal analogique délivré en entrée de la porte et appartenant à cette zone donnera un signal de sortie pouvant être interprété par l'étage suivant indifféremment comme **un** ou un **zéro** logique de manière non-déterministe.

⁸ Une des méthodes fréquemment utilisée consiste à utiliser plusieurs modèles de transistors correspondant à différents niveaux de réussite ou d'échec de la technologie et de procéder à des analyses Monte Carlo pour étudier l'impact des dispersions technologiques sur le circuit.

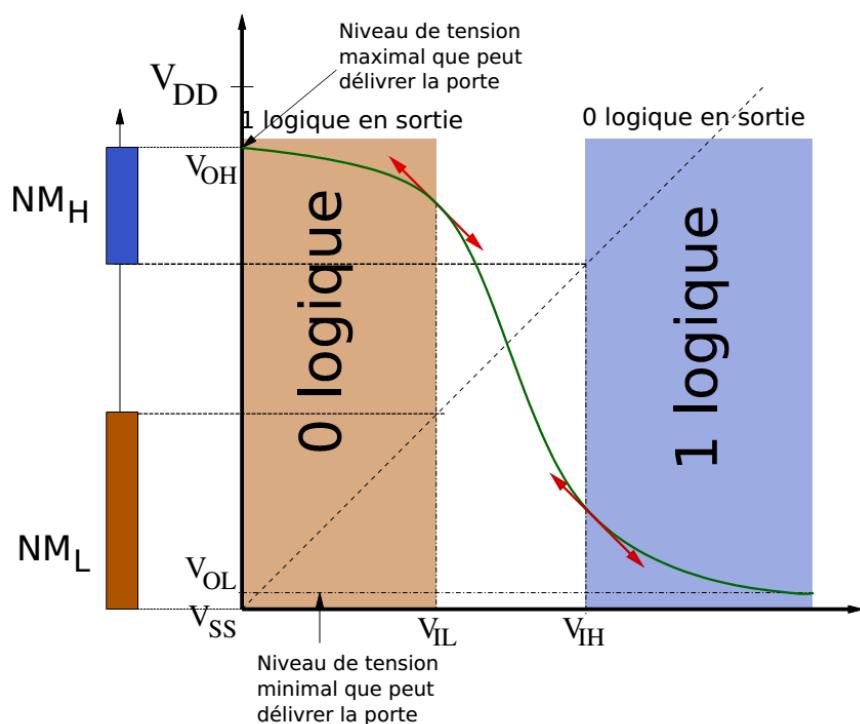


Figure 12: Niveaux de tensions et marges de bruit.

Si de plus, en sortie de cette porte on connecte la même porte, on obtient les marges de bruit.

Marges de bruit

La marge de bruit est la différence entre la valeur analogique du signal envoyé en entrée et la valeur limite pour laquelle ce signal est interprétable. En clair, pour un **zéro** logique, la marge de bruit est définie comme étant $NM_L = V_{IL} - V_{OL}$: c'est la différence entre la valeur de tension maximale que la porte considère comme étant un **zéro** logique soit V_{IL} et la valeur nominale du **zéro** logique que délivre la porte soit V_{OL} . Cette différence ou marge de bruit représente la quantité de bruit qui peut se superposer au signal d'entrée sans en affecter l'interprétation comme un **zéro** logique.

De même, la marge de bruit associée au **un** logique est définie comme étant la différence entre la valeur nominale du **un** logique que délivre la porte soit V_{OH} et la valeur minimale de tension que cette porte admet comme étant un **un** logique soit V_{OL} . Cela donne $NM_H = V_{OH} - V_{IH}$.

Ces deux marges de bruit sont représentées de manière graphique sur la barre verticale dans la partie gauche de la figure 12. On remarquera d'ailleurs que sur l'exemple choisi, la robustesse de la porte pour le **zéro** logique est supérieure à celle du **un** logique. Dans une utilisation optimale de la porte, on s'arrangerait donc, si nécessité forte d'une robustesse accrue, pour utiliser un code employant majoritairement les **zéros** logiques.

Directivité et entrée/sortance

La directivité : Elle est directement relié à l'une des bases de conception des circuits électroniques, à savoir l'indépendance entrée/sortie : un circuit électronique est d'autant plus directif que les circuits présents en sortie n'influencent pas le comportement du

système et ne provoquent de retour sur l'entrée (cela revient à obtenir un h_{12} de la matrice hybride du système nul).

Entrance : L'entrée (ou *fan-in* en anglais) est le nombre d'entrées indépendantes que supporte le circuit : c'est bien évidemment directement relié à la structure physique du circuit. Cela joue sur la capacité électrique équivalente que le circuit présente comme impédance d'entrée et également sur la complexité du circuit et donc sa rapidité. Par exemple, l'entrée d'un additionneur complet (circuit étudié plus loin) est trois (le circuit a 3 entrées A, B et C_{in}).

Sortance : La sortance (ou *fan-out* en anglais) représente le nombre maximal de portes qui peuvent être connectées en sortie d'un circuit sans dégradation de sa fonctionnalité et de ses performances. C'est plutôt un paramètre affectant le fonctionnement du circuit dans son aspect non-linéaire. Comme nous le verrons plus loin, chaque commutation $0 \rightarrow 1$ ou $1 \rightarrow 0$ crée une charge ou décharge d'une capacité. Chaque porte absorbe ainsi une certaine quantité de courant. La sortance d'un circuit représente donc le courant maximal total qu'il est capable de délivrer et donc le nombre de portes qu'il peut piloter. La sortance est un phénomène non-linéaire dans le sens où le retard créé par la sortance n'est pas un phénomène de type *RC* (charge-décharge classique d'une capacité) comme le montre la figure 13.

Supposons l'inverseur idéal, c'est à dire se comportant comme une source de courant idéale en sortie de valeur de courant I_{out} . La tension aux bornes de la capacité C_{in} de la porte suivante est égale à $V=Q(t)/C_{in}$ avec $Q(t)=I_{out} \times (t-t_0)$. Si l'on met n inverseurs en sortie de l'inverseur, le courant de charge chute de I_{out} à I_{out}/n , ce qui augmente le temps de montée $0 \rightarrow 1$ de manière radicale ! Ce phénomène est rarement pris en compte dans l'étude conventionnelle des circuits linéaires pour lesquels on suppose que le courant délivrable en sortie est infini (et donc chargement instantané des capacités en l'absence de résistances ...) : il est à rapprocher de la notion de *slew rate* des amplificateurs opérationnels.

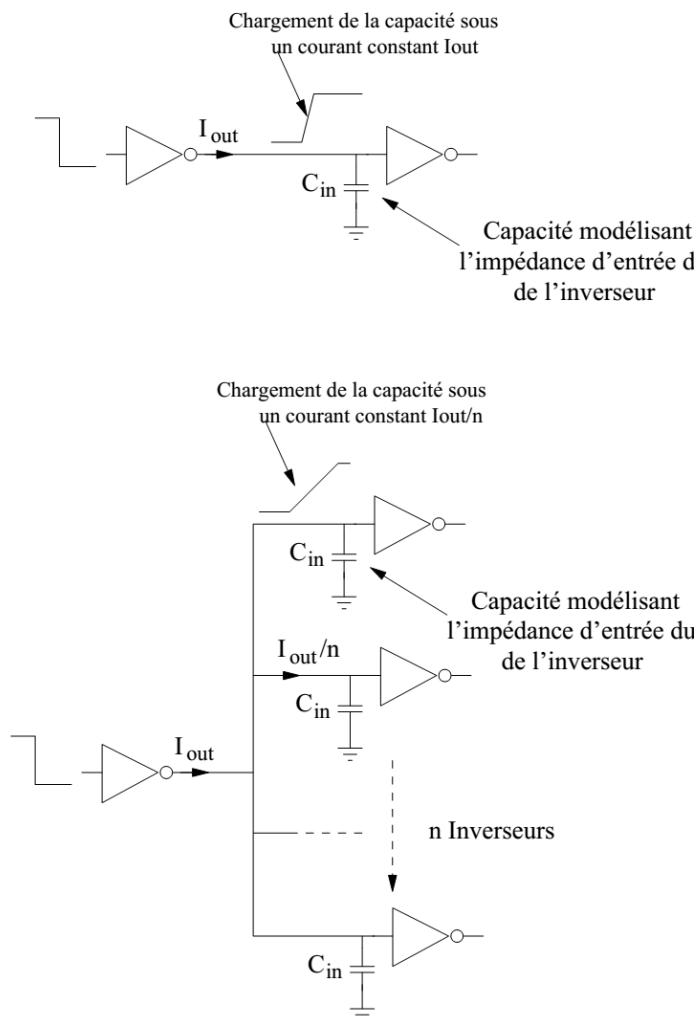


Figure 13: Sortance : un phénomène non-linéaire.

4.3.5 Performances

La notion de performance est rattachée à l'aspect dynamique du système et se décline principalement en deux critères principaux qui sont les délais au travers de la porte (qui vont fixer la vitesse à laquelle on peut demander à la porte de travailler) et la consommation en courant de la porte.

Cadence

Il est courant de réduire la performance d'un circuit à sa charge calculatoire, à savoir le *nombre d'opérations par seconde* qu'il est capable d'effectuer. Le critère principal qui affecte ce critère est le temps de propagation qui représente le temps que met l'information à se propager de l'entrée de la porte à l'entrée de la porte suivante. Ce temps se découpe en deux parties comme le montre la figure 14.

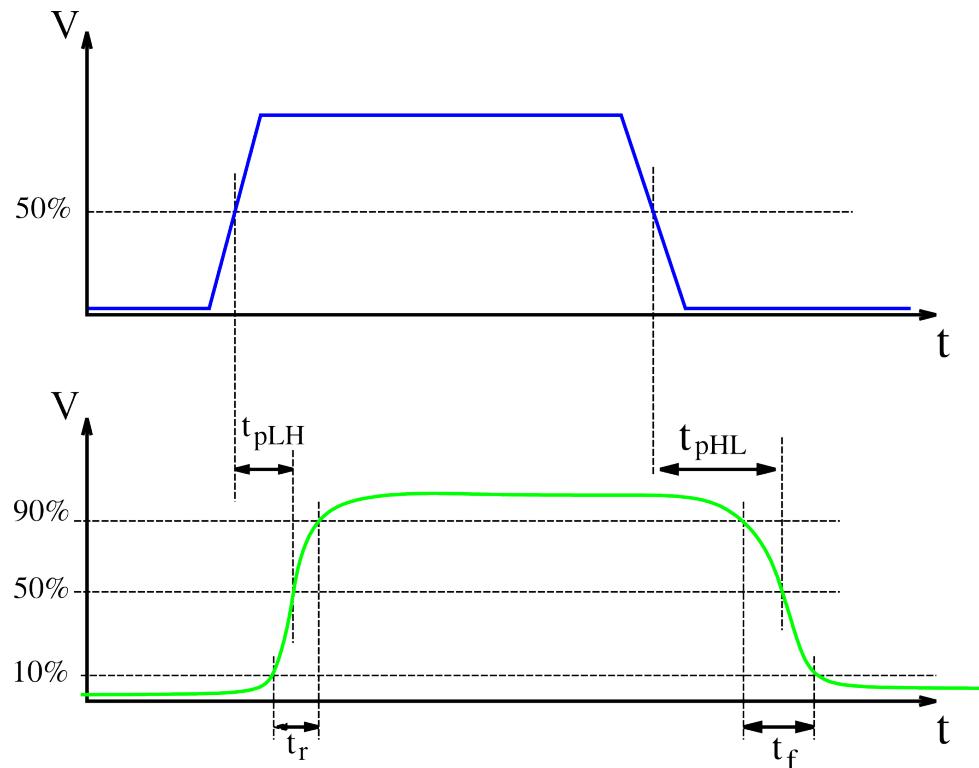


Figure 14: Temps de propagation.

Le temps de propagation se partage en deux catégories :

- le temps de propagation proprement dit, qui représente le temps que met l'information à se propager au travers de la porte (la référence étant 50% de la valeur maximale de l'information). Ce temps est une qualité intrinsèque de la porte en elle-même. Ces temps sont nommés t_{pHL} et t_{pLH} pour séparer les temps de propagation concernant les transitions 0→1 (ou LH pour Low-to-High) à la sortie d'une part, et les transitions 1→0 (ou HL pour High-to-Low) à la sortie d'autre part.
- les temps de montée et de descente qui représentent le temps que met l'information à se charger en entrée ou en sortie de la porte. Ce sont des temps de transition qui ne dépendent pas seulement de la porte en elle-même mais surtout de son environnement et des différentes capacités à charger en dehors du circuit. Ces temps sont dénommés t_r et t_f pour le temps de montée (rise) et le temps de descente (fall).

De manière générale on utilise la moyenne des deux temps de propagation pour caractériser un circuit :

$$t_p = \frac{t_{pHL} + t_{pLH}}{2}$$

Consommation

La consommation est un critère de performance puisqu'elle représente le courant et la puissance consommés par le circuit. Les circuits étant généralement alimentés en tension fixe, il suffit de déterminer le courant pour établir la consommation. Un circuit qui requiert beaucoup de courant aura une dissipation thermique élevée et donc une durée de vie amoindrie du fait du vieillissement accéléré des composants, un comportement plus bruité et plus bruyant, une fiabilité incertaine et surtout, dans un contexte de nomadisme

croissant, une durée de vie sur batterie réduite. Ce sont ces raisons qui rattachent la consommation à la notion de performance.

On étudiera la consommation selon deux approches : une approche où l'on sépare le comportement moyen du circuit du comportement de forte intensité et une seconde où l'on différencie l'aspect statique de l'aspect dynamique, les deux approches étant liées.

Puissance moyenne et puissance de pic

La puissance moyenne est la puissance moyenne d'un circuit calculée sur une certaine période. Son calcul permet de déterminer la taille des batteries pour un fonctionnement en autonomie par exemple. Elle se calcule de la manière suivante :

$$P_{moyenne} = \frac{1}{T} \int_0^T p(t) dt = \frac{V_{alim}}{T} \int_0^T I_{alim}(t) dt$$

La puissance de pic est la puissance maximale que peut afficher le circuit sur une utilisation. Elle correspond donc au courant maximal que peut demander le circuit et donc permet de dimensionner les alimentations du circuit. L'expression mathématique de cette puissance est $P_{pic} = V_{alim} \times i_{pic} = \max[(p(t)]_0^\infty$.

Puissance statique et puissance dynamique

Un circuit numérique a deux régimes de fonctionnement : celui où il maintient une valeur numérique pour ses sorties (état appelé régime statique) et celui où il modifie l'une ou plusieurs valeurs (état appelé régime dynamique).

Ces deux régimes font appel à deux types de courant différents :

- le régime statique n'affichera comme consommation que celle due aux courants statiques qui existent dans les différents chemins de conduction du circuit entre les rails d'alimentation ainsi que les courants de fuite. Bien évidemment, l'objectif de tout concepteur de circuit est de minimiser ces courants,
- le régime dynamique fait apparaître des courants transitoires entre les alimentations et des courants de charge et de décharge de capacités le temps du changement des états.

Énergie par commutation

Souvent les contraintes sur les délais et celles sur la puissance consommée sont antinomiques comme le montre la figure 15. Il existe un point optimal qui combine délais courts et consommation pas trop importante : c'est ce point que l'on cherchera à obtenir de manière générale (sont mises à part les applications hautes performances où l'énergie n'est pas un problème, et les applications ultra-nomades où la puissance calculatoire est sacrifiée). On a recours dès lors à la notion d'énergie par commutation afin de pouvoir caractériser de manière efficace les circuits.

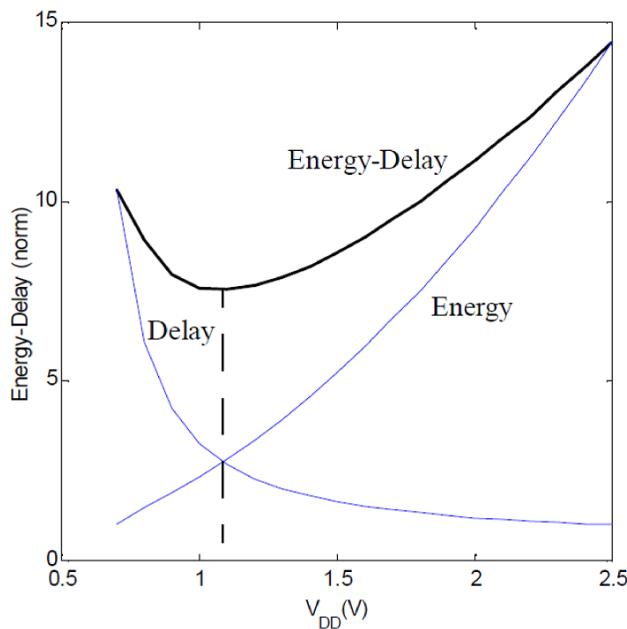


Figure 15: Technologie 0.25 microns : énergie et délais.

4.3.6 Étude au premier ordre

Le tracé de la fonction de transfert de l'inverseur donne le résultat déjà étudié de la figure 16. On y retrouve les deux zones de fonctionnement avec celle qui nous intéresse plus particulièrement dénommée "Tout Ou Rien" pour lequel on adopte le modèle de l'interrupteur précédemment développé (§4.1.3).

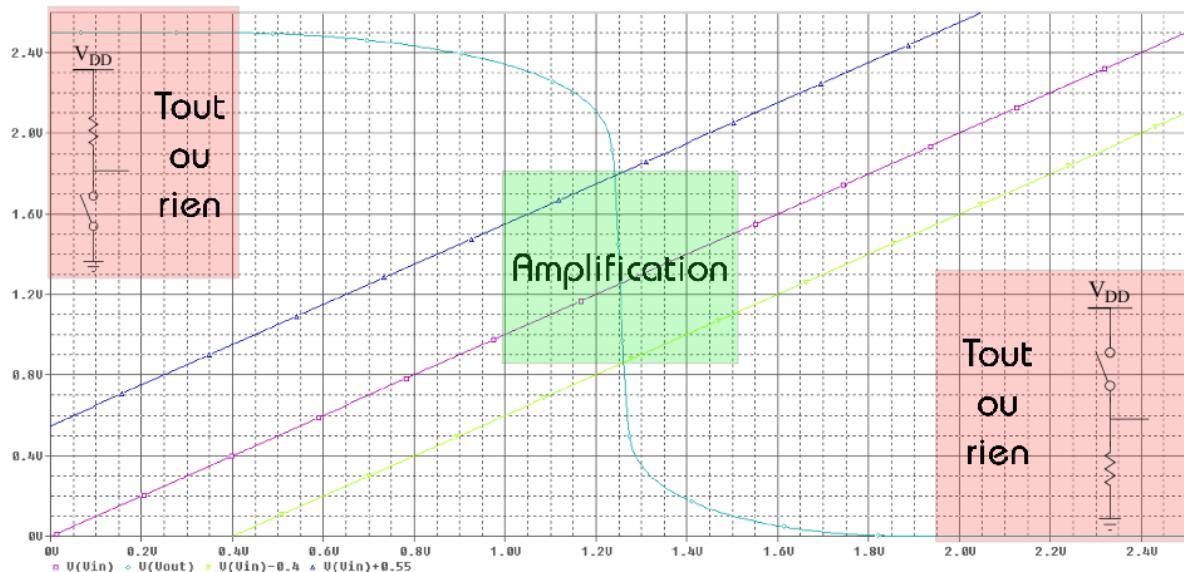


Figure 16: L'inverseur CMOS : deux zones de travail.

On peut déduire un grand nombre de propriétés de l'inverseur CMOS de ce simple modèle interrupteur :

- Les niveaux de tension le plus bas V_{OL} et le plus haut V_{OH} qui seront associés à un **zero**

logique ou à un **un logique** sont respectivement égaux à $0V$ et V_{DD} soit les deux tensions d'alimentation. L'inverseur CMOS est donc un dispositif qui fournit deux niveaux logiques égaux à ses tensions d'alimentation (donc aucune perte dans le circuit). Par ailleurs, ce résultat est indépendant des dimensions du transistor ("ratioless logic" en anglais).

- En état stationnaire établi, il existe toujours un chemin électrique avec une résistance finie entre la sortie et l'un des deux rails d'alimentation. Ainsi, un dimensionnement judicieux de l'inverseur CMOS permet d'obtenir une résistance de sortie faible, le rendant ainsi moins sensible au bruit et aux perturbations. Des valeurs typiques de résistance de sortie sont de l'ordre du $k\Omega$.
- Les niveaux de tension en sortie étant fournis à partir des rails d'alimentation, le courant délivré en sortie de l'inverseur CMOS est directement généré à partir de l'alimentation. Ainsi, la sortance de l'inverseur CMOS ("fan-out") est théoriquement infinie puisque l'inverseur CMOS peut piloter une infinité de portes connectées à sa sortie. En pratique, l'augmentation du temps de propagation avec la sortance limite le nombre d'éléments que l'on peut connecter en sortie.
- La résistance d'entrée de l'inverseur CMOS est très élevée (théoriquement infinie du fait de l'oxyde de grille) et donc le courant d'entrée est virtuellement nul. On peut donc connecter un grand nombre d'inverseurs CMOS en sortie de tout circuit numérique sans affecter sa sortance.
- En fonctionnement "Tout Ou Rien" il n'existe pas de chemin direct entre les deux rails d'alimentation, ce qui veut dire qu'en régime établi (niveau de sortie indifféremment haut ou bas) la porte ne consomme pas d'énergie statique (si l'on suppose les courants de fuite dans les transistors nuls).
- En supposant que la commutation des transistors soit instantanée, le fonctionnement dynamique de la porte inverseuse peut se résumer à la figure 17.

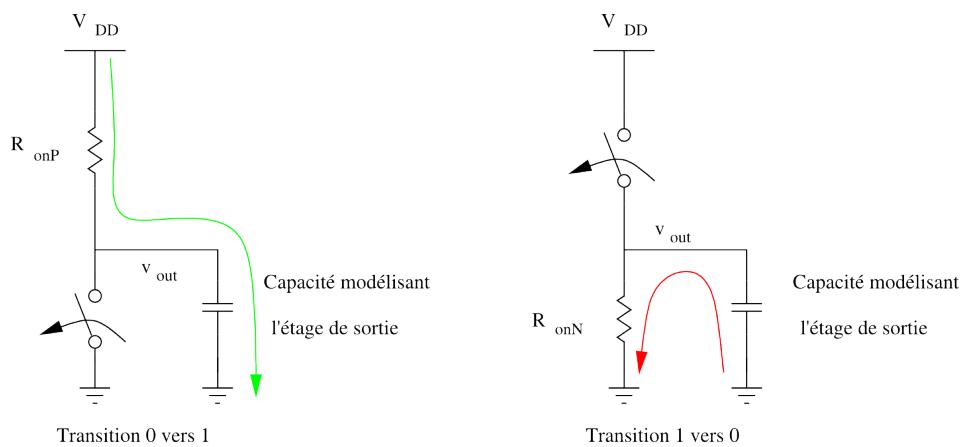


Figure 17: L'inverseur CMOS : sa dynamique au premier ordre .

Ainsi les changements d'état $0 \rightarrow 1$ et $1 \rightarrow 0$ sont respectivement :

- une charge de capacité au travers de la résistance équivalente du transistor P passant, et
 - une décharge de capacité au travers de la résistance équivalente du transistor N passant.
- Les temps de propagation, de montée et de descente de l'inverseur sont donc directement reliés aux capacités d'entrée des inverseurs CMOS et aux résistances équivalentes de leurs transistors. Obtenir une porte rapide signifie présenter une faible capacité d'entrée ou alors de faibles résistances de transistors. Opter pour cette dernière solution se traduit par un

grand rapport W/L des dimensions des transistors (ce qui augmente la capacité d'entrée du transistor et donc nuit à la sortance) et donc un courant de transistor élevé et une consommation élevée⁹.

4.3.7 Étude statique de l'inverseur CMOS numérique

Tensions de références et marges de bruit

L'analyse fine de la robustesse de l'inverseur CMOS passe par l'étude de la fonction de transfert entrée-sortie de l'inverseur déjà réalisée dans la partie 4.1. Le calcul de la fonction de commutation établie par l'équation 4.1 associé au tracé de la dérivée figure 9 pour déterminer les pentes à -1 permet d'établir rapidement les tension d'opération (V_{OL} , V_{OH} , V_{IL} , V_{IH}) et les marges de bruit associées (NM_L et NM_H). Les résultats ainsi obtenus tendent à montrer que l'inverseur CMOS est très proche de la porte idéale car V_{IH} et V_{IL} sont très proches de $V_{DD}/2$, ce qui donne des marges de bruits symétriques proches de $V_{DD}/2$.

Il existe cependant une méthode plus rapide qui donne des résultats de qualité équivalente (surtout appliquée sur les dernières technologies). Il s'agit de prendre la caractéristique entrée-sortie de l'inverseur CMOS, de tracer les deux asymptotes horizontales pour les tensions d'entrées égales à 0V et à V_{DD} , c'est à dire les droites $V=V_{OL}$ et $V=V_{OH}$ et la droite tangente au point de commutation. Cela donne le tracé de la figure 18 :

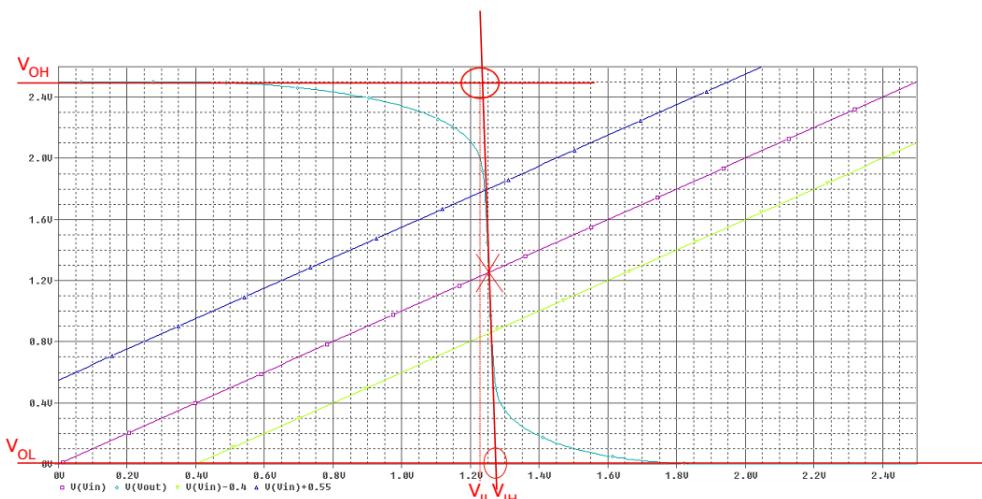


Figure 18: L'inverseur CMOS : détermination de la robustesse par méthode asymptotique.

Les intersections des asymptotes avec la droite de pente de gain donnent les valeurs V_{IL} et V_{IH} (respectivement 1.22V et 1.27V) : les marges de bruit se déduisent facilement. Si l'on compare les valeurs obtenues par cette méthode avec les résultats tirés du tracé de la dérivée de la courbe de transfert (regarder pour quelles valeurs de tension d'entrée on a la dérivée qui vaut -1 sur la figure 9 : $V_{IL}=1.18V$ et $V_{IH}=1.32V$), on s'aperçoit que les erreurs sont minimes.

Là où cette méthode offre un réel avantage par rapport à celle utilisant le tracé de la dérivée, c'est qu'elle ne nécessite pas de connaître le tracé exact de la fonction de transfert. En effet, graphiquement, on s'aperçoit que le calcul de la pente de la tangente à la courbe en V_M est égal à $\frac{V_{OH}-V_{OL}}{V_{IH}-V_{IL}}$ avec, pour l'inverseur CMOS $V_{OL}=0V$ et $V_{OH}=V_{DD}$. La valeur de cette

⁹ On retrouve une fois de plus le compromis vitesse/puissance qui veut que privilégier la vitesse d'un circuit se fasse au détriment de sa consommation.

pente est un gain g que l'on est capable de calculer formellement. En posant que V_M se trouve au milieu de $[V_{IL}, V_{IH}]$, on obtient les relations suivantes :

$$V_{IL} = V_M - \frac{V_{DD}}{2g}$$

$$V_{IH} = V_M + \frac{V_{DD}}{2g}$$

$$NM_L = V_{IL}$$

$$NM_H = V_{DD} - V_{IH}$$

Ces expressions peuvent encore se simplifier si $V_M = V_{DD}/2$. L'énorme avantage de la démarche est que l'on peut très facilement obtenir les critères de robustesse d'une technologie donnée de manière purement analytique, directement à partir des paramètres de la technologie.

Variations technologiques

Les variations survenant lors du process de réalisation des transistors MOS introduisent des écarts dans les valeurs des paramètres géométriques et physiques pour les transistors. Si la plupart des transistors d'une technologie donnée correspondent aux valeurs de référence, un certain nombre sont plus rapides et plus performants que la moyenne et un certain nombre présentent un comportement légèrement dégradé. Ceci va évidemment influer sur la robustesse de l'inverseur réalisé. Une étude approfondie et détaillée de ces modifications sort du cadre de ce cours et ne sera donc pas menée. Cependant, pour donner une exemple des études qu'il faudrait mener, la figure 19 montre de quelle manière la fonction de transfert d'un inverseur peut être affecté par la qualité de réalisation du transistor NMOS et du transistor PMOS la constituant.

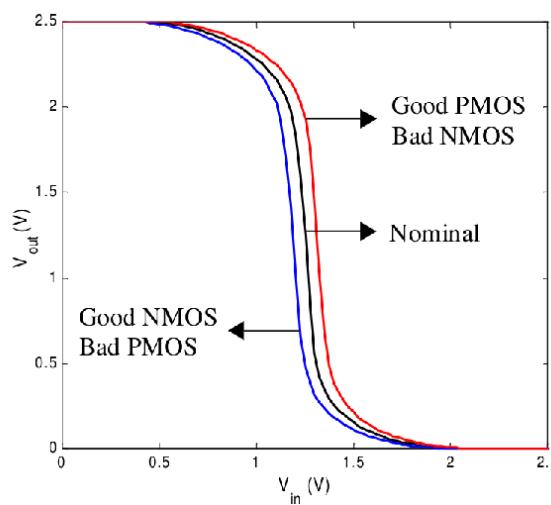


Figure 19: L'inverseur CMOS : influence de la qualité du process. Le bon transistor a un excès de 30nm sur sa largeur de grille, une réduction de 25nm sur sa longueur, une épaisseur d'oxyde réduite de 3nm et un courant de seuil diminué de 60mV. Le transistor MOS dégradé présente les caractéristiques inverses –tension de seuil plus élevée, largeur de canal réduite, ...

4.3.8 Performances de l'inverseur CMOS numérique

Deux éléments principaux sont à étudier pour établir les performances de l'inverseur CMOS : délais de propagation et consommation. Établir les délais de propagation et de commutation est assez aisé puisque, si l'on s'appuie sur la vision simpliste de la figure 17, il suffit d'établir les valeurs des capacités et des résistances équivalentes de l'inverseur CMOS puis d'appliquer les calculs classiques de charge et de décharge de capacité. L'étude de la consommation sera un peu plus délicate puisqu'il faudra séparer l'état dynamique de commutation de l'état statique (pour lequel l'approche au premier ordre d'interrupteur ouvert se révèlera insuffisante).

Délais de propagation

Calcul des capacités d'un inverseur CMOS : Comme on a pu le voir dans le chapitre 3, le calcul des valeurs de capacités d'un transistor est complexe du fait des nombreux effets capacitifs inhérents à la structure du transistor MOS allié à une dynamique de ces phénomènes. Le schéma de la figure 20 résume la complexité de la situation en présentant le cas classique d'un inverseur CMOS dont la charge est un autre inverseur CMOS : ne sont représentées que les capacités intervenant dans le transfert de l'information de l'entrée du premier inverseur CMOS à sa sortie.

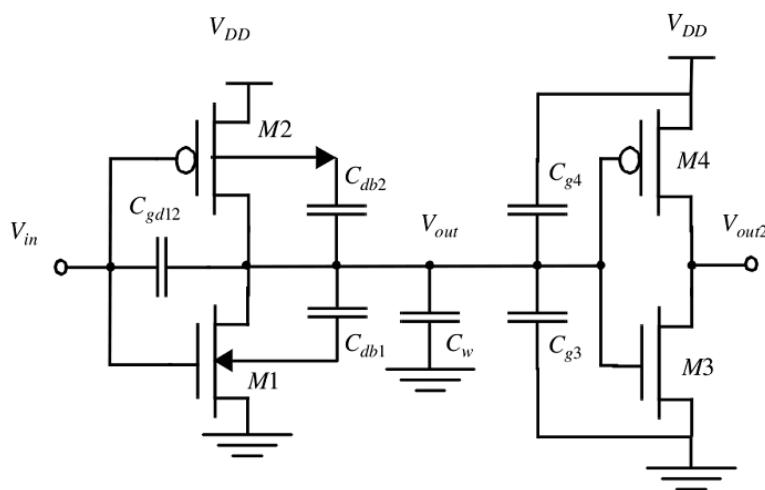


Figure 20: L'inverseur CMOS : les capacités en jeu lors de la transmission d'une information.

Le tableau 2 donne, pour information, un récapitulatif des formules permettant de calculer ces capacités et de leurs valeurs (technologie 0.25 microns) pour les transitions $0 \rightarrow 1$ et $1 \rightarrow 0$ ¹⁰.

¹⁰ Du fait de l'aspect dynamique des capacités, les valeurs ne sont, bien évidemment, pas identiques pour les deux transitions.

Capacité	Expression formelle	Valeur _{0→1} (fF)	Valeur _{1→0} (fF)
C_{gd1}	$2C_{GD0n}W_n$	0.23	0.23
C_{gd2}	$2C_{GD0p}W_p$	0.61	0.61
C_{db1}	$K_{eq} \cdot AD_n \cdot C_J + K_{eqwsn} \cdot PD_n \cdot C_{JSW}$	0.90	0.66
C_{db2}	$K_{eq} \cdot AD_p \cdot C_J + K_{eqwsp} \cdot PD_p \cdot C_{JSW}$	1.15	1.5
C_{g3}	$(C_{GD0n} + C_{GD0p})W_n + C_{ox}W_nL_n$	0.76	0.76
C_{g4}	$(C_{GD0p} + C_{GD0p})W_p + C_{ox}W_pL_p$	2.28	2.28
C_x	Valeurs issues d'une extraction de paramètres physiques	0.12	0.12
C_{gd1}	Σ	6.0	6.1

Tableau 2: Composantes de la capacité totale de charge à calculer pour les délais. C_{GDO} , C_{GSO} , A_D , P_D , C_{JSW} , ..., sont des paramètres physiques issus de la géométrie du transistor MOS.

Analyse formelle du problème :

Intéressons-nous dans un premier temps au calcul du délai de propagation à travers le calcul du temps de décharge de la capacité de sortie lors d'une transition 0→1 en entrée de l'inverseur. Le calcul serait le suivant :

$$t_p = \int_{V_{OH}}^{V_{OL}} \frac{C_L(v)}{i(v)} dv$$

avec C_L la capacité à décharger et $i(v)$ le courant de décharge de la capacité. Ces deux grandeurs étant, comme on l'a vu précédemment non-linéaires, un calcul exacte formel explicite de cette grandeur est impossible. Afin d'en obtenir une expression approché, nous allons utiliser le modèle interrupteur de la figure 17.

Toute la difficulté réside dès lors dans le calcul de la résistance équivalente du transistor NMOS quand celui-ci est passant. La première méthode est de calculer ce que vaut la résistance équivalente en intégrant le rapport V/I sur la plage où le transistor NMOS est passant, à savoir sur la gamme $[V_{DD}/2, V_{DD}]$, soit :

$$R_{eqn} = \frac{1}{V_{DD}/2} \int_{V_{DD}/2}^{V_{DD}} \frac{V}{I_{Dsatn}(1+\lambda V)} dv \approx \frac{3}{4} \cdot \frac{V_{DD}}{I_{Dsatn}} \left(1 - \frac{7}{9} \frac{V_{DD}}{2} \right)$$

avec $I_{Dsatn} = K_n \left((V_{DD} - V_T) V_{Dsatn} - \frac{V_{Dsatn}^2}{2} \right)$

Dans ce calcul la tension d'entrée à considérer est bien évidemment V_{DD} puisque nous étudions une transition 0→1 et la tension V_{Dsatn} est la tension limite à partir de laquelle le transistor NMOS rentre en régime saturé (on peut fixer au pire V_{Dsatn} à $V_{DD}/2$).

Une manière plus simple de procéder pour le calcul de cette résistance équivalente est de faire une moyenne plutôt qu'une intégration. Le transistor NMOS étant en régime saturé, la résistance varie de manière quasi linéaire avec la tension (cf. figure 21) il suffit donc de prendre les deux valeurs extrêmales des résistances (pour les tensions V_{DD} et $V_{DD}/2$)

et d'en faire la moyenne.

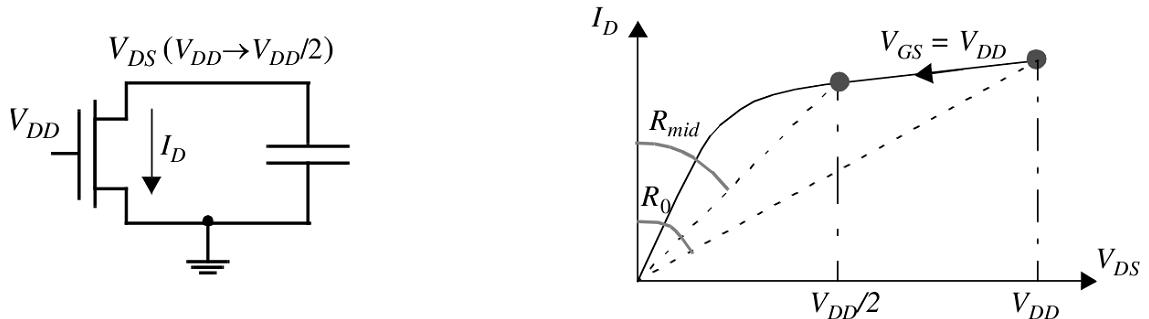


Figure 21: Calcul de la résistance du transistor passant : une moyenne.

Cela revient à faire le calcul suivant :

$$R_{eqn} = \frac{1}{2} \left(\frac{V_{DD}}{I_{Dsatin}(1+\lambda V_{DD})} + \frac{V_{DD}/2}{I_{Dsatin}(1+\lambda V_{DD}/2)} \right) \approx \frac{3}{4} \cdot \frac{V_{DD}}{I_{Dsatin}} \left(1 - \frac{5}{6} V_{DD} \right)$$

Bien entendu, les deux méthodes donnent un résultat similaire.

Une fois la difficulté du calcul de la résistance équivalente levée (on procède de la même manière pour le calcul de la résistance R_{eqp} du transistor PMOS pour les transitions 1→0), le calcul du temps de propagation se résume à une simple analyse d'un réseau RC pour lequel il est bien connu que :

$$t_{pHL} = \ln(2) \cdot R_{eqn} \cdot C_L = 0.69 \cdot R_{eqn} \cdot C_L$$

Le calcul du temps de propagation moyen donne :

$$t_p = \frac{t_{pHL} + t_{pLH}}{2} = 0.69 \cdot C_L \frac{R_{eqn} + R_{eqp}}{2}$$

Une simulation en technologie 0.25μm donne le résultat de la figure 22 :

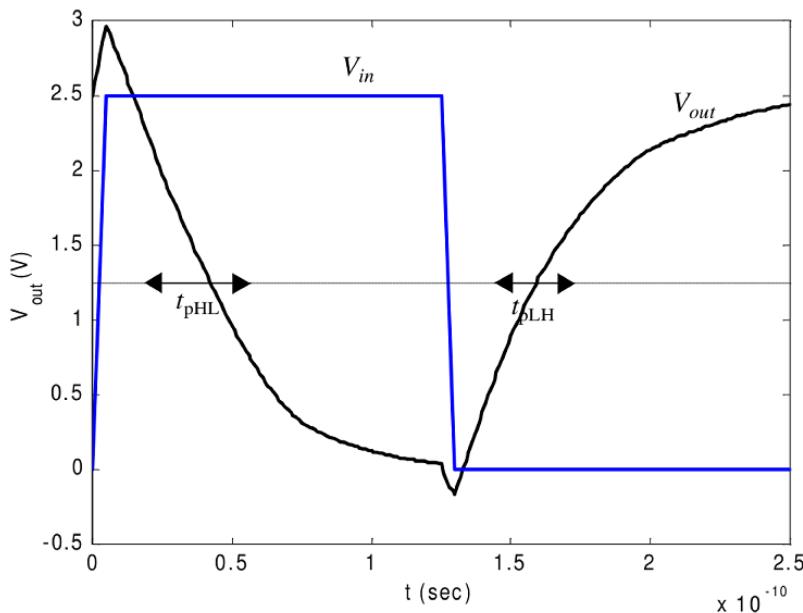


Figure 22: Simulation en technologie 0.25μm : temps de propagation.

L'implémentation des résultats précédents et l'utilisation du tableau 2 permet d'obtenir les résultats suivants :

$$t_{pHL} = 0.69 \times 8.6k\Omega \times 6.1fF = 36ps$$

$$t_{pLH} = 0.69 \times 6.9k\Omega \times 6.0fF = 29ps$$

$$t_p = \frac{t_{pHL} + t_{pLH}}{2} = 32.5 psec$$

Une comparaison entre ces résultats et ceux issus de la simulation permet d'établir une surprenante précision dans les résultats issus du calcul formel basé sur une approche du premier ordre. L'expression formelle permet surtout de dégager les règles suivantes pour l'aide au concepteur dans sa démarche de réduction des délais :

- **Réduction de C_L** : élément fondamental du délai, réduire la capacité est peu aisée car il faut jouer à la fois sur taille de la grille des transistors, les interconnexions et la sortance...
- **Augmentation du rapport W/L** : cette technique, la plus performante, est à manier avec précaution car elle joue également sur les capacités internes du circuit et la sortance.

Conclusion

Ainsi se termine la partie consacrée à la robustesse de l'inverseur CMOS : on a pu mettre en pratique les notions de robustesse et les appliquer à l'inverseur CMOS. Les formules établies dans cette partie, les notions abordées et les tendances dégagées peuvent être facilement extrapolées à des circuits beaucoup plus complexes que notre inverseur CMOS. La partie suivante sera consacrée au second aspect de la performance : la consommation. Là encore, on cherchera à mettre en application les notions théoriques vues précédemment et surtout à dégager des tendances générales qui peuvent être appliquées à tout type de circuit numérique, aussi complexe soit-il.

4.3.9 Consommation statique de l'inverseur CMOS numérique

Des deux types de consommation, nous allons nous intéresser au régime statique en premier. L'analyse sommaire du paragraphe 4.3.6 a mis en évidence une absence de chemin direct entre les deux rails d'alimentation en régime établi. Il existe cependant une consommation statique dans l'inverseur CMOS due aux courants de fuite des transistors MOS (jonctions des diodes en inverse) en état bloqué¹¹.

4.3.10 Consommation dynamique de l'inverseur CMOS numérique

Consommation moyenne

La consommation moyenne d'un inverseur MOS est à calculer à partir des transitions 0→1 et 1→0. Reprenons l'exemple de la figure 17. Nous avons un cycle de charge et de décharge complet.

Calculons l'énergie nécessaire à une charge de capacité (depuis V_{DD} et au travers du transistor PMOS):

¹¹ Les courants de fuite ont tendance à augmenter avec la réduction des dimensions la technologie, ce qui en fait un problème majeur actuellement. Si, à température ambiante, ces courants de fuite sont somme toute négligeables, l'augmentation de la température au sein d'un circuit peut multiplier ces courants par un facteur 100, ce qui en fait dès lors un problème important et demeure aujourd'hui un enjeu de recherche important.

$$E_{V_{DD}} = \int_0^\infty i_{V_{DD}}(t)V_{DD}dt = V_{DD} \int_0^\infty C_L \frac{dv_{out}}{dt} dt = C_L V_{DD} \int_0^{V_{DD}} dv_{out} = C_L \cdot V_{DD}^2$$

Maintenant, calculons l'énergie stockée **effectivement** dans la capacité :

$$E_C = \int_0^\infty i_{V_{DD}}(t)V_{out}dt = \int_0^\infty C_L \frac{dv_{out}}{dt} v_{out} dt = C_L \int_0^{V_{DD}} v_{out} dv_{out} = \frac{1}{2} \cdot C_L \cdot V_{DD}^2$$

On s'aperçoit que seulement la moitié de l'énergie tirée de l'alimentation est stockée dans la capacité : le reste a été perdu par dissipation thermique dans la résistance équivalente du transistor PMOS. Encore plus important, cette dissipation **ne dépend pas** de la valeur de la résistance et donc de la taille du transistor. Durant la décharge de la capacité, l'énergie est enlevée de la capacité et dissipée dans le transistor NMOS. Ainsi chaque cycle 0→1→0 nécessite une énergie de $C_L V_{DD}^2$. Ainsi, chaque transition étant faite un certain nombre de fois, noté f_{0-1} , nous obtenons la relation suivante :

$$P_{dyn_moyenne} = C_L V_{DD}^2 \cdot f_{0-1}$$

Cette règle, obtenue pour un simple inverseur CMOS, s'applique à tout type de circuit jusqu'au microprocesseur : la puissance consommée par un circuit augmente linéairement avec sa fréquence de fonctionnement et diminue de manière quadratique avec la diminution de sa tension d'alimentation¹². Ce phénomène est à la base de nombreuses stratégies matérielles de gestion d'énergie mais également de techniques de codage qui vise à minimiser le nombre de transitions !

Consommation de pic

L'analyse sommaire du paragraphe 4.3.6 peut laisser supposer, à tort, qu'il n'existe jamais de chemin direct entre les deux rails d'alimentation. Lors d'une transition (quelqu'elle soit), l'inverseur CMOS passe par l'état du milieu (V_M) pour lequel les deux transistors sont en état de saturation. Il existe donc un court instant pendant la transition pour lequel un courant de court-circuit i_{sc} part de V_{DD} vers 0 comme le montre la figure 23.

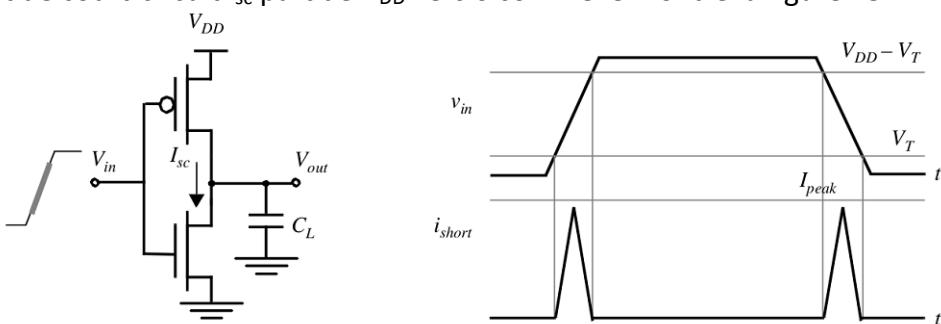


Figure 23: Courant de court-circuit lors des commutations.

En supposant que les temps de court-circuits t_{sc} sont identiques pour les deux transitions, on peut écrire que l'énergie consommée lors d'un cycle de commutations :

$$E_{dyn_court-circuit} = 2I_{short} \cdot V_{DD} \cdot t_{sc} / 2 = I_{short} \cdot V_{DD} \cdot t_{sc}$$

¹² C'est d'ailleurs pour cela que sur de nombreux ordinateurs portables, les processeurs gèrent la réduction de leur fréquence de fonctionnement et de leur tension d'alimentation selon la charge processeur demandée afin d'optimiser l'autonomie. C'est également pour cela que les circuits numériques actuels ont plusieurs horloges différentes et plusieurs alimentations différentes pour les différentes parties qui les composent.

Le temps de court-circuit dépend bien entendu du temps montée/descente, de l'inverseur CMOS et en supposant que le signal d'entrée présente une rampe linéaire de transition, on peut écrire :

$$t_{sc} = \frac{V_{DD} - 2V_T}{V_{DD}} \times \frac{t_{r(f)}}{0.8}$$

En faisant apparaître le cadencement des opérations, il vient :

$$P_{dyn_court-circuit} = I_{short} \cdot V_{DD} \cdot t_{sc} \cdot f_{0-1}$$

La puissance dissipée par court-circuit est très affectée par tout ce qui touche aux délais et aux capacités.

Puissance totale dissipée par l'inverseur CMOS

La puissance totale dissipée est la somme de ces trois contributions :

$$P_{totale} = P_{dyn_moyenne} + P_{dyn_court-circuit} + P_{stat}$$

dont typiquement $P_{dyn_moyenne}$ représente 90% de la puissance totale, $P_{dyn_court-circuit}$ 8% et P_{stat} 2%. Toutefois, si la puissance dynamique a tendance à diminuer avec l'évolution des technologies, la puissance statique, elle, a malheureusement tendance à prendre de plus en plus d'importance avec les réductions des tailles des transistors.

4.3.11 Produit Energie-Délai (EDP, Energy-Delay Product)

Le critère permettant de quantifier au mieux les performances d'un circuit numérique est la notion d'énergie consommée par opération, car elle intègre à la fois la notion de puissance dissipée et de temps de réaction de la porte. Son mode de calcul est le suivant :

$$PDP = P_{dyn_moyenne} \cdot t_p$$

(sachant que nous négligeons les puissances statiques et dynamiques de court-circuit par rapport à la puissance moyenne, et que pour rappel $f_{0-1} = 1/2t_p$), soit

$$PDP = V_{DD}^2 C_L / 2$$

soit l'énergie nécessaire par commutation. Cependant cette métrique en tant qu'outil d'optimisation des portes est insuffisante : en effet, de manière évidente, il suffit de diminuer la tension d'alimentation du circuit pour augmenter la performance du circuit – ce qui n'est pas tout à fait exacte si l'on englobe à la fois l'aspect temporel et puissance.

Une meilleure métrique est le produit Énergie-Délai (EDP pour Energy-Delay Product) qui permet d'intégrer tous les aspects temporels que seule l'énergie par opération ne permet pas de considérer. La définition en est la suivante :

$$EDP = PDP \times t_p = \frac{C_L V_{DD}^2}{2} t_p$$

Le temps de propagation moyen peut être approximé par la formule suivante :

$$t_p = \frac{\alpha V_{DD} C_L}{V_{DD} - V_{Te}}$$

avec α paramètre technologie, $V_{te} = V_T + V_{Dsat}/2$, et l'hypothèse que les transistors NMOS et PMOS ont un comportement similaire (tensions de seuil et de saturation). Dès lors on

obtient la métrique suivante :

$$EDP = \frac{C_L^2 V_{DD}^3}{2(V_{DD} - V_{Te})}$$

On peut, avec cette formule, tracer des graphes tels que celui de la figure 24.

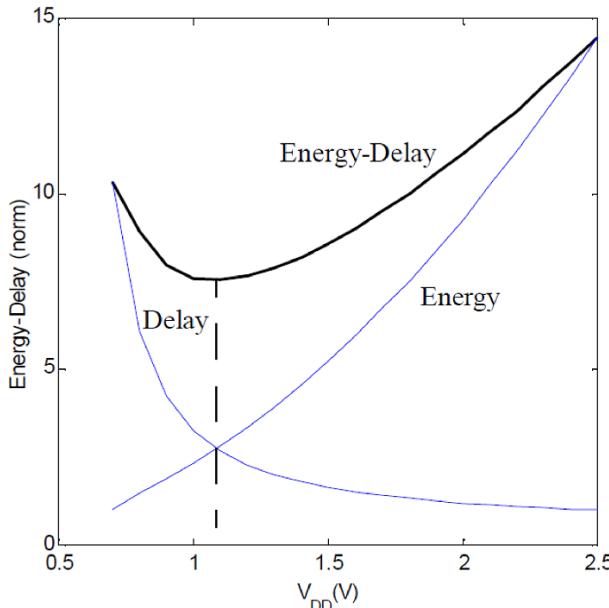


Figure 24: Produit Énergie-Délai.

On y voit clairement les deux tendances opposées de la puissance consommée et de la vitesse du circuit en fonction de la tension d'alimentation, avec ce qui semble être le meilleur compromis (si l'on considère que puissance et délai ont un poids égal de priorité) autour de 1.2V.

Bien évidemment cette formule donne des tendances générales et aide le concepteur : elle n'est en aucun cas la recette ultime puisqu'elle ne prend pas en compte les avantages que peuvent apporter par exemple des architectures telles que la mise en pipeline ou le recours à la concurrence.

4.4 Conclusion

Ce chapitre consacré à l'inverseur CMOS a permis l'étude du circuit élémentaire qu'est l'inverseur CMOS, à la fois d'un point de vue analogique où il peut servir d'amplificateur et être à la base de filtre, mais également du point de vue numérique où il est une porte de base présente dans la quasi-totalité des circuits numériques existants. De par sa dualité analogique-numérique, il a permis de mettre en évidence les relations qui existent entre le monde analogique et le monde numérique, les relations entre les performances statiques et dynamiques en tension du circuit analogique et celle attendues en terme de vitesse, consommation et délai pour le circuit numérique.

La partie analogique aura également permis de mettre en application et d'illustrer les notions vues en cours d'électronique analogique sur l'analyse d'un circuit, le régime statique, le régime dynamique, l'établissement d'un point de fonctionnement et l'optimisation des performances dynamiques du circuit autour de ce point de fonctionnement.

Ce chapitre clôture enfin l'étude de l'électronique analogique et a permis d'introduire

l'électronique numérique qui sera développée de manière plus conventionnelle dans les chapitres suivants sous la forme électronique numérique combinatoire, électronique numérique séquentielle (du combinatoire avec de la mémoire) et enfin architectures de microprocesseurs (l'agencement intelligent de circuits séquentiels et combinatoires afin d'obtenir des systèmes pseudo-intelligents).

References

- [9] Rabaey Jan M., *Digital Integrated Circuits, 2nd Edition*, Prentice Hall.
- [10] Tsividis Yannis, *Operation and modeling of the MOS transistor*, WCB/McGraw-Hill, 621.381 528 TSI
- [11] Razavi Behzad, *Design of analog CMOS integrated circuits*, McGraw-Hill, Publication Boston, MA, 2001, 621.39 RAZ
- [12] Sedra Adel S. et al., *Microelectronic circuits*, Oxford University Press, New York Oxford, 2004, 621.381 5 SED
- [13] Berkeley University of California, *Cours EE141 :Introduction to Digital Integrated Circuits*, http://bwrc.eecs.berkeley.edu/Classes/icdesign/ee141_s02/notes.html, Chapitre 5.

5 Logique Combinatoire

La logique combinatoire traite les informations binaires (logiques) "0" et "1" avec des opérateurs booléens simples.

Codage des nombres signés

Pour coder des nombres positifs ou négatifs, plusieurs méthodes existent. Les codes "valeur absolue et bit de signe", "complément à 1", et complément à 2 sont les codages les plus pertinents.

En considérant des mots binaires sur 4 bits, nous représentons ici les valeurs des codages :

par valeur absolue et bit de signe, en rouge

par complément à 1, en vert

par complément à 2, en bleu

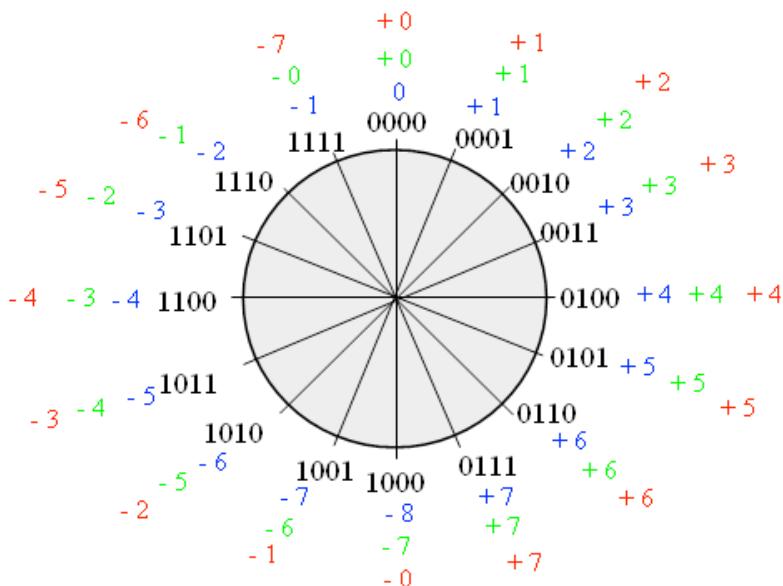


Figure 1: Nombres signés selon différents codages.

Le codage valeur absolue et bit de signe consiste à prendre le bit à gauche non plus comme le MSB mais comme un indicateur de signe : "0" signifie + et "1" signifie -.

Le **complément à 1** a pour principe de complémer tous les bits d'une valeur positive pour en calculer son négatif. Exemple: la valeur -2 est : 2 (0010) complémenté: 1101.

Nous constatons que ces 2 codages ont l'inconvénient d'avoir 2 zéros (et les nombres positifs et négatifs ne sont pas symétriques par rapport au 0 pour le codage valeur absolue et bit de signe).

Le codage utilisé en logique combinatoire est **le complément à 2**. Il se calcule en ajoutant 1 au complément à 1 (CA1).

Exemple : pour trouver -2: CA1(2) = CA1(0010) + 1 = 1101 + 1 = 1110

Ce codage n'est pas intuitif, mais il a l'avantage de ne pas nécessiter une logique spécifique, car une soustraction revient à une addition modulo 16.

5.1 Portes logiques combinatoires

Les portes logiques CMOS sont composées de transistors de type N et P. Elles sont caractérisées par des délais de transition, qui sont différents selon que l'entrée passe de "0" à "1" ou de "1" à "0". Ces délais sont notés tp_{LH} et tp_{HL} , et définissent ensemble la période (donc fréquence) maximale d'un circuit. Lorsqu'il y a plusieurs portes logiques entre l'entrée et la sortie, ces délais sont cumulatifs.

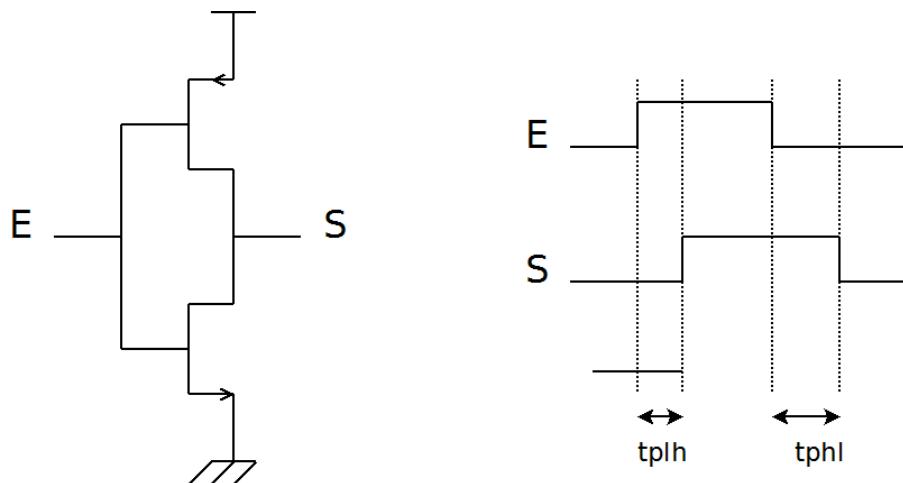


Figure 2 : Temps de propagation

Pour plus de simplicité, nous dessinons classiquement les schémas logiques au niveau portes, mais il est important de connaître la structure au niveau transistors.

Les portes les plus simples sont :

- l'inverseur (ou NOT)
- la porte ET (ou AND)
- la porte OU (ou OR)

L'ensemble des symboles associés à ces portes se trouve figure ci-dessous¹.

¹ L'industrie électronique utilisant de manière préférentielle (voire exclusive) les symboles américains, nous ne nous servirons dans ce chapitre que des symboles américains.

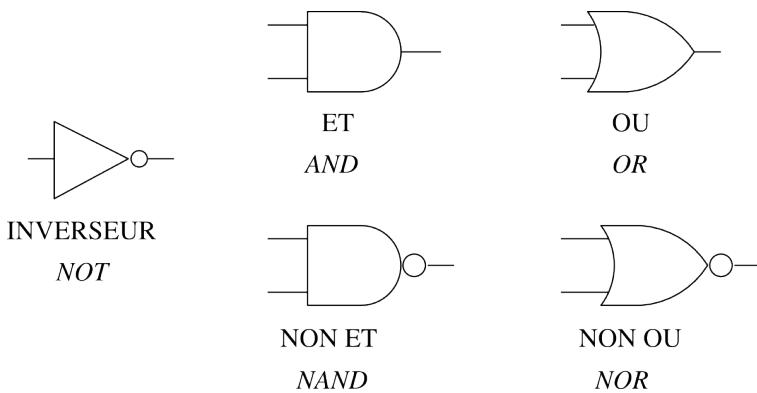


Figure 3 : Symboles des portes élémentaires.

A l'aide de ces opérateurs, nous pouvons écrire des équations du type $S = \overline{CA} + \overline{CB} + BA$ (voir le chapitre de pré-requis pour plus de précisions sur l'algèbre de Boole).

Cependant, la technologie CMOS très utilisée aujourd'hui ne permet pas de réaliser directement ces portes simples. Pour démontrer cela, nous allons étudier la structure des portes de base CMOS: inverseur, NON-ET, NON-OU.

5.1.1 Porte inverseuse (NOT)

C'est la porte la plus simple composée d'un transistor de type N et d'un transistor de type P, en montage push-pull. Comme leur structure est complémentaire et qu'ils ont une entrée commune, ils vont conduire ou être bloqués de manière complémentaire. La structure est présentée à gauche dans la figure :

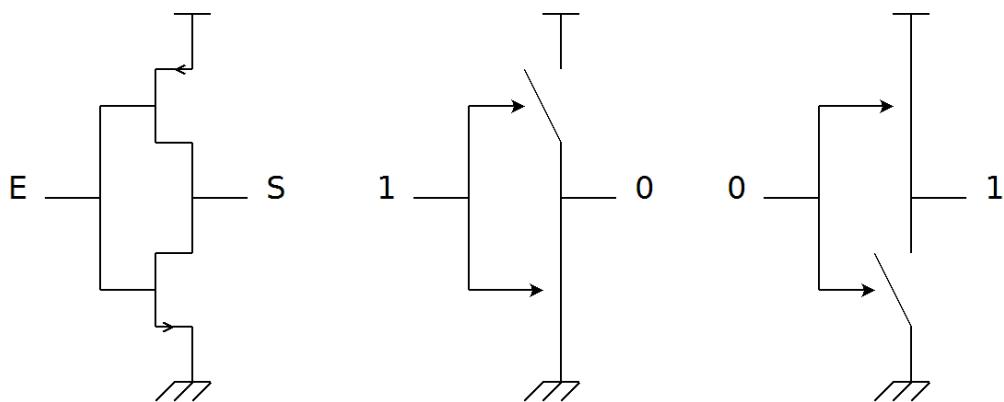


Figure 4 : Fonctionnement simplifié de l'inverseur.

Si $E = 1$, le transistor N conduit, le transistor P est bloqué, nous retrouvons donc le schéma (idéal) équivalent du milieu de la figure ci-dessus. La sortie S est donc tirée à l'état bas: $S = 0$. De même si $E = 0$, le transistor P conduit, le transistor N est bloqué, $S = 1$.

La table de vérité se trouve résumée :

E	S
0	1
1	0

Table 1: Table de vérité d'une porte inverseuse.

5.1.2 Porte NON-ET (NAND)

Avec la même hypothèse (un "1" sur un transistor N est équivalent à un court-circuit, un "0" sur un transistor N est équivalent à un circuit ouvert; un "0" sur un transistor P est équivalent à un court-circuit, un "1" sur un transistor P est équivalent à un circuit ouvert), nous pouvons réfléchir sur la structure minimale à 4 transistors de la figure ci-dessous.

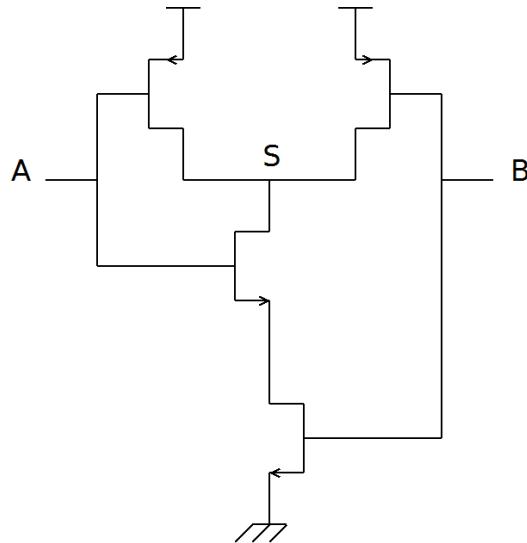


Figure 5 : Structure niveau transistor d'une porte NON-ET.

Si $A = 0$ OU $B = 0$ (ou les deux), S est relié à V_{DD} (alimentation haute) donc $S = 1$. A l'inverse, il faut que - à la fois - A ET B soient à 1 pour que S soit à 0 :

A	B	S
0	0	1
0	1	1
1	0	1
1	1	0

Table 2: Table de vérité d'une porte NON ET.

Nous reconnaissons ici la table de vérité d'une porte NON-ET (NAND) qui est une structure de base. En technologie CMOS, la porte ET est donc composée d'une porte NON-ET suivie d'un inverseur. Pour rappel, la table de vérité de la porte ET (AND) se trouve ci-dessous.

A	B	S
0	0	0
0	1	0
1	0	0
1	1	1

Table 3 Table de vérité d'une porte ET.

Nous pouvons remarquer que le fonctionnement dual des transistors N et P entraine une antisymétrie dans la structure: les 2 transistors P sont en parallèle alors que les 2 transistors N sont en série. Nous parlons alors de dualité entre le plan N et le plan P.

Nous pouvons aussi remarquer qu'en connectant A et B, nous obtenons une porte inverseuse. La démonstration se fait en une ligne : $\overline{A} \cdot A = \overline{A}$. ou encore $\overline{A} \cdot A = \overline{A} + \overline{A} = \overline{A}$ selon le moment où nous appliquons le théorème d'idempotence et éventuellement le théorème de De Morgan.

5.1.3 Porte NON-OU (NOR)

En dessinant le schéma symétrique du précédent, nous obtenons celui de la figure ci-dessous.

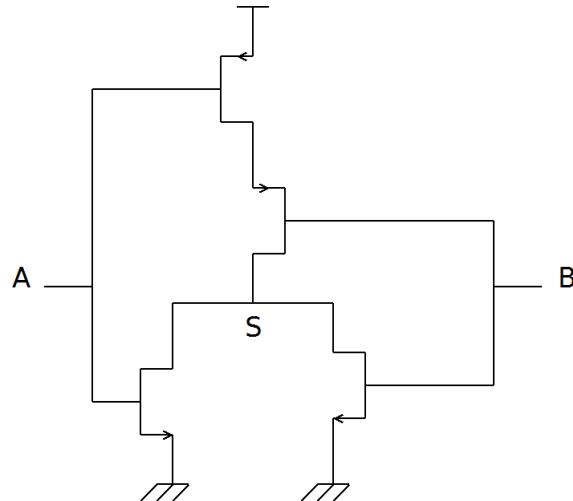


Figure 6 : Structure niveau transistorde la porte NON-OU.

Si $A = 1$ OU $B = 1$ (ou les deux), S est relié à Gnd (alimentation basse) donc $S = 0$. A l'inverse, il faut que - à la fois - A ET B soient à 0 pour que S soit à 1. En résumé :

A	B	S
0	0	1
0	1	0
1	0	0
1	1	0

Table 4: Table de vérité de la porte NON OU.

Nous reconnaissons ici la table de vérité d'une porte NON-OU (NOR) qui est une structure de base. Nous remarquons qu'en connectant A et B, nous obtenons une porte inverseuse. Une fois de plus, la démonstration est triviale et se fait de deux manières (selon le moment où nous appliquons l'idempotence et éventuellement De Morgan) : $\overline{A} + \overline{A} = \overline{A}$ où $\overline{A} + \overline{A} = \overline{A} \cdot \overline{A} = \overline{A}$.

Nous pouvons aussi remarquer que la table est exactement l'inverse de la précédente, ce qui permet de vérifier une partie du théorème de De Morgan :

$$NAND = \overline{NOR} \text{ c'est-à-dire } \overline{A \cdot B} = \overline{\overline{A} + \overline{B}} = \overline{A} + \overline{B}$$

De même que pour la porte ET, la porte OU s'obtient en technologie CMOS par l'adjonction d'un inverseur CMOS en sortie de la porte NOR. Pour rappel, voici la table de vérité de la porte OU (OR) :

A	B	S
0	0	0
0	1	1
1	0	1
1	1	1

Table 5: Table de vérité de la porte OU.

5.1.4 Porte complète

Les portes NON-ET et NON-OU, outre leur structure transistor simple, ont pour avantage, par rapport aux portes ET et OU, d'être des portes complètes. Une porte complète et une porte dont le type permet de réaliser *toutes* les opérations logiques qui sont au nombre de trois : ET, OU et inversion. Nous avons déjà vu que les portes NON-ET et NON-OU permettent de réaliser la porte inverseuse.

La porte NON-ET définie comme $\overline{A \cdot B}$ réalise intrinsèquement une fonction OU. Il suffit d'ajouter une porte inverseuse (réalisée avec des portes NON-ET) sur les entrées pour lever le problème de l'inversion des entrées. Enfin, une porte NON-ET utilisée en inverseur que nous ajoutons à la sortie d'une porte NON-ET permet de réaliser la fonction ET car $\overline{\overline{A \cdot B}} = A \cdot B$.

De même, la porte NON-OU définie comme $\overline{A + B}$ réalise intrinsèquement une fonction ET. Il suffit d'ajouter une porte inverseuse (réalisée avec des portes NON-OU) sur les entrées pour lever le problème de l'inversion des entrées. Enfin, une porte NON-OU utilisée en inverseur que nous ajoutons à la sortie d'une porte NON-OU permet de réaliser la fonction OU car $\overline{\overline{A + B}} = A + B$ d'après le théorème d'involution.

Ces deux portes sont donc des portes complètes, qui outre leur structure MOS simple, permet de réaliser la totalité des fonctions logiques avec *un seul type* de porte. Les équivalences sont rappelées dans la figure ci-dessous.

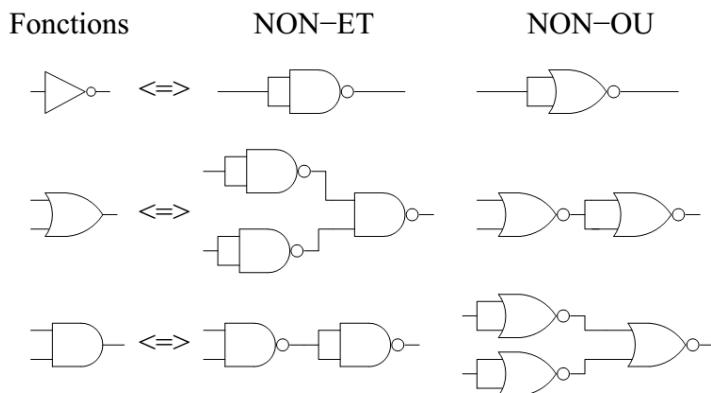


Figure 7 : Equivalences de portes.

5.1.5 Porte simples

D'autres portes, non élémentaires du point de vue transistors mais simples de part leur fonction sont largement utilisées. Ainsi, elles ont leur propre implémentation physique spécifique en technologie MOS qui est produite (au contraire des autres fonctions que nous chercherons généralement à réaliser avec des combinaisons des portes élémentaires).

OU EXCLUSIF (XOR)

Comme son nom l'indique, il faut que A et B soient exclusivement différents pour que la sortie soit vraie (= 1) comme le montre la table de vérité :

A	B	S
0	0	0
0	1	1
1	0	1
1	1	0

Table 6: Table de vérité du OU exclusif.

Cette porte est classiquement composée de 2 inverseurs, 2 ET et 1 OU: $S = \overline{A} \cdot B + A \cdot \overline{B}$ noté également $S = A + B$.

Le OU EXCLUSIF étant dans les faits une comparaison (différence ou égalité), cette porte est très utilisée dans les comparateurs ou les additionneurs. Le complément de cette porte est le NON-OU EXCLUSIF (XNOR). Les symboles de ces 2 portes sont rappelés :



Figure 8 : Symboles des portes XOR et XNOR.

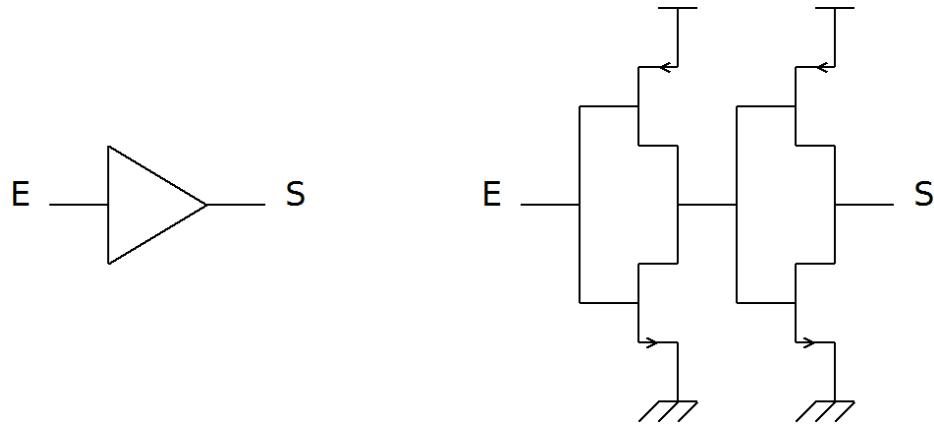
BUFFER

Le buffer est un suiveur logique, classiquement composé de 2 inverseurs successifs. Sa fonction principale est l'adaptation d'impédance, la création d'arbres de signaux de contrôle (comme un arbre d'horloge) et parfois la numérisation² d'un signal dans des systèmes mixtes (mélangeant analogique et numérique).

La table de vérité du buffer et le symbole sont rappelés ci-dessous :

² En se servant des tensions V_{IH} et V_{IL} du buffer comme valeur de référence, nous pouvons réaliser une conversion analogique numérique sommaire avec simplement un buffer : toute tension supérieure à V_{IH} est transformée en un logique et toute tension inférieure à V_{IL} est transformée en un zéro logique.

A	S
0	0
1	1

Table 7: Table de vérité du Buffer.*Figure 9 : Symbole du Buffer.*

Porte 3 états

Une porte 3 états est une porte qui permet d'implémenter un troisième état appelé état de haute impédance noté Z . Lorsqu'une porte est dans l'état de haute impédance, cela signifie que son impédance de sortie est infinie : tout se passe comme si la sortie de cette porte présentait une impédance infinie avec l'entrée, et les rails d'alimentation. Cet état est très utilisé dans le partage de ligne (bus de communication type I2C et autres) car il permet à plusieurs signaux d'entrée d'utiliser le même fil pour communiquer : en effet, si une entrée veut émettre sur le fil commun, il suffit de mettre tous les autres accès en état de haute impédance : l'élément de base de cette logique est la **porte de transmission 3 états** dont la table de vérité et la structure sont :

E	A	S
0	0	Z
0	1	Z
1	0	0
1	1	1

Table 8: Table de vérité de la table de transmission 3 états.

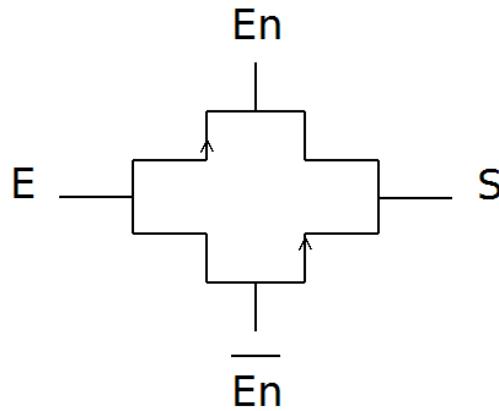


Figure 10 : Architecture de la porte de transmission 3 états.

Une implémentation plus performante de cette fonctionnalité est celle du **buffer 3 états**. La table de vérité reste inchangée mais la structure du circuit est plus complexe :

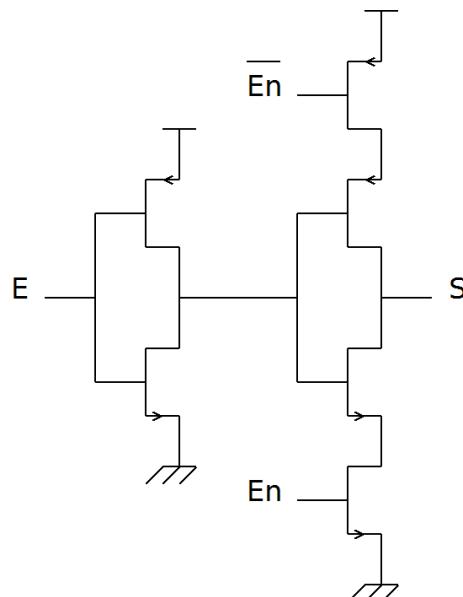


Figure 11 : Architecture du buffer 3 états.

La grande différence avec la porte de transmission porte sur les signaux : lorsque la porte est passante, dans le cas de la transmission, le signal d'entrée est transféré en sortie alors que pour le buffer, nous recopions la valeur logique du signal d'entrée en sortie à partir des alimentations. Concrètement, cela signifie que dans le cas du buffer, nous reconstruisons le signal au contraire de la porte de transmission qui a tendance à le dégrader encore plus qu'il ne l'est en entrée.

5.2 Synthèse combinatoire

Pour aller d'un cahier des charges à une équation logique et un schéma électrique, il y a une méthode de synthèse à suivre.

Nous allons appuyer l'explication sur la synthèse d'un additionneur 1 bit avec retenue entrante.

Dans un premier temps, il faut retranscrire le problème dans une table de vérité.

5.2.1 Table de vérité

Elle permet de poser clairement le problème, en détaillant les entrées et les sorties (et donc la totalité des états possibles), et en écrivant leurs relations.

Pour l'additionneur 1 bit, nous voulons additionner A, B et C (booléens) et obtenir comme résultat S (booléen) avec une retenue sortante Co (faisant office de poids plus fort que S), Co étant la contraction de *carry out*.

Le symbole de l'additionneur est rappelé :

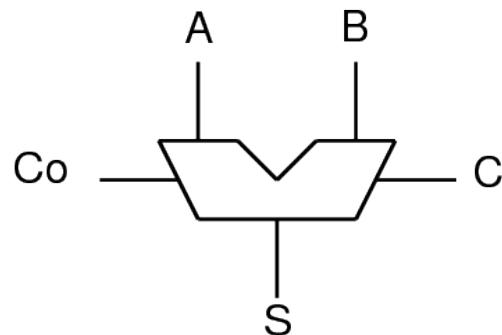


Figure 12 : Symbole de l'additionneur complet 1 bit.

Nous écrivons la table 19, avec les entrées à gauche, les sorties à droite :

C	B	A	Co	S
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

Table 9: Table de vérité de l'additionneur complet 1 bit.

Une fois cette table faite, nous pouvons écrire facilement les équations complètes de S et Co. Les équations complètes sont des équations brutes, non optimisées algébriquement (dites "non réduites"). En lisant la table ligne à ligne, il est possible de faire une synthèse sur les "1" (équations pour S = 1), ou sur les "0" (nous calculons alors \bar{S}). Il peut être plus rapide de chercher l'équation de \bar{S} si il y a beaucoup de résultats faux (à "0") dans la table.

Synthèse sur les "1"³ :

$$S = \overline{C}\overline{B}A + \overline{C}B\overline{A} + C\overline{B}\overline{A} + CBA = A + B + C$$

$$C_o = \overline{C}B.A + C\overline{B}.A + C.B.\overline{A} + CBA$$

Synthèse sur les "0" :

$$\begin{aligned}\bar{S} &= \bar{C}\bar{B}\bar{A} + \bar{C}B.A + C.\bar{B}.\bar{A} + C.B.\bar{A} \\ \bar{Co} &= \bar{C}\bar{B}\bar{A} + \bar{C}\bar{B}.A + \bar{C}.B.\bar{A} + C.\bar{B}.\bar{A}\end{aligned}$$

5.2.2 Tableau de Karnaugh

La synthèse d'équations logiques par l'algèbre de Boole ne permet pas toujours d'obtenir la solution optimale (c'est-à-dire l'équation la plus réduite). Des tableaux de Karnaugh permettent donc de faire des résolutions visuelles.

Ce tableau se construit en mettant les opérandes sur les colonnes et les lignes, et le résultat à l'intérieur du tableau. Les opérandes sont marqués en **binaire réfléchi**, donc 1 seul bit change à la fois.

Exemple pour $S = f(A,B,C)$, nous définissons le tableau comme indiqué sur la table :

Z YX	00	01	11	10
0				
1				

Table 10 Écriture d'un tableau de Karnaugh à 3 entrées.

Nous écrivons donc le tableau pour **chaque** sortie.

Ex: soit l'équation $S = \bar{C}BA + \bar{C}BA + \bar{C}BA + CBA$

Nous remplissons le tableau avec les valeurs de S :

C	BA	00	01	11	10
0	0	0	1	1	1
1	0	0	1	1	0

Ensuite, si nous choisissons une résolution par les 1 (ce qui est classiquement fait), nous faisons des regroupements de 1 logiques par paquets de puissance 2 (donc nous regroupons 1, 2, 4, ou 8 bits).

Nous pouvons prendre un bit plusieurs fois dans des regroupements, cela correspond à une factorisation.

Chaque regroupement donne une expression plus réduite issue d'une factorisation : nous ignorons le bit qui change d'état dans un regroupement. Ainsi, pour le premier regroupement (rouge): C = 0, B change d'état, et A = 1. Donc pour toute combinaison C = 0 et A = 1 (quel que soit B), S = 1. Nous trouvons donc l'équation réduite (minimale) :

$$S = \bar{C}A + \bar{C}B + BA$$

Rq: Toute autre factorisation ne réduirait pas le nombre de portes logiques nécessaires.

Pour minimiser cette équation algébriquement, il aurait fallu écrire :

$$S = \bar{C}BA + \bar{C}BA + \bar{C}BA + CBA$$

$$S = \bar{C}BA + \bar{C}BA + \bar{C}BA + CBA + \bar{C}BA + \bar{C}BA \Rightarrow \text{phase non intuitive}$$

$$S = (\overline{C}\overline{B}A + \overline{C}BA) + (\overline{C}\overline{B}\overline{A} + \overline{C}BA) + (CBA + \overline{C}BA) \Rightarrow \text{phase non intuitive}$$

$$S = (\overline{C}A(B + \overline{B})) + (\overline{C}B(A + \overline{A})) + (BA(C + \overline{C}))$$

$$S = \overline{C}A + \overline{C}B + BA$$

Le recours aux tableaux de Karnaugh permet de faire apparaître de manière automatique les termes redondants⁴ nécessaires pour une simplification optimale des équations.

Il y a une limite à l'application simple du tableau de Karnaugh (basé sur le binaire réfléchi) : au-delà de 4 variables en entrée, le codage et les regroupements sont un peu plus complexes.

5.2.3 Les "don't care"

Il se peut que des valeurs en entrée ne soient jamais utilisées : nous pouvons dans ce cas considérer qu'ils génèrent un "1" ou un "0" en sortie. Leur incidence importe peu, puisque ces entrées ne sont pas sensées apparaître. Nous les notons donc X ("don't care"), et nous pouvons leur attribuer la valeur que nous souhaitons (celle qui permettra d'optimiser le résultat).

Exemple : réalisation d'un détecteur de nombre premier

Réalisons la synthèse d'un détecteur de nombres premiers sur les nombres allant de 0 à 5.

Donc la sortie S vaut 1 si E est égal 1,2,3,5⁵.

Comme les entrées varient de 0 à 5, nous allons coder les entrées sur 3 bits : E2 E1 E0. Nous obtenons la table de vérité:

E3	E2	E1	S
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	1
1	0	0	0
1	0	1	1
1	1	0	X
1	1	1	X

Table 11: Table de vérité d'un détecteur de nombre premier.

Les valeurs 6 (110) et 7 (111) qui peuvent être codées sur 3 bits ne sont pas considérées, donc la valeur de S associée importe peu.

Nous allons alors dessiner le tableau de Karnaugh de S :

⁴ Les "1" sur plusieurs groupements sont les éléments que nous doublons ou triplons (voire plus) par idempotence dans la résolution analytique à la main.

⁵ 0 n'est pas considéré comme un nombre premier et 1 est considéré comme un nombre premier en accord avec la majorité des scientifiques débattant sur ce sujet épineux.

E3 \ E2E1	00	01	11	10
0	0	1	1	1
1	0	1	X	X

Table 12: Tableau de Karnaugh simple du décodeur de nombre premier.

Si nous ne considérons pas les X (ce qui revient à considérer X=0), nous obtenons :

$$S = \overline{E2} \cdot E1 + \cancel{\overline{E3} \cdot E1} + \overline{E3} \cdot E2$$

N.B : le regroupement rouge est inutile car il prend en compte des 1 déjà considérés par ailleurs (de plus, le terme associé au groupement rouge disparaît si nous appliquons les théorèmes de simplification à la formule précédente).

Ce résultat n'est pas optimal (au sens minimisation du nombre de portes et complexité de circuit minimale).

En affectant "1" aux deux "X", nous pourrons faire des regroupements de "1" plus intéressants comme le montre le tableau :

E3 \ E2E1	00	01	11	10
0	0	1	1	1
1	0	1	X	X

Table 13: Tableau de Karnaugh optimal du décodeur de nombre premier.

Nous obtenons l'équation compacte finale suivante : $S = E1 + E2$

5.2.4 Implémentation physique

Une fois l'équation obtenue, nous pouvons dessiner le schéma électrique du circuit, composé de 2 portes OU EXCLUSIF, 3 portes ET et deux portes OU :

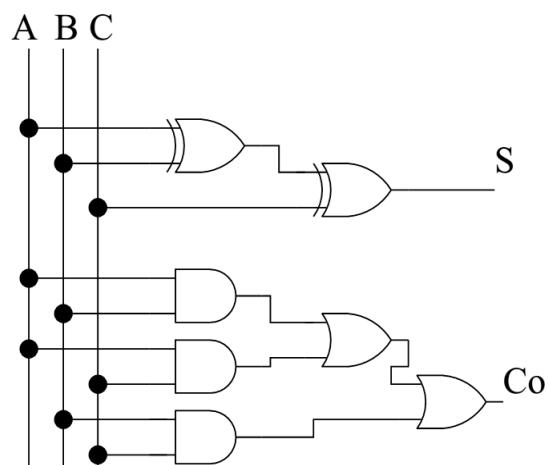


Figure 13 : Implémentation physique de l'additeur complet 1 bit en portes élémentaires.

L'implémentation ci-dessus a deux inconvénients majeurs :

- Plusieurs types de portes sont nécessaires
- Le nombre de portes utilisées n'est pas minimal (donc délais plus grand)

En utilisant exclusivement des portes NON-ET, nous obtenons la solution :

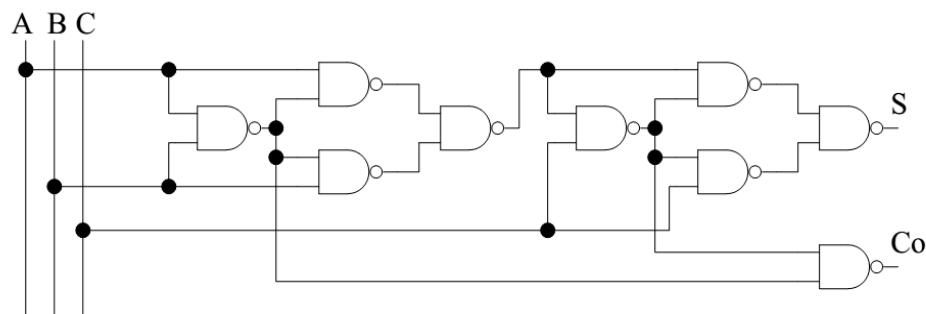


Figure 14 : Implémentation de l'additionneur complet en portes NON-ET.

5.3 Blocs combinatoires

5.3.1 Aiguilleurs

Pour aiguiller les signaux numériques, des multiplexeurs (ou démultiplexeurs pour la fonction inverse) sont utilisés. Un multiplexeur permet de sélectionner un signal parmi N –à l'aide d'une adresse de sélection- et de la connecter sur la sortie.

La figure ci-dessous donne un exemple pour un multiplexeur 4 vers 1 (4 entrées vers 1 sortie). Deux entrées de sélection sel1, sel0 sont alors nécessaires.

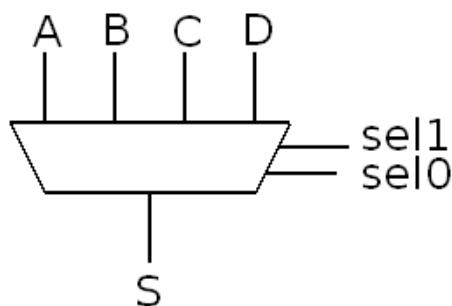


Figure 15 : Symbole du multiplexeur 4 vers 1.

Nous obtenons la table de vérité simplifiée :

sel1	sel0	S
00		S = A
01		S = B
10		S = C
11		S = D

Table de vérité simplifiée du multiplexeur 4 vers 1.

5.3.2 Additionneur et soustracteur N bits

Additionneur N bits

En connectant plusieurs additionneurs 1 bit, nous pouvons obtenir un additionneur N bits. Nous observons que le résultat ne sera obtenu qu'après la propagation des signaux de retenue (C vers Co) à travers tous les additionneurs 1 bit. Cette structure est simple, mais n'est pas la plus rapide (cf. figure ci-dessous). Elle a pour avantage d'être facilement scalable c'est à dire que le passage d'un additionneur N bits à un additionneur (N+1)bits se fait simplement par l'adjonction d'un additionneur 1bit. Il existe d'autres architectures (moins facilement scalables) qui permettent d'accélérer les propagations de retenue (anticipation, saut, pré-calcul, ...).

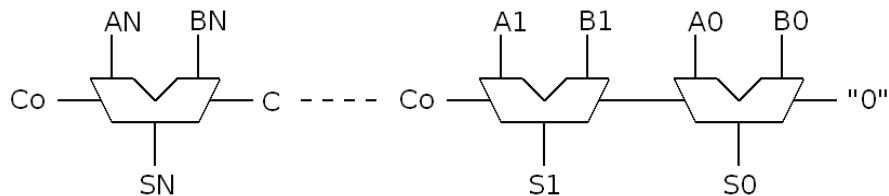


Figure 16 : Additionneur N bits obtenu par la mise en cascade de N additionneurs 1 bits.

Soustracteur N bits

Pour faire un soustracteur $S=A-B$, il faut faire $S=A+\overline{B}+1$ (voir codage des nombres signés). En utilisant des inverseurs sur les entrées $BN \dots B0$ et en forçant "1" sur la retenue entrante C, nous effectuons bien $A+\overline{B}+1$.

A l'aide de multiplexeurs 2 vers 1, il est possible de concevoir un additionneur/soustracteur, donc le signal de configuration M détermine l'opération à effectuer.

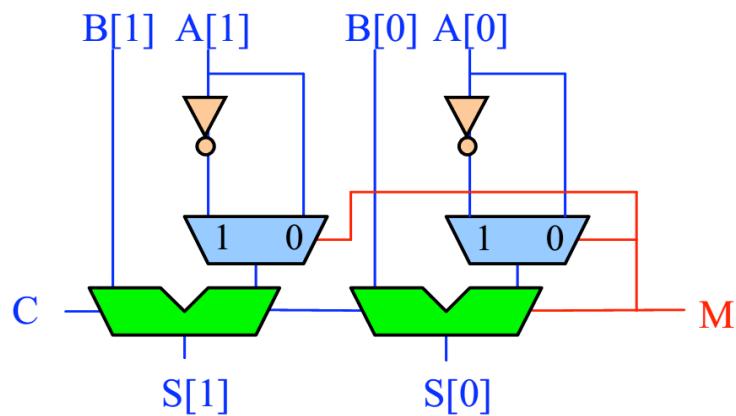


Figure 17 : Exemple d'additionneur/soustracteur 2 bits.

6 La logique séquentielle.

6.1 Définitions

Un système est dit *séquentiel*, lorsque la ou les sorties, à un instant donné (t), dépendent :

- de la combinaison des entrées à l'instant (t), et
- de l'état qu'avaient les sorties à l'instant ($t-1$)

Le schéma bloc de la figure 1 représente un circuit séquentiel comme un circuit composé d'un élément de mémorisation et d'une partie de logique combinatoire.

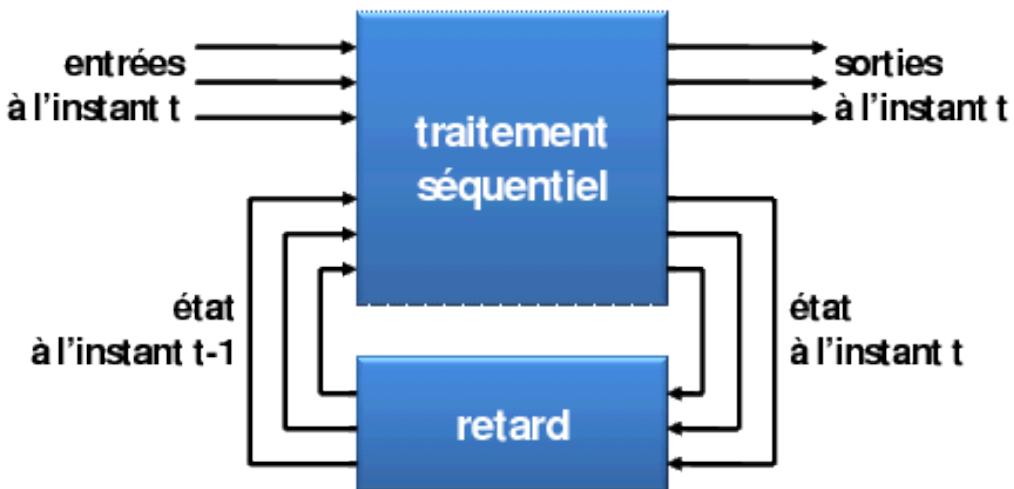


Figure 1: Schéma bloc d'un circuit séquentiel.

Ainsi le fonctionnement d'un système séquentiel passe par une séquence d'états. Cette notion de séquencement impose donc que le système doit posséder des éléments de *mémorisation*. Selon l'état dans lequel se trouve le système, les sorties peuvent être différents pour un même ensemble d'entrées. Il n'est donc pas possible de représenter le système directement par un tableau de vérité ou de Karnaugh car il y aurait plusieurs états possibles dans une même case. Cette approche ne devient possible que lorsque l'on intègre la notion d'états dans ces tableaux.

On distingue les circuits séquentiels *asynchrones* pour lesquels il n'existe pas de référence de temps (c'est à dire que l'action des entrées est prise en compte dès leur changement d'état et il n'y a que le retard intrinsèque des portes logiques qui intervient), des circuits *synchrones* pour lesquels on applique un signal spécial synchronisant que l'on appelle une *horloge* (ici, les entrées sont insensibles aux signaux qui leur sont appliqués sauf pendant un court intervalle de temps déterminé par ce signal d'horloge) – cf Figure 1a.

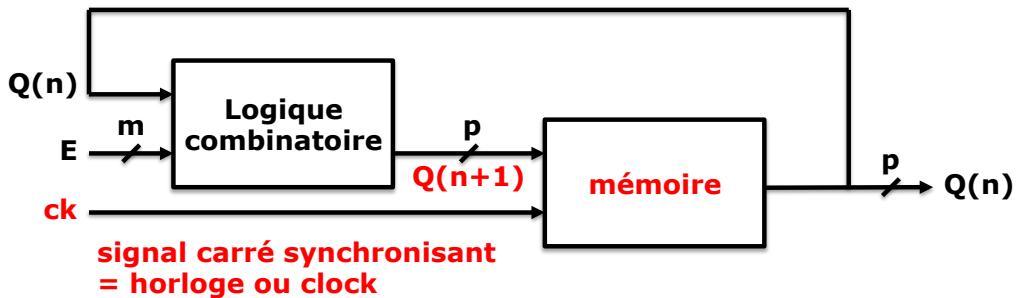


Figure 1a: Schéma bloc d'un circuit séquentiel synchrone.

Dans un circuit séquentiel qui est dans un état S_{t-1} , une modification des entrées peut faire passer le circuit dans un nouvel état S_t (on parle de *transition* de S_{t-1} vers S_t). Ce changement s'effectuera après une durée déterminée par une constante de temps (pour les systèmes asynchrones) ou par l'horloge (pour les systèmes synchrones). L'état du système comprend l'ensemble des variables de sortie et d'entrée.

6.2 Bascules

Une bascule est un circuit électronique réalisant la fonction de mémorisation. Elles représentent la brique de base de tout circuit séquentiel. Plusieurs bascules existent, dont la bascule SR, la bascule D et la bascule JK.

6.2.1 Bascule RS

La bascule RS se compose de deux portes logiques NOR ou de deux portes logiques NAND dont les sorties sont rétro-couplées sur les entrées. Les sorties Q et \bar{Q} représentent deux sorties complémentaires (\bar{Q} est l'inverse de Q), S (Set) permet la mise à 1 de Q , et R (Reset) permet la mise à 0 de Q .

Dans la structure à base de portes NOR (figure 2) :

- **Mise à 0** : $R=1, S=0$. Lorsque R vaut 1, Q vaut 0 quelle que soit la valeur de l'autre entrée sur la porte NOR. S valant 0, la sortie de l'autre porte NOR \bar{Q} vaudra 1.
- **Mise à 1** : $R=0, S=1$. Lorsque S vaut 1, \bar{Q} vaut 0 quelle que soit la valeur de l'autre entrée sur la porte NOR. R valant 0, la sortie de l'autre porte NOR Q vaudra 1.
- **Mémorisation** : $R=0, S=0$. Comme une porte NOR avec l'une des entrées à zéro est équivalent à une fonction d'inversion, Q vaudra l'inverse de \bar{Q} , et \bar{Q} vaudra l'inverse de Q . La bascule est dans un état stable, imposé par la dernière opération (mise à 0 ou mise à 1).
- **Etat interdit** : $R=1, S=1$. Dans cet état, $Q=\bar{Q}=0$. Cet état est possible électriquement, mais la complémentarité des sorties Q et \bar{Q} n'est plus assurée (cela peut poser des problèmes pour les portes logiques en aval de la bascule).

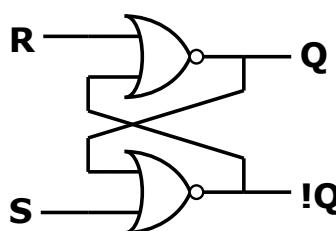


Figure 2: Bascule RS en porte NOR.

Dans la structure à base de portes NAND (cf. figure 3), les entrées S et R sont complémentées en \bar{S} et \bar{R} , qui sont donc des signaux actifs à l'état bas. Cela signifie que pour une activation de l'opération set (mise à 1), $\bar{S}=0$; et pour une activation de l'opération reset (mise à 0), $\bar{R}=0$. Le fonctionnement est le suivant :

- **Mise à 0** : $\bar{R}=0, \bar{S}=1$. Lorsque \bar{R} vaut 0, \bar{Q} vaut 1 quelle que soit la valeur de l'autre entrée sur la porte NAND. \bar{S} valant 1, la sortie de l'autre porte NAND Q vaudra 0.
- **Mise à 1** : $\bar{R}=1, \bar{S}=0$. Lorsque \bar{S} vaut 0, Q vaut 1 quelle que soit la valeur de l'autre entrée sur la porte NAND. \bar{R} valant 1, la sortie de l'autre porte NAND \bar{Q} vaudra 0.
- **Mémorisation** : $\bar{R}=1, \bar{S}=1$. Comme une porte NAND avec l'une des entrées à 1 est équivalent à une fonction d'inversion, Q vaudra l'inverse de \bar{Q} , et \bar{Q} vaudra l'inverse de Q. La bascule est dans un état stable, imposé par la dernière opération (mise à 0 ou mise à 1).
- **Etat interdit** : $\bar{R}=0, \bar{S}=0$. Dans cet état, $Q=\bar{Q}=1$. Cet état est possible électriquement, mais la complémentarité des sorties Q et \bar{Q} n'est plus assurée (cela peut poser des problèmes pour les portes logiques en aval de la bascule).

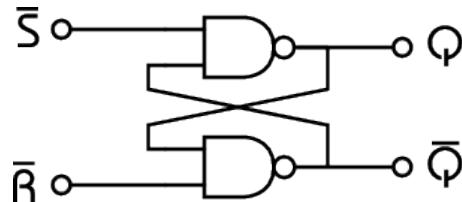


Figure 3: Porte $\bar{R}\bar{S}$ en bascule NAND.

Dans le chronogramme exemple donné figure 4, les opérations de set, reset et mémorisation sont indiquées pour une structure de type NOR, avec passage par l'état interdit.

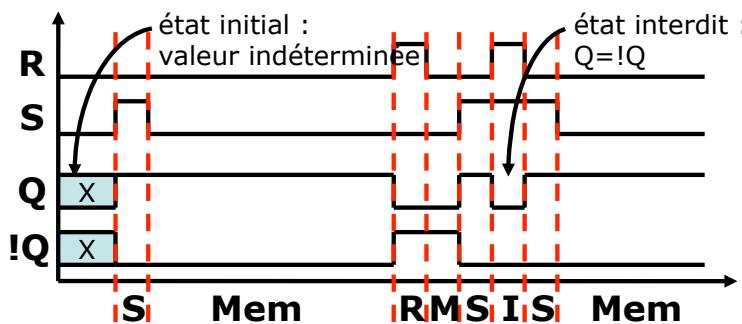


Figure 4: Chronogramme d'une bascule RS.

Q_n est fonction de l'état précédent de \bar{Q}_{n-1} (ou Q_{n-1}) pour la combinaison ($R=0, S=0$). Le passage de S de l'état 0 à l'état 1 met Q à 1 ; le retour à 0 ne change pas la valeur de Q (mémorisation) ; Q et \bar{Q} sont complémentaires. Le passage de l'état 0 à l'état 1 de R met \bar{Q} à 1 et Q à 0 ; le retour à 0 ne change pas la valeur de Q (mémorisation) ; Q et \bar{Q} sont complémentaires. L'état $S=1, R=1$ est interdit (à éviter) car Q et \bar{Q} ne sont plus complémentaires : cette combinaison ne doit pas être utilisée.

Le fonctionnement peut être résumé par le tableau de vérité 1, où Q_{n-1} représente

l'état antérieur, et Q_n l'état présent.

S	R	Q_n	\bar{Q}_n
0	0	Q_{n-1}	\bar{Q}_{n-1}
0	1	0	1
1	0	1	0
1	1	interdit ($=0=\bar{Q}$)	interdit ($=0=Q$)

Tableau 1: Tableau de vérité de la bascule RS.

6.2.2 Verrou-D (D-latch)

Pour contrôler les changements d'état par rapport à un signal indépendant et synchronisant (l'horloge), on rajoutera à la structure de base un étage supplémentaire en amont pour générer les signaux aux entrées des portes NOR. Ainsi la structure devient celle de la figure 5 :

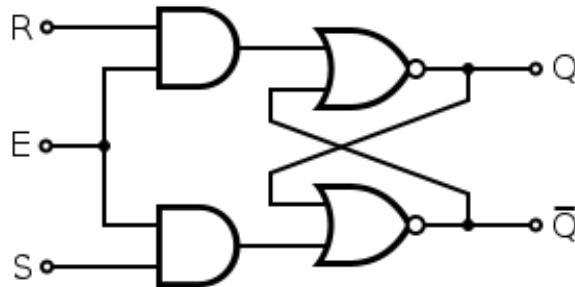


Figure 5: Bascule RS à activation.

Si l'on appelle les signaux internes aux entrées des portes NOR R' ($=R \cdot E$) et S' ($=S \cdot E$), le fonctionnement décrit précédemment pour la structure NOR sera atteint lorsque $E=1$. Si $E=0$, les signaux $R'=S'=0$ et la bascule est en mode mémorisation ($Q_n=Q_{n-1}$). On obtient la tableau de vérité 2.

E	S	R	Q_n	\bar{Q}_n
0	X	X	Q_{n-1}	\bar{Q}_{n-1}
1	0	0	Q_{n-1}	\bar{Q}_{n-1}
1	0	1	0	1
1	1	0	1	0
1	1	1	interdit ($=0=\bar{Q}$)	interdit ($=0=Q$)

Tableau 2: Tableau de vérité bascule RS à activation.

Cette structure est appelée un latch (ou verrou) et fonctionne selon l'état (0 ou 1) du signal de contrôle E. Elle est communément utilisée dans la configuration D-latch, où la donnée D remplacera les signaux R ($=\bar{D}$) et S ($=D$) tel que $R'=\bar{D} \cdot E$ et $S'=D \cdot E$ (cf. figure 6).

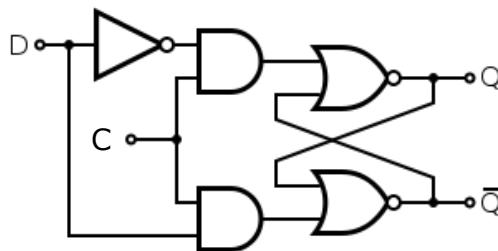


Figure 6: Verrou D.

Il est à noter que la configuration interdite est implicitement éliminée de cette structure comme R' et S' , lorsqu'elles ne valent pas 0, sont en fonction des valeurs complémentées de D.

Dans un souci de minimisation du nombre de transistors (en logique CMOS, une porte ET est constituée en fait d'une porte NAND suivie d'une porte inverseuse, soit 6 transistors). La structure réelle est celle de la figure 7.

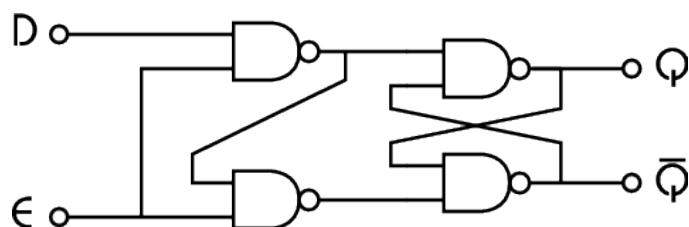


Figure 7: Structure réelle d'un verrou D.

Ici, l'étage de sortie est la structure à base de NAND. Ainsi, si l'on se réfère à la structure de la bascule RS à base de portes NAND, les signaux internes $\bar{S}' = \overline{D \cdot E}$ et $\bar{R}' = \overline{(\overline{D} \cdot E) \cdot E} = \overline{(\overline{D} \cdot E)}$, tel que $S' = D \cdot E$ et $R' = \overline{(\overline{D} \cdot E)}$, comme dans la structure précédente (cf. tableau 3).

E	D	$\overline{S'}$	$\overline{R'}$	Q_n	$\overline{Q_n}$
0	X	X	X	Q_{n-1}	$\overline{Q_{n-1}}$
1	0	1	0	0	1
1	1	0	1	1	0

Tableau 3: Tableau de vérité d'un verrou D.

Un exemple d'utilisation est donné dans le chronogramme de la figure 8. Ici, le signal E est remplacé par un signal horloge (C, comme *Clock*). On peut constater que lorsque C=0, la sortie Q maintient la valeur établie lorsque C est passé de 1 à 0, et que lorsque C=1, la sortie Q suit l'entrée D. On dit du verrou dans ce cas qu'il est *transparent ou passant*.

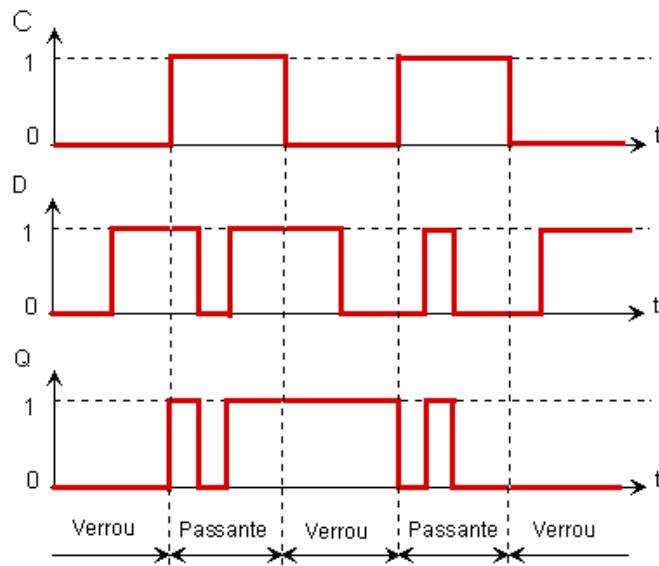


Figure 8: Chronogramme classique d'un verrou D.

Le symbole qui représentera le D-latch est celui de la figure 9.

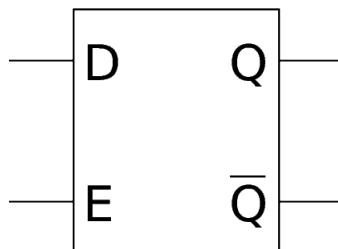


Figure 9: Symbole du verrou D.

6.2.3 Bascule-D (flip-flop)

La transparence du verrou posera des problèmes si l'on l'utilise comme élément de mémorisation dans des systèmes synchrones. En effet, le système synchrone avancera parmi ses états en fonction de la période de l'horloge – il doit donc exister un évènement du signal synchronisant qui déclenchera la transition et non pas un état de ce même signal. Ainsi la bascule-D ne sera pas transparent comme le D-latch pendant la moitié de la période de l'horloge, mais fonctionnera sur un *front* de l'horloge. Ainsi pour une application de mémorisation, le signal d'horloge permet d'autoriser la prise en compte des entrées à un instant donné (front montant ou front descendant). Il devient ainsi possible de reboucler la sortie de la bascule à son entrée (en général à travers de la logique combinatoire comme illustré dans le principe général de la logique séquentielle en début de ce chapitre).

Une bascule-D est constituée généralement de deux verrous synchronisés par un même signal d'horloge mais en opposition de phase. Dans le schéma figure 10, les verrous L1 et L2 sont identiques (noter que l'entrée G, ou Gate, représente l'entrée de contrôle E). Dans le fonctionnement illustré, le verrou L1 sera transparent sur l'état 0 de CLK, et la sortie L1 suit ainsi la valeur de D entre le front descendant de CLK et le front montant. Le verrou L2 sera transparent sur l'état 1 de CLK, et la sortie Q suit donc la sortie de L1 entre le front

montant de CLK et le front descendant. Or, comme le verrou L1 est en état de mémorisation lorsque $CLK=1$, la sortie sera stable. Il s'agit d'une bascule-D qui fonctionne sur front montant, car l'échantillonnage de D sur L1 s'effectue sur le front montant (lorsque L1 passe de l'état transparent à l'état de mémorisation).

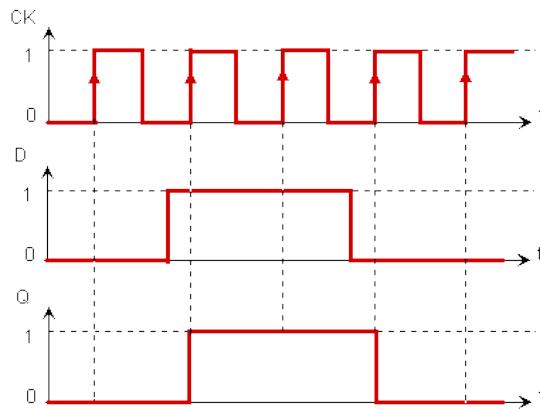
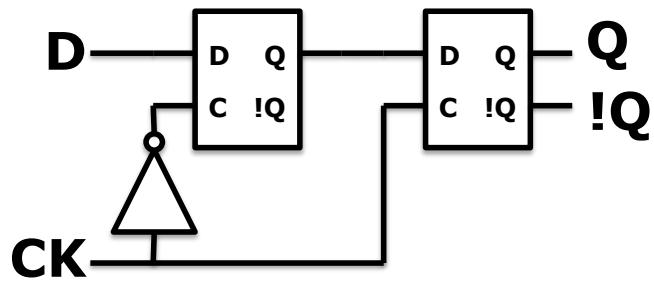


Figure 10: Chronogramme du fonctionnement de la bascule D

Le tableau de vérité 4 montre la dépendance de la sortie Q sur l'évènement du front montant de l'horloge CLK. Il est à noter que la bascule D peut également fonctionner sur front descendant en transmettant CLK non inversé à L1 et CLK inversé à L2.

CLK	D	Q_n	\bar{Q}_n
\uparrow	0	0	1
\uparrow	1	1	0
$0, 1, \downarrow$	X	Q_{n-1}	\bar{Q}_{n-1}

Tableau 4: Tableau de vérité de la bascule D synchrone sur front montant.

6.2.4 Bascule JK

La bascule JK est une bascule universelle dont la structure est donnée figure 11. Elle se comporte comme une bascule typique, sans l'ambiguïté pour la combinaison 1-1 (l'état interdit dans la bascule RS). En effet, pour cette combinaison, il y a inversion des sorties.

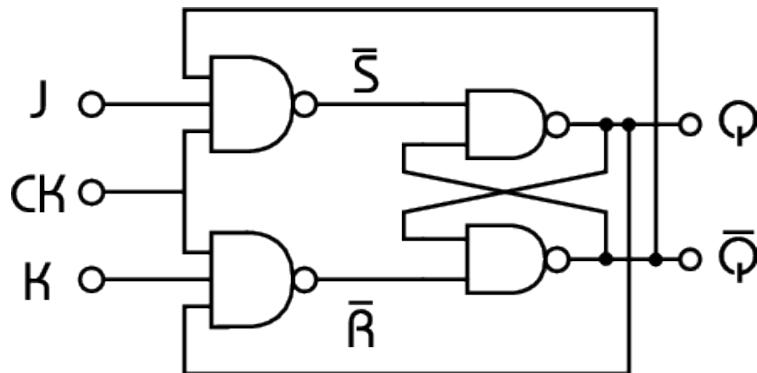


Figure 11: Structure d'une bascule JK.

Dans cette structure, le dernier étage est identique à la bascule RS à base de portes NAND, commandé par les signaux internes \bar{S} et \bar{R} . Le fonctionnement peut ainsi s'expliquer en générant les équations booléennes de \bar{S} et \bar{R} en fonction de J, K, CLK et Q :

$$\bar{S} = \overline{(\bar{Q}_{n-1} \cdot J \cdot CLK)}$$

$$\bar{R} = \overline{(Q_{n-1} \cdot K \cdot CLK)}$$

Ainsi on peut considérer, pour l'état $CLK=1$ (il s'agit donc d'une structure de type verrou, transparent dans cet état), que

$$\begin{aligned}\bar{S} &= Q_{n-1} && \text{si } J = 1 \\ &= 1 && \text{si } J = 0 \\ \bar{R} &= \bar{Q}_{n-1} && \text{si } K = 1 \\ &= 1 && \text{si } K = 0\end{aligned}$$

On peut établir le tableau de vérité (tableau 5).

CLK	J	K	\bar{S}	\bar{R}	Q_n	\bar{Q}_n	Fonction
0	X	X	1	1	Q_{n-1}	\bar{Q}_{n-1}	Mémorisation
1	0	0	1	1	Q_{n-1}	\bar{Q}_{n-1}	Mémorisation
1	0	1	1	\bar{Q}_{n-1}	0	1	Mise à 0
1	1	0	Q_{n-1}	1	1	0	Mise à 1
1	1	1	Q_{n-1}	\bar{Q}_{n-1}	\bar{Q}_{n-1}	Q_{n-1}	Toggle (inversion)

Tableau 5: Tableau de vérité de la bascule JK.

Pour la combinaison $J=0$ $K=1$, $R=\bar{Q}_{n-1}$. La sortie Q_n sera égale à 0 dans tous les cas car si $Q_{n-1}=0$, $\bar{R}=1$ et l'état est maintenu ; et si $Q_{n-1}=1$, $\bar{R}=0$ et la sortie Q est mise à zéro.

Pour la combinaison $J=1$ $K=0$, $\bar{S}=Q_{n-1}$. La sortie Q_n sera égale à 1 dans tous les cas car si $Q_{n-1}=1$, $\bar{S}=1$ et l'état est maintenu ; et si $Q_{n-1}=0$, $\bar{S}=0$ et la sortie Q est mise à un.

Pour la combinaison $J=1$ $K=1$, $\bar{S}=Q_{n-1}$ et $\bar{R}=\overline{Q_{n-1}}$. La sortie Q_n sera inversée par rapport à sa valeur précédente. En effet, si $Q_{n-1}=0$, $\bar{S}=0$ et $\bar{R}=1$ et la sortie Q est mise à un ($Q_n=1$). Si $Q_{n-1}=1$, $\bar{S}=1$ et $\bar{R}=0$ et la sortie Q est mise à zéro ($Q_n=0$). La deuxième sortie respecte la complémentarité.

Il est également possible d'effectuer l'analyse des combinaisons des entrées JK pour les *transitions* voulues de la sortie Q, de l'état antérieur Q_{n-1} vers l'état présent Q_n (tableau 6).

Q_{n-1}	Q_n	J	K
0	0	0	X
0	1	1	X
1	0	X	1
1	1	X	0

Tableau 6: Tableau de transitions de la bascule JK.

Pour la transition $Q_{n-1}=0 \rightarrow Q_n=0$, il s'agit de la mémorisation ($\bar{R}=1$, $\bar{S}=1$). Pour obtenir $\bar{R}=1$, lorsque $Q_{n-1}=0$ et CLK=1, **K=X**. Pour obtenir $\bar{S}=1$, lorsque $\overline{Q_{n-1}}=1$ et CLK=1, **J=0**.

Pour la transition $Q_{n-1}=0 \rightarrow Q_n=1$, il s'agit de la mise à un ($\bar{R}=1$, $\bar{S}=0$). Pour obtenir $\bar{R}=1$, lorsque $\overline{Q_{n-1}}=0$ et CLK=1, **K=X**. Pour obtenir $\bar{S}=0$, lorsque $\overline{Q_{n-1}}=1$ et CLK=1, **J=1**.

Pour la transition $Q_{n-1}=1 \rightarrow Q_n=0$, il s'agit de la mise à zéro ($\bar{R}=0$, $\bar{S}=1$). Pour obtenir $\bar{R}=0$, lorsque $Q_{n-1}=1$ et CLK=1, **K=1**. Pour obtenir $\bar{S}=1$, lorsque $\overline{Q_{n-1}}=0$ et CLK=1, **J=X**.

Pour la transition $Q_{n-1}=1 \rightarrow Q_n=1$, il s'agit de la mémorisation ($\bar{R}=1$, $\bar{S}=1$). Pour obtenir $\bar{R}=1$, lorsque $Q_{n-1}=1$ et CLK=1, **K=0**. Pour obtenir $\bar{S}=1$, lorsque $\overline{Q_{n-1}}=0$ et CLK=1, **J=X**.

Il est à noter que la bascule JK expliquée ici fonctionne sur *état* de CLK. Il existe bien évidemment une structure qui fonctionne sur front de l'horloge, au même titre que la bascule-D par rapport au D-latch.

Le symbole qui représentera la bascule JK est le suivant :

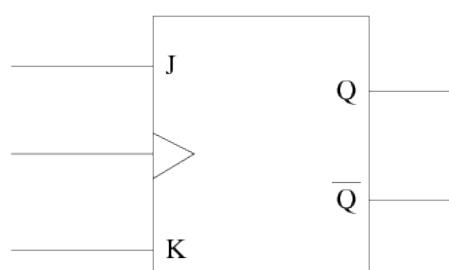


Figure 12: Symbole de la bascule JK.

6.3 Fonctions complexes

6.3.1 Compteurs

Les compteurs, décompteurs et diviseurs (synchrones ou asynchrones) sont des cas spécifiques (et très utilisés) de systèmes séquentiels. En effet, ces circuits doivent générer des séquences de nombres, et comporteront autant d'états que de nombres qu'ils peuvent générer en sortie. Le cas le plus fréquent est le compteur modulo $N=2^n$, où le circuit compte de 0 jusqu'à $N-1$, et n représente le nombre de bits (et de bascules). Par exemple un compteur binaire 8 bits est aussi un compteur modulo 256, soit 256 étapes de 0 à 255.

La démarche générale pour construire un tel système est la suivante :

- identifier les états et d'y affecter un code tel que chaque état possède un code unique. Dans le cas des compteurs / décompteurs / séquenceurs, les combinaisons des sorties sont intrinsèquement uniques et peuvent donc servir comme code d'état. Dans le cas où les sorties ne possèdent pas de valeur unique pour chaque état, il est nécessaire d'utiliser des codes spécifiques internes.
- identifier la séquence des états, ainsi que les conditions nécessaires pour effectuer les transitions entre les états.
- traduire ces conditions en logique combinatoire booléenne pour piloter les entrées de chaque bascule (qui représentent la partie "mémoire" du système).

Ainsi, connaissant l'état à l'instant n , il est nécessaire d'être capable de déclencher l'état $n+1$ attendu. L'état précédent devient donc une entrée du système.

Prenons l'exemple d'un compteur modulo 4 (2 bits). La séquence attendue est représentée dans le tableau d'états (ou tableau de comptage) tableau 6 et dans le chronogramme de fonctionnement figure 13:

Etat / Sortie	Q_1	Q_0
0	0	0
1	0	1
2	1	0
3	1	1

Tableau 7 : Tableau d'états du compteur modulo 4

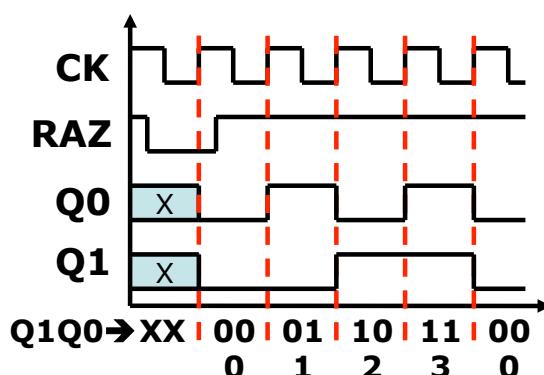


Figure 13: Chronogramme de fonctionnement du compteur modulo-4

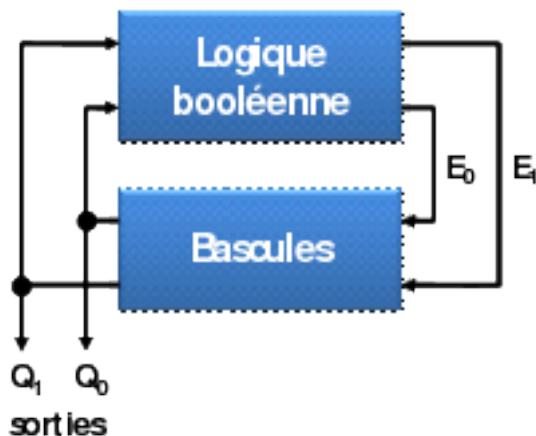


Figure 13a: Structure générale d'un séquenceur d'état 2 bits.

Dans le schéma général de la figure 13a, les sorties $Q_{1(n)}$ $Q_{0(n)}$ de l'état présent sont interprétées par un bloc de logique booléenne afin de générer les nouvelles entrées $E_{1(n+1)}$ $E_{0(n+1)}$ des bascules leur permettant de générer les sorties $Q_{1(n+1)}$ $Q_{0(n+1)}$ de l'état futur. La valeur des entrées doit correspondre au type des bascules (D pour les bascules D, J et K pour les bascules JK).

Comme indiqué en début de ce chapitre, il existe deux types de circuits séquentiels – synchrones et asynchrones.

Les compteurs et décompteurs asynchrones utilisent le principe des diviseurs de fréquence par 2 (où la sortie de la bascule générant le bit de poids inférieur sert également comme signal d'horloge pour la bascule qui suit), montés en cascade. Les bascules doivent ainsi réaliser la fonction "toggle" (ou commutation) – pour une bascule-D, cela signifie que l'entrée D de chaque bascule sera reliée à la sortie de la même bascule \bar{Q} ; pour une bascule JK, cela signifie que les entrées J et K de chaque bascule seront reliées à 1.

Les principaux inconvénients des (dé)compteurs asynchrones (présence d'états indésirables, lenteur) proviennent de l'asynchronisme. L'horloge n'est appliquée qu'à la première bascule (générant le bit de poids le plus faible). Ainsi, il ne peut y avoir de transitions simultanées sur l'ensemble des sorties. Des états indésirables apparaissent pendant le temps de propagation total de l'information de l'horloge à la dernière sortie. Ce temps de propagation maximal correspond à $n \cdot t_p$ (n étant le nombre de bascules, et t_p étant le temps de propagation (t_{PLH} ou t_{PHL}) d'une bascule. On ne peut donc pas les utiliser à des fréquences élevées.

Les compteurs et décompteurs synchrones éliminent l'accumulation des temps de propagation des bascules, en utilisant le même signal d'horloge pour toutes les entrées d'horloge des bascules. Ils peuvent être réalisés à l'aide de bascules D ou JK. Les entrées des bascules sont alors appelées *report* (*retention*). Lorsque la retenue est active, la bascule changera d'état au prochain front d'horloge.

Pour la réalisation en bascules D, le tableau des transitions est donnée dans la tableau 7.

Etat (n)	$Q_{1(n)}$	$Q_{0(n)}$	Etat (n+1)	$Q_{1(n+1)}$	$Q_{0(n+1)}$	D_1	D_0
0	0	0	1	0	1	0	1
1	0	1	2	1	0	1	0
2	1	0	3	1	1	1	1
3	1	1	0	0	0	0	0

Tableau 7: Tableau de transition en bascule D.

Il est ainsi possible d'écrire les équations logiques booléennes permettant de générer D_1 et D_0 :

$$D_0 = \overline{Q_{0(n)}}$$

$$D_1 = Q_{1(n)} \oplus Q_{0(n)}$$

Soit la réalisation physique de la figure 14.

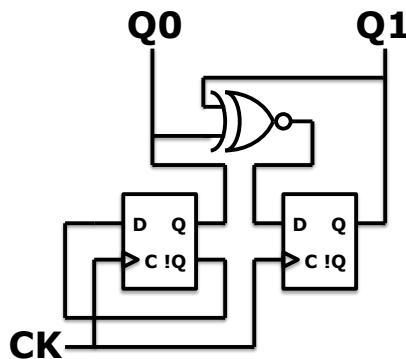


Figure 14: Réalisation physique du compteur modulo-4 (2-bits)

Pour la réalisation en bascules JK, le tableau des transitions est donnée dans la tableau 8.

Etat / Sortie	$Q_{1(n)}$	$Q_{0(n)}$	$Q_{1(n+1)}$	$Q_{0(n+1)}$	J_1	K_1	J_0	K_0
0	0	0	0	1	0	X	1	X
1	0	1	1	0	1	X	X	1
2	1	0	1	1	X	0	1	X
3	1	1	0	0	X	1	X	1

Tableau 8: Tableau de transition en bascule JK.

Il est ainsi possible d'écrire les équations logiques booléennes permettant de générer J_1K_1 et J_0K_0 :

$$J_0 = 1$$

$$K_0 = 1$$

$$J_1 = Q_{0(n)}$$

$$K_1 = Q_{0(n)}$$

Afin de réaliser des compteurs modulo-x, où x représente un nombre quelconque, la même démarche s'applique. Par exemple, pour un compteur modulo-10 :

- on définit le nombre d'états = 10
- on définit le nombre de bits nécessaires (4) pour représenter les 10 états. En effet, $2^3 < 10 < 2^4$ et 4 bits (et bascules) sont nécessaires. Les 6 états superflus ne seront pas utilisés dans le compteur.
- On établit le tableau des états, en commençant par les états n (de 0 à 9) et les codages Q₃-Q₀ associés. Puis l'on établit les états n+1 pour chaque état n (par exemple, pour un état n de 1, l'état suivant n+1 sera de 2), et les codages Q₃-Q₀ associés.
- Pour une réalisation en bascules D, les sorties Q_(n+1) sont directement les entrées D_(n) de la même bascule.

Etat (n)	Q _{3(n)}	Q _{2(n)}	Q _{1(n)}	Q _{0(n)}	Etat (n+1)	Q _{3(n+1)} =D _{3(n)}	Q _{2(n+1)} =D _{2(n)}	Q _{1(n+1)} =D _{1(n)}	Q _{0(n+1)} =D _{0(n)}
0	0	0	0	0	1	0	0	0	1
1	0	0	0	1	2	0	0	1	0
2	0	0	1	0	3	0	0	1	1
3	0	0	1	1	4	0	1	0	0
4	0	1	0	0	5	0	1	0	1
5	0	1	0	1	6	0	1	1	0
6	0	1	1	0	7	0	1	1	1
7	0	1	1	1	8	1	0	0	0
8	1	0	0	0	9	1	0	0	1
9	1	0	0	1	0	0	0	0	0

Tableau 8a: Tableau de transition du compteur modulo-10 en bascule D

- Afin d'établir les relations combinatoires pour les entrées D_{x(n)}, on procède à l'établissement des tableaux de Karnaugh :

		Q1Q0			
		00	01	11	10
D0	00	1	0	0	1
	01	1	0	0	1
	11	X	X	X	X
	10	1	0	X	X
		$D_0 = \overline{Q}_0$			

		Q1Q0			
		00	01	11	10
D2	00	0	0	1	0
	01	1	1	0	1
	11	X	X	X	X
	10	0	0	X	X
		$D_2 = Q_2 (\overline{Q}_1 + \overline{Q}_0) + \overline{Q}_2 Q_1 Q_0$			

		Q1Q0			
		00	01	11	10
D1	00	0	1	0	1
	01	0	1	0	1
	11	X	X	X	X
	10	0	0	X	X
		$D_1 = \overline{Q}_3 (Q_1 \oplus Q_0)$			

		Q1Q0			
		00	01	11	10
D3	00	0	0	0	0
	01	0	0	1	0
	11	X	X	X	X
	10	1	0	X	X
		$D_3 = Q_3 \overline{Q}_1 \overline{Q}_0 + Q_2 Q_1 Q_0$			

Tableau 8b: Tableaux de Karnaugh des entrées des bascules D du compteur modulo-10

- Enfin, on reprend le schéma général du séquenceur d'état (figure 13a) en l'étendant à 4 bits, et en réalisant les relations combinatoires pour générer les entrées des 4 bascules D

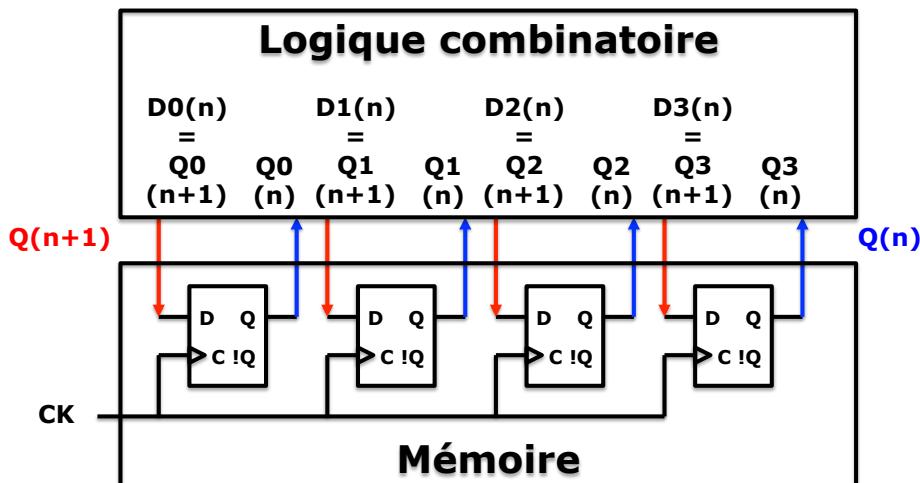


Figure 14b: Réalisation physique du compteur modulo-10

6.3.2 Registres

Un registre est un élément de mémorisation permettant de stocker (en général de manière temporaire) une donnée de taille quelconque. Il s'agit de structures à base de bascules-D mises en parallèle (voir figure 15).

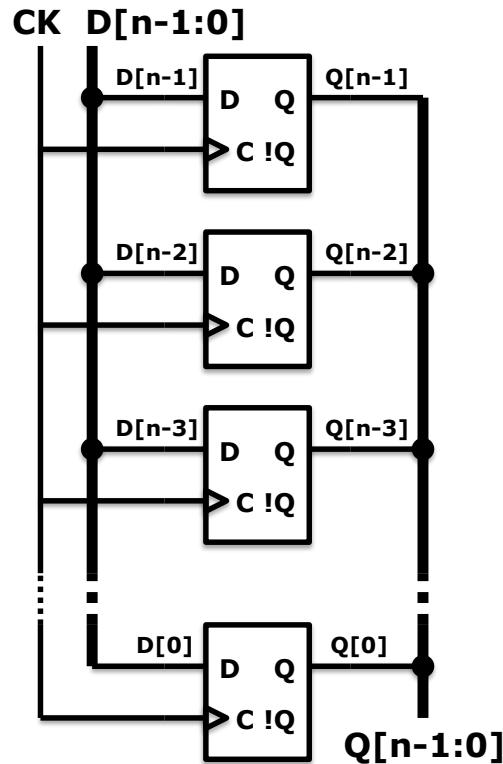


Figure 15: Structure d'un registre

Dans cette structure, un seul signal d'horloge est utilisé pour synchroniser l'ensemble des bascules-D. Un *bus* d'entrée D, comportant n connexions d'un bit, délivre les bits individuels constituant la donnée à stocker dans chaque bascule. La donnée est échantillonnée sur le front montant de l'horloge pour être stockée dans le registre ; un bus de sortie Q, comportant le même nombre de connexions d'un bit, rend disponible la donnée qui a été stockée, synchronisée sur le front de l'horloge. Moyennant les ressources matérielles, cette structure peut être étendue à un nombre très élevé de bits.

Cependant, cette structure ne peut servir que comme "tampon" pour les données – toute donnée D arrivant sur le bus est copiée sur le front montant suivant sur la sortie Q. Il n'y a aucun moyen de contrôler le stockage des données. Afin de rendre possible le contrôle sur le chargement des données, et ainsi permettre l'isolation des ressources de mémorisation du bus d'entrée, il est possible d'utiliser la structure illustrée figure 16.

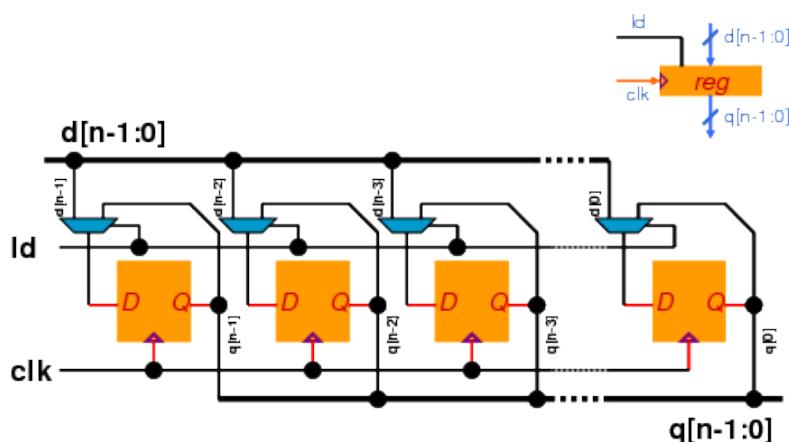


Figure 16: Structure d'un registre à préchargement.

Ici, un multiplexeur, contrôlé par le signal externe LD, est utilisé soit pour connecter le bus d'entrée aux entrées des bascules (fonction de chargement), soit pour isoler des bascules du bus et maintenir la donnée stockée dans le registre (fonction de mémorisation).

6.4 Généralisation : machines d'états

Les machines d'états permettent de décrire des systèmes séquentiels dont l'évolution est plus complexe que les compteurs ou les registres. En effet, dans le schéma général d'un système séquentiel figure 17, et d'après la définition donnée en début de ce chapitre, la sortie du système est fonction des entrées et de son état précédent. Les systèmes séquentiels que nous avons étudiés jusqu'à présent (compteurs) n'avaient pas d'entrées permettant d'intervenir sur la séquence ou sur l'état des sorties.

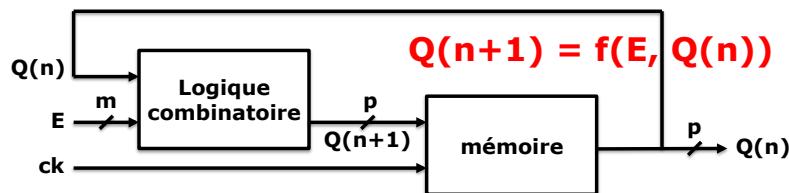


Figure 17: Structure générale d'une machine séquentielle

Le formalisme des machines d'états se retrouve dans des applications diverses : circuits numériques, automatismes industriels, processeurs ou microcontrôleurs, programmes informatiques ... et plus largement, dans les applications telles que les ascenseurs, les feux de circulation, les distributeurs automatiques, les parseurs linguistiques, les protocoles de communication.

6.4.1 Diagrammes d'états

Afin de concevoir ces machines d'états, il est utile de recourir à l'établissement de diagrammes d'états. En effet, toute l'information présente dans un tableau d'état peut être représentée sous forme graphique par un diagramme d'état, et vice versa. Un diagramme d'état ne contient pas plus d'information qu'un tableau d'état mais facilite la compréhension du comportement du circuit.

Dans un diagramme d'état :

- les états sont identifiés par des cercles, avec leur nom et/ou leur code binaire associé écrit dans le cercle
- les transitions entre les états sont identifiées par des flèches entre les cercles
- les conditions pour toute transition (c'est à dire les valeurs nécessaires de variables d'entrée) sont placées à côté des flèches de transition

Pour réaliser les machines d'états, qu'ils soient d'ailleurs en matériel ou en logiciel, il existe deux architectures différentes : la machine de **Moore** et la machine de **Mealy**. L'impact sur la diagramme d'états est le suivant :

- pour les machines de Moore (où si les sorties ne dépendent que de l'état présent), la

valeur des signaux de sortie est placée à l'intérieur des cercles

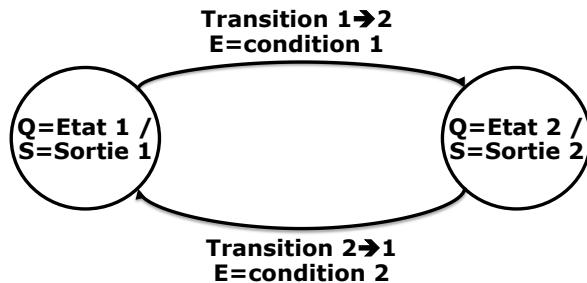


Figure 18a: Diagramme d'états pour une machine de Moore

- pour les machines de Mealy (où les sorties dépendent de l'état présent et des entrées), la valeur des signaux de sortie est placée à côté des flèches de transition - on les sépare des conditions de transition par une barre oblique

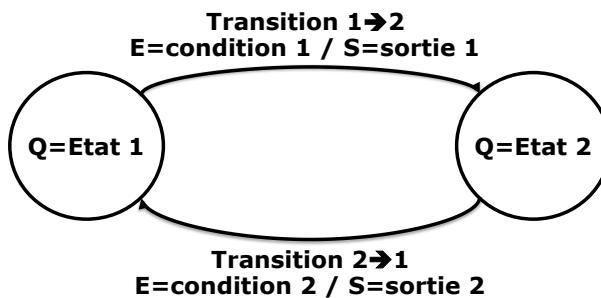


Figure 18b: Diagramme d'états pour une machine de Mealy

6.4.2 Machine de Moore

Dans une machine de Moore, les sorties S dépendent de l'état présent $Q_{(n)}$: les sorties changent de manière synchrone sur un front d'horloge et il n'y a pas de connexion directe entre les entrées E et les sorties S. L'état futur $Q_{(n+1)}$ dépend des entrées E et de l'état présent $Q_{(n)}$.

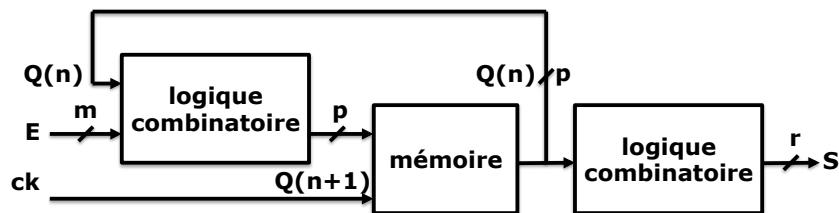


Figure 19a: Structure générale d'une machine de Moore

Un exemple de machine séquentielle est un détecteur de séquence. Ici, nous cherchons à détecter la séquence 1-0. Cette détection ne peut évidemment pas se faire avec un circuit purement combinatoire, car il faut mémoriser le passage du 1 avant le passage du 0.

On procède par l'établissement du diagramme d'états :

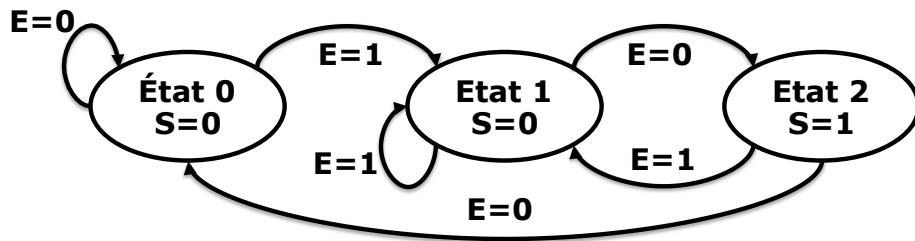


Figure 19b: Diagramme d'état pour un détecteur de séquence 1-0 en machine de Moore

- l'état 0 correspond à l'état initial du système, ou l'état lorsqu'aucun élément de la séquence n'a été détecté. Dans la machine de Moore, la sortie associée à cet état est égale à 0. Les deux états possibles de l'entrée E sont E=0 ou E=1. Si E=0, aucun élément de la séquence n'a encore été détecté et le système reste dans l'état 0. Si E=1, le premier élément de la séquence a été détecté et le système passe à l'état 1.
- Dans l'état 1, on a détecté le premier élément de la séquence. L'état de la sortie est toujours à 0 car on attend le deuxième élément de la séquence. Si E=0, le deuxième élément de la séquence a été détecté et le système passe à l'état 2. Si E=1, il s'agit encore du premier élément de la séquence et le système reste dans l'état 1.
- Enfin, l'état 2 correspond à la détection des deux éléments de la séquence. La sortie vaut 1. Si E=0, le système retournera à l'état 0 ; si E=1, le premier élément de la séquence a été détecté et le système retournera à l'état 1.

Le nombre d'états N définit le nombre de bascules nécessaires à la machine ($\log_2(N)$). Ici, il faudra 2 bascules pour représenter les 3 états du système. Le diagramme d'états permet l'établissement du tableau des états comme vu précédemment. On commence par les états n (de 0 à 2), les codages Q_1-Q_0 associés et l'entrée E. Chaque combinaison utilisée doit être décrite, ce qui nécessite six lignes du tableau. Puis l'on établit les états n+1 pour chaque état n, les codages Q_1-Q_0 associés et la valeur de la sortie S. Comme pour l'exemple précédent du compteur, pour une réalisation en bascules D, les sorties $Q_{(n+1)}$ sont directement les entrées $D_{(n)}$ de la même bascule.

Etat (n)	$Q_{1(n)}$	$Q_{0(n)}$	E	Etat (n+1)	$Q_{1(n+1)} = D_{1(n)}$	$Q_{0(n+1)} = D_{0(n)}$	S
0	0	0	0	0	0	0	0
0	0	0	1	1	0	1	0
1	0	1	0	2	1	0	0
1	0	1	1	1	0	1	0
2	1	0	0	0	0	0	1
2	1	0	1	1	0	1	1

Tableau 9: Tableau des états pour le détecteur de séquence 1-0 en machine de Moore

Une analyse habituelle par tables de Karnaugh permet d'établir les équations combinatoires définissant les entrées des bascules ainsi que la sortie :

$$D_1 = \overline{Q_1} Q_0 \overline{E}$$

$$D_0 = E$$

$$S = Q_1 \overline{Q_0}$$

Et il en résulte le schéma suivant :

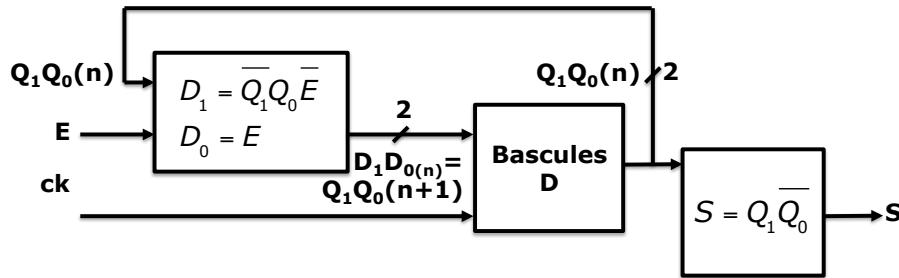


Figure 20: Schéma du détecteur de séquence 1-0 en machine de Moore

6.4.3 Machine de Mealy

Dans une machine de Mealy, les sorties S dépendent de l'état présent $Q_{(n)}$ et des entrées E : les sorties changent de manière asynchrone en raison de la connexion directe entrée-sortie. L'état futur $Q_{(n+1)}$ dépend des entrées E et de l'état présent $Q_{(n)}$.

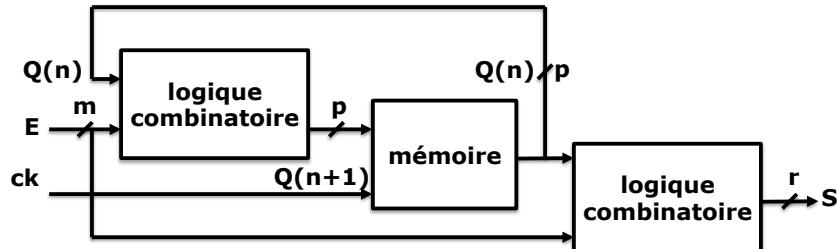


Figure 21a: Structure générale d'une machine de Mealy

Pour la réalisation du détecteur de séquence 1-0, on procède par l'établissement du diagramme d'états :

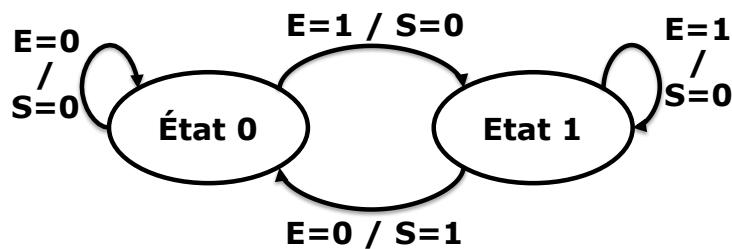


Figure 21b: Diagramme d'état pour un détecteur de séquence 1-0 en machine de Moore

Ici on notera que comme les sorties dépendent des états et des entrées, elles sont

donc associées aux **transitions** (combinaison d'un état et d'une condition sur les entrées) et changent de manière asynchrone. Il s'agit là de la différence fondamentale entre les machines Moore et Mealy et représente la conséquence directe de l'arrivée des entrées E sur la logique combinatoire permettant de générer les sorties S.

En ce qui concerne le détecteur de séquence, il n'existe que deux états :

- l'état 0 correspond à l'état initial du système, ou l'état lorsqu'aucun élément de la séquence n'a été détecté. Si $E=0$, aucun élément de la séquence n'a encore été détecté et le système reste dans l'état 0. La transition état 0 – état 0 est accompagnée par une sortie égale à 0. Si $E=1$, le premier élément de la séquence a été détecté et le système passe à l'état 1, accompagné par une sortie toujours égale à 0.
- Dans l'état 1, on attend le deuxième élément de la séquence. Si $E=0$, le deuxième élément de la séquence a été détecté et le système retourne à l'état 1 en activant la sortie à 1. Si $E=1$, il s'agit encore du premier élément de la séquence et le système reste dans l'état 1 en attendant le deuxième élément, avec la sortie toujours à 0.

Ici, il faudra une seule bascule pour représenter les 2 états du système. Le diagramme d'états permet l'établissement du tableau des états comme vu précédemment. On commence par les états n (de 0 à 1), le codage Q_0 associé et l'entrée E. Puis l'on établit les états n+1 pour chaque état n, le codage Q_0 associé et la valeur de la sortie S.

Etat (n)	$Q_{0(n)}$	E	Etat (n+1)	$Q_{0(n+1)} = D_{0(n)}$	S
0	0	0	0	0	0
0	0	1	1	1	0
1	1	0	0	0	1
1	1	1	1	1	0

Tableau 10: Tableau des états pour le détecteur de séquence 1-0 en machine de Mealy

Une analyse habituelle par tables de Karnaugh permet d'établir les équations combinatoires définissant les entrées des bascules ainsi que la sortie :

$$\begin{aligned} D_0 &= E \\ S &= Q_0 \bar{E} \end{aligned}$$

6.4.3 Comparaison

Les machines de Moore et de Mealy peuvent être fonctionnellement équivalentes. Cependant, comme nous avons pu voir dans l'exemple donné, la machine de Mealy permet une description plus complexe avec un nombre d'états réduit (2 états au lieu de 3 dans l'exemple, ce qui permet de réduire le nombre d'éléments, et donc la surface occupée ainsi que la consommation du circuit). La différence fondamentale réside dans le fait que la machine de Mealy prend en compte immédiatement un changement en entrée, et donc un temps de réponse réduit par rapport à son équivalent de Moore car la machine de Moore n'a pas de chemin combinatoire liant l'entrée et la sortie et doit passer par des éléments

commandés par l'horloge synchronisant. Cela constitue par ailleurs l'avantage de la machine de Moore par rapport à Mealy : les signaux de sortie sont garantis synchrones dans la machine de Moore, alors que le chemin direct dans la machine de Mealy peut conduire à des aléas de fonctionnement qu'il faudra alors gérer.

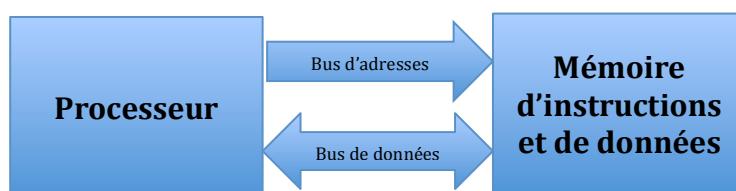
7 Architecture des microprocesseurs

7.1 Structure d'un ordinateur

Un ordinateur est une machine "universelle" capable d'exécuter un algorithme quelconque. La plupart des ordinateurs reposent sur le modèle suivant: ils sont composés d'un **processeur** (Central Processing Unit en anglais, ou CPU, composée lui-même d'un **œur de calcul** associé à une **unité de contrôle**), d'une **mémoire**, d'**entrées** et de **sorties**. Ces éléments sont reliés par un système **d'interconnexion**:

- La mémoire permet de stocker les **données** et les **instructions**. Chaque emplacement mémoire possède une adresse unique et contient, par convention, 1 octet (8 bits).
- Les instructions commandent le fonctionnement du processeur via son unité de contrôle et forment un programme décrivant un algorithme;
- Le système d'interconnexion repose sur un bus bidirectionnel permettant les transferts des données et des instructions. Le bus contient un champ « **adresse** » permettant d'identifier l'emplacement mémoire sélectionné pour un transfert et un champ « **données** » permettant le transfert de la donnée vers/depuis l'emplacement sélectionné. La taille du champ adresse définit la capacité de la mémoire du processeur: pour une architecture 32 bits, le processeur peut adresser 2^{32} emplacements dans sa mémoire et à donc une capacité de stockage de 4Go.

a) Architecture Von Neumann



b) Architecture Harvard

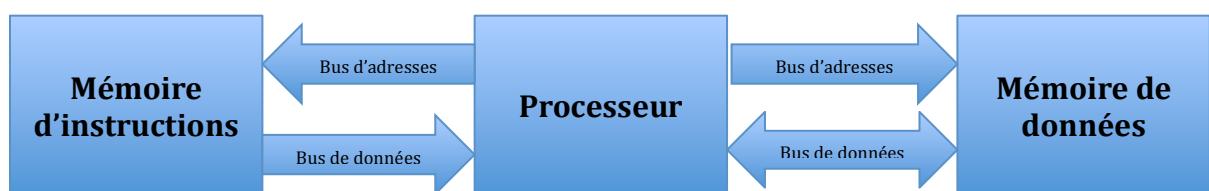


Figure 1: Architectures Von Neumann et Harvard

La Figure 1 représente les deux principales architectures aujourd'hui utilisées :

- L'architecture Von Neumann dispose d'un seul système d'interconnexion reliant le processeur à la mémoire. La mémoire contient donc les instructions et les données. Dans cette architecture, les performances sont limitées car les instructions et les données sont transférées séquentiellement.
- L'architecture Harvard sépare les instructions des données, ce qui permet de réaliser les transferts en parallèle et donc d'augmenter les performances du processeur. Cette architecture dispose de deux systèmes d'interconnexion, ce qui augmente la complexité du processeur.

La Figure 2 représente un exemple d'architecture de type Harvard. Ce processeur dispose de 8 emplacements mémoire (R0 à R7) et le taille des instructions est de 8 bits (cf les ports s0 à s7 issus du registre d'instructions). Ces ports permettent de générer des signaux électriques de commande ; ils sont transmis i) à l'unité de calcul (Unité Arithmétique et Logique, ALU, dans la figure) pour sélectionner l'opération à effectuer et ii) à la mémoire de données pour sélectionner le sens des transferts (lecture / écriture). La séquence des signaux électriques est définie par les instructions exécutées et qui sont stockées dans la mémoire d'instructions.

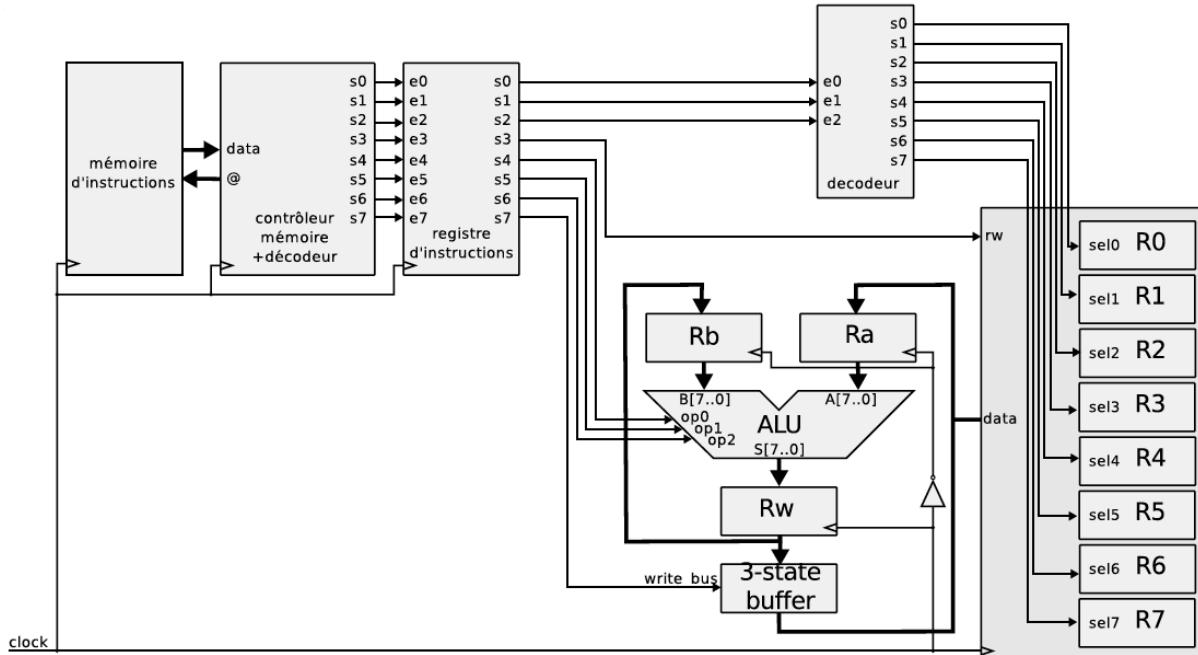


Figure 2: Exemple d'architecture de processeur suivant le modèle Harvard

7.2 Jeux d'instructions

Le jeu d'instructions décrit l'ensemble des opérations élémentaires que le processeur peut exécuter. Il est l'interface entre le code « haut-niveau » écrit dans un langage de programmation indépendant du processeur et l'exécution du code « bas-niveau » spécifique au processeur cible.

7.2.1 Compilation et assemblage

La compilation est le processus permettant de transformer un programme dans un langage évolué vers un programme exécutable par un processeur. Le résultat de la compilation est un programme sous la forme assembleur, au sein duquel chaque ligne correspond à une opération élémentaire disponible dans le processeur. L'assemblage permet de transformer le code assembleur en un code machine, c'est-à-dire un code binaire directement exécutable par le processeur, et de le stocker dans la mémoire d'instructions.

La Figure 3 représente un exemple de compilation et d'assemblage du programme « a = a + b » sur le processeur illustré dans la Figure 2. Dans ce programme, a et b sont des variables que nous supposons stockées dans les emplacements R0 et R1 de la mémoire de données du processeur. L'exécution de l'addition se déroule en trois cycles: 1) la donnée a est transférée vers l'unité de calcul sans qu'aucun traitement ne soit réalisé (instruction « LOAD R0 ») ; 2) la donnée b est transférée vers l'unité de calcul qui réalise en même temps l'addition avec la donnée précédemment chargée (instruction « ADD R1 ») ; 3) Le résultat de

l'opération est transféré vers la mémoire de données (instruction « STORE R0 »). Le code machine est ensuite obtenue et correspond à la suite d'instructions « B8, 89, 30 » en hexadécimal (soit «1011 1000, 1000 1001, 0011 0000 » en binaire).

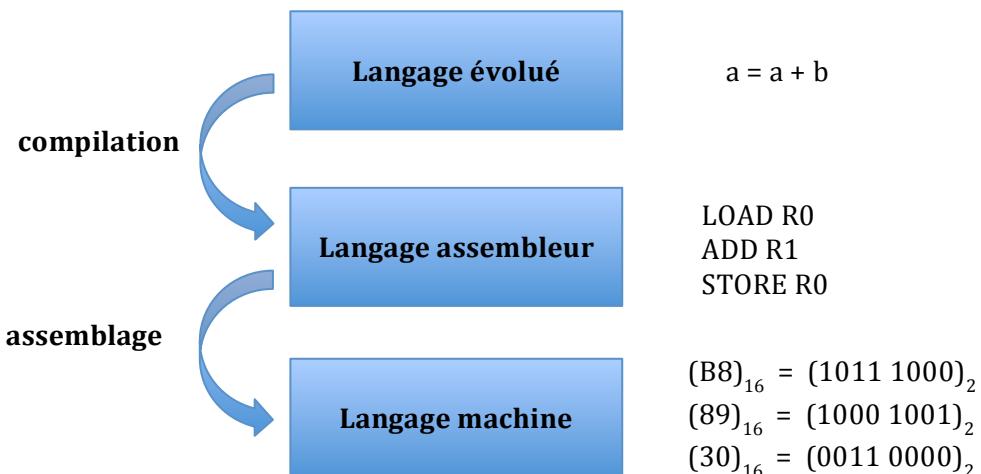


Figure 3: Compilation d'un programme dans un langage évolué vers un code machine

7.2.2 Structure des instructions

La taille de l'instruction (qui est de 8 bits dans l'architecture exemple) est importante pour la complexité du processeur car elle détermine notamment :

- Le nombre et la taille des opérandes contenues dans une instruction;
- la taille des adresses et donc l'espace mémoire adressable ;
- le nombre d'opérations élémentaires du processeur ;

La taille des instructions est contrainte par le coût : elle a un impact direct sur la taille du bus d'interconnexion (nombre de fils), des registres et des unités de calcul.

Une instruction de taille N bits est composée de deux parties (Figure 4) : m bits indiquent l'opération à effectuer (opcode) et N-m bits indiquent les opérandes de l'instruction (les sources pour les emplacements des données à traiter et la destination pour l'emplacement où le résultat doit être stocké).

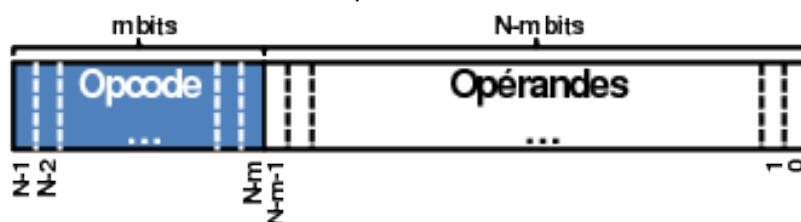


Figure 4: décomposition d'une instruction

Le jeu d'instructions d'un processeur est construit à partir de sa structure interne et des signaux de contrôle issus du registre d'instruction (le mot de contrôle). Les 8 signaux de contrôle (s0 à s7) du processeur de la Figure 2 sont listés dans la Figure 5 :

- s2, s1 et s0 permettent de sélectionner un emplacement dans la mémoire (via le bus d'adresse de la mémoire de données) ;
- s3 indique le sens du transfert sur le bus de données : 1 pour une lecture de la mémoire (le contenu de l'emplacement mémoire sélectionné est transféré sur le bus), 0 pour une écriture (la donnée sur le bus est stockée dans l'emplacement mémoire sélectionné) ;
- s6, s5 et s4 permettent de sélectionner l'opération réalisée par l'unité de calcul. Les

opérandes sont les données contenues dans les registres locaux Ra et Rb et le résultat est stocké dans Rw. Ra est une copie de la donnée présente sur le bus. Rb est une copie de la donnée stockée dans Rw.

- s7 permet de transférer le contenu de Rw sur le bus de communication.

s7	s6	s5	s4	s3	s2	s1	s0
Transfert unité de calcul → mémoire de données	Sélection de l'opération		Transfert mémoire de données → unité de calcul	Sélection de l'emplacement mémoire			
0 écriture sur le bus 1 pas d'écriture sur le bus	000 Rw = Ra + Rb 001 Rw = Ra - Rb 010 Rw = Ra x Rb 011 Rw = Ra 100 Rw = Ra / Rb 101 Rw = !Ra 110 Rw = Ra + Rb + 1 111 Réserve		1 lecture de la mémoire (écriture sur le bus) 0 écriture dans la mémoire (lecture du bus)	000 R0 001 R1 010 R2 011 R3 100 R4 101 R5 110 R6 111 R7			

Figure 5: signaux de contrôle du registre d'instructions

A partir des combinaisons possibles des signaux de contrôle, le jeu d'instructions du processeur est défini. Dans cet exemple, il est composé de 7 instructions (Figure 6), chacune étant codée sur 8 bits et permettant de définir un seul opérande Rx (source ou destination). La plupart des processeurs modernes disposent d'un jeu d'instructions plus complet et permettent de réaliser des calculs logiques (e.g. NAND, NOR, décalage à gauche) et de contrôler l'exécution du programme (e.g. appels de fonctions, conditions if/then/else, boucles).

Instruction (assembleur)	Description	Code machine en binaire	Code machine en hexadécimal (donné pour Rx = R0)
LOAD Rx	Transfert d'une donnée de la mémoire vers le registre Rw de l'unité de calcul	1011 1xxx	B8
STORE Rx	Transfert du registre Rw vers la mémoire	0011 0xxx	30
ADD Rx	Addition entre Rw et une donnée de la mémoire puis stockage du résultat dans Rw	1000 1xxx	88
SUB Rx	Soustraction ...	1001 1xxx	98
MULT Rx	Multiplication ...	1010 1xxx	A8
DIV Rx	Division ...	1100 1xxx	D8
ADDINC Rx	Addition et incrémentation ...	1110 1xxx	E8

Figure 6: jeu d'instructions

Des architectures plus complexes permettent de définir dans une même instruction plusieurs sources et destinations, ce qui réduit le nombre d'instructions nécessaires pour l'exécution d'un programme mais augmente la complexité du processeur. Dans l'exemple de la Figure 7, une instruction est codée sur 15 bits : 7 bits permettent de sélectionner l'opération et 9 bits permettent de sélectionner jusqu'à 3 opérandes (Rd pour la destination, Rs1 pour la source 1 et Rs2 pour la source 2). NB : le processeur permettant d'exécuter une telle instruction est différent de celui représenté dans la Figure 2.

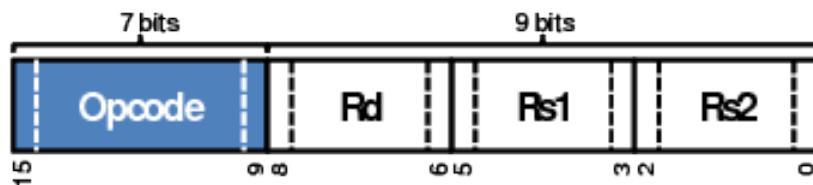


Figure 7: Exemple d'instruction avec trois opérandes (une destination Rd et deux sources Rs1 et Rs2)

7.2.3 Types d'adressage

Il existe plusieurs façons de transmettre des données à l'unité de calcul:

- **Adressage Immédiat** : La méthode la plus simple car la donnée est spécifiée directement dans l'instruction (Figure 8). Les valeurs possibles sont limitées par le nombre de bits dédiés aux opérandes dans l'instruction. Dans l'exemple de la Figure 7, 9 bits sont dédiés aux opérandes, c'est-à-dire que seules 2^9 valeurs différentes peuvent être transmises.



Figure 8: Adressage immédiat

- **Adressage Direct** : l'adresse de la donnée est transmise dans l'instruction (Figure 9). La taille de l'adresse est contrainte par le nombre de bits de l'instruction dédié aux opérandes : dans l'exemple précédent, 2^9 adresses différentes peuvent être transmises. Ce mode d'adresage est celui utilisé dans le processeur de la Figure 2 : les 3 bits s0, s1 et s2 du mot de contrôle (Figure 5) permettent de sélectionner une donnée parmi 2^3 emplacements dans la mémoire de données (R0 à R7).

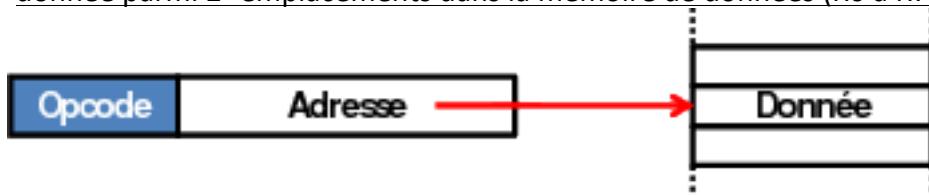


Figure 9: Adressage direct

- **Adressage indirect** : l'adresse de la donnée est contenue dans l'adresse spécifiée dans l'instruction (Figure 10). Ce mode d'adresage est plus flexible que l'adresage direct car la donnée peut se trouver à n'importe quel endroit de la mémoire (il n'est pas nécessaire d'utiliser un emplacement mémoire spécifique).

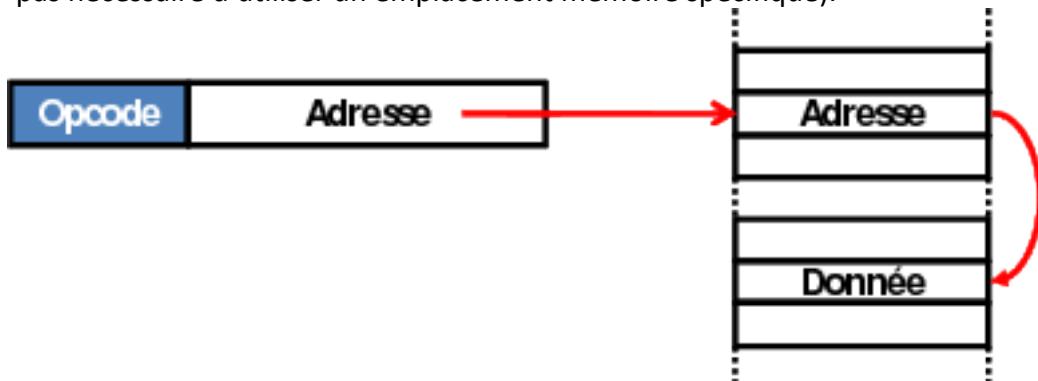


Figure 10: Adressage indirect

7.2.4 Types de jeu d'instruction (RISC et CISC)

Deux grandes familles de jeu d'instruction existent : le CISC (Complete Instruction Set Computer) et le RISC (Reduced Instruction Set Computer).

CISC :

Le jeu d'instruction CISC permet de rendre indépendant l'architecture du processeur du logiciel. Il s'agit d'un jeu d'instruction riche, avec une grande variété des modes d'adressage, permettant la construction d'instructions complexes et de longueurs variables. Cette richesse d'instructions permet de simplifier l'étape de compilation et d'améliorer les performances du compilateur. On rapproche ainsi le langage haut-niveau du langage machine. Le code machine généré est réduit, et par conséquent la place occupée en mémoire par le programme.

Cependant, la richesse du jeu d'instructions résulte également en des instructions de longueur variable, souvent avec un seul opérande en mémoire. Il existe également un grand nombre de modes d'adressage, nécessaire pour refléter la richesse des structures de données en langage haut niveau. Les accès mémoire nécessitent ainsi une gestion complexe et coûteuse, et la complexité des instructions et des modes d'adressage impose un séquenceur microprogrammé.

RISC :

En langage machine, les types de fonction exécutables par un processeur sont l'affectation, l'itération, l'appel de procédure et le branchement (conditionnel ou non). Dans les programmes classiques, l'affectation représente 60% des instructions. Les appels de procédure représentent 20% des instructions d'un langage haut-niveau, mais 60% du temps d'exécution des instructions machines et 70% des accès mémoire. Ces chiffres reflètent l'importance des paramètres et de leur passage entre la mémoire et le processeur. De plus, 75% des références mémoires portent sur des données simples (constantes ou variables scalaires). Ainsi, une architecture RISC vise une optimisation globale matérielle et logicielle par rapport à l'architecture CISC, pour tirer le meilleur parti de cette synergie. Pour atteindre cette optimisation, l'architecture RISC réalise trois modifications principales :

- Diminution de la partie contrôle. Le séquenceur est réalisé en logique câblée, et sera donc plus rapide.
- En conséquence de la diminution de la surface du séquenceur, une place plus importante peut être attribuée aux registres. Cette approche donne donc la possibilité de séparer les données et les instructions même au niveau interne du composant.
- Enfin, les modes d'adressage et les instructions sont simplifiés (avec une longueur fixe). Le compilateur est plus simple et l'accélération de l'exécution est plus efficace.

7.3 Exécution des instructions

7.3.1 Fonctionnement général

Le programme est stocké dans la mémoire d'instructions que le processeur doit donc parcourir en suivant la séquence prévue lors de la compilation. Les interactions sont gérées par le contrôleur mémoire qui, pour une architecture Harvard, possède un accès direct à la mémoire via un bus d'adresse et un bus de données (Figure 11-a). Le contrôleur mémoire

dispose d'un registre interne appelé PC (program counter, ou compteur ordinal) qui contient l'adresse en mémoire de la prochaine instruction à exécuter. Le processeur envoie cette adresse à la mémoire (port @ dans la figure), qui renvoie alors le contenu de l'emplacement mémoire correspondant (c'est-à-dire l'instruction) sur le port data. L'instruction reçue est décodée puis transférée au registre d'instructions pour être exécutée (voir Section 7.2). En parallèle, le registre PC est mis à jour et l'instruction suivante est récupérée, etc. Les instructions étant stockées de manière contigüe dans la mémoire, le déroulement logique d'un programme s'appuie principalement sur l'incrémentation de la valeur contenue dans PC.

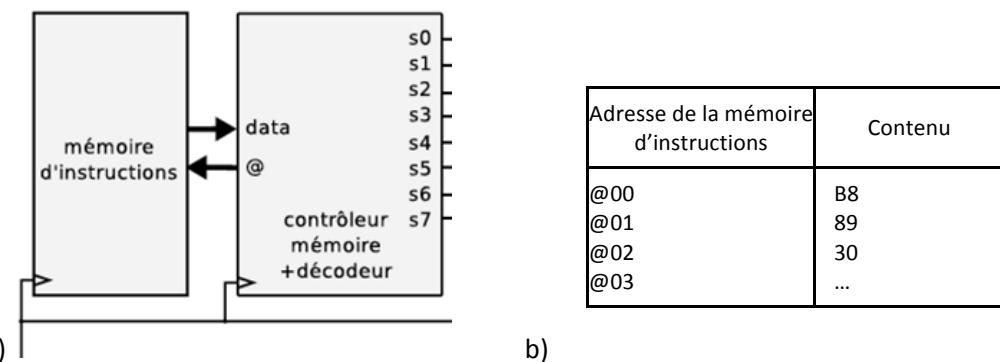


Figure 11: le contrôleur mémoire permet de gérer l'exécution du programme contenu dans la mémoire d'instructions : a) extrait de la Figure 2 et b) exemple de contenu dans la mémoire.

7.3.2 Exemple de déroulement de l'exécution d'un programme

Dans cet exemple, nous considérons le code machine correspondant au programme « $a = a + b$ » de la Figure 3 stocké dans la mémoire d'instructions à partir de l'adresse de base, c'est-à-dire @00 (Figure 11-b). L'exécution du programme étant cadencée par l'horloge du processeur, il est possible de décomposer, cycle par cycle, les valeurs sur les ports @ et data du contrôleur mémoire :

- Cycle 1 : @ = 00 ; data = inconnu (data dépend de la dernière instruction exécutée)
- Cycle 2 : @ = 01 ; data = B8 (la mémoire envoie l'instruction 1 cycle après la requête)
- Cycle 3 : @ = 02 ; data = 89 ...
- Cycle 4 : @ = 03 ; data = 30 ...

Le programme s'arrête lorsqu'une instruction de fin de programme est exécutée. Le temps de calcul peut donc s'exprimer en nombre de cycles d'horloge ou être exprimée sous la forme d'une durée si la fréquence est connue. La fréquence maximale d'un processeur est fortement dépendante de sa structure et des paramètres technologiques.

References

- [14] Cazes A., Delacroix J., *Architectures des machines et systèmes informatiques*, DUNOD, ISBN 2100493833, 004.22 CAZ. La plupart des figures de ce chapitre sont issues de ce livre.
- [15] Tanenbaum Andrew, *Architecture de l'ordinateur*, Pearson Education, ISBN 2744073776, 004.22 TAN
- [16] Stallings William, *Organisation et architecture de l'ordinateur*, 6e Edition, Prentice Hall, ISBN 2-7440-7007-6, 004.22 STA

ANNEXE

7.5 Rappel de physique du semiconducteur

7.5.1 Potentiel de surface

Considérons deux matériaux différents, typiquement deux semiconducteurs, ou un semiconducteur et un métal. Sans polarisation externe, lorsque les deux matériaux sont mis en contact (cf. figure 161), du fait que leur concentration est normalement différente en J_1 et J_2 , un flot de porteurs traverse le plan de jonction des deux matériaux.

De ce fait, les porteurs quittant un matériau laissent en place une charge électrique équivalente opposée, et créent un champ électrique à travers le plan de jonction qui tend à inhiber le mouvement des électrons et des trous. Ainsi, si un électron quitte J_1 pour aller en J_2 , il laisse une charge positive en J_1 telle que la contribution de cette charge positive au champ électrique est dans une direction qui attire les électrons.

Il est donc facile d'admettre qu'un état d'équilibre s'établisse, tel que le champ électrique généré s'oppose à tout mouvement de porteurs. Le potentiel électrique correspondant est par définition le potentiel de contact entre J_1 et J_2 . Ce potentiel de contact est une caractéristique propre des matériaux considérés. Ainsi (Si étant du silicium intrinsèque) : $\Phi_{Cu,Si} = 0.0V$, $\Phi_{Al,Si} = 1.35V$, $\Phi_{Au,Si} = -0.3V$, ...

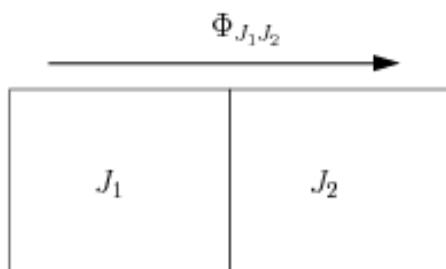


Figure 161: Régimes de fonctionnement de la capacité MOS.

7.5.2 Potentiel de Fermi

Dans le cas particulier où J_1 est un semiconducteur extrinsèque et J_2 est un semiconducteur intrinsèque, le potentiel de contact prend le nom de potentiel de Fermi, et il s'exprime formellement par

$$\Phi_F = Vt \times \log \frac{n_i}{n_o} = Vt \times \log \frac{p_o}{n_i}$$

Ainsi, si J_1 est un matériau de type P, le potentiel de Fermi a pour valeur

$$\Phi_{Fp} \approx +Vt \times \log \frac{NA}{n_i}$$

et pour un matériau de type N

$$\Phi_{Fn} \approx -Vt \times \log \frac{ND}{n_i}$$

Typiquement, le potentiel de Fermi a comme ordre de grandeur quelques dixièmes de Volt à température ambiante.

Tension de bande plate

Initialement (c'est à dire sans aucun apport de tension externe), la capacité MOS n'est pas neutre en terme de potentiel électrostatique. En effet, de par la superposition de matériaux de nature différente, il existe un potentiel de contact qui comme nous le savons ne dépend que des terminaisons, ce potentiel de contact est noté $\Phi_{MS} = \Phi_{substrat} - \Phi_{grille}$.

D'autre part, des charges parasites sont inévitablement présentes au sein de la structure MOS. Ces charges ont des origines très diverses : procédés de fabrication, défauts d'interface, contaminations accidentelles, introduction volontaire (...). Elles sont regroupées sous le terme générique "charge d'interface effective surfacique" Q'_0 et localisées à l'interface oxyde/semiconducteur. Elles génèrent un potentiel électrostatique

$$\Phi_0 = \frac{Q'_0}{C_{ox'}}$$

Ainsi, par définition, la tension de bande plate V_{FB} est la tension qu'il faut placer aux bornes du condensateur MOS pour neutraliser les deux potentiels électrostatiques précédemment définis

$$V_{FB} = \Phi_{MS} - \frac{Q'_0}{C_{ox'}}$$

7.5.3 Étude dans le cas général

Lorsque l'on applique une tension V_{GB} quelconque aux bornes du condensateur MOS, par influence électrostatique, des charges apparaissent dans le semiconducteur. Ces charges sont situées au voisinage de l'interface isolant semiconducteur (au dessus du trait d'axe sur la figure 11), et au-delà, le substrat peut être considéré comme neutre avec une densité d'électrons n_0 et une densité de trous p_0 telles que $p_0 = n_0 + N_A$. Si nous considérons un point de coordonnée x dans le substrat, et notons $\Phi(x)$ son potentiel par rapport au substrat, en terme de densité d'électrons, nous avons $n(x) = n_0 e^{\Phi(x)/V_t}$, et en terme de densité de trous, $p(x) = p_0 e^{-\Phi(x)/V_t}$.

Ainsi, la densité de charge s'écrit $\rho(x) = q(p(x) - n(x) - N_A)$, partant de l'équation de Poisson

$$\frac{d^2\Phi}{dx^2} = -\frac{q}{\epsilon_{SI}} (p_0 e^{-\Phi(x)/V_t} - n_0 e^{\Phi(x)/V_t} - N_A)$$

on peut déterminer la charge surfacique dans le semiconducteur :

$$Q_C' = \pm \sqrt{2qN_A\epsilon_{SI}} \sqrt{V_t e^{-\Phi_s/V_t} + \Phi_s - V_t + e^{-2\Phi_s/V_t} (V_t e^{\Phi_s/V_t} - \Phi_s - V_t)}$$

V_{GB} : Le signe négatif est à utiliser lorsque Φ_s est positif, Φ_s étant défini comme le potentiel de surface comme indiquée sur la figure 11. Quant à la charge de grille, elle est simplement liée au potentiel aux bornes de l'oxyde par la relation $Q_G' = C_{ox'} \Phi_{ox}$.

Si maintenant on considère d'une part, la boucle de tension formée par la tension externe, le potentiel aux bornes de l'oxyde et le potentiel de surface telle que $V_{GB} = \Phi_{ox} + \Phi_s + \Phi_{MS}$ et d'autre part la neutralité de la charge de la structure MOS qui implique $Q_G' + Q_C' + Q_0' = 0$, on obtient un jeu de quatre équations contenant quatre inconnues (Φ_{ox} , Φ_s , Q_G' , Q_C') qui ne peut être résolu dans le cas général que numériquement.

Pour obtenir un formalisme analytique simple, on peut distinguer essentiellement trois modes de fonctionnement limites, correspondants à différentes valeurs de tension V_{GB} :

- le mode déserté avec V_{GB} "tout juste" supérieure à V_{FB}
- le mode inversé avec V_{GB} "nettement" supérieure à V_{FB}
- le mode accumulé avec V_{GB} inférieure à V_{FB} .