

COURS DE TELECOMMUNICATION

IUT DE VILLETANEUSE
Département Génie Télécom et Réseaux
A. DUPRET, A. FISCHER

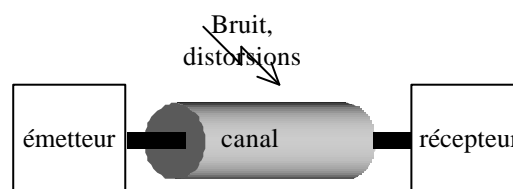
CODAGE DE L'INFORMATION : **CANAL DE TRANSMISSION**

INTRODUCTION AU TRAITEMENT DE L'INFORMATION

Généralités

Lorsqu'il s'agit de transmettre des informations, plusieurs cas peuvent se présenter :

D'une part, il faut soit effectuer une liaison point à point entre un émetteur et un récepteur (téléphonie), soit diffuser l'information à partir d'un émetteur vers plusieurs récepteurs (radio/télé diffusion). En fonction du type de liaison (point à point ou diffusion), des contraintes économiques et matérielles, s'effectue le choix du media de transmission (câble coaxial ou paires torsadées, fibres optiques, vide ou air pour les communications Hertziennes, etc...). Le media de transmission, au quel s'ajoutent les perturbations et déformations (bruits, diaphonie, distorsions...) de l'information, est appelé canal.



Il est fréquent de devoir transmettre plusieurs informations simultanément au travers d'un même canal. Pour ce faire, il est nécessaire de recourir au multiplexage.

CODAGE DE L'INFORMATION : **MODULATIONS ANALOGIQUES**

GENERALITES

Au cours du développement des dispositifs de télécommunication, il est rapidement apparu indispensable de coder l'information à transmettre, soit pour adapter l'information au canal de transmission (fibre optique, câble coaxial, faisceaux hertziens...), soit pour transmettre simultanément plusieurs signaux informatifs sur un seul et même canal. De ce fait, le codage de l'information s'est révélé être un point-clef qui fait aujourd'hui encore l'objet de recherches et de normalisation.

L'une des formes de codage de l'information parmi les plus simples et les plus anciennes consiste à effectuer une translation en fréquence du signal informatif. Ce type de codage est appelé modulation. Il est d'usage de distinguer trois types de modulations analogiques :

Modulation d'amplitude	AM (Amplitude Modulation)	
Modulation de phase	PM (Phase Modulation)	} Ces deux modulations sont des modulations angulaires
Modulation de fréquence	FM (Frequency Modulation)	

Le fonctionnement de ces trois modulations repose sur la modification d'une des caractéristiques (fréquence, phase ou amplitude) d'un signal sinusoïdal haute fréquence qui est transmis tel quel en l'absence de signal informatif. Ce signal prend la dénomination de porteuse (carrier en anglais) et sa fréquence est appelée fréquence porteuse. Elle sera notée f_p (la pulsation porteuse $\omega_p = 2 \cdot \pi \cdot f_p$).

Par ailleurs, dans tous les cas, nous considérerons que le signal à transmettre ($u(t)$ signal informatif ou signal utile) est à spectre fini. A titre d'exemple, sont présentées ci-dessous quelques signaux courants, le type de modulation utilisé et la bande passante du signal informatif :

signal\caratéristique	type de modulation	bande passante de $u(t)$	porteuse f_p	puissance
radio grandes ondes	AM	4,5 kHz	150→285 kHz	1→2 MW
radio locale	FM	15 kHz	88→108 MHz	1→2 W
télévision (bandes IV/V)	AM (image)	6 MHz	470→860 MHz	10→50 W

1. MODULATION D'AMPLITUDE

Généralités

Des techniques de modulations analogiques, la modulation d'amplitude fut la première employée. Comme nous le verrons par la suite, elle se caractérise par une grande simplicité de mise en œuvre. On la trouve fréquemment pour les transmissions hertziennes (stations radiophoniques grandes ondes par exemple).

Description du signal AM. Cas d'un signal sinusoïdal

Signal AM dans le domaine temporel

Comme son nom l'indique, un signal $s(t)$ (courant ou tension) modulé en amplitude est un signal constitué par une porteuse sinusoïdale de fréquence f_p dont l'amplitude A_p est modifiée suivant une loi linéaire par le signal informatif $u(t)$. Si nous prenons le cas d'un signal modulant sinusoïdal, l'expression de $s(t)$ est donc :

$s(t) = A_p \cdot \cos(\omega_p t) + k \cdot A_m \cdot \cos(\omega_m t) \cdot \cos(\omega_p t)$ où k est le facteur de proportionnalité du modulateur. k est parfois appelé sensibilité du modulateur. La grandeur de k dépend des grandeurs de A_m et A_p .

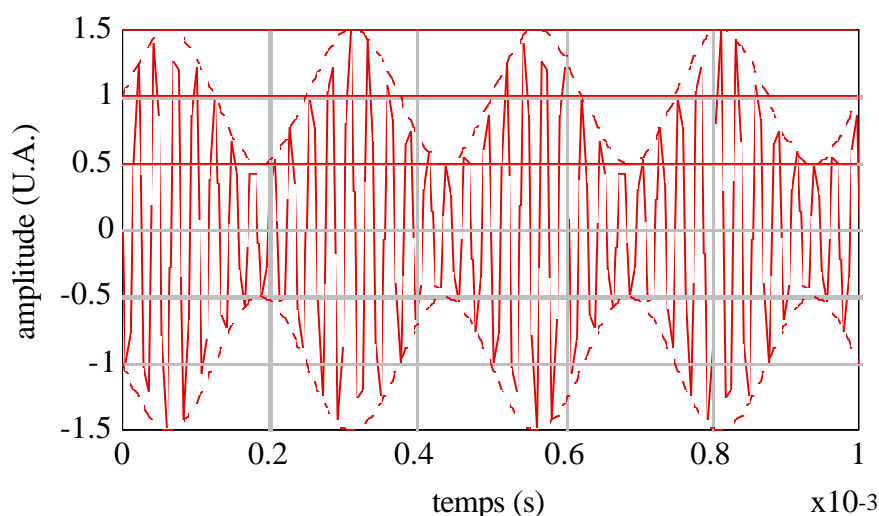
$$\text{soit : } s(t) = A_p \cdot \left(1 + \frac{k \cdot A_m}{A_p} \cdot \cos(\omega_m t) \right) \cdot \cos(\omega_p t) = A_p \cdot (1 + m \cdot \cos(\omega_m t)) \cdot \cos(\omega_p t)$$

Dans le cas général l'expression d'un signal modulé en amplitude est :

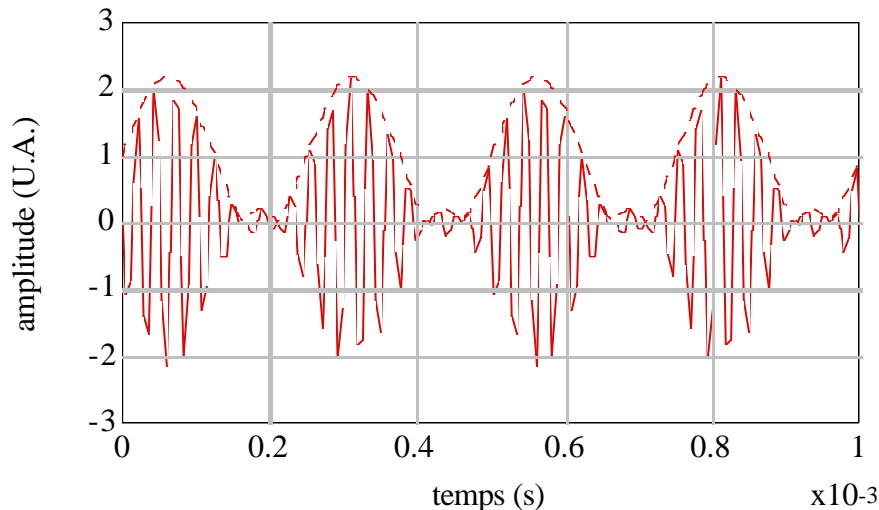
$$s(t) = A_p \cdot (1 + m \cdot u(t)) \cdot \cos(\omega_p t).$$

où m est un paramètre essentiel appelé taux de modulation. Il est d'usage d'exprimer m en %.

La représentation temporelle de $s(t)$ est illustrée par la figure 1.



signal sinusoïdal modulé en amplitude. La porteuse est à $f_p=162$ kHz, l'amplitude $A_p=1$ (U.A.) et le taux de modulation est $m=50\%$. Le signal modulant $u(t)$ est à une fréquence de 4 kHz.



signal sinusoïdal modulé en amplitude. La porteuse est à $f_p=162$ kHz, l'amplitude $A_p=1$ (U.A.) et le taux de modulation est $m=120\%$. Le signal modulant $u(t)$ est à une fréquence de 4 kHz.

La courbe en pointillés est appelée l'enveloppe de $s(t)$. Dans le cas où m est inférieur à 1 ($m < 100\%$), l'enveloppe de $s(t)$ est identique à $u(t)$.

m est défini par le rapport entre l'excursion en amplitude et l'amplitude en l'absence de signal informatif ($u(t)=0$) :

$$m = \frac{\Delta s}{2 \cdot A_p} \quad \text{ou encore :} \quad m = \frac{s_{\text{Max}} - s_{\text{min}}}{s_{\text{Max}} + s_{\text{min}}}$$

Signal AM dans le domaine fréquentiel

Le signal modulé $s(t)$ prend la forme :

$$s(t) = A_p \cdot (1 + m \cdot u(t)) \cdot \cos(\omega_p t)$$

Supposons que le signal utile soit de la forme $u(t) = \cos(\omega_m t)$. En développant l'équation précédente nous avons successivement :

$$s(t) = A_p \cdot (1 + m \cdot \cos(\omega_m t)) \cdot \cos(\omega_p t)$$

soit :

$$s(t) = A_p \cdot \cos(\omega_p t) + m \cdot A_p \cdot \cos(\omega_m t) \cdot \cos(\omega_p t)$$

$$s(t) = A_p \cdot \left(\cos(\omega_p t) + \frac{m}{2} \left(\cos((\omega_p + \omega_m)t) + \cos((\omega_p - \omega_m)t) \right) \right)$$

Le spectre du signal est donc composé de 3 raies. L'une d'amplitude A_p à la fréquence f_p , les 2 autres sont d'amplitude $\frac{m}{2} A_p$ aux fréquences $f_p - f_m$ (appelée bande latérale inférieure ou, dans la littérature anglo-saxonne Lower Side Band : LSB) et $f_p + f_m$ (appelée bande latérale supérieure ou Upper Side Band : USB). La bande passante requise pour transmettre le signal $u(t)$ en préservant son intégrité est appelée bande de fréquence B et vaut donc :

$$B_{pAM} = B_f = 2 f_m.$$

remarque : il aurait été possible de passer par la transformée de Fourier de $s(t)$. On aurait eu :

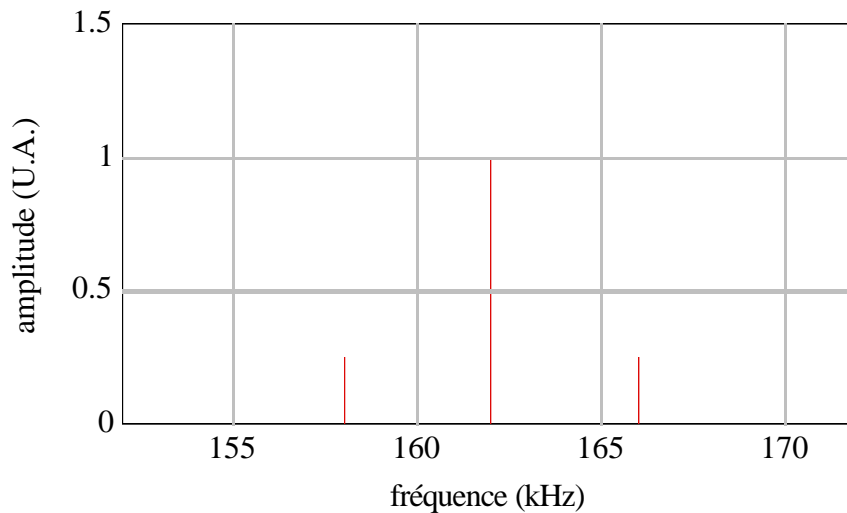
$$\underline{S}(\omega) = TF(s(t)) = TF\left((1 + m \cdot \cos(\omega_m t)) \cdot A_p \cdot \cos(\omega_p t)\right) = TF\left((1 + m \cdot \cos(\omega_m t)) * TF(A_p \cdot \cos(\omega_p t))\right)$$

$$\underline{S}(\omega) = \frac{A_p}{2} \cdot \left(d(\omega - \omega_p) + \frac{m}{2} (d(\omega - \omega_p - \omega_m) + d(\omega - \omega_p + \omega_m)) \right) \\ + \frac{A_p}{2} \cdot \left(d(\omega + \omega_p) + \frac{m}{2} (d(\omega + \omega_p - \omega_m) + d(\omega + \omega_p + \omega_m)) \right)$$

Cela fait apparaître un spectre de part et d'autre de $0 \text{ rad}\cdot\text{s}^{-1}$ (un pour les pulsations négatives, un pour les pulsations positives). Lorsqu'on effectue une mesure, les deux spectres sont superposés. On obtient alors :

$$\underline{S}(\omega) = A_p \cdot \left(d(\omega - \omega_p) + \frac{m}{2} (d(\omega - \omega_p - \omega_m) + d(\omega - \omega_p + \omega_m)) \right)$$

La représentation spectrale de $|\underline{S}(\omega)|$, module de la transformée de Fourier de $s(t)$, appelée spectre de $s(t)$, est donnée par la figure ci-dessous :



spectre du signal $s(t)$ (porteuse à 162 kHz, signal modulant à 4 kHz, $m=50\%$)

Puissance

La puissance nécessaire à la transmission du signal est obtenue en élevant le signal au carré. S'il s'agit d'une tension (d'un courant), il suffit de diviser (multiplier) par la résistance de charge, souvent constituée par l'antenne de l'émetteur. A un facteur constant près, la modulation AM en courant ou en tension est donc identique.

Calculons la puissance moyenne du signal $s(t)$. Nous nommerons P_p la puissance en l'absence de signal modulant ($u(t)=0$).

Pour cela nous prendrons le cas d'un signal modulant $u(t)$ sinusoïdal. Le signal $s(t)$ est périodique de période T . On a alors :

$$P = \alpha \int_T s^2(t) \cdot dt \text{ où } \alpha \text{ est une impédance si } s(t) \text{ est un courant et une admittance si } s(t)$$

est une tension.

$$P = \alpha \int_T \left[(1 + m \cdot \cos(\omega_m t)) (A_p \cdot \cos(\omega_p t)) \right]^2 \cdot dt$$

en développant, on obtient :

$$P = a \int_T \left(1 + 2 \cdot m \cdot \cos(\mathbf{w}_m t) + m^2 \cdot \cos^2(\mathbf{w}_m t) \right) \cdot A_p^2 \cdot \cos^2(\mathbf{w}_p t) \cdot dt$$

Les termes en $\cos^2(x)$ peuvent être mis sous la forme $(1 + \cos(2x))/2$, puis en développant l'équation précédente et en remarquant que les termes en $\cos(2x)$ de l'intégrale sont nuls, on obtient :

$$P = a \frac{A_p^2}{2} \frac{(2 + m^2)}{2}$$

$$P = a \left(\frac{A_p}{\sqrt{2}} \right)^2 \left(1 + \frac{m^2}{2} \right)$$

Le terme en $P = a \frac{A_p^2}{2}$ représente la puissance en l'absence de signal modulant. C'est la puissance qui est constamment nécessaire pour transmettre au moins la porteuse. Le terme en $P = a \frac{m^2}{2} \frac{A_p^2}{2}$ est la puissance effectivement utilisée pour le signal informatif $u(t)$. Notons que le signal $u(t)$ est "présent" deux fois : dans la bande latérale inférieure d'une part mais également dans la bande inférieure.

On peut donc écrire :

$$P_t = P_p \left(1 + \frac{m^2}{2} \right) \text{ et } P_{LSB} = P_{USB} = P_p \frac{m^2}{4}.$$

Si l'on considère maintenant le courant envoyé, par exemple à une antenne d'impédance R , il est possible d'établir une relation entre puissances et courants, donc :

$$\frac{P_t}{P_p} = \frac{\frac{1}{2} R \cdot I_t^2}{\frac{1}{2} R \cdot I_p^2} = \left(\frac{I_t}{I_p} \right)^2 = \left(1 + \frac{m^2}{2} \right) \text{ d'où : } I_t = I_p \sqrt{1 + \frac{m^2}{2}}$$

Par un raisonnement identique : $U_t = U_p \sqrt{1 + m^2/2}$

Cas de plusieurs signaux modulants :

Si une porteuse est modulée par plusieurs signaux sinusoïdaux de fréquences $(f_{m1}, f_{m2}, f_{m3}, \dots, f_{mn})$, chacun étant caractérisé par sa propre amplitude, à chacun correspond un taux de modulation $(m_1, m_2, m_3, \dots, m_n)$. L'opération de modulation AM va alors revenir à la superposition de signaux aux fréquences $(f_p \pm f_{m1}, f_p \pm f_{m2}, f_p \pm f_{m3}, \dots, f_p \pm f_{mn})$ plus la porteuse à la fréquence f_p . Par un raisonnement identique au précédent, on définit un taux de modulation total m_t tel que :

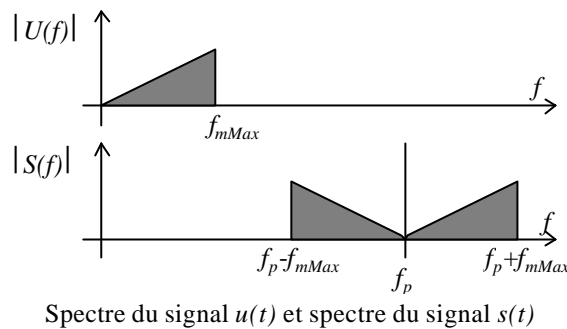
$$m_t^2 = \sum_{i=1}^n m_i^2 \text{ d'où } m_t = \sqrt{\sum_{i=1}^n m_i^2}$$

Cas d'un signal modulant quelconque :

Dans le cas où $u(t)$ est un signal quelconque, il nous faut reprendre quelques définitions. En général, $u(t)$ est un signal aléatoire. On doit alors faire appel à des notions relativement complexes pour déterminer la représentation spectrale de $u(t)$ et de $s(t)$. Il faut alors utiliser le

concept de Densité Spectrale de Puissance, c'est-à-dire, la transformée de Fourier de la fonction d'auto-corrélation du signal.

Retenons que le spectre présente alors l'allure suivante :



Modulateurs

La génération d'un signal AM sort du domaine de ce cours mais il suffit de savoir qu'il existe un grand nombre de multiplicateurs électroniques. En effet les diodes et les transistors bipolaires ont des caractéristiques courant/tension exponentielles. Une multiplication revient alors à une conversion exponentielle puis l'addition des deux signaux (c-à-d., la somme de deux courants), suivie de la conversion logarithmique. Les modulateurs les plus fréquemment rencontrés sont :

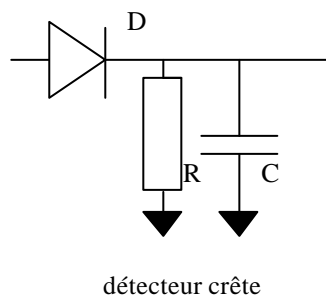
- les modulateurs en anneaux constitués de 4 diodes et de transformateurs (ces multiplieurs sont principalement employés en HF),
- multiplieurs à découpage,.
- multiplieurs de Gilbert, 2 ou 4 quadrants, employant des transistors bipolaires.

Démodulation

La démodulation est l'opération qui consiste à retrouver le signal $u(t)$ à partir du signal $s(t)$. Dans le cas de la modulation d'amplitude, 2 techniques peuvent être utilisées.

Démodulation par détection d'enveloppe

Lorsque le taux de modulation m est inférieure à 1, l'enveloppe de $s(t)$ est identique au signal modulant $u(t)$. Le procédé le plus simple consiste à extraire l'enveloppe de $s(t)$. Pour cela, il suffit de remarquer que l'enveloppe est constituée tous les $T_p = 1/f_p$ par le maximum du signal $s(t)$. La réalisation électronique d'un tel dispositif est simple puisqu'il s'agit d'un détecteur crête dont le schéma de principe est donné ci-dessous.



Le principe en est simple : lorsque la diode est passante, le condensateur est chargé à la tension du signal d'entrée. Dès que le signal d'entrée décroît la diode se bloque, car la tension aux bornes du condensateur devient supérieure à la tension du signal d'entrée, le condensateur ne pouvant se décharger que "lentement" dans la résistance. Tant que la diode est bloquée, la diode se décharge lentement dans la résistance. Ce type de démodulation ne peut-être appliquée quand le taux de modulation m reste supérieur à 1.

Démodulation cohérente ou par détection synchrone.

Pour aborder l'autre technique de démodulation consiste à remarquer que le spectre du signal modulé $s(t)$ comprend le spectre du signal $u(t)$ transposé en fréquence. La démodulation va donc consister à effectuer un changement de fréquence sur le signal $s(t)$ de façon à faire réapparaître le signal modulant $u(t)$ aux basses fréquences tandis que le signal de la porteuse sera rejeté en haute fréquence. Pour cela $s(t)$ est multiplié par un signal de même fréquence et de même phase que la porteuse. Le résultat de la multiplication est au signal intermédiaire $i(t)$.

De ce fait :

$$i(t) = s(t) \times \cos(\omega_p t)$$

$$i(t) = \left[(1 + m \cdot u(t)) \times A_p \cos(\omega_p t) \right] \times \cos(\omega_p t)$$

En développant le terme en $\cos^2(x)$, on obtient :

$$i(t) = (1 + m \cdot u(t)) \times A_p \frac{1 + \cos(2\omega_p t)}{2}$$

En filtrant le signal $i(t)$ à l'aide d'un filtre passe-bas, le terme en $\cos(2\omega_p t)$ est supprimé et il ne reste plus que le terme en $u(t)$ auquel s'ajoute un terme continu. Le signal multipliant $s(t)$ étant rigoureusement en phase avec la porteuse, on parle alors de détection synchrone

Voyons maintenant le cas où le signal modulé n'est pas rigoureusement en phase avec la porteuse du signal modulé :

$$i(t) = s(t) \times \cos(\omega_p t + \mathbf{j}) \quad \text{où } \mathbf{f} : \text{déphasage entre la porteuse et le terme en } \cos(\omega_p t).$$

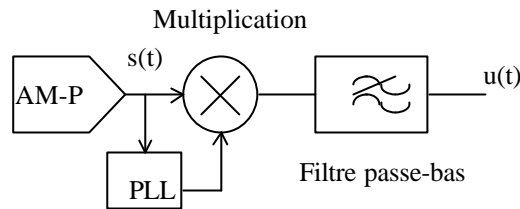
$$i(t) = \left[(1 + m \cdot u(t)) \times A_p \cos(\omega_p t) \right] \times \cos(\omega_p t + \mathbf{j})$$

En développant le produit $\cos(\omega_p t) \cos(\omega_p t + \mathbf{f})$, on obtient :

$$i(t) = (1 + m \cdot u(t)) \times A_p \frac{\cos(\mathbf{j}) + \cos(2\omega_p t + \mathbf{j})}{2}$$

Après le filtrage, le déphasage \mathbf{f} entre la porteuse et le terme en $\cos(\omega_p t)$ se traduit par le fait que le terme en $u(t)$ est multiplié par un terme en $\cos(\mathbf{f})$. Si la phase est constante mais aléatoire, le terme démodulé peut être nul ou proche de zéro (phase \mathbf{f} voisine de $\mathbf{p}/2 + k\mathbf{p}$). Si la phase varie dans le temps, le signal $u(t)$ se trouve modulé par un signal parasite. Cette technique de démodulation impose donc de disposer d'un signal rigoureusement en phase avec la porteuse. Or, en pratique, il n'existe pas d'oscillateur suffisamment stable en fréquence pour rester constamment synchrone avec la porteuse du signal émis. Pour réaliser une démodulation

synchrone, la technique employée consiste à générer un signal en phase avec la porteuse à partir du signal modulé. Cette technique fait appel à un dispositif appelé PLL (Phase Lock Loop ou Boucle à Verrouillage de Phase) qui permet de synthétiser un signal en phase avec la porteuse. Le schéma de démodulation revient à celui présenté ci-dessous :



Principe de la démodulation synchrone

Principe de la PLL :

Comme son nom l'indique, la PLL est un système asservi en phase. En d'autre terme, il s'agit d'un dispositif qui compare la phase du signal d'entrée à celle du signal produit par son oscillateur commandé en tension (VCO, Voltage Controlled Oscillator). Si l'écart de phase entre les signaux d'entrée et sortie n'est pas nulle, la tension d'entrée du VCO varie ce qui entraîne une variation de fréquence du signal présent en sortie du VCO. La fréquence du VCO ayant changée, la phase du signal de sortie peut "rattraper" celle du signal d'entrée.

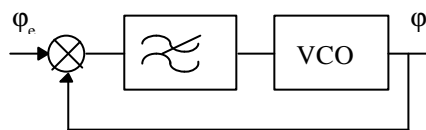


Schéma de principe de la PLL

Un des avantages d'un tel système est de ne pas être tributaire de la forme des signaux. De plus, la théorie des systèmes asservis indique qu'un tel dispositif se comporte comme un filtre passe-bas. Le système à PLL va donc mettre un certain temps à réagir une variation de la phase du signal d'entrée. De ce fait, si le signal d'entrée subit de fortes perturbations rapides, le signal de sortie restera constant. A la limite le filtrage passe-bas joue comme un effet mémoire : si j_e disparaît un court instant, j_s conservera une valeur constante pendant quelque temps et le signal démodulé sera peu affecté.

Modulation d'amplitude sans porteuse

Principe

Nous avons vu que la modulation d'amplitude se traduit par la transmission d'une raie à fréquence de la porteuse. Cette raie contenant la porteuse ne contient aucune information utile, la puissance utilisée pour la transmettre n'est apparemment pas indispensable. La modulation sans porteuse, appelée AM-P (ou DSB-SC : Double Side Band Suppressed Carrier), consiste à ne transmettre que les bandes latérales. Lors de la démodulation, il faudra pouvoir reconstituer la porteuse pour restituer le signal modulant $u(t)$.

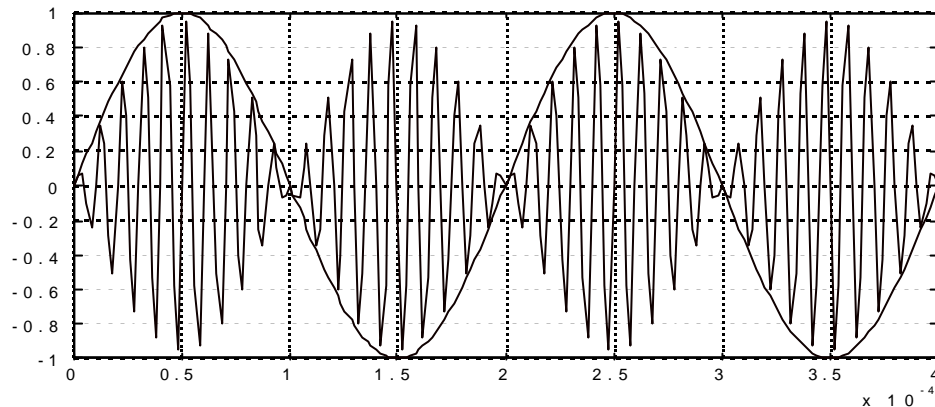
En modulation AM nous avons : $s(t) = A_p \cdot \cos(w_p t) + A_p \cdot m \cdot u(t) \cdot \cos(w_p t)$. La suppression de la porteuse, dans le spectre uniquement, revient à supprimer le terme. La

génération d'un signal AM-P revient donc à produire un signal $s(t)$ de la forme :

$$s(t) = A_p \cdot u(t) \cdot \cos(\omega_p t).$$

Représentation temporelle d'un signal AM-P

La représentation temporelle de $s(t)$ est simple à tracer à partir du signal modulant $u(t)$ puisque les valeurs de $s(t)$ sont toutes comprises entre $u(t)$ et $-u(t)$. Dans le cas où $u(t)$ est un signal sinusoïdal, nous avons :



signal sinusoïdal modulé en amplitude sans porteuse.

La porteuse est à $f_p=100$ kHz, l'amplitude $A_p=1$ (U.A.). Le signal modulant $u(t)$ est une sinusoïde de fréquence 4 kHz.

L'enveloppe de $s(t)$ est constituée par la valeur absolue de $u(t)$. A chaque changement de signe de $u(t)$, la porteuse change de signe. Une autre façon de voir les choses est de dire qu'à chaque changement de signe de $u(t)$, la phase de la porteuse varie de π .

Spectre d'un signal AM-P

En suivant un raisonnement analogue à celui qui nous avait permis de déterminer le spectre du signal AM, c'est-à-dire en prenant un signal $u(t) = \cos(\omega_m t)$ nous déduisons le spectre du signal AM-P :

$$s(t) = A_p \cdot \cos(\omega_m t) \cdot \cos(\omega_p t)$$

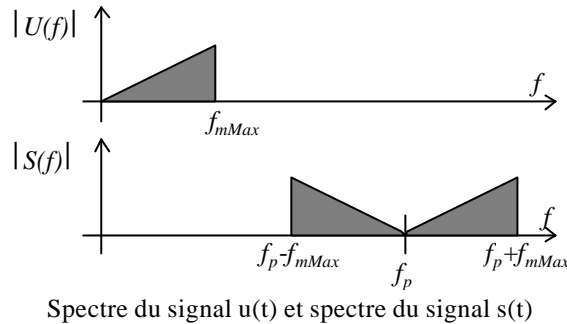
soit :

$$s(t) = A_p \cdot \frac{m}{2} \left(\cos((\omega_p + \omega_m)t) + \cos((\omega_p - \omega_m)t) \right)$$

Le spectre du signal est donc composé de 2 raies. Toutes les 2 sont d'amplitude $\frac{m}{2} A_p$ aux fréquences $f_p - f_m$ (appelée Lower Side Band LSB) et $f_p + f_m$ (appelée Upper Side Band USB). La bande passante requise pour transmettre le signal $u(t)$ en préservant son intégrité est donc :

$$B_{pAM} = B_f = 2f_m.$$

Dans le cas général, le spectre d'un signal modulé AM-P est le suivant :



Puissance

L'amélioration du rendement en puissance est sensible puisque toute la puissance émise est consacrée au signal. Toutefois, il faut noter que le signal utile est transmis deux fois : une dans la LSB et l'autre dans la USB, chaque bande contenant : $a \frac{A_p^2}{4}$.

Démodulation d'un signal AM-P

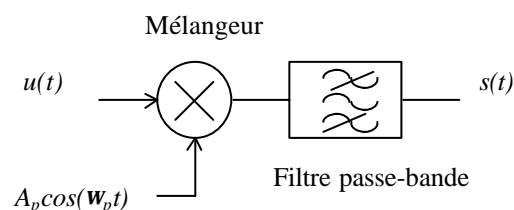
L'enveloppe de $s(t)$ n'étant pas une fonction bijective du signal modulant, la démodulation du signal AM-P ne peut se faire que par démodulation synchrone. Du fait de la complexité de produire un signal rigoureusement en phase avec la porteuse et ses conséquences sur la qualité signal démodulé, la modulation AMP-P n'est pas utilisée pour la transmission de signaux audio.

Modulation à Bande Latérale Unique

Il a été vu précédemment que, pour les modulations AM et AM-P, les deux bandes latérales sont porteuses de la même information, $u(t)$. Il est donc envisageable de n'en transmettre qu'une des deux, ce qui permet d'une part de réduire la bande de fréquence allouée pour transmettre le signal, et d'autre part, de réduire la puissance à émettre pour transporter la même quantité d'information. Ce type de modulation est appelé Modulation à Bande Latérale Unique ou BLU (SSB : Single Side Band).

Principe de la modulation en Bande Latérale Unique

La technique la plus simple et la plus communément employée pour obtenir un signal en modulation SSB consiste à réaliser une AM-P puis à filtrer l'une ou l'autre des deux bandes. Pour conserver l'USB (respectivement la LSB), il serait théoriquement possible de n'utiliser qu'un filtre passe-haut (respectivement passe-bas). Néanmoins, en pratique des filtres passe-bande sont employés afin de réduire la puissance de bruit. Le schéma de principe d'un modulateur SSB est donné ci-dessous :



Principe de la modulation SSB

Prenons un signal $u(t) = \cos(\omega_m t)$ dont nous déduirons le spectre du signal SSB. Nous appellerons $i(t)$ le signal intermédiaire modulé AM-P :

$$i(t) = A_p \cdot \cos(\omega_m t) \cdot \cos(\omega_p t)$$

soit :

$$i(t) = A_p \cdot \frac{m}{2} \left(\cos((\omega_p + \omega_m)t) + \cos((\omega_p - \omega_m)t) \right)$$

Pour obtenir une SSB-USB, on introduit un filtre passe-bas chargé de supprimer toutes les composantes fréquentielles de $i(t)$ supérieures à f_p . Après filtrage il ne reste que :

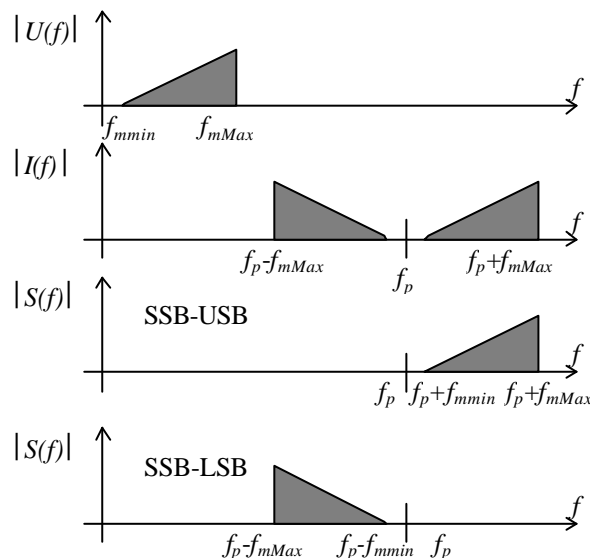
$$s(t) = A_p \cdot \frac{m}{2} \cdot \cos((\omega_p + \omega_m)t)$$

Le spectre du signal est donc composé d'une seule raie. En fait, on constate que dans le cas de la SSB-USB, le spectre du signal $s(t)$ est simplement formé du spectre du signal $u(t)$ décalé en fréquence de f_p (Attention, ce n'est pas le cas pour la SSB-LSB puisque les composantes hautes fréquences de $u(t)$ deviennent les composantes basses pour $s(t)$).

La bande passante requise pour transmettre le signal $u(t)$ en préservant son intégrité est donc :

$$B_{pAM} = B_f = f_m.$$

Dans le cas général, le spectre d'un signal modulé SSB est le suivant :



Spectre du signal $u(t)$ et spectre du signal $s(t)$

Notons que l'introduction du filtre passe-bande entraîne des contraintes sur la forme de $u(t)$. En effet, il ne sera pas possible de réaliser des filtres dont les pentes soient infiniment raides. De ce fait, comme le filtre passe-bande ne doit laisser passer que des signaux dans la bande $f_p < B < f_p + f_{mMax}$, il est nécessaire que $u(t)$ ne contiennent pas de composantes très basses fréquences.

Nous avons vu que la modulation d'amplitude SSB revient à effectuer un changement de fréquence. Le procédé de démodulation sera donc, comme pour la modulation AM-P, une démodulation synchrone. Nous raisonnerons en nous appuyant sur le cas d'un signal $u(t)$ sinusoïdal. Nous avons donc :

$i(t) = s(t) \times \cos(\omega_p t + \mathbf{j})$ où \mathbf{f} : déphasage entre la porteuse et le terme en $\cos(\omega_p t)$.

$$i(t) = A_p \cos((\omega_p - \omega_m)t) \times \cos(\omega_p t + \mathbf{j})$$

En développant le produit $\cos((\omega_p - \omega_m)t) \cos(\omega_p t + \mathbf{f})$, on obtient :

$$i(t) = A_p \frac{\cos(\omega_m t + \mathbf{j}) + \cos(\omega_m t + 2\omega_p t + \mathbf{j})}{2}$$

Après filtrage passe-bas de $i(t)$, on retrouve bien le signal $u(t)$ à une erreur de phase près.

Or, il se trouve que l'oreille humaine est insensible aux déphasages. Le fait de renoncer à une erreur de phase nulle permet de simplifier la réalisation des démodulateurs. La SSB est donc utilisée en téléphonie où ses avantages (bande minimum, pas de signaux aux très basses fréquences, puissance utile optimisée) ne sont pas contre balancés par son unique défaut à savoir l'erreur de phase.

Modulation à Bande Latérale Résiduelle

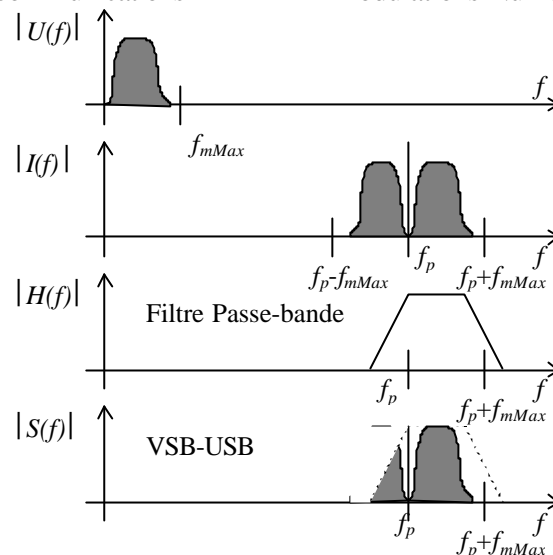
La transmission des signaux vidéo a introduits des contraintes supplémentaires par rapport au cas des signaux audio. En effet, pour une image télévision, le signal vidéo se caractérise par une bande passante importante (5 à 6 MHz selon les normes). Le spectre de ce signal comporte d'importantes composantes basses-fréquences qu'il n'est pas question de filtrer. Enfin, le signal reçu doit être conforme.

Il n'est donc pas possible d'utiliser des modulations SSB, ni AM-P. Recourir à une modulation AM limiterait considérablement le nombre de canaux utilisables et nécessiterait des émetteurs sensiblement plus puissants.

Il faut donc trouver un compromis entre AM et SSB, c'est la Modulation à Bande Latérale Résiduelle (VSB : Vestigial Side Band).

Pour obtenir un signal VSB, la technique employée consiste à réaliser une modulation AM suivie d'un filtre qui atténue sensiblement les composantes fréquentielles inférieures à f_p . L'une des bandes (souvent la LSB) se trouve donc partiellement atténuée.

Le spectre du signal VSB est donc le suivant :



La démodulation du signal VSB est réalisée à l'aide d'un simple détecteur crête lorsque le signal AM est caractérisé par un taux de modulation $m \ll 1$.

Conclusion

En résumé :

type de modulation	Δf	puissance totale	puissance utile	démodulation
AM	$2 f_{mMax}$	$A_p^2/2 (1+2 m^2/4)$	$2 \text{ fois } A_p m^2/4$	simple (détection crête)
AM-P	$2 f_{mMax}$	$A_p^2/2$	$2 \text{ fois } A_p^2/4$	très complexe (synchrone)
SSB	f_{mMax}	$A_p^2/4$	$A_p^2/4$	complexe (anisochrone)
VSB	$> f_{mMax}$	$> A_p^2/2 (1+m^2/4)$	$A_p m^2/4$	simple (détection crête)

2. MODULATION DE FREQUENCE (F.M.)

Nous avons vu que le principe de la modulation AM repose sur la modification de l'amplitude de la porteuse sans modification de la fréquence. Une autre forme de modulation consiste à garder l'amplitude de la porteuse constante mais à faire varier sa fréquence.

Notion de fréquence et de phase instantanée

Considérons les signaux sinusoïdaux :

$$r(t) = A_s \cdot \cos(\omega t) \text{ et } s(t) = A_s \cdot \cos(\omega t + j).$$

Supposons que ω et j soient constantes dans le temps. Le déphasage de $s(t)$ par rapport à $r(t)$ est constant et les deux signaux sont caractérisés par une fréquence identique. Si j est variable avec le temps, le déphasage entre $s(t)$ et $r(t)$ varie dans le temps. Ce changement de phase se traduit également par un changement de fréquence du signal $s(t)$ car pour passer d'une phase j à $j + dj$ en un temps dt , il est nécessaire que la pulsation du signal varie. On appelle alors pulsation instantanée la grandeur :

$$\omega_i = \frac{d}{dt}(\omega \cdot t + j)$$

$$\text{dans le cas où } \omega \text{ est constante, alors : } \omega_i = \omega + \frac{dj}{dt} \text{ ou encore : } f_i = \frac{\omega}{2\pi} + \frac{1}{2\pi} \frac{dj}{dt}.$$

De même, sera appelée phase instantanée la grandeur : $j_i(t) = \omega \cdot t + j(t)$.

Principe de la modulation de fréquence

Généralités

La modulation de fréquence ou F.M. (Frequency Modulation), très souvent employé en VHF (Very High Frequencies) ou en UHF (Ultra High Frequencies) pour la transmission des signaux audio, consiste à utiliser une porteuse dont la fréquence varie en fonction du signal modulant. Il est alors possible d'écrire :

$f_s = f_p + f(t)$ où $f(t)$ est une fréquence variant linéairement avec le signal modulant $u(t)$. c'est-à-dire :

$f(t) = n \cdot u(t)$ avec n facteur de proportionnalité du modulateur (en Hz/A_m) qui prend parfois la dénomination de sensibilité du modulateur.

La fréquence de la porteuse étant constante, la phase instantanée de $s(t)$ est donnée par :

$$j_s(t) = \int (\omega_p + 2\pi \cdot f(t)) \cdot dt$$

$$j_s(t) = \omega_p t + 2\pi \cdot \int f(t) \cdot dt. \text{ ou encore : } j_s(t) = \omega_p t + 2 \cdot \pi \cdot n \cdot \int u(t) \cdot dt$$

L'expression du signal $s(t)$ modulé en fréquence est alors :

$$s(t) = A_p \cos\left(\omega_p t + 2 \cdot \pi \cdot n \cdot \int u(t) \cdot dt\right).$$

Les techniques de modulations de fréquence sortent du domaine de ce cours. Ces techniques font appel à des opérateurs électroniques nommés VCO (Voltage Controlled Oscillator ou Oscillateur Contrôlé en Tension). Néanmoins, pour mémoire, nous pouvons citer deux techniques très répandues. La première, utilisée en BF (jusqu'à plusieurs dizaines de MégaHertz dans les circuits intégrés), est fondée sur l'emploi d'oscillateur à relaxation et consiste à charger un condensateur jusqu'à une tension de référence à l'aide d'une source de courant commandée en tension (VCCS). Lorsque la tension du condensateur atteint la tension de référence, le condensateur est déchargé et l'opération recommence. La seconde technique, employée en HF (typiquement VHF et UHF), repose sur un oscillateur LC quasi sinusoïdal. Le condensateur employé est en fait une diode spéciale, appelée VARICAP ou varactor, qui polarisée en inverse, présente une capacité variable en fonction de la tension inverse de diode. La fréquence de résonance de l'oscillateur est modifiée en fonction de la tension inverse appliquée à la diode VARICAP.

Représentation temporelle du signal FM.

Prenons le cas d'un signal modulant sinusoïdal : $u(t) = A_m \cos(\omega_m t)$ avec $\omega_m \ll \omega_p$. L'expression du signal $s(t)$ est alors :

$$s(t) = A_p \cos\left(\omega_p t + 2 \cdot \mathbf{p} \cdot \mathbf{n} \cdot \int A_m \cos(\omega_m t) \cdot dt\right)$$

$$s(t) = A_p \cos\left(\omega_p t + \frac{2 \cdot \mathbf{p} \cdot \mathbf{n}}{\omega_m} \cdot A_m \sin(\omega_m t)\right).$$

En reprenant l'expression de la pulsation (ou de la fréquence) instantanée, il est possible d'en déduire la fréquence instantanée f_s du signal FM $s(t)$:

$$f_s(t) = f_p + \mathbf{n} \cdot A_p \cos(\omega_m t).$$

L'excursion en fréquence Δf_s de $s(t)$ est donc :

$$\Delta f_s = \mathbf{n} A_m.$$

L'expression de la phase instantanée est immédiate :

$$\mathbf{j}_s(t) = \omega \cdot t + \frac{2\mathbf{p}\mathbf{n}}{\omega_m} \cdot A_p \sin(\omega_m t), \text{ ou également : } \mathbf{j}_s(t) = \omega \cdot t + \frac{A_p \cdot \mathbf{n}}{f_m} \cdot \sin(\omega_m t)$$

L'excursion en phase $\Delta \mathbf{j}_s$ de $s(t)$ est donc :
$$\Delta \mathbf{j}_s = \frac{2\mathbf{p}\mathbf{n} \cdot A_p}{\omega_m} = \frac{\mathbf{n} \cdot A_p}{f_m}.$$

On en déduit alors la relation qui lie l'excursion de phase à l'excursion de fréquence :

$$\Delta \mathbf{j}_s = \frac{\Delta f_s}{f_m}.$$

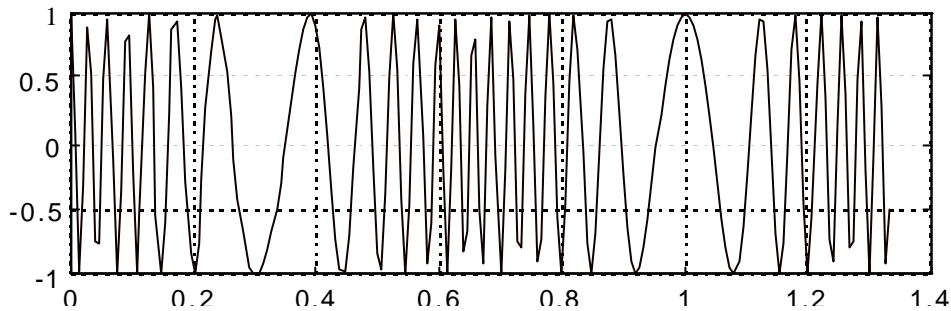
Il est d'usage d'appeler indice de modulation le terme \mathbf{d} : $\mathbf{d} = \Delta \mathbf{j}_s = \frac{\Delta f_s}{f_m}$. Il est alors

possible de réécrire le signal modulé FM $s(t)$ sous la forme :

$$s(t) = A_p \cos\left(\omega_p t + \mathbf{d} \cdot \sin(\omega_m t)\right)$$

Il est essentiel de remarquer que l'indice de modulation \mathbf{d} , contrairement au taux de modulation AM m , dépend à la fois de la fréquence et de l'amplitude du signal modulant $u(t)$.

L'allure d'un signal FM, typiquement un signal de bande intermédiaire (FI-FM) audio est donné ci-dessous :



Signal FM. Porteuse $f_p = 10,7\text{MHz}$, signal modulant sinusoïdal à $f_s = 15\text{ kHz}$, indice modulation $d = 10$.

Représentation fréquentielle d'un signal FM, cas d'un signal modulant sinusoïdal

Pour déterminer le spectre de $s(t)$, nous prendrons $u(t)$ sinusoïdal. Nous avons alors :

$$s(t) = A_p \cos(\omega_p t + d \cdot \sin(\omega_m t)).$$

A l'aide de la formule de Moivre, il est possible de mettre l'expression précédente sous la forme :

$$s(t) = A_p \frac{e^{j(\omega_p t + d \cdot \sin(\omega_m t))} + e^{-j(\omega_p t + d \cdot \sin(\omega_m t))}}{2}.$$

$$s(t) = A_p \frac{e^{j\omega_p t} e^{jd \cdot \sin(\omega_m t)} + e^{-j\omega_p t} e^{-jd \cdot \sin(\omega_m t)}}{2}.$$

Trouver le spectre de $s(t)$ revient à calculer $\underline{S}(\omega)$ la transformée de Fourier de $s(t)$. Pour cela il faut commencer par calculer la transformée de Fourier d'un des deux produits d'exponentielles. De ce fait, il faut développer en série de Taylor le terme $e^{jd \cdot \sin(\omega_m t)}$. Il est alors possible de démontrer que :

$$e^{jd \cdot \sin(\omega_m t)} = \sum_{k=-\infty}^{+\infty} e^{jk \cdot \omega_m t} \cdot J_k(d).$$

Où la fonction $J_k(x)$ est appelée fonction de Bessel de première espèce et $J_k(x)$ est définie par :

$$J_k(x) = \frac{1}{2\pi} \int_0^{2\pi} \cos(k \cdot q - x \sin(q)) dq.$$

Les fonctions de Bessel

Les fonctions de Bessel sont les solutions particulières de l'équation différentielle :

$$y'' + \frac{y'}{x} + \left(1 - \frac{n^2}{x^2}\right) \cdot y = 0 \text{ dont les solutions sont de la forme :}$$

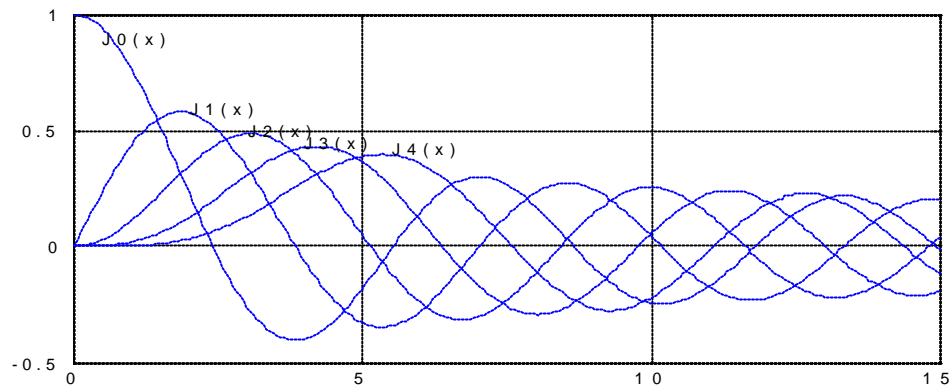
$$J_n(x) = \sum_{k=0}^{\infty} \frac{(-1)^k \left(\frac{x}{2}\right)^{2k+n}}{k! (k+n)!}$$

Les fonctions de Bessel vérifient plusieurs propriétés intéressantes notamment :

$$J_n(x) = (-1)^n J_{-n}(x).$$

Il est simple de vérifier que : $\sum_{k=-\infty}^{+\infty} J_k^2(d) = 1.$

La représentation des cinq premières fonctions de Bessel est donnée ci-dessous :



Par ailleurs, étant donné que $e^{jd \sin(x)}$ est une fonction périodique, il est possible de la développer en série de Fourier. On a alors :

$$e^{jd \cdot \sin(x)} = \sum_{k=-\infty}^{+\infty} e^{jk} J_k(d) \quad \text{avec} \quad C_k(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{jx \sin(q)} e^{-jkq} dq.$$

avec l'expression développée en série de Taylor, nous obtenons : $C_k(x) = J_k(x)$.

Le spectre $\underline{S}(w)$ du signal $s(t)$ est alors obtenu en réalisant le produit de convolution de la Transformée de Fourier de $e^{jw_p t}$ avec $\sum_{k=-\infty}^{+\infty} e^{jk \cdot w_m \cdot t} \cdot J_k(d)$. $\underline{S}(w)$ est un spectre composé

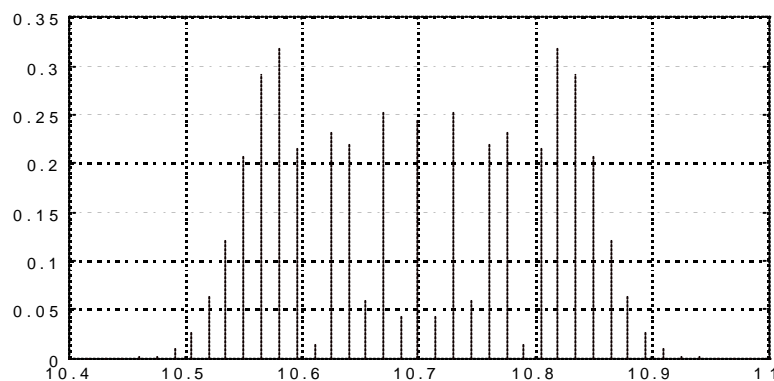
d'une infinité de raies centrées sur w_p et espacées les unes des autres de w_m :

$$\underline{S}(w) = A_p \sum_{k=-\infty}^{+\infty} J_k(d) \delta(w_p - k \cdot w_m)$$

Puisque le spectre est composé d'une infinité de raies d'amplitude $A_p J_n(d)$ aux fréquences : $f = f_p \pm n f_m$, en toute rigueur, pour transmettre un signal FM, il est nécessaire de disposer d'une bande infinie.

Etant donné que $J_n(x) = (-1)^k J_{-n}(x)$, les raies à $f_p + n f_m$ ont la même amplitude que les raies à $f_p - n f_m$. Toutefois, pour les termes impairs, les composantes $f_p - (2n+1)f_m$ sont en opposition de phase avec les composantes $f_p + (2n+1)f_m$.

La figure ci-dessous représente le spectre d'un signal FM $s(t)$ dont le signal modulant $u(t)$ est sinusoïdal.



Spectre d'un signal modulé en fréquence par un signal sinusoïdal, $d=10, f_m=150$ kHz, $f_p=10,7$ MHz.

Remarquez que l'encombrement est largement supérieur à l'excursion en fréquence ($\Delta f=150$ kHz)

Puissance

Nous avons vu précédemment que :

$$\sum_{k=-\infty}^{+\infty} J_k^2(d) = 1$$

De ce fait, la puissance du signal FM $s(t)$ se répartit dans l'ensemble des raies et est égale à la puissance de la porteuse en l'absence du signal modulant $u(t)$. Cela revient à :

$$P_s = A \frac{A_p^2}{2} \forall d.$$

L'interprétation physique de ce résultat est simple : seule la fréquence du signal varie, et non son amplitude. Etant donné que l'amplitude de la raie à la fréquence de la porteuse est donnée par $J_0(d)$, certains indices de modulation d conduisent à une absence de raie à la fréquence de la porteuse.

Théoriquement, le signal FM nécessite une bande infinie. En pratique, seules les N raies de part et d'autre de f_p (soit au total $2N+1$ raies) qui contribuent à 99% de la puissance sont prises en compte lors de la détermination de la bande de fréquence de $s(t)$. Pour cela, il est d'usage de considérer trois cas :

1) $d \ll 1$, La bande de fréquence est alors $B = 2f_m$.

Le spectre du signal FM $s(t)$ est voisin de celui d'un signal AM et l'on parle de modulation à bande étroite (Narrow Band Modulation).

2) $0,3 \leq d \leq 20$, La bande de fréquence est alors $B = 2Nf_m$ où N représente les N raies dont l'amplitude est supérieure à 0,01 de A_p .

3) $d > 20$, l'approximation $B = 2[d f_m + f_m]$ est alors utilisée.

Très souvent, on a recours à une règle empirique dite règle de Carson : $B = 2[d + 1]f_m$.

Cas de plusieurs signaux modulants

Il est important de noter que le principe de superposition ne s'applique pas aux signaux modulés FM. En effet, le terme correspondant au signal modulant est sous le cosinus, fonction non-linéaire ! Dans le cas d'une modulation par plusieurs signaux modulant $u_1(t)$, $u_2(t)$, etc., d'indices de modulation respectivement d_1 , d_2 , etc., le signal modulé $s(t)$ est de la forme :

$$s(t) = A_p \cos\left(\omega_p t + d_1 \sin(\omega_{m1} t) + d_2 \sin(\omega_{m2} t) + \dots + d_n \sin(\omega_{mn} t)\right).$$

Nous avons vu que, pour un signal FM modulé par un signal sinusoïdal, il était tout d'abord nécessaire de calculer la transformée de Fourier de $e^{jd \sin(\omega_m t)}$, soit :

$$TF\left(e^{jd \sin(\omega_m t)}\right) = \sum_{k=-\infty}^{+\infty} d(\omega - k \cdot \omega_m) \cdot J_k(d)$$

Dans le cas de plusieurs signaux modulant, il faut calculer la Transformée de Fourier de :

$$e^{j \sum_{i=1}^n d_i \sin(\mathbf{w}_{m_i} t)} = \prod_{i=1}^n e^{j d_i \sin(\mathbf{w}_{m_i} t)}.$$

La transformée de Fourier d'un produit de deux fonctions donne le produit de convolution des transformées de Fourier, donc :

$$\begin{aligned} TF \left[e^{j \sum_{i=1}^n d_i \sin(\mathbf{w}_{m_i} t)} \right] &= TF \left[\prod_{i=1}^n e^{j d_i \sin(\mathbf{w}_{m_i} t)} \right] \\ &= \sum_{k_1=-\infty}^{+\infty} d(\mathbf{w} - k \mathbf{w}_{m_1}) J_{k_1}(d_1) * \left(\dots \left(\sum_{k_{n-1}=-\infty}^{+\infty} d(\mathbf{w} - k \mathbf{w}_{m_{n-1}}) J_{k_{n-1}}(d_{n-1}) * \left(\sum_{k_n=-\infty}^{+\infty} d(\mathbf{w} - k \mathbf{w}_{m_n}) J_{k_n}(d_n) \right) \right) \dots \right) \end{aligned}$$

Le spectre du signal modulé contient alors une raie à la fréquence de la porteuse f_p , des raies à $f_p \pm f_{m1}$, à $f_p \pm 2f_{m1}$ etc.... dues au seul premier signal modulant, des raies à $f_p \pm f_{m2}$, à $f_p \pm 2f_{m2}$ etc.... dues au second signal modulant ainsi que des raies dues deux signaux modulants, au trois signaux modulants etc....

Prenons l'exemple d'un signal $s(t)$ modulé par :

$u_1(t) = \cos(\mathbf{w}_{m1}t)$, $u_2(t) = \cos(\mathbf{w}_{m2}t)$ d'indices de modulation respectifs $\delta_1 = 0,25$ $\delta_2 = 0,5$.

Le spectre de $s(t)$ avec le seul signal modulant $u_1(t)$ contient 3 raies aux fréquences f_p et $f_p \pm f_{m1}$.

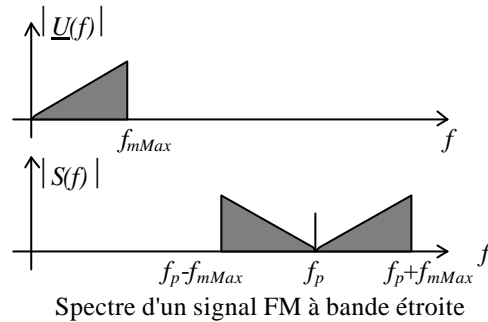
Le spectre de $s(t)$ avec le seul signal modulant $u_2(t)$ contient 5 raies aux fréquences f_p , $f_p \pm f_{m2}$ et $f_p \pm 2f_{m2}$.

Le spectre de $s(t)$ avec les deux signaux modulants est alors composé de 15 raies comme suit :

fréquence	amplitude
f_p	$A_p J_0(\delta_1) J_0(\delta_2)$
$f_p \pm f_{m1}$	$A_p J_1(\delta_1) J_0(\delta_2)$
$f_p \pm f_{m2}$	$A_p J_0(\delta_1) J_1(\delta_2)$
$f_p \pm 2f_{m2}$	$A_p J_0(\delta_1) J_2(\delta_2)$
$f_p \pm f_{m1} \pm f_{m2}$	$A_p J_1(\delta_1) J_1(\delta_2)$
$f_p \pm f_{m1} \pm 2f_{m2}$	$A_p J_1(\delta_1) J_2(\delta_2)$

Cas d'un signal modulant quelconque

Dans le cas d'une modulation FM à bande étroite (Narrow Band), le spectre du signal FM est très voisin de celui d'un signal AM :

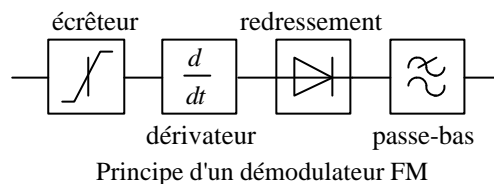


En dehors du cas d'une modulation FM à bande étroite, le spectre du signal FM ne peut se déduire simplement à partir de celui du signal modulant.

Technique de démodulation de signaux FM

Démodulation par dérivation et détection d'enveloppe

Pour démoduler un signal modulé FM, il est nécessaire de recourir à dispositif dont la tension (ou le courant) de sortie, varie linéairement avec la fréquence. L'opérateur mathématique qui permet cette opération est évidemment la dérivation. En récupérant l'enveloppe du signal dérivé on obtient le signal modulant. L'un moyen les plus simples revient alors au synoptique suivant :



Afin d'avoir des signaux qui varient proportionnellement à la fréquence, il est nécessaire d'employer un dérivateur. L'écriteur permet d'avoir des signaux d'amplitude constante avant le dérivateur et donc de s'affranchir des éventuelles variations d'amplitude du signal modulé. Il ne reste plus ensuite qu'à effectuer une détection crête.

Remarque :

Formellement le signal dérivé est :

$$\frac{d}{dt} \left(A_p \cdot \cos \left(\omega_p t + d \cdot \int u(t) \cdot dt \right) \right) = -A_p \cdot \sin \left(\omega_p t + d \cdot \int u(t) \cdot dt \right) \times \left(\omega_p + d \cdot u(t) \right)$$

Il s'agit donc d'un signal modulé en fréquence et en amplitude. La détection d'enveloppe permet de ne récupérer que le signal modulé en amplitude sans tenir compte de la fréquence instantanée du signal dérivé.

Démodulation par utilisation d'une Boucle à verrouillage de phase

Nous avons vu, lors de l'étude des démodulateurs AM, que la PLL est un dispositif maintenant constant l'écart de phase entre le signal qu'elle délivre et le signal d'entrée.

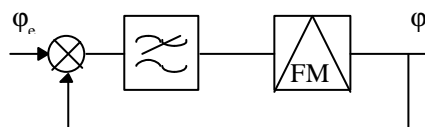
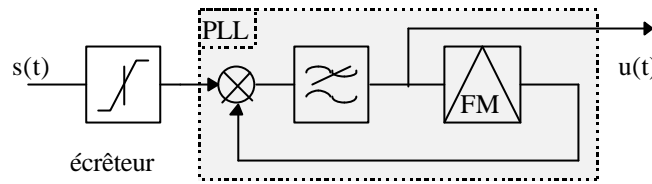


Schéma de principe de la PLL

Nous avons vu qu'après filtrage, le signal $\phi_e - \phi_s$ sert d'entrée au VCO (VCO qui revient à un modulateur FM de pulsation porteuse ω_p). Or : $j_e - j_s = c^{te}$ ce qui implique : $\frac{d}{dt}(j_e - j_s) = 0$, ou encore $f_e - f_s = 0$.

La fréquence instantanée du signal d'entrée est donc identique à la fréquence instantanée du signal produit par le VCO. Etant donné que la fréquence instantanée du signal d'entrée de la PLL est une fonction linéaire du signal modulant $u(t)$, le signal de commande du modulateur FM de la PLL suit exactement le signal $u(t)$.

Le schéma du démodulateur FM à PLL est donc le suivant :



Démodulateur FM à PLL

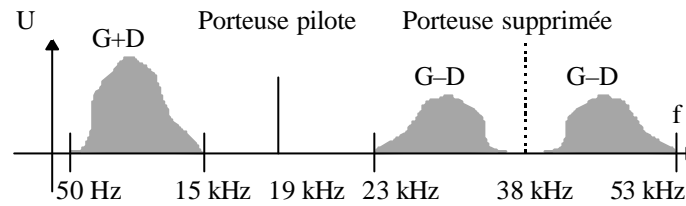
Application de la modulation de fréquence

Comme nous l'avons vu les principaux avantages de la modulation FM réside dans le fait que la puissance est constante quel que soit le signal modulant, le signal démodulé dépend très peu du niveau du signal modulé présent à l'entrée du démodulateur. Lorsque les indices de modulations sont élevés, un signal FM est également plus robuste vis-à-vis des perturbations qu'un signal AM. En revanche la bande de fréquence est toujours supérieure à celle d'un signal AM.

De ce fait, la modulation FM est utilisée lorsque le signal reçu est caractérisé par un affaiblissement variable (radiocommunications mobiles en milieu urbain), que la largeur de bande est élevée et qu'une bonne qualité de signal démodulé est nécessaire. On rencontre donc les signaux FM à des hautes fréquences (VHF, UHF) dans les liaisons par satellites, dans la radio diffusion FM (stéréo) ou dans les téléphones mobiles.

Cas particulier : Modulation AM-FM Application à la transmission radiophonique FM stéréophonique.

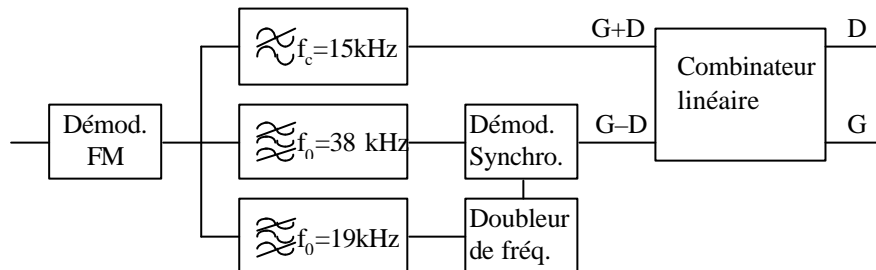
L'émission d'un signal audio stéréophonique doit produire un signal compatible avec des récepteurs non stéréo. La solution retenue consiste à émettre simultanément un signal G+D contenant le message de la voie droite plus celui de gauche et un signal G-D contenant le message de la voie droite moins celui de gauche. Le signal G+D est à basse fréquence alors que le signal G-D est modulé AM-P par une porteuse à 38 kHz. Une fréquence pilote à 19 kHz est transmise simultanément. Le spectre du signal modulant est alors le suivant :



Signal modulant en radiodiffusion FM stéréophonique.

La restitution du signal de la voie droite (ou gauche), implique :

- un filtrage passe-bas qui permet de ne conserver que le signal $G+D$.
- un filtrage passe-bande suivi d'un doubleur de fréquence pour restituer la porteuse à 38 kHz.
- un démodulateur AM précédé d'un filtre passe-bande (démodulation synchrone) pour restituer le signal $G-D$.
- enfin, un sommateur et un soustracteur pour obtenir G ou D .



Principe du démodulateur FM stéréophonique.

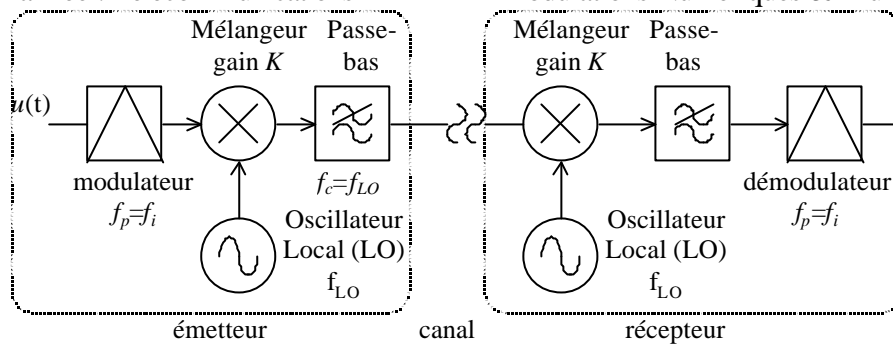
Technique de fréquence intermédiaire

Très souvent, le signal modulant est un signal basses-fréquences alors que la bande de fréquences allouée pour la transmission du signal est à très haute fréquence. Techniquement, il s'avère difficile de moduler ou de démoduler des signaux par des porteuses à très haute fréquence. Il est fréquent d'utiliser un premier mélangeur à "moyenne fréquence" (IF) avant le second mélangeur dont la fréquence porteuse est à très haute fréquence. Cette technique est employée notamment pour les récepteurs radio AM/FM. Les fréquences intermédiaires sont alors normalisées :

FI-AM : 455 kHz,

FI-FM : 10,7 MHz.

Le schéma de principe d'un modulateur et d'un démodulateur est alors le suivant :



Principe des modulateurs/démodulateur en fréquence intermédiaire

Le signal issu du modulateur et celui arrivant au démodulateur sont en fréquence intermédiaire (FI)

3. MODULATION DE PHASE (FM)

Principe

Dans le cas de la modulation de phase, c'est la phase instantanée qui varie linéairement en fonction du signal modulant. Donc, par rapport à un signal FM, l'expression du signal $s(t)$ modulé ΦM est :

$s(t) = A_p \cos(w_p t + \Delta j(t))$ avec $\Delta j(t) = q \times A_p \cos(w_m t)$ où q est le coefficient de proportionnalité du modulateur (en rd/V ou rd/A). L'excursion en phase est : $\Delta j_s = q \times A_p$.

La fréquence instantanée est donc : $\Delta f(t) = \frac{1}{2\pi} \frac{d}{dt} \Delta j(t) = -\frac{q \times A_p w_m}{2\pi} \sin(w_m t)$.

L'excursion en fréquence est donc : $\Delta f_s = q \times A_p \times f_m$.

Par analogie avec la modulation de fréquence, l'indice de modulation est donc :

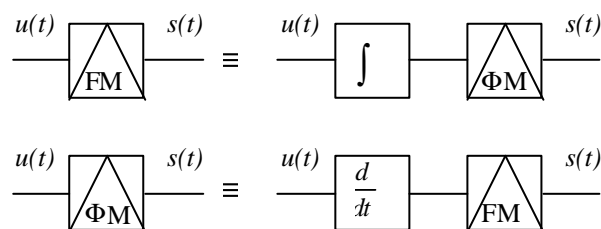
$d = \frac{\Delta f_s}{f_m} = q \times A_p$. Notons, que contrairement à ce qui se produit pour un signal modulé

FM, l'indice de modulation en ΦM est indépendant de la fréquence du signal modulant.

En faisant apparaître l'indice de modulation, l'expression du signal $s(t)$ peut se mettre sous la forme : $s(t) = A_p \cos(w_p t + d \cdot \cos(w_m t))$. Le spectre d'un signal modulé ΦM est donc identique à celui d'un signal modulé FM.

Relation entre FM et ΦM

Dans le cas particulier d'un signal modulant sinusoïdal, la différence entre FM et ΦM ne se traduit que par une variation de phase de $\pi/2$ du signal modulant. En général, à un terme de fréquence près, le signal ΦM est un signal FM dont le signal modulant a été dérivé. Il est donc possible de faire un lien direct en modulation FM et modulation ΦM .



Equivalence entre FM et ΦM

Démodulation FM

La technique de démodulation d'un signal ΦM revient à démoduler un signal FM puis à faire suivre le démodulateur par un intégrateur. Cependant, pour pouvoir démoduler sans ambiguïté un signal ΦM , il est nécessaire que l'excursion de phase ne dépasse pas π . Donc la modulation ΦM n'est employée qu'avec de faibles indices de modulation ($d < \pi$). De ce fait, la modulation ΦM est une modulation à bande étroite.

RESUME DES MODULATIONS ANGULAIRES

Voici un tableau récapitulatif des principales caractéristiques de signaux $s(t)$, modulés FM ou ΦM par un signal $u(t)=A_m \cos(\omega_m t)$ sinusoïdal :

	indice de modulation	Phase instantanée	Fréq. instantanée	DF	Df_s
Signal FM	$d = \frac{\Delta f_s}{f_m}$	$w_p t + d \sin(w_m t)$	$f_p + n A_m \cos(w_m t)$	$DF = d$	$Df_s = n A_m$
Signal ΦM	$d = J \times A_m$	$w_p t + d \cos(w_m t)$	$f_p - q A_m f_m \sin(w_m t)$	$DF = d$	$Df_s = q A_m f_m$

COMPARAISON DES DIFFERENTES MODULATIONS ANALOGIQUES

Pouvoir quantifier la qualité du signal après sa transmission et sa démodulation est l'un des points essentiels des techniques de communication. Il est relativement intuitif de comprendre que le signal reçu après une transmission à longue distance va être considérablement atténué et comme par ailleurs, à ce signal vont se superposer des parasites, le signal arrivant au démodulateur. Enfin, le signal subit différent traitement (modulation, amplification, démodulation...) qui vont entraîner des déformations.

Il est d'usage de caractériser le bruit par sa puissance. De ce fait, l'influence du bruit sur un signal est donné par le rapport signal bruit qui est le rapport de la puissance du signal non bruité sur la puissance de bruit (SNR : Signal to Noise Ratio). Cette quantité est fréquemment exprimée en décibels ($10 \times \log\left(\frac{S}{N}\right)$). Plus le SNR_{dB} est élevé, meilleure est la qualité du signal récupéré.

La mesure de la déformation d'un signal est donné par le taux de distorsion harmonique total (THD : Total Harmonic Distorsion). La mesure du THD consiste à déterminer les harmoniques du signal après traitement au signal original, lequel est un signal sinusoïdal pur.

Nous reviendrons ultérieurement sur la détermination des performances des différentes modulations. Retenons simplement les points suivants :

	Bande de fréquence	Complexité	Amélioration après démodulation	Sensibilité aux perturbations
AM	$2 f_m$	minime	jamais en détection crête	importante
FM	$2 N(d) f_m$	complexe	si $\delta > 0,475$	faible
ΦM	$2 N(d) f_m$	complexe	si $\delta > 0,775$	faible

MODULATIONS MULTIPLES

La combinaison de plusieurs modulations peut être envisagée, soit pour s'adapter à des contraintes spécifiques (amélioration du S/N, simplification des étages HF...), soit afin de transmettre plusieurs informations simultanément.

4. MODULATIONS ANALOGIQUES DE SIGNAUX **NUMERIQUES**

Généralités

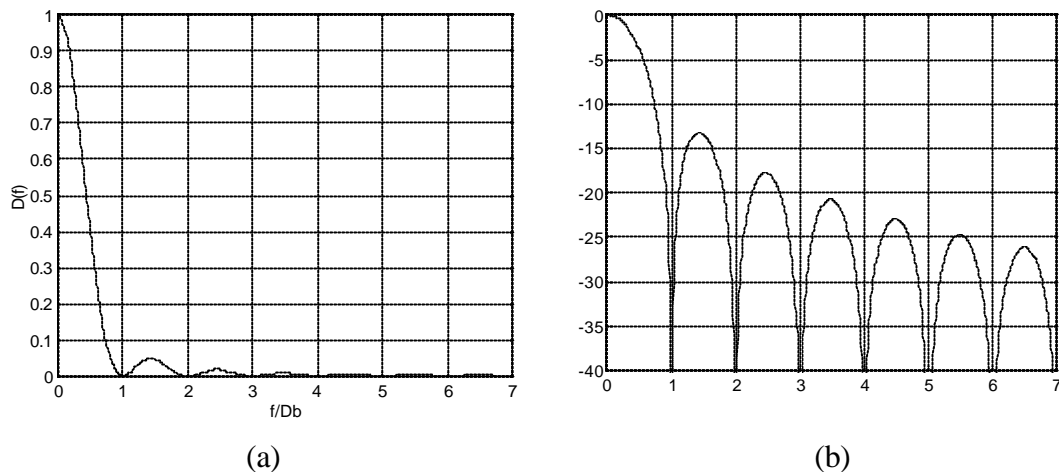
Dans un nombre croissant de situations, il est nécessaire de transmettre des signaux numériques, en général sous la forme d'une séquence binaire. Les signaux numériques présentent en effet plusieurs propriétés intéressantes pour les télécommunications : souplesse des traitements, signal à états discrets donc moins sensibles aux bruits (il suffit de seuiller le signal) et simple à régénérer, utilisation de codes correcteurs d'erreur, cryptage de l'information. En revanche, nous verrons ultérieurement qu'à quantité d'informations transmise identique, un signal numérique nécessite une bande de fréquence nettement plus importante.

Pour des raisons identiques à celles évoquées dans le cas de signaux analogiques, ces signaux numériques modulent une porteuse sinusoïdale afin de présenter, soit des caractéristiques compatibles avec le canal de transmission utilisé (exemple des modems (modem : MODulator–DEModulator)), soit pour transmettre plusieurs signaux simultanément. Toutefois, c'est l'explosion de la téléphonie mobile et de la télévision numérique qui suscite une étude de ce type de modulations.

Cas des signaux binaires

Un signal numérique est caractérisé par son débit D_b (en bit/s), par les amplitudes A_1 et A_0 représentant respectivement le "1" et le "0". Le signal étant aléatoire, pour la représentation spectrale du signal, seule la densité spectrale de puissance du signal à un sens. Par abus de langage, nous appellerons parfois spectre la . La densité spectrale de puissance, $D(f)$, d'un signal binaire aléatoire à valeur moyenne nulle ($A_1 = -A_0$ et la probabilité d'avoir un « 1 » est identique à celle d'avoir un « 0 ») est de la forme :

$$D(f) = \frac{1}{D_b} \left(A \frac{\sin\left(\frac{P f}{D_b}\right)}{\frac{P f}{D_b}} \right)^2 = \frac{\sin^2\left(\frac{P f}{D_b}\right)}{D_b}$$



Densité spectrale de puissance d'un signal binaire aléatoire (en en échelle linéaire (a) et dB (b)).
Les fréquences sont normalisées par rapport au débit binaire

Signaux *m*-aires

En fait, le propre des signaux numériques est qu'ils sont caractérisés par des états quantifiés qui ne sont pas nécessairement binaires (2 valeurs discrètes, "-1" et "1" ou "0" et "1"), mais peuvent prendre m valeurs discrètes (par exemple, pour $m=4$ valeurs discrètes, "-3", "-1", "1" ou "3"). On parle alors d'un alphabet *m*-aire. En pratique, les valeurs de m seront de la forme 2^n (par exemple : 2, 4, 16...) car les symboles sont créés à partir d'une suite de n bits. Un mot de n bits forme alors un symbole *m*-aire. A partir du flux binaire de débit D_b , on forme donc un flux de symboles de débit D_s . Le débit de symbole est alors donné par :

$$D_s = \frac{D_b}{n}$$

Le symbole *m*-aire est simplement construit à l'aide d'un convertisseur numérique analogique de résolution n bits.

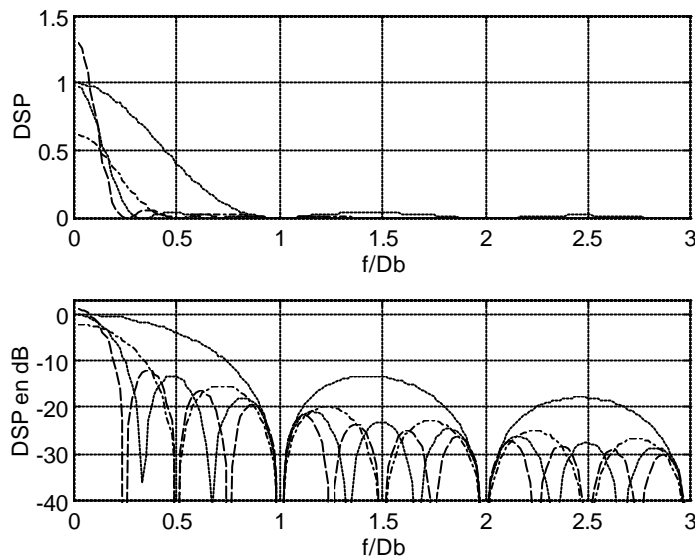
Lorsque les valeurs discrètes du symbole *m*-aire sont de la forme $(2k+1)A/m$ (par exemple $m=4$ valeurs discrètes, "-3/4", "-1/4", "1/4" ou "3/4"), la densité spectrale de puissance est alors donnée par :

$$D(f) = \frac{m^2 - 1}{3D_s} \left(A \frac{\sin\left(\frac{pf}{D_s}\right)}{\frac{pf}{D_s}} \right)^2 = \frac{m^2 - 1}{3D_s} A^2 \operatorname{sinc}^2\left(\frac{pf}{D_s}\right),$$

soit en reprenant le débit binaire :

$$D(f) = \frac{n(m^2 - 1)}{3D_b m^2} \left(A \frac{\sin\left(\frac{npf}{D_b}\right)}{\frac{npf}{D_b}} \right)^2 = \frac{n(m^2 - 1)}{3D_b m^2} A^2 \operatorname{sinc}^2\left(\frac{npf}{D_b}\right)$$

L'allure de la densité spectrale de puissance est alors :

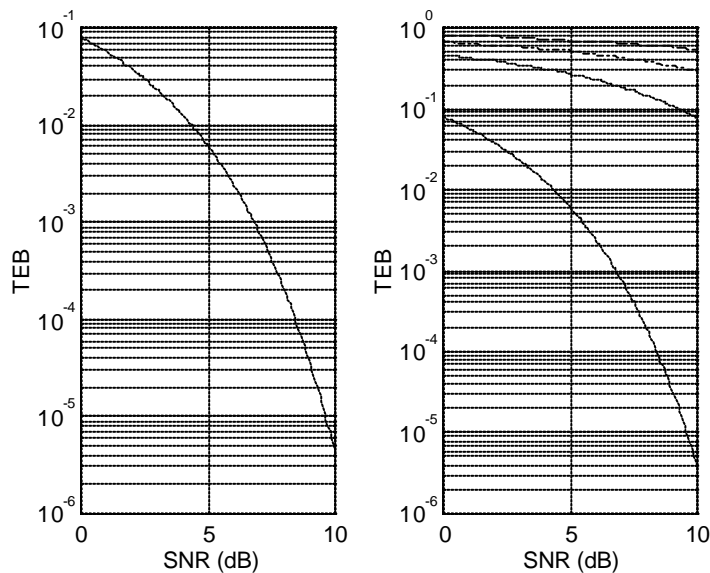


densité spectrale de puissance de signaux m-aires aléatoires (en en échelle linéaire (a) et dB (b)).

Les fréquences sont normalisées par rapport au débit binaire, $m=2, 4, 8, 16$

En conservant le même débit binaire, il est alors possible de réduire la bande passante nécessaire à la transmission du signal. Réciproquement, à bande passante donnée, le fait d'utiliser un signal m-aire permet d'augmenter le débit binaire.

Cet avantage se paye toutefois par l'augmentation du taux d'erreur binaire à rapport signal sur bruit donné.



Taux d'erreurs en fonction du rapport signal/bruit pour $m=4, 8, 16$

Les différentes modulations de signaux numériques

Modulations « historiques »

Ces modulations sont citées essentiellement pour mémoire. Elles ne sont plus guère employées sur les systèmes modernes.

Modulation OOK

On–Off Keying : Modulation AM en tout ou rien. La porteuse est émise ou non selon la valeur du signal numérique ("0" pas de porteuse, "1" porteuse émise). Le "spectre" d'un signal OOK revient à celui d'un signal binaire aléatoire (donc en $\sin(x)/x$ au quel s'ajoute une raie due à la valeur moyenne non nulle du signal RZ) translaté à la fréquence de porteuse. Il existe donc une raie à f_p . L'absence de porteuse lors de l'émission de « 0 » rend difficile la synchronisation.

Modulation FSK

Frequency–Shift Keying : Modulation par saut de fréquence. La fréquence du signal modulé peut prendre 2 valeurs différentes selon la valeur du bit à transmettre. Lorsque ce type de modulation comporte des discontinuités de phase (modulation obtenue en commutant d'un oscillateur de fréquence f_{p1} vers un oscillateur de fréquence f_{p2} en fonction de la valeur du bit à transmettre), le spectre est simplement la superposition des spectres de signaux modulés OOK à f_{p1} et f_{p2} .

Lorsqu'il n'y a pas de discontinuité de phase (modulation obtenue en utilisant un VCO), nous retrouvons une modulation FM dont le signal modulant est un signal binaire aléatoire. La densité spectrale de puissance est alors bien plus complexe à établir. Retenons que

Ce type de modulation a été utilisé, en particulier, dans les premières générations de modems

Modulation PSK

Phase–Shift Keying : Modulation par saut de phase. Ce type de modulation, du fait de sa grande résistance aux bruits, est utilisé en particulier pour les signaux GPS. La phase de la porteuse prend la valeur 0 ou π , selon la valeur du bit à transmettre. Etant donné que le signal modulé est un signal sinusoïdal multiplié par +1 ou -1, le "spectre" d'un signal PSK revient à celui d'un signal binaire aléatoire (donc en $\sin(x)/x$) translaté à la fréquence de porteuse.

Diagramme de constellation

Comme nous l'avons vu précédemment, lorsqu'un signal est modulé AM-P, le fait de le démoduler par multiplication avec un signal déphasé de $\pi/2$ puis filtrage passe-bas, donne un signal nul. Il est possible d'exploiter cette propriété pour transmettre simultanément deux signaux sur deux porteuses de même fréquence f_p mais déphasées l'une par rapport à l'autre de $\pi/2$; nous dirons deux porteuses en quadrature. Hormis la FSK, toutes les modulations sont basées sur le principe de la modulation AM-P de signaux m -aires par deux porteuses en quadrature.

Le signal modulé s'écrit alors sous la forme :

$$s(t) = i(nT) \cos(\omega_p t) + q(nT) \sin(\omega_p t),$$

avec $i(nT)$ et $q(nT)$ deux signaux m -aires de débit de symbole $1/T$.

La densité spectrale de puissance de ces différentes modulations aura l'allure d'un $\sin x/x$.

Chacun des signaux codés par ces modulations à signaux modulants discrets peut être représenté par un diagramme de constellation. Celui-ci consiste à représenter dans le plan (I,

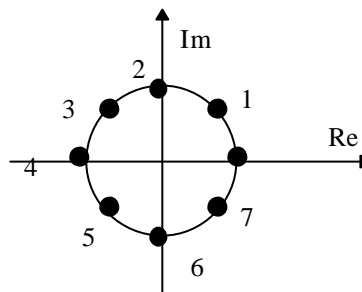
Q) (en phase et en quadrature avec la porteuse en $\cos(\omega t)$) les positions respectives des m signaux aux instant $n T$.

ASK-m

Amplitude-Shift Keying : Modulation par saut d'amplitude. L'amplitude de la porteuse peut prendre m valeurs différentes. Ce type de modulation est utilisé pour des signaux m-aires avec $m=4$.

PSK-m

Phase-Shift Keying : Modulation par saut de phase. La phase du signal modulé peut prendre m valeurs différentes. Ce type de modulation est utilisé pour des signaux m-aires mais le plus souvent dans les cas $m=2$ et $m=4$. La PSK- m est la plus robuste vis-à-vis du bruit. Typiquement, ce type de modulation est mis en œuvre pour les télécommunications par satellite.

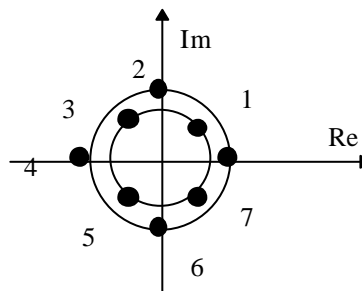


Constellation en PSK-8

Il existe de nombreuses variantes de PSK : DPSK, ODPSK...

A+PSK

Amplitude and Phase-Shift Keying : Modulation par saut d'amplitude et de phase. La phase du signal modulé peut prendre m' valeurs différentes et l'amplitude prend m'' valeurs différentes. Ce type de modulation est utilisé pour des signaux m-aires mais le plus souvent dans les cas $m=8$ ou $m=16$.

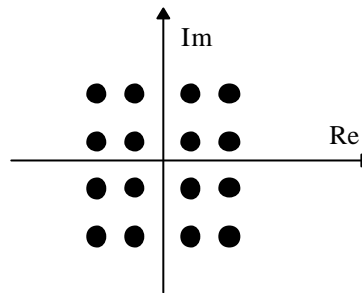


Constellation en A+PSK

QAM

Quadratic Amplitude Modulation : un signal en phase (In Phase) et un signal en Quadrature de Phase avec la porteuse sont modulés en ASK par m' ($m' = \frac{1}{2}\sqrt{m}$) et m'' ($m'' = 1/2$). Le signal QAM est obtenu en sommant ces deux signaux. Notons que le signal peut également s'écrire sous une forme qui rend son appellation immédiate :

$s(t) = A_p a_{m1}(t) \cos(\omega_p t) + A_p a_{m2}(t) \sin(\omega_p t)$. Il s'agit là d'un signal composé de la somme d'un signal modulé ASK en phase avec la porteuse (composante I pour In phase), avec un signal modulé ASK en quadrature avec la porteuse (composante Q pour Quadrature)



Constellation en QAM

MSK

Minimum Phase-Shift Keying : Modulation par saut de phase. La phase du signal modulé peut prendre m valeurs différentes. Ce type de modulation est utilisé pour des signaux m -aires mais le plus souvent dans les cas $m=2$ et $m=4$. Dans cette modulation, la phase varie continûment car la modulation revient à une modulation de fréquence. Le diagramme de constellation est alors identique à celui d'une PSK- m .

La MSK est également très robuste au bruit. C'est la modulation employée en téléphonie GSM et DCS.

FSK- m

Frequency-Shift Keying : Modulation par saut de fréquence. La fréquence du signal modulé peut prendre m valeurs différentes. Ce type de modulation est utilisé pour des signaux m -aires mais le plus souvent dans le cas $m=2$.

C'est l'exception à la règle car :

$$s(t) = A_p \cos(\omega_p t + m(nT) \Delta \omega t)$$

et non :

$$s(t) = i(nT) \cos(\omega_p t) + q(nT) \sin(\omega_p t).$$

Autres modulations numériques

Modulation par un code (CDMA et étalement de spectre)

De même que deux porteuses en quadratures peuvent transmettre deux signaux distincts, il est possible de moduler le signal binaire utile, de débit D , par un signal autre signal binaire pseudo-aléatoire et de débit D_a , D_a grand devant D . En prenant des signaux pseudo aléatoires dits orthogonaux entre eux, il est possible de transmettre, sans brouillage, plusieurs signaux utiles simultanément.

Le spectre du signal modulé est alors très large ce qui permet de rendre le signal robuste vis-à-vis des brouillages apparaissant sur une bande étroite de fréquences ou dus aux échos. Ce type de modulation a été retenu pour les téléphones cellulaires de 3^{ème} génération (UMTS).

Modulation fréquences orthogonales (OFDM)

De même que deux porteuses en quadratures dans le domaine temporel, il est possible de générer des porteuses orthogonales dans le domaine spectral. Cette solution a été retenue pour la diffusion TV numérique terrestre. Le signal modulé est alors robuste vis-à-vis des brouillages dus aux échos.

Généralités sur une liaison numérique

Synoptique

Filtrage optimal

Afin de limiter la bande passante et d'éviter l'Interférence Entre Symboles (IES) induite par le canal (filtrage, distorsions, etc...), il est nécessaire de filtrer le signal après modulation puis avant de le démoduler. En l'absence de bruit et de distorsion, pour que le signal vérifie le critère de Nyquist (IES nulle), le filtre optimal est un filtre dit en cosinus surélevé.

En pratique, il faut recourir à des filtres adaptatifs en réception afin de garantir une IES nulle, le filtre d'émission restant un filtre dit en cosinus surélevé.

Récupération de rythme. Synchronisation

Codes correcteurs d'erreur

MODULATIONS ANALOGIQUES D'IMPULSIONS ET MULTIPLEXAGE TEMPOREL

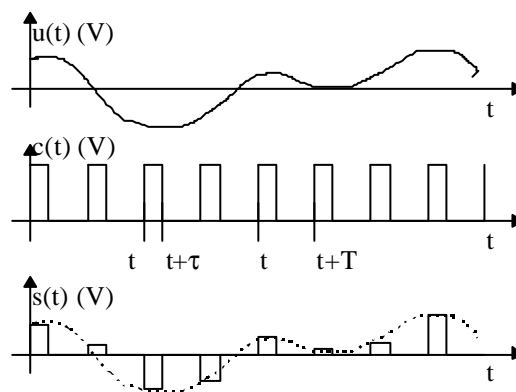
Plutôt que d'employer une porteuse sinusoïdale, il est possible de ne transmettre qu'un train d'impulsions dont les caractéristiques (amplitude, durée, décalage temporel par rapport à une référence) vont permettre le codage de l'information puis son multiplexage avec d'autres signaux.

PAM

principe

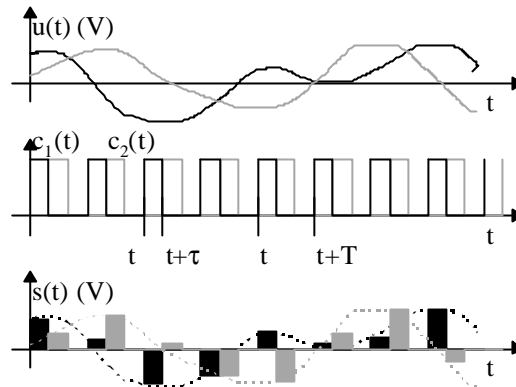
La modulation d'impulsion en amplitude ou PAM (Pulse Amplitude Modulation) est d'un principe simple. Il consiste à utiliser un échantillonneur-bloqueur pour obtenir le signal $s(t)$ à partir du signal informatif $u(t)$, échantillonné à intervalles de temps réguliers.

Le signal $s(t)$ est obtenu en multipliant le signal $u(t)$ par un signal carré $c(t)$ de période T_e dont le taux de remplissage r est inférieur à $1/2$ ($r = \tau/T_e$). L'allure du signal $s(t)$ est alors illustrée par la figure ci-dessous :



Signal modulé PAM.

L'un des intérêt d'une telle modulation est de permettre le multiplexage temporel en multipliant les différents signaux $u_1(t), u_2(t), \dots, u_n(t)$ ($n=8$) respectivement par les signaux carrés $c_1(t), c_2(t)=c_1(t+\tau), \dots, c_n(t)=c_1(t+n\tau)$.



Multiplexage temporel de signaux modulés PAM.

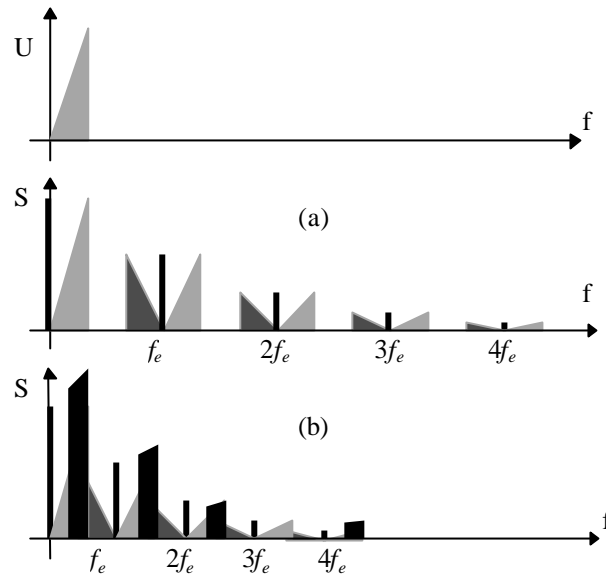
En téléphonie, jusqu'à 30 voies sont ainsi échantillonnées et multiplexées.

spectre

Le spectre du signal $s(t)$ est alors donné par le produit de convolution de $\underline{U}(j\omega)$, la transformée de Fourier de $u(t)$, avec $\underline{C}(j\omega)$, la transformée de Fourier de $c(t)$. Or, le spectre de $c(t)$ est composé de raies aux fréquences n/T et d'amplitude :

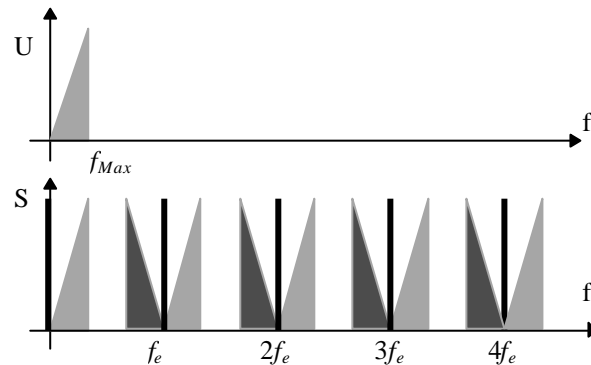
$$C_n = A \frac{\frac{t}{T} \sin\left(n \frac{t}{T}\right)}{n \frac{t}{T}} = A \cdot r \cdot \text{sinc}(nr).$$

De ce fait, le spectre de $s(t)$ est un spectre où le spectre de $u(t)$ se retrouve symétriquement de part et d'autre des fréquences $n/T_e = n f_e$. F_e est appelée fréquence d'échantillonnage. Si le spectre de $u(t)$ est plus large que $f_e/2$, il y a recouvrement du spectre de $u(t)$ centré sur la raie $n f_e$ avec celui centré sur $(n+1)f_e$. Pour qu'il n'y ait pas de repliement de spectre, il est impératif que la fréquence d'échantillonnage soit supérieure à deux fois la fréquence maximum du signal $u(t)$. C'est le théorème de Shannon valable dès lors que l'on travaille avec des signaux échantillonnés.



spectre de signaux échantillonnés, sans, puis avec repliement de spectre.

En pratique, on cherche à multiplier le nombre voies transmises simultanément ce qui conduit à rechercher τ petit devant T_e . Le signal $c(t)$ se rapproche alors d'un peigne de Dirac et le spectre du signal échantillonné est alors de la forme :



Spectre d'un signal échantillonné par un peigne de Dirac

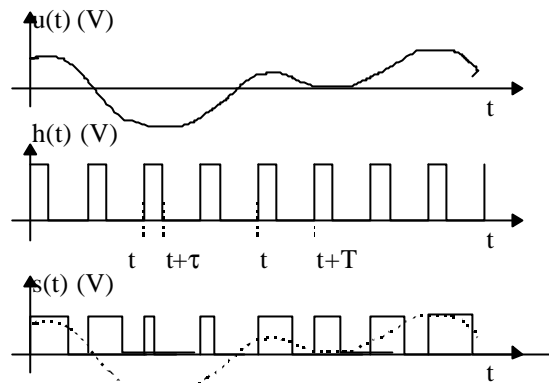
Retenons que lorsque les signaux téléphoniques sont échantillonnés, ils le sont à une fréquence de 8 kHz, ce qui est une normalisation internationale.

La démodulation du signal PAM est très simple puisqu'il suffit de filtrer les composantes HF dues à l'échantillonnage. En revanche la difficulté liée à ce type de multiplexage vient de la synchronisation des différents signaux.

La PAM n'est pas utilisée tel quel en téléphonie car les signaux ainsi codés sont relativement sensibles aux perturbations, en particulier la diaphonie temporelle. La PAM n'a pas d'application directe mais elle constitue une étape pour la réalisation d'autres modulations.

Modulation par durée d'impulsion ou PDM

La modulation par durée d'impulsion ou PDM (Pulse Duration Modulation) consiste à échantillonner $u(t)$. A chaque front montant de l'horloge, le signal modulé est un signal carré dont le rapport cyclique ρ varie en modifiant la durée τ en fonction de l'amplitude du signal $u(t)$. Le signal prend alors la forme suivante :

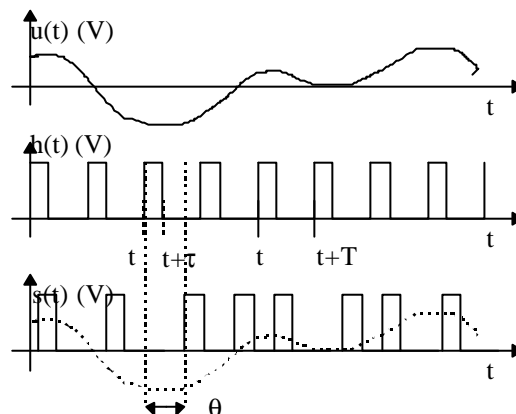


Signal modulé PDM.

Par rapport à la modulation PAM, la modulation PDM présente l'avantage d'une bonne immunité aux bruits. La modulation PDM peut être réalisée en utilisant une modulation PAM suivie d'une conversion amplitude-temps.

Modulation d'impulsion en position PPM

La modulation par durée d'impulsion ou PPM (Pulse Position Modulation) consiste à échantillonner $u(t)$. A chaque front montant de l'horloge, le signal modulé $s(t)$ est un train d'impulsions. Chaque impulsion est retardée de θ , par rapport à l'horloge, en fonction de l'amplitude du signal $u(t)$. Le signal prend alors la forme suivante :



Signal modulé PPM.

La modulation PPM peut-être obtenue simplement en réalisant une opération XOR entre un signal modulé PDM et une horloge dont le taux de remplissage est $\rho=1/2$. La puissance nécessaire à la transmission du signal PPM est indépendante du signal modulant. Les signaux L'encombrement spectral est cependant relativement important.

Conclusion

De part leurs caractéristiques, les modulations PDM et surtout PPM (et sont extension PFM) sont surtout utilisées pour les transmissions de signaux analogiques sur fibres optiques, en télécommande IR ou en télémétrie.

5. MODULATIONS NUMERIQUES

GENERALITES

Les modulations numériques impliquent nécessairement l'échantillonnage des signaux analogiques puis leur codage sous la forme d'un mot écrit sous forme binaire. Comme précédemment, l'échantillonnage implique le respect du théorème de Shannon ($f_{Max} < 2 f_e$). Par ailleurs, le codage sous forme d'un mot binaire se traduit par une erreur de quantification. En effet, un mot de n bits ne peut prendre que 2^n valeurs discrètes. Cette discrétisation introduit une erreur entre le signal original et le signal numérisé. Cet écart entre les deux signaux est nommé bruit de quantification. Enfin, du fait même du codage, le signal numérisé occupe un spectre nettement plus important que le signal analogique original. Toutefois, les problèmes de distorsion et de bruit peuvent être traités de façon largement simplifiés ce qui permet régénérer les signaux au cours des transmissions. Enfin, la simplicité (relative, ne nous y trompons pas) et la souplesse de mise en œuvre, conduisent à la généralisation des codages numériques.

QUANTIFICATION

Rappelons tout d'abord qu'un signal binaire de n -bits ne peut prendre que 2^n valeurs différentes c'est-à-dire 2^n valeurs discrètes. Par ailleurs, les convertisseurs analogiques-numériques sont caractérisés par une dynamique finie V . De ce fait, les valeurs extrêmes du signal numérisé seront comprises entre V_{refmin} et V_{refMax} et $V_{refMax} - V_{refmin} = V$. Le pas de quantification est alors la quantité q . Deux valeurs numériques consécutives du signal numérique se traduisent par un écart de q . Par ailleurs, on parle de quantification uniforme lorsque q est constante et définie par : $q = \frac{V}{2^n}$.

La dynamique limitée et l'existence d'un pas de quantification entraînent deux types de distorsion du signal numérisé par rapport au signal analogique original.

Distorsion par écrêtage

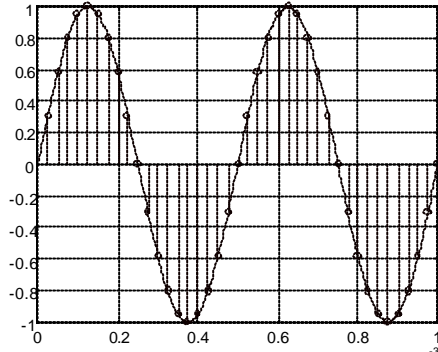
Lors de la numérisation d'un signal analogique $u(t)$, celui-ci ne pourra être convenablement numérisé que si ses valeurs extrêmes sont comprises entre les deux valeurs extrêmes admises par le convertisseur : V_{refmin} et V_{refMax} . Si le signal $u(t)$ est en dehors de cette plage, le signal numérique $u_q(t)$ est écrêté. Cet écrêtage introduit une distorsion harmonique.

Bruit de quantification pour une quantification uniforme

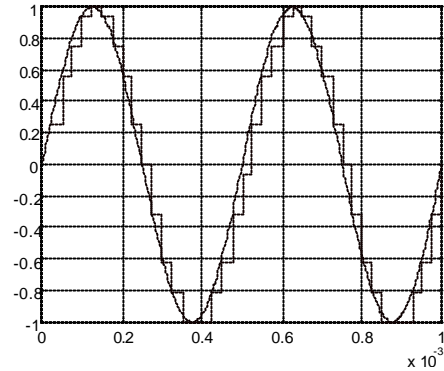
Il serait possible de concevoir un dispositif pour lequel la quantification serait naturelle. Cependant, en règle générale, la première étape de la numérisation est l'échantillonnage. Nous appellerons $u^*(t)$ le signal $u(t)$ échantillonné à une fréquence f_e et $u_q(t)$ le signal quantifié avec un pas uniforme q . Le pas de quantification entraîne un écart systématique entre les signaux $u_q(t)$ et $u^*(t)$. Cet écart inévitable entre les deux signaux est équivalent à un bruit dont il faut évaluer l'impact sur la qualité de la numérisation à l'aide du rapport signal/bruit $(S/N)_q$.

Cas général

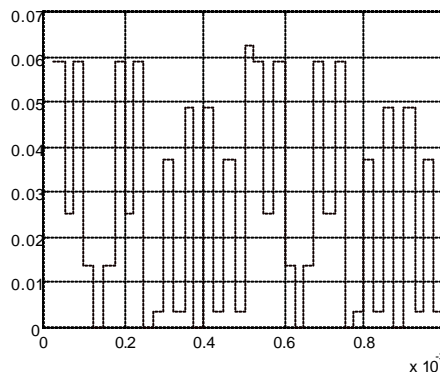
Si $u(t)$ est un signal aléatoire, dont les valeurs extrêmes U_{min} et U_{Max} sont comprises dans $[V_{min}, V_{Max}]$ (le signal numérisé n'est pas échantillonné), alors le signal $b_q(t) = u^*(t) - u_q(t)$ est un signal en marche d'escalier à valeur moyenne nulle et dont les valeurs extrêmes sont : $-q/2$ et $q/2$.



signal 1 kHz échantillonné



signal 1 kHz échantillonné et quantifié

écart entre le signal $u^*(t)$ et $u_q(t)$

Dans le cas où q est petit ($2^n \gg 2$), il est possible d'admettre que la distribution de $b_q(t)$ est uniforme sur $]-q/2, q/2]$. Il est alors relativement simple de montrer que le spectre de la puissance de ce signal $b_q(t)$, la densité spectrale de puissance, est constante entre 0 et $f_e/2$. La puissance de bruit de quantification, P_q , est alors donnée par :

$$P_q = a \frac{q^2}{12}, \text{ où } a \text{ vaut } I/R \text{ si } u(t) \text{ est une tension et } R \text{ si } u(t) \text{ est un courant.}$$

Moyennant certaines hypothèses, vérifiées par la très grande majorité des signaux, la puissance P_s , du signal $u(t)$ est donnée par :

$$P_s = a \int_{U_{min}}^{U_{Max}} u(t) \cdot dt = a \cdot U_{eff}^2.$$

Le rapport signal/bruit dû à la quantification, $(S/N)_q$, est : $\left(\frac{S}{N}\right)_q = 12 \times \left(\frac{U_{eff}}{q}\right)^2$, soit en

dB :

$$\left(\frac{S}{N}\right)_{q(dB)} = 10 \times \log \left(12 \times \frac{U_{eff}^2}{q^2} \right).$$

Cette formule met en évidence un aspect qu'il est facile à admettre : plus q est faible (donc plus n est grand), plus la quantification est fine et meilleur est le rapport signal-bruit.

Cas d'un signal sinusoïdal

Dans le cas où $u(t)$ est un signal sinusoïdal $u(t)=A \cos(\omega t)$, on a :

$$P_s = a \int_{U_{\min}}^{U_{\max}} u(t) \cdot dt = a \cdot \frac{A^2}{2} \text{ et donc : } \left(\frac{S}{N} \right)_q = 6 \times \left(\frac{A}{q} \right)^2.$$

Pour un convertisseur centré sur $[-A, A]$, le pas de quantification, q , vaut : $q = \frac{2A}{2^n}$.

Le rapport signal sur bruit est donc : $\left(\frac{S}{N} \right)_q = 6 \times \left(\frac{2^n}{2} \right)^2 = \frac{3}{2} \times 2^{2n}$. En clair, lorsque la

résolution du convertisseur augmente d'un bit, le rapport signal/bruit augmente de 6 dB.

En fait, dans le cas du signal sinusoïdal, il est plus intéressant de définir les distorsions introduites par la quantification en utilisant le taux de distorsion d_q . P_q représente directement la puissance de l'écart à la sinusoïde échantillonnée, donc, le taux de distorsion est :

$$d_q = \sqrt{\frac{P_q}{P_s}} = \sqrt{\frac{2}{3}} \times \frac{1}{2^n}. \text{ Il est d'usage d'exprimer } d_q \text{ en pour-cent. A titre d'exemple, le tableau}$$

ci-dessous donne le taux de distorsion :

$n (2^n)$	3 (8)	4 (16)	5 (32)	6 (64)	7 (128)	8 (256)	10 (1024)	12 (4096)
$d_q (\%)$	10,2%	5,1%	2,55%	1,28%	0,64%	0,32%	0,0797	0,02

Rapport signal/bruit de quantification instantané. Cas général.

Jusqu'à présent, nous avons défini le rapport signal/bruit en raisonnant sur les puissances moyennes. Or, dans le cas de signaux pour lesquels les valeurs instantanées sont importantes, il devient nécessaire de travailler avec les puissances instantanées. Par analogie avec $(S/N)_q$, cela conduit au rapport signal/bruit instantané $(S/N)_q(t)$ donné par :

$$\left(\frac{S}{N} \right)_q(t) = 12 \times \frac{u^2(t)}{q^2}.$$

En supposant que le signal $u(t)$ soit toujours dans $[-A, A]$ et en prenant un convertisseur centré sur $[-A, A]$, étant donné que le pas de quantification, q , vaut : $q = \frac{2A}{2^n}$, l'expression

$(S/N)_q(t)$ est :

$$\left(\frac{S}{N} \right)_q(t) = 3 \times \frac{u^2(t)}{A^2} \times 2^{2n}, \text{ soit en dB : } \left(\frac{S}{N} \right)_q(t)_{dB} = 20 \log \left| \frac{u(t)}{A} \right| + n \times 6 \text{ dB} + 4,7 \text{ dB}.$$

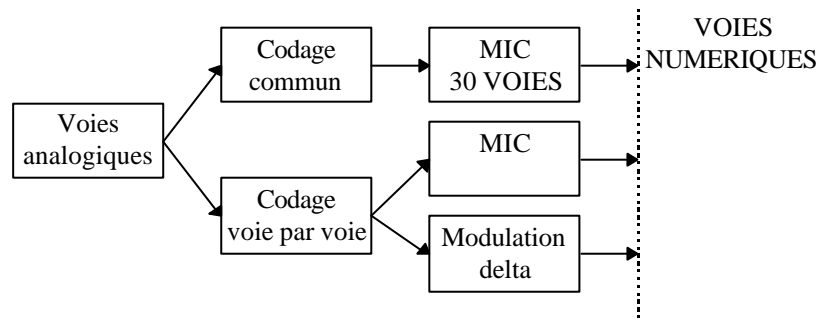
Encore une fois, augmenter la résolution du convertisseur d'un bit permet de gagner 6 dB.

NUMERISATION : CAS DES VOIES TELEPHONIQUES

En téléphonie, plusieurs impératifs se combinent. D'une part il est nécessaire de préserver une bonne qualité des signaux transmis. En particulier, il est indispensable que signal/bruit $(S/N)_q$ reste élevé pour des signaux de faibles amplitudes auxquels l'oreille humaine est sensible et qui constituent une part importante du message. D'autre part, il est essentiel de n'utiliser qu'un nombre de bits relativement faible (8 en pratique) afin de ne pas trop augmenter la bande des signaux à transmettre, ce qui revient à limiter le débit binaire, D^* .

Dans le cas de signaux téléphoniques numérisés, deux techniques ont été adoptées pour préserver un bon rapport signal tout en limitant le débit binaire D^* . Afin d'éviter d'avoir à coder le signal $u(t)$ sur un nombre important de bits, une première solution consiste à coder les variations du signal entre deux instants successifs (codage delta). L'autre solution consiste à utiliser une quantification non-linéaire (codage MIC). Ces deux solutions ont été retenue dans le réseau téléphonique commuté (RTC) dont les équipements se répartissent en deux grandes familles :

- ceux réalisant le codage commun des voies (l'échantillonnage est effectuée voie par voie puis les signaux sont modulé PAM puis multiplexés avant d'être numérisés),
- ceux réalisant le codage voie par des voie (le codage numérique est effectuée voie par voie puis les signaux codés en binaire sont multiplexés).

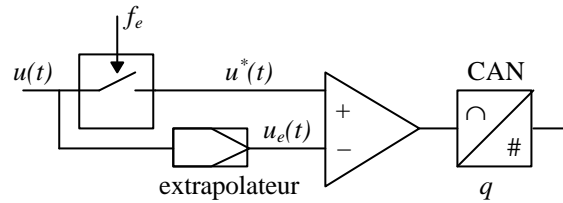


Principe de numérisation des voies téléphoniques

MODULATIONS NUMERIQUES DIFFERENTIELLES**Modulation delta****Principe**

Une des techniques de codage consiste à coder à chaque instant nT_e , pour un signal $u(t)$, l'écart entre le signal échantillonné $u^*(t)$ et un signal $u_e(t)$, estimation de $u(t)$, obtenue par extrapolation $u(t-T_e)$. En modulation delta, $u_e(t)$ est un signal en escalier qui ne peut varier que d'un pas de quantification q entre deux instants d'échantillonnage. C'est alors tout simplement l'écart entre $u(t)$ et $u_e(t-T_e)$ qui est codé par un bit. Si $u(t)-u_e(t-T_e) > 0$ alors le bit transmis vaut "1" et $u_e(t) = u_e(t-T_e) - q$, sinon, le bit transmis vaut "0" et $u_e(t) = u_e(t-T_e) + q$.

Un synoptique de la modulation delta est donnée ci-dessous :



Principe de la modulation différentielle

Les deux seuls paramètres de cette modulation sont la fréquence d'échantillonnage et le pas de quantification q . Le débit d'information D^* vaut alors f_e .

Saturation de pente

Voyons maintenant quelle sera la pente maximum que pourra avoir $u_e(t)$. Le signal estimé ne peut varier que de q pendant un instant T_e . La pente maximum de ce signal est donc $q f_e$. Si l'on considère un signal sinusoïdal $u(t) = A \cos(\omega t)$, la pente S_l de $u(t)$ est maximale en $t=0$ et est donnée par :

$$S_l = \left. \frac{d}{dt} u(t) \right|_{t=0}$$

$$S_l = A \cdot \omega$$

De ce fait, à fréquence $f = \omega/2\pi$ maximum donnée, l'amplitude du signal devra être :

$$A = \frac{q \cdot f_e}{2 \cdot \pi \cdot f}.$$

Cette formule appelle deux remarques :

1. Pour avoir une bonne résolution (ou une distorsion faible) la fréquence d'échantillonnage f_e en modulation delta doit être grande devant f_{Max} . En pratique, il faut utiliser une fréquence f_e valant 16 kHz ou 32 kHz (soit respectivement 5 et 10 fois f_{Max} pour les signaux téléphoniques). Il faut cependant noter que ces débits sont relativement modestes.

2. Le codage sera d'autant plus efficace si l'amplitude des signaux décroît lorsque les fréquences augmentent, ce qui est le cas pour les signaux téléphoniques.

Bruit de quantification

La distorsion de quantification est inférieure à q car $|u(t) - u_e(t)| < q$.

En faisant certaines hypothèses, vérifiées par des mesures, (en particulier que $u(t) - u_e(t)$ soit uniformément distribué jusqu'à f_e), il est possible de montrer que la puissance du bruit de quantification P_q est :

$$P_q = a \frac{q^2}{3} \quad (\text{rappel : } a = 1/R \text{ ou } R). \text{ Or, on a vu que } f_{Max} \text{ doit être nettement inférieure à } f_e.$$

De ce fait, comme la puissance de bruit est nulle au-delà de f_{Max} , l'expression de la puissance du bruit de quantification peut être estimée par :

$$P_q = a \frac{q^2}{3} \times \frac{f_{Max}}{f_e}.$$

Le rapport signal sur bruit de quantification $(S/N)_q$ est alors :

$$\left(\frac{S}{N}\right)_q = \frac{3}{2} \times \left(\frac{A}{q}\right)^2 \times \frac{f_{Max}}{f_e}.$$

A q fixé, le signal/bruit croît avec le carré de l'amplitude et augmente linéairement avec la fréquence du signal $u(t)$. Lorsque q et f_{Max} sont fixés, la seule solution consiste à augmenter la fréquence d'échantillonnage.

Codage delta auto-adaptatif

Pour améliorer la qualité du codage en conservant la même fréquence d'échantillonnage, il peut être intéressant de faire varier le pas de quantification. Puisque le codage delta est un codage différentiel, il est intéressant de faire varier le pas q en fonction de la pente de $u(t)$. La variation du pas q peut être réalisée par des moyens analogiques ou numériques. Dans ce dernier cas, c'est l'analyse de la séquence de bits issus du comparateur qui induit le changement de q (en particulier, les longues séquences de "1" ou de "0"). Ce type de codage est appelé PACN (codage delta à pente asservie et à codage numérique).

Autres applications

En dehors de la téléphonie, le codage delta est également très bien adapté à la transmission des images. En effet, dans les cas les plus fréquents, les signaux codant des images sont caractérisés par des spectres dont l'amplitude décroît avec la fréquence.

MODULATION MIC ADAPTATIVE

Comparaison entre modulations MIC et Delta.

Le système Delta est beaucoup plus simple à mettre en œuvre que le système MIC car le codage ne se fait que sur un bit (un seul comparateur suffit). La modulation delta est d'autant plus efficace que les échantillons successifs d'un signal ne sont pas indépendants. Ce type de modulation est donc bien adaptée à la parole ou à la transmission d'image. En revanche, la PCM est nettement plus efficace pour transmettre des signaux dont les échantillons sont aléatoires et donc proches d'un bruit blanc gaussien. La modulation PCM est donc d'un emploi plus générale mais plus lourd que la modulation delta.

L'amélioration du rapport signal/bruit de 6dB nécessite une augmentation d'un bit en modulation PCM et de doubler la fréquence d'échantillonnage en modulation delta. De ce fait, La modulation MIC permet d'obtenir un meilleur rapport signal/bruit que la modulation delta dès que le débit D^* est supérieur à 48 kbits.s⁻¹.

En revanche, du fait même du principe de codage différentiel, le codage delta est beaucoup moins sensible aux erreurs de transmissions (erreurs de retranscription sur les signaux binaires). Ainsi, un taux d'erreur de 10^{-2} n'affecte pas sensiblement la restitution d'un signal modulé delta alors qu'un taux d'erreur de 10^{-3} rend inopérant un signal modulé MIC.

Dans le cas d'une transmission vocale, il est possible de réduire le débit binaire en préservant la qualité du message en combinant le principe de la modulation différentielle avec une prédiction supérieur à zéro et une adaptation dynamique du pas de quantification. Ainsi, l'UIT-T a défini un procédé de modulation par impulsion et codage différentiel adaptatif ADPCM qui fait l'objet de la recommandation G 721. Les principales caractéristiques sont les suivantes :

- $u - u_e$ est quantifié sur 15 niveaux (par 4 bits) selon une loi non-uniforme
- le débit est de 32 kbits.s⁻¹ (qualité identique à la modulation MIC à 64 kbits.s⁻¹),
- u_e est estimé par un algorithme complexe.

A débit donné, le rapport signal/bruit de l'ADPCM est systématiquement meilleure d'environ 10 dB.

TRANSMISSIONS NUMERIQUES

MULTIPLEXAGE DES SIGNAUX NUMERIQUES

Généralités

Quelle que soit la solution adoptée pour le codage, soit par voie ou commun, son résultat est une suite binaire. L'étape suivante consiste en général à fournir un train à plus grand débit par multiplexage temporel de plusieurs trains numériques. Pour former ce train, deux approches sont possibles

Multiplexage tramé

Le multiplexage par répartition dans le temps de N signaux numériques de durées élémentaires t est réalisé en affectant à chaque impulsion incidente un intervalle de temps bien, déterminé d'une durée voisine de t/N et en entrelaçant les impulsions ainsi réduites en largeur.

Le problème essentiel qui se pose dans ce type de multiplex est la distribution du rythme.

Multiplexage non tramé - Commutation de paquets adressés

Lorsque l'on utilise un support de transmission pour lequel la notion de bande passante est essentielle, il est intéressant d'utiliser au mieux ce support.

Ceci revient à réaliser un multiplexage auto-adaptatif à la quantité d'information à acheminer.

Dans ce type de multiplexage, une communication donnée n'a pas d'intervalle de temps affecté en propre, elle s'insère dans le flot des autres communications acheminées sur le support dès qu'elle trouve un intervalle de temps libre.

Le multiplexage non tramé est asynchrone.

Multiplexage tramé

principe

La trame du train multiplexé est divisée en intervalles de temps bien définis. Chaque intervalle de temps est affecté en propre un train composant incident. Le multiplexage est réalisé par entrelacement symbole par symbole.

Rythme

Pour pouvoir être multiplexés, les signaux numériques doivent être synchronisés, c'est-à-dire que leurs instants significatifs doivent avoir une relation de phase voulue. Lorsque le signal de rythme restitué (instants de test) s'accompagne d'une modulation de phase parasite, il y a de la gigue (en anglais : jitter).

Il faut donc s'affranchir des variations du temps de propagation et de la gigue et positionner les impulsions : ces fonctions sont réalisées dans une mémoire tampon ; celle-ci sert également à permettre l'insertion des bits de remplissage.

Pour obtenir un multiplexage sans pertes d'information, deux procédés sont possibles :

Multiplexage synchrone

Le réseau est synchronisé sur une même horloge. Le multiplexage est simple à réaliser mais la distribution du rythme pose des problèmes délicats lorsqu'il s'agit d'un réseau universel couvrant une grande étendue géographique.

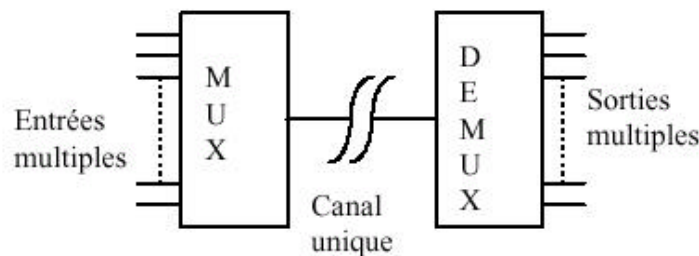
Multiplexage asynchrone

Ou plus exactement, multiplexage de signaux plésiochrones.

Les signaux numériques plésiochrones sont tels que leurs instants significatifs se présentent à la même cadence nominale, toute variation de cette cadence étant maintenue dans des limites spécifiées.

6. LES TECHNIQUES DE MULTIPLEXAGES

Les techniques décrites dans les cours précédents faisaient référence à des communications point à point, c'est à dire d'un émetteur vers un récepteur sans préjugé du canal de transmission. Mais que faire lorsque plusieurs utilisateurs utilisent le même support physique, c'est à dire le même canal de transmission ? Comment partager une onde radio, ou une fibre optique, ou un câble électrique ? c'est la réponse à laquelle ce cours essaye de répondre. En effet, à partir du moment où plusieurs utilisateurs se partagent un seul support de transmission, il est nécessaire de définir le principe de partage du canal de transmission. L'opération qui consiste à permettre à plusieurs signaux d'entrées d'accéder à un canal de transmission unique s'appelle le multiplexage. L'opération inverse s'appelle le démultiplexage, et elle permet en sortie du canal unique de séparer les différents signaux en leur affectant à chacun une sortie.

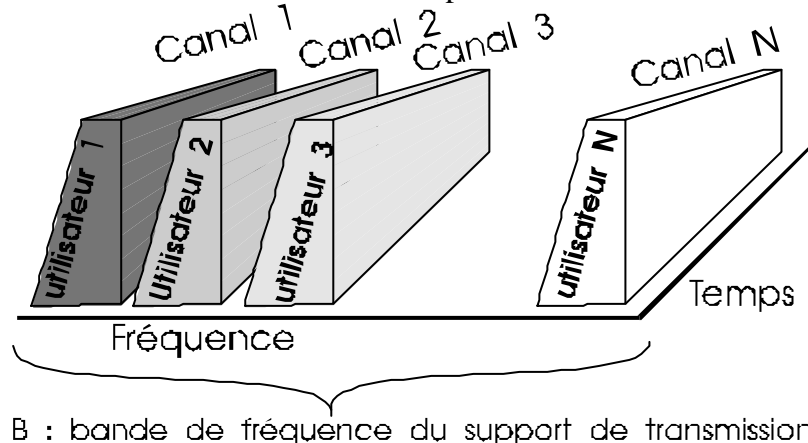


LE MULTIPLEXAGE EN FREQUENCE (FDM: FREQUENCY DIVISION MULTIPLEXING)

Historiquement c'est l'une des techniques les plus anciennes. On peut considérer qu'elle est utilisée en radio où les différentes stations utilisent des fréquences porteuses différentes. Elle a été aussi utilisée dans les années 60 en téléphonie fixe.

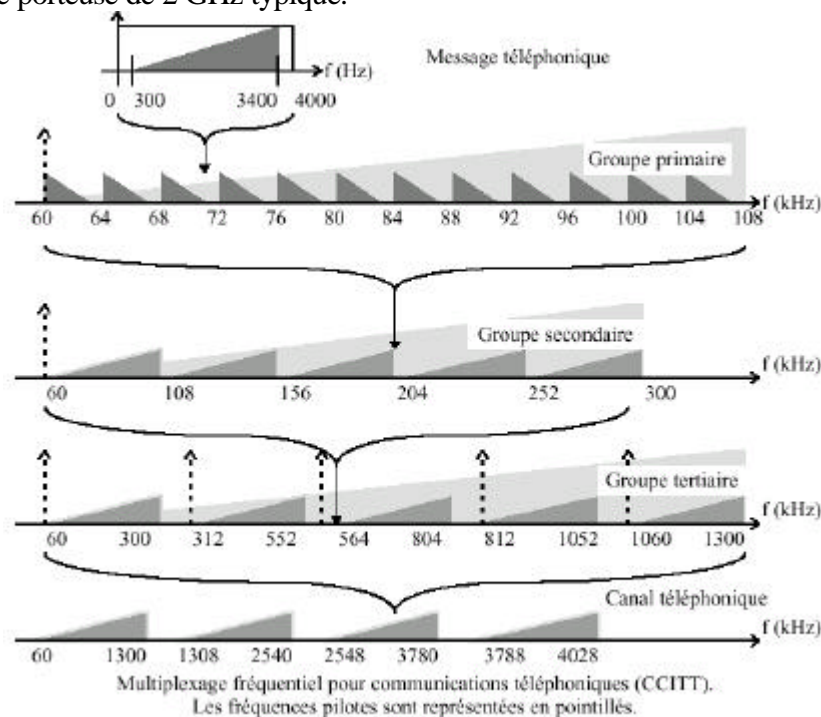
PRINCIPE:

La technique de multiplexage en fréquence consiste à partager la bande de fréquence du support de transmission en différents canaux caractérisés par des fréquences différentes. Chaque utilisateur se voit alors attribuer un canal c'est à dire une fréquence sur laquelle il peut émettre en continu c'est à dire sans contrainte temporelle.



EXEMPLE : MODULATION AM-AM. APPLICATION AUX COMMUNICATIONS TELEPHONIQUES

Pour transmettre plusieurs conversations téléphoniques sur un même câble, les compagnies téléphoniques ont utilisé la FDM. Dans ce cas, le procédé de modulation employé est une modulation SSB multiple qui a fait l'objet d'une recommandation par le UIT-T (Union Internationale des Télécommunications - T) et se présente de la manière suivante : un message téléphonique occupe une bande comprise entre 300 et 3400 Hz. Le message correspondant est utilisé pour réaliser une modulation d'amplitude sans porteuse à bande latérale unique (AM SSB-USB) et avec une occupation spectrale de 4 kHz. La porteuse de ce dernier message est un multiple de 4KHz compris entre 60 kHz et 108 kHz. Dans la plage 60 à 108 KHz il est donc possible de placer 12 canaux qui forment le groupe primaire. Un ensemble de cinq groupes primaires chacun modulant une porteuse comprise entre 60 et 300kHz forme le groupe secondaire (AM SSB-USB) et contient donc 60 messages. De même ensemble de cinq groupes secondaires, séparés les uns des autres par une bande de fréquence inutilisée de 8 à 12 kHz, forme un groupe tertiaire qui comporte 300 messages. Enfin, un canal téléphonique est constitué de 4 groupes tertiaires séparés les uns les autres par 8Khz; et d'occupation spectrale 4,9MHz. Notons que simultanément, sont transmises des fréquences pilotes qui permettent de simplifier les démodulations. finalement le canal téléphonique sert de modulant a une porteuse de 2 GHz typique.

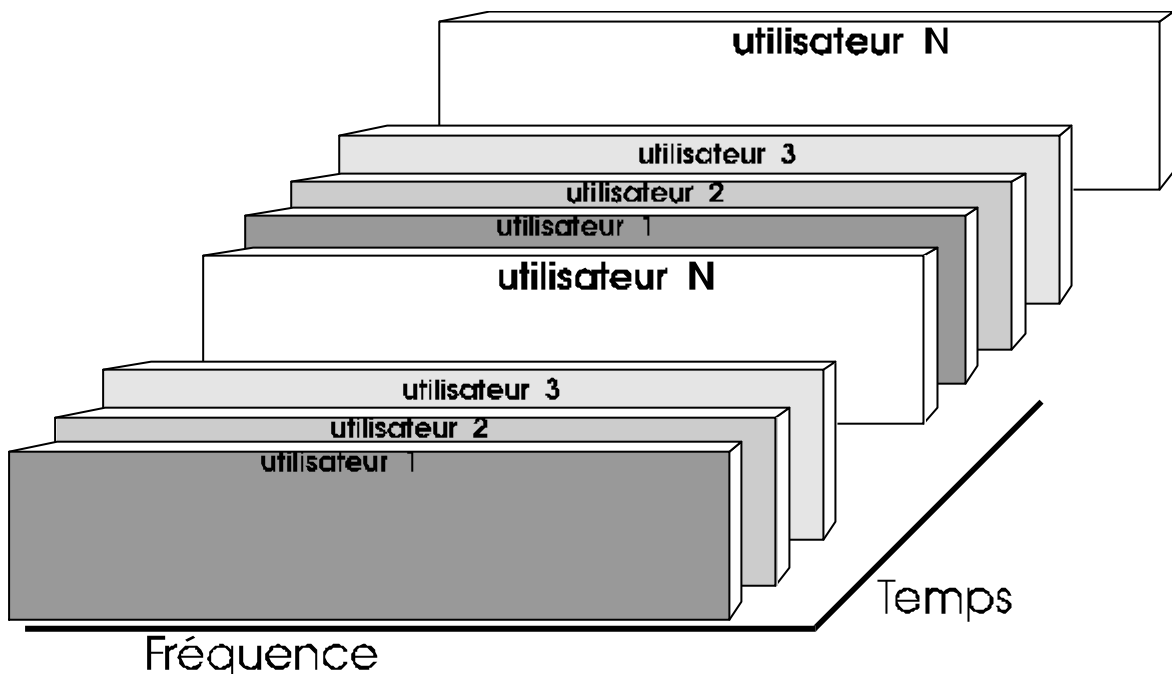


LE MULTIPLEXAGE TEMPOREL : (TDM : TIME DIVISION MULTIPLEXING)

On l'appelle encore accès multiple à répartition en temps

PRINCIPE :

Les utilisateurs se partagent le canal en temps. Le temps est découpé en intervalles de longueur fixe appelé IT (Intervalles de temps). Périodiquement, chaque utilisateur peut émettre une partie de son message. Pour émettre la suite de son message chaque utilisateur doit attendre une période afin que l'intervalle de temps qui lui a été attribué soit à nouveau disponible. L'émetteur n'émet donc pas en continu, mais les messages sont compressés et insérés dans l'IT alloué.

**EXEMPLE1: MULTIPLEXAGE TEMPOREL DE SIGNAUX PAM**

Le signal issu d'une modulation PAM n'occupe pas le canal de transmission pendant la durée comprise entre deux impulsions successives. Il est donc possible d'y intercaler d'autres impulsions PAM provenant d'autres signaux. En faisant cela on réalise ce qu'on appelle du multiplexage temporel.

Le nombre de signaux PAM que l'on peut entrelacer dépend de la largeur τ des impulsions PAM. En général, un modulateur PAM transmet ces impulsions à une paire de fils torsadés

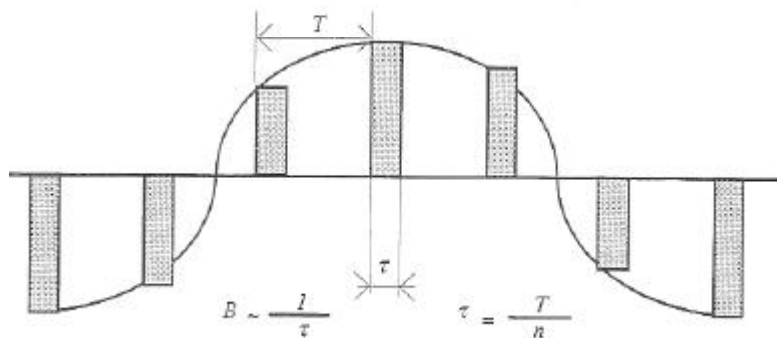
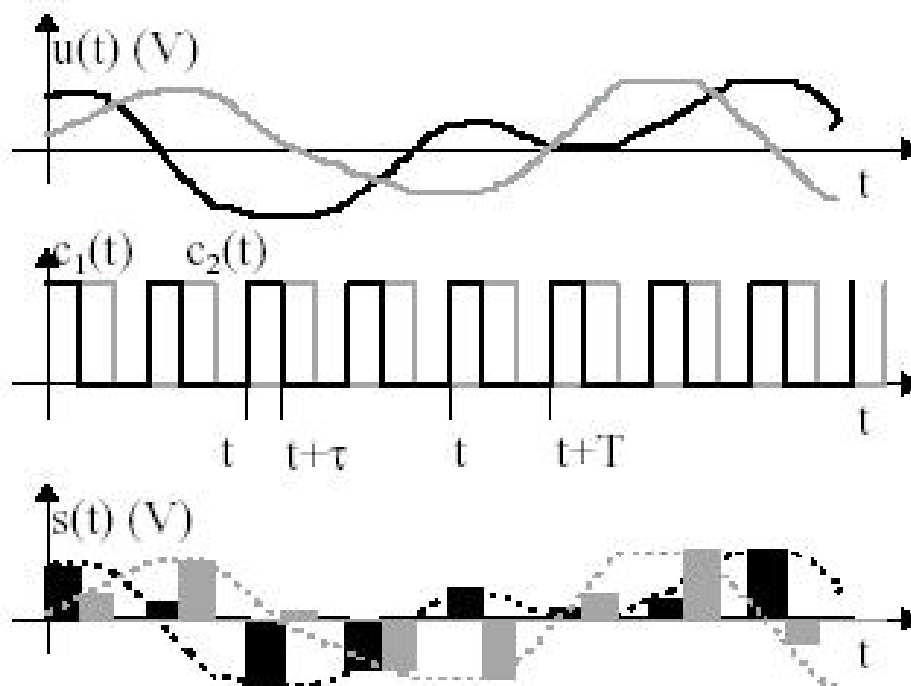


Figure 3. Durée de l'impulsion PAM

d'environ 2km, ce qui définit une bande passante d'environ $B=260\text{KHz}$. On a donc $1/\tau \geq B$, ce qui donne une durée d'impulsion PAM légèrement inférieure à $\tau=4\mu\text{s}$. Le nombre de signaux modulés PAM que l'on peut donc intercaler dans une fenêtre de $125\mu\text{s}$ est donc d'environ 31. Au total 32 voies sont ainsi multiplexées dans une fenêtre temporelle de $125\mu\text{s}$.



Multiplexage temporel de signaux modulés PAM.

Figure 4 : principe du multiplexage temporel sur des signaux PAM

EXEMPLE 2 : TRAME PRIMAIRE TN1 DE LA TECHNOLOGIE PDH

Une fois qu'un échantillon MIC a été quantifié, il peut être transmis. Et un nouvel échantillon peut alors être échantillonné, compressé et quantifié. L'ensemble de ces opérations doit cependant être limité dans le temps. En effet puisqu'on échantillonne à 8000 Hz, la durée de "fabrication" des mots de 8 bits codant l'échantillon ne peut dépasser 125µs. En réalité les mots de 8 bits sont produits très rapidement et on peut même insérer entre deux mots successifs provenant de la même voie, des mots provenant d'autres voies. On réalise ainsi une transmission multiplexée temporellement de plusieurs voies. (Recommandation UIT-T Rec. G711)

En Europe le système qui réalise cette fonction est appelé MIC E1 (Modulateur d'impulsions codées). C'est le premier niveau de la hiérarchie de multiplexage. Dans ce système 30 voies sont multiplexées temporellement dans une "fenêtre" de 125µs.

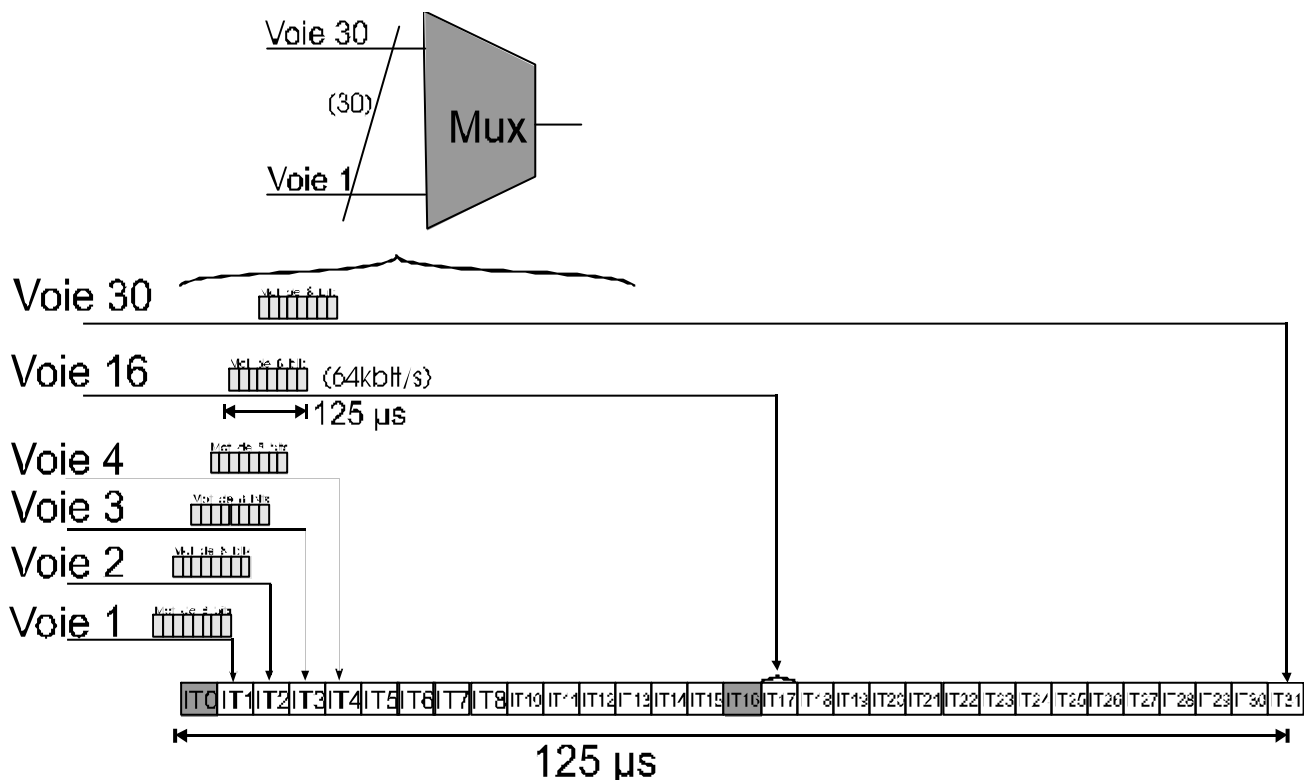


Fig 8 : Structure d'une trame MIC

La fenêtre de 125µs appelée trame est donc découpée en 32 morceaux appelés "Intervalle de temps" ou "IT", qui vont de IT0 à IT31. Les 32 intervalles de temps sont occupés de la manière suivante ; Les échantillons compressés et quantifiés en provenance des voies 1 à 15, sont placés respectivement dans les IT1 à IT15. De même pour les voies 16 à 30 placées dans les IT17 à IT31. Seuls les IT0 et IT 16 sont particuliers car ils ne sont pas utilisés pour transmettre des voies mais pour la signalisation.

L'ETALEMENT DE SPECTRE**HISTORIQUE**

Au début des années 20, Goldsmith utilise la FM pour étaler le spectre d'une modulation AM afin de réduire l'effet du bruit et des trajets multiples sur une communication. Dès les années 1940, Wiener et Shannon ont développé des théories de traitement du signal dans le but d'utiliser la technique d'étalement de spectre à des fins de cryptage dévolu aux transmissions militaires. On ne parle pas encore de multiplexage par codage. Dans plusieurs pays, l'armée met au point différents types d'étalement de spectre pour la guerre électronique. Les techniques restent très confidentielles. Les systèmes de l'époque, à tube, sont trop encombrants, chères et complexes pour être utilisés en télécommunication civile. L'utilisation de cette technique dans la radio-localisation (GPS) lui permet d'entrer enfin dans le domaine public. Fin des années 90, l'étalement de spectre apparaît comme une solution de multiplexage d'avenir qui permet d'envisager les radiocommunications de masse.

PRINCIPE DE L'ETALEMENT DE SPECTRE PAR CODAGE :

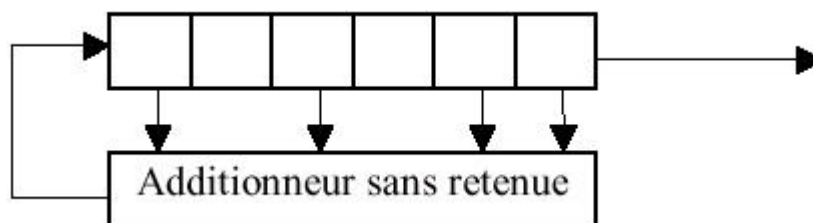
Il existe différentes techniques d'étalement analogique, mais à l'heure actuelle, on ne retient que celles qui sont basées sur une séquence NRZ binaire pseudo-aléatoire. Cette séquence aléatoire est ensuite multipliée par le signal à transmettre. C'est une séquence de '0' et de '1' qui, sur une courte échelle de temps paraît aléatoire. Mais en réalité cette séquence se répète sans arrêt.



Exemple de séquence binaire se répétant 3 fois

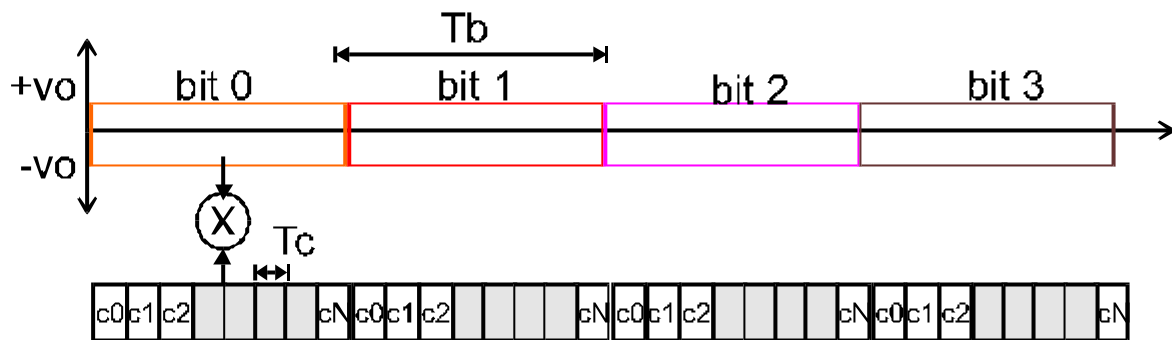
La génération de la séquence est souvent réalisée au moyen d'un registre à décalage. Le code est caractérisé par deux éléments :

- la valeur initial du registre à décalage
- l'équation de l'additionneur (pour l' exemple ci-dessus : $X^6 + X^5 + X^3 + X + 1 = 0$)



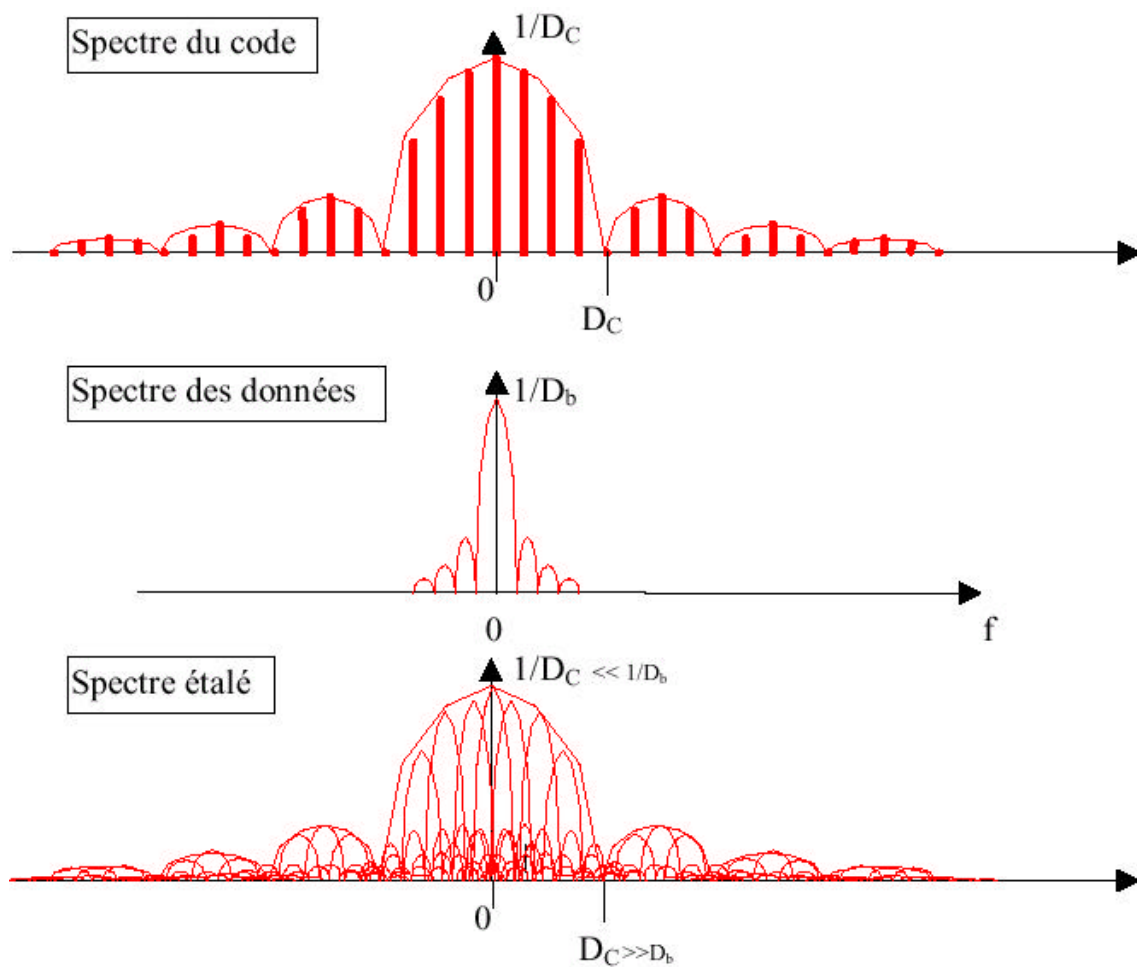
Structure d'un générateur de séquence pseudo aléatoire.

Chaque bit de données codé $+1$ (ou $+v_0$) pour l'état logique Haut ou -1 (ou $-v_0$) pour l'état logique bas est ensuite multiplié par le code codé de la même manière.



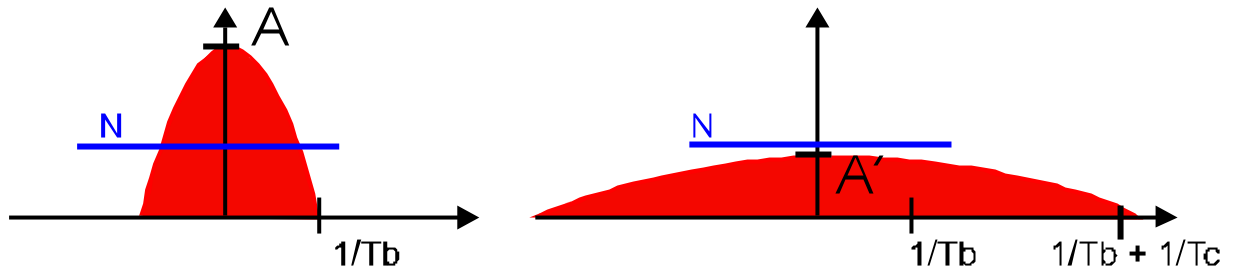
SPECTRE

La multiplication dans le domaine temporel du code par les données, se traduit dans le domaine fréquentiel par un produit de convolution.



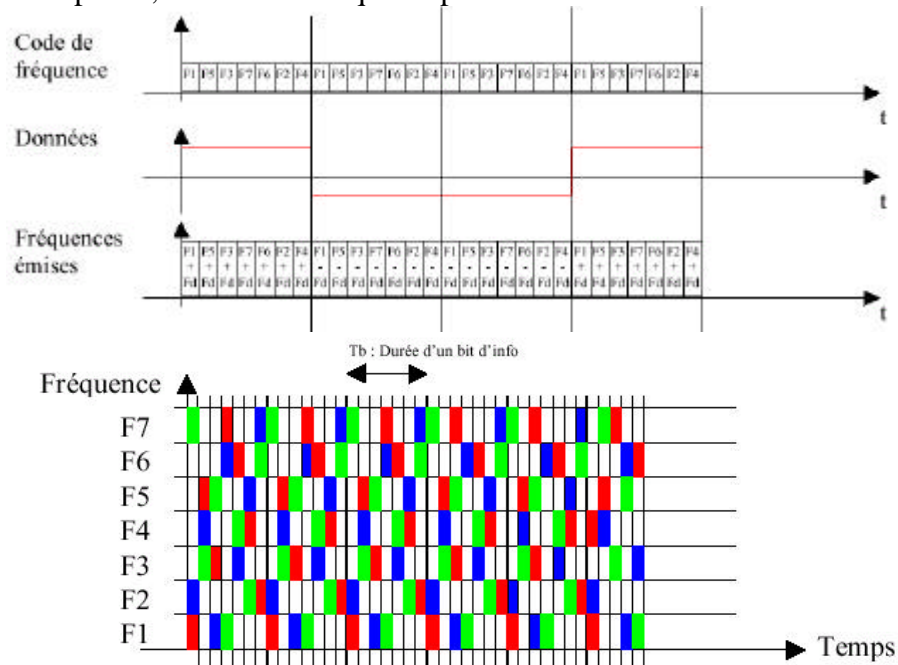
NIVEAU DE BRUIT

Si les données ont une amplitude spectrale A suffisante pour sortir du bruit N , il en est en revanche différemment du spectre étalé. En effet l'amplitude de celui ci est $A' < A$. Il y a conservation de l'énergie contenue dans le spectre. Si donc la bande spectrale s'élargit l'amplitude diminue en proportion. Le niveau du spectre étalé résultant peut donc être inférieur au niveau du bruit. C'est sur ce principe que fonctionnait certaine communication cryptée des années 40. Le signal noyé dans le bruit ne peut être récupéré qu'en connaissant le code.



PRINCIPE DE L'ETALEMENT DE SPECTRE PAR SAUT DE FREQUENCE :

Le principe est beaucoup plus simple à comprendre. On réalise en générale une FSK dont la fréquence porteuse change selon une loi pseudo-aléatoire. Chaque bit est découpé dans le temps en L morceaux, chaque morceau est modulé à un instant différent et une fréquence différente. Cela ressemble à du FD-TDMA, mais au lieu de transmettre un paquet de données par IT, on ne transmet qu'une partie d'un seul bit.



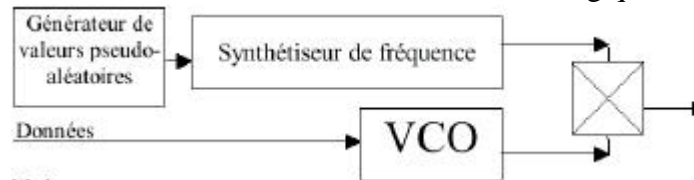


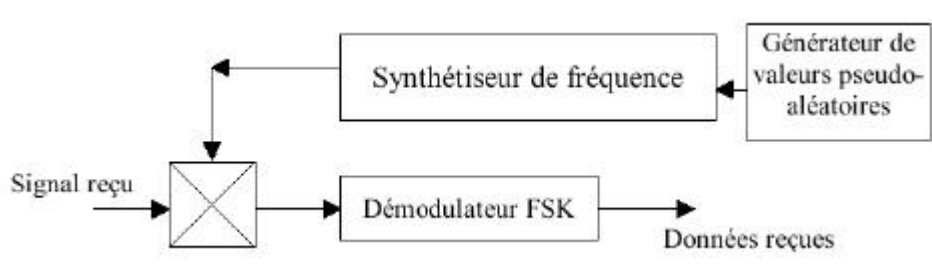
Schéma de principe du modulateur d'étalement de spectre par saut de fréquence

La modulation par saut de fréquence est appelée FHSS : Frequency Hopping Spread Spectrum.

- Fast FHSS si la fréquence des sauts est rapide par rapport au débit binaire
- Slow FHSS si plusieurs bits sont transmis sur chacune des fréquences.

DEMODULATION :

La démodulation consiste à multiplier le signal reçu par la même séquence de saut de fréquence suivant la même loi pseudo-aléatoire qu'à la modulation. La difficulté de démodulation réside encore dans la synchronisation du code de réception par rapport à celui qui a servi à émettre.



LE MULTIPLEXAGE PAR CODE

PRINCIPE DU MULTIPLEXAGE PAR CODAGE

Le multiplexage par code (CDMA : Code Division Multiplexing Access) utilise la technique d'étalement de spectre à la différence près qu'on transmet plusieurs signaux simultanément. Pour cela on associe à chaque signal un code particulier par lequel il est multiplié.

MATRICE DE CODAGE – CODES ORTHOGONAUX

Pour pouvoir démoduler il est nécessaire d'utiliser des codes ayant les particularités suivantes :

- Un code de N bits multiplié par lui même bit à bit et sommé (produit de convolution numérique) donne N.
- Un code de N bits multiplié par tout autre code que lui même donne zéro. On parle de codes orthogonaux entre eux.

Pour définir des codes orthogonaux on définit un ensemble de matrices construites de la façon suivante :

$$H_1=1, \quad H_2=\begin{bmatrix} H_1 & H_1 \\ H_1 & -H_1 \end{bmatrix}=\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_3=\begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix}=\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}, \quad H_4=\begin{bmatrix} H_3 & H_3 \\ H_3 & -H_3 \end{bmatrix}=\dots$$

Si on considère la matrice H3, elle fournit 4 codes de N=4 bits :

$$\begin{array}{lcl} C_1 & \rightarrow & \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \\ C_2 & \rightarrow & \begin{bmatrix} 1 & -1 & 1 & -1 \end{bmatrix} \\ C_3 & \rightarrow & \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix} \\ C_4 & \rightarrow & \begin{bmatrix} 1 & -1 & -1 & 1 \end{bmatrix} \end{array}$$

Le code C1 est donc constitué de la séquence binaire "1111" (c'est à dire 4 bits à l'états haut consécutifs $C_1^1=1, C_1^2=1, C_1^3=1, C_1^4=1$), le code C2 lui est constitué de la séquence "1 -1 1 -1" (c'est à dire un bit à l'état haut puis un bit à l'état bas puis un bit à l'état haut et pour finir un bit à l'état bas $C_2^1=1, C_2^2=-1, C_2^3=1, C_2^4=-1$). Pour les codes suivants on a : $C_3^1=1, C_3^2=1, C_3^3=-1, C_3^4=-1$ et $C_4^1=1, C_4^2=-1, C_4^3=-1, C_4^4=1$

On peut vérifier l'orthogonalité de ces codes :

$$C_1 \times C_1 = 1 \times 1 + 1 \times 1 + 1 \times 1 + 1 \times 1 = 4 = N$$

$$C_2 \times C_2 = 1 \times 1 + -1 \times -1 + 1 \times 1 + -1 \times -1 = 4 = N$$

$$C_3 \times C_3 = 1 \times 1 + 1 \times 1 + -1 \times -1 + -1 \times -1 = 4 = N$$

$$C_4 \times C_4 = 1 \times 1 + -1 \times -1 + -1 \times -1 + 1 \times 1 = 4 = N$$

En revanche :

$$C_1 \times C_2 = 1 \times 1 + 1 \times -1 + 1 \times 1 + 1 \times -1 = 0$$

$$C_1 \times C_3 = 1 \times 1 + 1 \times 1 + 1 \times -1 + 1 \times -1 = 0$$

$$C_1 \times C_4 = 1 \times 1 + 1 \times -1 + 1 \times -1 + 1 \times 1 = 0$$

$$C_2 \times C_3 = 1 \times 1 + -1 \times 1 + 1 \times -1 + -1 \times -1 = 0$$

$$C_2 \times C_4 = 1 \times 1 + -1 \times -1 + 1 \times -1 + -1 \times 1 = 0$$

$$C_3 \times C_4 = 1 \times 1 + 1 \times -1 + -1 \times -1 + -1 \times 1 = 0$$

Les codes C sont donc bien orthogonaux entre eux.

MODULATION

Supposons 4 signaux S1, S2, S3, S4, multiplexés par codage. Pendant l'intervalle de temps $[t, t+4T_c]$ le bit B1 est présent sur la voie S1, B2 sur S2, B3 sur S3 et le bit B4 est présent sur la voie S4. Le multiplexeur CDMA émet en même temps les 4 séquences :

$$B_1 C_1^1 = B_1, B_1 C_1^2 = B_1, B_1 C_1^3 = B_1, C_1^4 = B_1$$

$$B_2 C_2^1 = B_2, B_2 C_2^2 = -B_2, C_2^3 = B_2, B_2 C_2^4 = -B_2$$

$$B_3 C_3^1 = B_3, B_3 C_3^2 = B_3, C_3^3 = -B_3, C_3^4 = -B_3$$

$$B_4 C_4^1 = B_4, B_4 C_4^2 = -B_4, B_4 C_4^3 = -B_4, B_4 C_4^4 = B_4$$

C'est à dire qu'entre l'instant $[t, t+T_c]$ l'émetteur envoie :

$$B_1 C_1^1 + B_2 C_2^1 + B_3 C_3^1 + B_4 C_4^1 = B_1 + B_2 + B_3 + B_4$$

pendant l'intervalle de temps $[t+T_c, t+2T_c]$

$$B_1C_1^2 + B_2C_2^2 + B_3C_3^2 + B_4C_4^2 = B_1 - B_2 + B_3 - B_4$$

pendant l'intervalle de temps $[t+2T_c, t+3T_c]$

$$B_1C_1^3 + B_2C_2^3 + B_3C_3^3 + B_4C_4^3 = B_1 + B_2 - B_3 - B_4$$

pendant l'intervalle de temps $[t+3T_c, t+4T_c]$

$$B_1C_1^4 + B_2C_2^4 + B_3C_3^4 + B_4C_4^4 = B_1 - B_2 - B_3 + B_4$$

DEMODULATION

Quatre récepteurs indépendants R_1, R_2, R_3, R_4 disposant respectivement et exclusivement du code C_1, C_2, C_3, C_4 reçoivent le signal et font le produit de convolution par leur code respectif. Ainsi le récepteur R_1 calcule le produit de convolution suivant:

$$S = C_1^1(B_1 + B_2 + B_3 + B_4) + C_1^2(B_1 - B_2 + B_3 - B_4) + C_1^3(B_1 + B_2 - B_3 - B_4) + C_1^4(B_1 - B_2 - B_3 + B_4) =$$

Si on refait le calcul sans préjugé des codes utilisés pourvu qu'ils soient orthogonaux on obtient:

$$S = C_1^1(B_1C_1^1 + B_2C_2^1 + B_3C_3^1 + B_4C_4^1) + C_1^2(B_1C_1^2 + B_2C_2^2 + B_3C_3^2 + B_4C_4^2) + C_1^3(B_1C_1^3 + B_2C_2^3 + B_3C_3^3 + B_4C_4^3) + C_1^4(B_1C_1^4 + B_2C_2^4 + B_3C_3^4 + B_4C_4^4)$$

qui peut se factoriser différemment :

$$S = B_1(C_1^1C_1^1 + C_1^2C_1^2 + C_1^3C_1^3 + C_1^4C_1^4) + B_2(C_1^1C_2^1 + C_1^2C_2^2 + C_1^3C_2^3 + C_1^4C_2^4) + B_3(C_1^1C_3^1 + C_1^2C_3^2 + C_1^3C_3^3 + C_1^4C_3^4) + B_4(C_1^1C_4^1 + C_1^2C_4^2 + C_1^3C_4^3 + C_1^4C_4^4)$$

Le premier terme est le produit de convolution du code C_1 par lui-même, alors que les suivants sont le produit de convolution de C_1 par C_2 , C_1 par C_3 et respectivement C_1 par C_4 . La définition de l'orthogonalité des codes implique que le premier produit est égal à 4 alors que les suivants sont nuls. On trouve donc $S = 4B_1$.

Similairement on trouve que R_2 détecte B_2 , R_3 détecte B_3 et R_4 détecte B_4 . Le bit suivant de chaque voie peut alors être multiplexé par code et l'opération recommence.

7. QUANTIFICATION, COMPRESSION, SYNCHRONISATION ET COMMUTATIONS DES SYSTEMES DE TRANSMISSION

Introduction :

Le but de ce chapitre est de fournir une vision d'ensemble des techniques de transmissions et de commutations, en retraçant un bref historique de leur évolution et en expliquant leur principe.

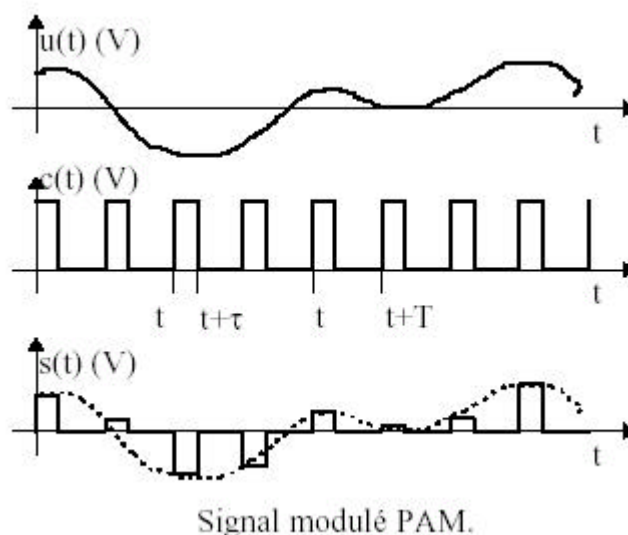
MULTIPLEXAGE D'IMPULSION CODEE : MIC

1-1 Modulation par modulation d'amplitude (PAM)

La modulation par impulsion d'amplitude (Pulse amplitude modulation PAM) consiste à mesurer le signal à intervalle régulier, avec une certaine fréquence d'échantillonnage F_e . Les échantillons sont donc des impulsions dont l'amplitude est la même que celle du signal initial, et appliqué directement sur le canal de transmission.

Figure 1. Principe de la modulation par impulsion d'amplitude.

Nous vous appellerons signal PAM, le signal constitué par ces échantillons. Nous allons voir qu'il est très facile de reconstruire le signal initial à partir du signal PAM, pour peu que certaines conditions soient remplies.



Le signal modulé PAM est le produit dans le domaine temporel du signal initial $u(t)$ et d'un signal qui est un peigne d'impulsions $c(t)$. Le résultat dans le domaine fréquentiel est alors

le produit de convolution de la transformée de Fourier (TF) de $U(t)$ et de la TF $c(t)$ du peigne d'impulsion.

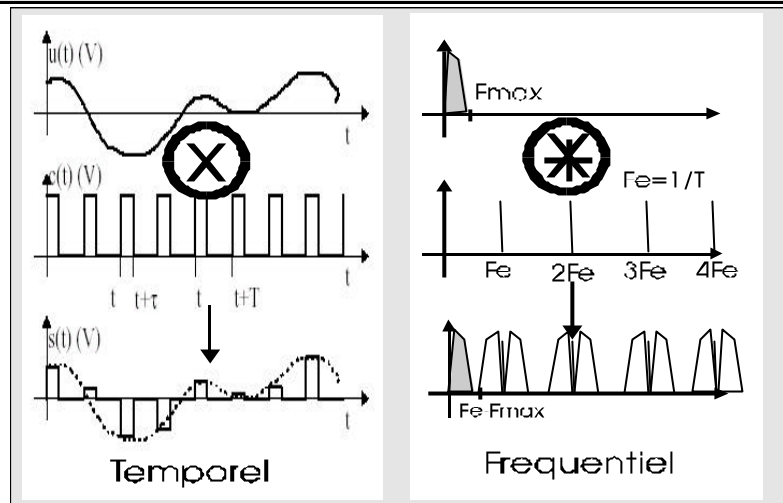


Figure 2 : Modulation PAM dans le domaine des Fréquence

Le résultat est une série infinie d'images constituées du spectre du signal initial centré autour des harmoniques de la fréquence d'échantillonnage. Pour éviter que les divers spectres ne se recouvrent, il faut que la fréquence d'échantillonnage soit supérieure à deux fois la fréquence maximale du spectre à transmettre. En effet le spectre de la première images est compris entre $F_e - F_{max}$ et $F_e + F_{max}$. Pour qu'il n'y ait pas recouvrement avec le spectre précédent ou le suivant il faut donc que $F_e > 2 F_{max}$. On retrouve ainsi le théorème d'échantillonnage de Shannon.

Le théorème de l'échantillonnage de Shannon permet d'affirmer qu'un signal échantillonné à une fréquence double de sa fréquence maximum peut être complètement reproduit.

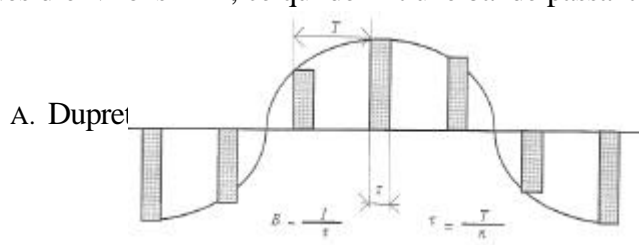
En téléphonie on considère que le spectre de la voix humaine est compris dans une bande passante comprise entre 300 et 3400 Hertz. On choisit de filtrer le signal ainsi obtenu par des filtres d'ordre 6 dont la pente est très raide afin de n'en laisser aucune trace à 4000HZ. Dans ces conditions, on peut échantillonner le signal vocal au double de cette fréquence c'est-à-dire à 8000 hertz. Dans la modulation PAM on a donc 8000 échantillons par seconde soit un échantillon tous les 125 μ s.

1-2 Multiplexage temporel :

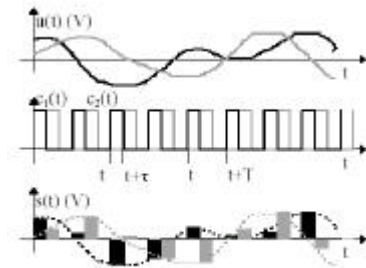
Le signal ainsi obtenu n'occupe pas le canal pendant la durée comprise entre deux impulsions successives. Il est donc possible d'y intercaler d'autres impulsions PAM provenant d'autres signaux. En faisant cela on réalise ce qu'on appelle du multiplexage temporel.

Figure 3. Durée de l'impulsion PAM

Le nombre de signaux PAM que l'on peut entrelacer dépend de la largeur τ des impulsions PAM. En général, un modulateur PAM transmet ses impulsions à une paire de fils torsadés d'environ 2km, ce qui définit une bande passante d'environ $B=260$ KHz. On a donc



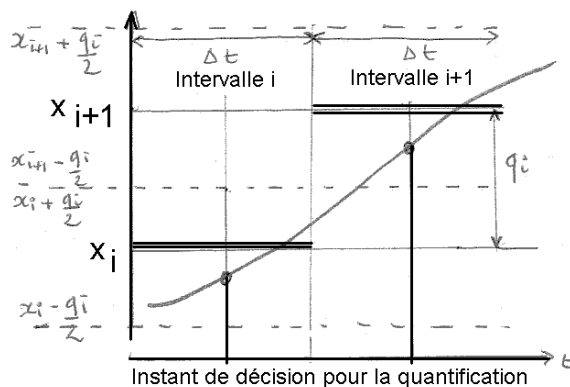
$1/\tau \geq B$, ce qui donne une durée d'impulsion PAM légèrement inférieure à $\tau = 4\mu\text{s}$. Le nombre de signaux modulés PAM que l'on peut donc intercaler dans une fenêtre de $125\mu\text{s}$ est donc d'environ 31. Au total 32 voies sont ainsi multiplexées dans une fenêtre temporelle de $125\mu\text{s}$.



Multiplexage temporel de signaux modulés PAM.

Figure 4 : principe du multiplexage temporel

Un système de transmission PAM est temporel. Bien que le signal soit échantillonné il reste analogique puisque l'information du signal PAM réside dans l'amplitude des impulsions. Ce système de transmission subit donc les mêmes défauts que les systèmes analogiques c'est à dire une atténuation avec la distance, et une sensibilité aux bruits.

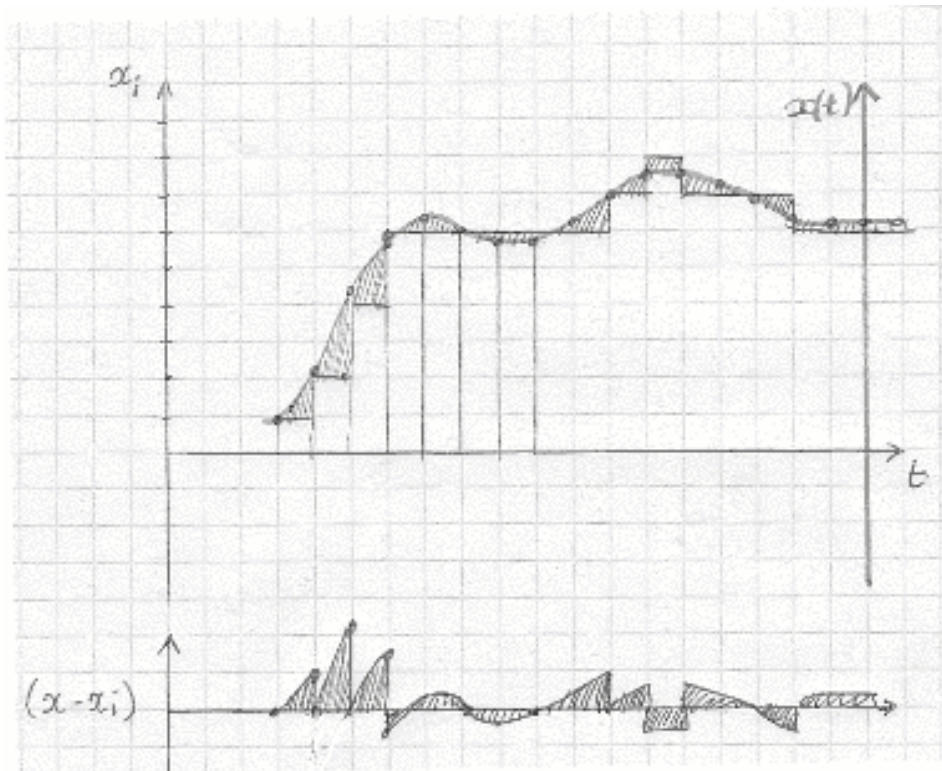
**Figure 5 : instant de décision de la quantification**

Quantification :

Pour éviter ces bruits, on peut numériser le signal, on parle de quantification. Cette opération est réalisée par un convertisseur analogique-numérique, et tous les échantillons contenus dans un même intervalle de quantification (appelé : pas de quantification q_i) sont représentés par la même valeur centrale x_i . Ainsi tous les échantillons $s_e(t)$ compris entre $x_i - q_i/2$ et $x_i + q_i/2$ prennent la valeur x_i . Les x_i valeurs sont des valeurs discrètes que l'on peut représenter par un nombre entier. En téléphonie on utilise $N=2^n$ valeurs discrètes ($N=256$ pour $n=8\text{bits}$). Si le signal varie entre $+V_{\text{max}}$ et $-V_{\text{max}}$, et que les pas d'échantillonnage sont tous identiques :

$$q_i = q = \frac{2V_{\text{max}}}{N}$$

Le débit des échantillons étant $D_e = 1/T_e = 8000$ éch/s, ($T_e = 125\mu\text{s}$) et chaque échantillon étant codé sur $n=8$ bits, le débit total est alors :

Bruits de quantification :**Figure 6: Bruit de quantification**

Puisque
tous les
échantillons
 $S_e(t)$
vérifient :

$$x_i - \frac{q_i}{2} \leq S_e(t) \leq x_i + \frac{q_i}{2}$$

,
l'erreur
faite lors
de la
numérisati
on
introduit
un bruit
($x - x_i$). La
valeur
moyenne

de ce bruit (calculée sur un grand nombre d'échantillons) est nulle puisque statistiquement il y a autant de chance de numériser en faisant une approximation par valeur supérieure que par valeur inférieure. ($\langle (x - x_i) \rangle = 0$).

Par contre la puissance moyenne du bruit calculée sur un échantillon n'est pas nulle en général.

$$b_{qi} = \langle (x - x_i)^2 \rangle = \frac{1}{q_i} \int_{x_i - \frac{q_i}{2}}^{x_i + \frac{q_i}{2}} (x - x_i)^2 dx$$

Si les q_i sont tous égaux ($q_i = q$) alors le calcul du bruit peut se simplifier et :

$$b_{qi} = \frac{1}{q} \left[\frac{1}{3} (x - x_i)^3 \right]_{x_i - \frac{q}{2}}^{x_i + \frac{q}{2}} = \frac{1}{12} q^2$$

Remarque 1 : Le bruit est d'autant plus grand que les pas sont grands. (pour diminuer le bruit il faut évidemment diminuer le pas en échantillonnant avec $N = 2^n$ niveaux le plus grand possible).

Remarque 2 : Le rapport signal sur bruit S/N qui évalue la qualité du signal transmis, diminue si l'amplitude du signal à transmettre S_e diminue :

$$\left(\frac{S}{N} \right)_q = \frac{S_{e\text{Efficace}}^2}{12 q^2} \quad \left(\frac{S}{N} \right)_{q(dB)} = 10 \cdot \text{Log} \left(12 \frac{S_{e\text{Efficace}}^2}{q^2} \right)$$

En ce sens la loi de quantification linéaire n'est pas optimale. Il faut donc quantifier différemment pour que les signaux de fortes amplitudes aient le même rapport signal sur bruit de quantification que les signaux de faibles amplitudes.

Loi de compression. Loi A et Loi μ

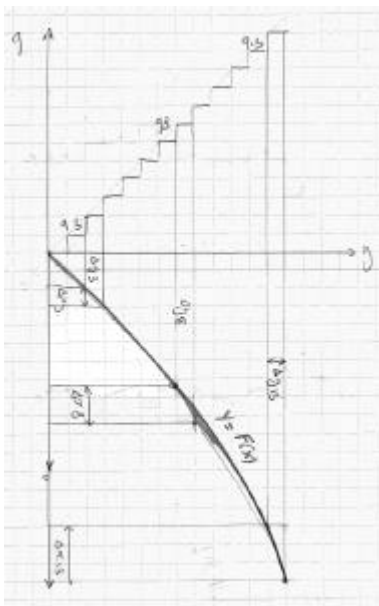
En téléphonie, on estime avoir une qualité suffisante lorsque le rapport signal à bruit est constant quelque soit l'amplitude des signaux.

Pour cela on peut essayer de trouver un moyen de coder les signaux de grande amplitude avec des pas de quantifications « grands », et les signaux de faibles amplitudes avec des pas de quantification plus faibles. On garde ainsi le même nombre $N=2^n$, en introduisant une loi non-linéaire qui répartit les échantillons sur les différents niveaux en minimisant le bruit.

On peut présenter le problème comme une quantification linéaire sur une échelle $Y(t)$, qui elle, est une fonction non linéaire des échantillons d'entrée $X(t)$. Le problème se résume donc à trouver une loi $Y=F(x)$.

Un échantillon $S_e(t)=x$ est converti par la loi de compression F en $Y=F(x)$, puis quantifié.

$$X \rightarrow Y = F(X) \rightarrow q_i \text{ (quantification)}$$



$$\Delta X_i = \Delta Y_i \left(\frac{dx}{dy} \right)_i$$

Les pas Δx deviennent :

Tous les pas q_i sont de même largeur $2V_{\max}/N$ et égaux aux ΔY_i :

$$q = q_i = \Delta Y_i = \Delta X_i \left(\frac{dy}{dx} \right)_i = 2 \frac{V_{\max}}{N}$$

En revanche les pas $q'_i = \Delta X_i$ sont de largeurs différentes et s'écrivent pour l'intervalle x_i de largeur Δx_i :

$$q'_i = \Delta X_i = \frac{2V_{\max}}{N} \left(\frac{dx}{dy} \right)_i$$

A partir de cette relation nous allons remonter à la forme de la fonction $y=F(x)$ qui permet le meilleur rapport signal à bruit :

Pour cela il faut exprimer le rapport S/N :

Si P_s est la puissance du signal, dont les échantillons se répartissent avec la loi $p(x)$ sur les différents niveaux, et si B_q est la puissance de bruit de quantification, le rapport signal à bruit s'écrit :

$$\left(\frac{S}{N} \right)_q = \frac{P_s}{B_q} \quad \text{et} \quad P_s = \int_{-1V_{\max}}^{1V_{\max}} x^2 p(x) dx \quad \text{et} \quad B_q = \int_{-1V_{\max}}^{1V_{\max}} p(x) \frac{q_i^2}{12} dx = \frac{1}{3N} \int_{-1}^1 p(x) \left(\frac{dx}{dy} \right)^2 dx$$

$$\left(\frac{S}{N}\right)_q = \frac{\int_{-V_{\max}}^{V_{\max}} p(x)x^2 dx}{\frac{1}{3N} \int_{-V_{\max}}^{V_{\max}} p(x) \left(\frac{dx}{dy}\right)^2 dx}$$
 d'où : une façon de simplifier ce calcul est de trouver une fonction pour remplacer $\left(\frac{dx}{dy}\right)$ qui soit directement proportionnelle à x:

donc : $\left(\frac{dx}{dy}\right) = kx$ où k est une constante que nous déterminerons ultérieurement. Il suffit pour trouver la fonction $Y=F(X)$ de résoudre l'équation : $\frac{dx}{x} = k dy \rightarrow A + \ln(x) = k y \rightarrow y = \frac{A + \ln(x)}{k}$

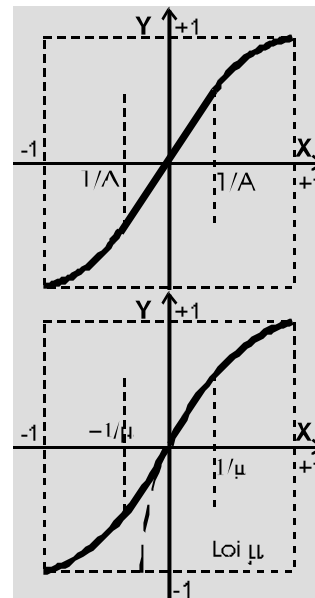
Il ne reste qu'à déterminer les constantes A et k. Auparavant il faut noter que cette fonction idéale n'est pas définie pour $x=0$.

En Europe, la loi A fait une interpolation linéaire pour logarithmique pour x grand :

- $y = \frac{Ax}{1 + \ln A}$ pour $0 < x < 1/A$
- $y = \frac{1 + \ln Ax}{1 + \ln A}$ pour $x > 1/A$ avec $A=87,5$.

En Amérique la loi μ , loi logarithmique décalée :

- $y = \frac{\ln(1 + \mu x)}{\ln(1 + \mu)}$ pour $x > 0$ avec $\mu=255$. Pour courbes en Y sont complétées par des courbes symétriques.



x petit, et

x < 0, les

d) Approximation par segments, codage SABCWXYZ.

Comme on cherche à transmettre un signal numérique il est inutile d'utiliser une fonction continue pour la loi de compression, une approximation par des segments de droite est suffisante. Dans le cas de la loi A, la courbe de compression (logarithmique) est approchée par 8 segments de droite pour les valeurs positives, et 8 pour les valeurs négatives. Ces segments peuvent être numérotés, et 3 bits suffisent pour ce faire. Un bit supplémentaire permet de coder le signe. Sur un mot de 8 bits, il reste 4 bits pour coder la position de l'échantillon sur le segment considéré.

Ainsi dans un codage SABCWXYZ le premier bit, S, code le signe de l'échantillon, les bits A,B,C codent le numéro d'un des 8 segments utilisés. Les quatre bits restants, (W, X, Y, Z), sont utilisés pour préciser l'emplacement de l'échantillon sur le segment.

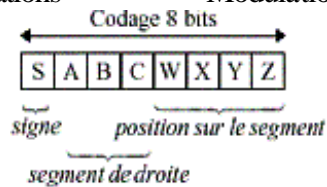


Figure 7: Codage des échantillons

Notons que le rapport des pentes de deux segments consécutif est de 2. De ce fait, le pas de quantification q augmente d'un facteur 2 lorsqu'on passe d'un segment à l'autre.

Dans le cas de la loi μ , la courbe logarithmique est approchée par 15 segments de droite pour les valeurs positives et n'est pas compatible avec le système de la loi A. Pour le reste, la loi μ est analogue à loi A.

Trame primaire TN1– Intervalle temporel (IT)

Une fois que l'échantillon a été quantifié, il peut être envoyé. Et un nouvel échantillon peut alors être échantillonné, compressé et quantifié. L'ensemble de ces opérations doit cependant être limité dans le temps. En effet puisqu'on échantillonne à 8000 Hz, la durée de "fabrication" des mots de 8 bits codant l'échantillon ne peut dépasser 125 μ s. En réalité les mots de 8 bits sont produits très rapidement et on peut même insérer entre deux mots successifs provenant de la même voie, des mots provenant d'autres voies. On réalise ainsi une transmission multiplexée temporellement de plusieurs voies. (Recommandation UIT-T Rec. G711)

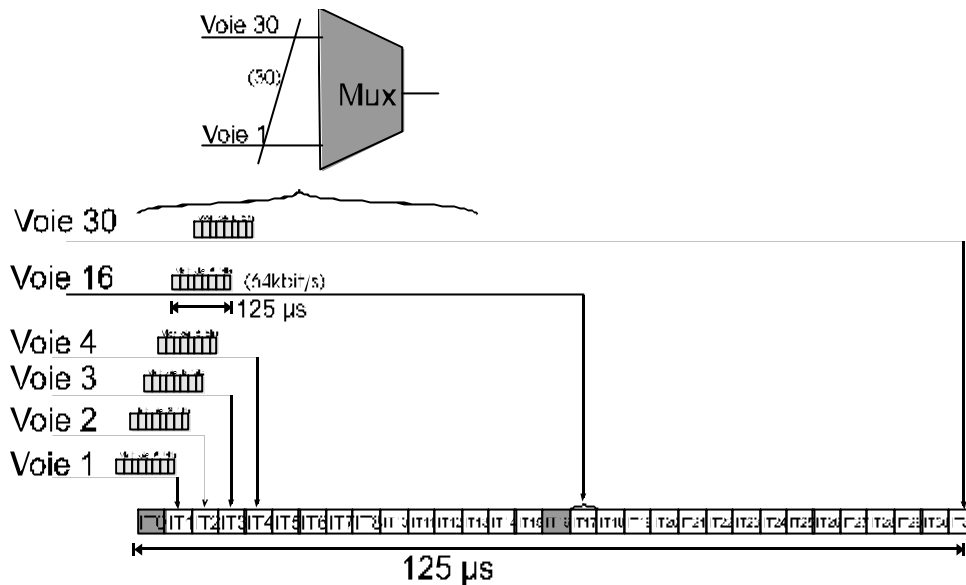
En Europe le système qui réalise cette fonction est appelé MIC E1 (Modulateur d'impulsions codées). C'est le premier niveau de la hiérarchie de multiplexage. Dans ce système 30 voies sont multiplexées temporellement dans une "fenêtre" de 125 μ s.

Fig 8 : Structure d'une trame MIC

La fenêtre de 125 μ s appelée trame est donc découpée en 32 morceaux appelés "Intervalle de temps" ou "IT", qui vont de IT0 à IT31. Les 32 intervalles de temps sont occupés de la manière suivante ; Les échantillons compressés et quantifiés en provenance des voies 1 à 15, sont placés respectivement dans les IT1 à IT15. De même pour les voies 16 à 30 placées dans les IT17 à IT31. Seuls les IT0 et IT 16 sont particuliers car ils ne sont pas utilisés pour transmettre des voies mais pour la signalisation.

Lorsque une trame est émise par une machine A (l'émetteur) vers une machine B (le récepteur) deux principes importants doivent être respectés: 1- La synchronisation 2- Le verrouillage.

Synchronisation : Les machines A et B doivent être synchronisées, c'est à dire que



leurs horloges HA et HB doivent être identiques afin que les bits soient lus à des instants "corrects". En effet si le récepteur cherche à détecter l'état d'un bit alors qu'un front montant ou descendant est présent une erreur peu se produire. Nous reviendrons plus loin sur la synchronisation.

Verrouillage : Le récepteur B doit savoir où commence la trame au bit près. Sans cela il n'y a aucun moyen de distinguer où commence une IT et où s'arrête une autre. A cela 2 conséquences; 1- Il faut indiquer le début de la trame, et 2- si B n'est pas synchronisé sur A, il est important que A et B puissent se transmettre un signal de perte de verrouillage les avertissant de l'état de leur synchronisation. Les IT0 et IT16 jouent cet office, et permettent à l'émetteur A et au récepteur B de "s'entendre". Nous allons voir plus précisément comment.

Mot de verrouillage de Trame (MVT)

Pour pouvoir être reconnue, et pour que son origine soit repérée, la trame a donc la particularité suivante :

Dans les trames paires, l'IT0 est remplie avec VER un mot de verrouillage de trame (MVT) (VER1=x0011011, 1B ou 9B en hexadécimal). Dès que le récepteur B détecte ce mot il peut se verrouiller.

Pour les trames impaires, le mot de verrouillage de trame de l'IT0 est remplie par VER2=xB₂Axxxxx. Dès que le verrouillage est constaté B renvoie à A dans une trame impaire le mot VER2 avec B₂=1 et A=0 . En revanche si A=1, c'est une alarme de perte de verrouillage de trame. Par sécurité le verrouillage n'est considéré comme perdu qu'après 3 alarmes successives.

L'IT16 contient des informations de signalisation supplémentaires. En effet il est nécessaire de transmettre les informations sur l'activité des 30 voies. Si une voie n'est pas

utilisée parce que le téléphone est "raccroché", par exemple à la fin d'une conversation, il est utile de signaler cet état.

L'IT16 ne contient que 8 bits, ce qui est insuffisant pour signaler l'état des 30 voies. Il est donc nécessaire d'utiliser plusieurs IT16 successifs dans plusieurs trames successives. Il a été choisi arbitrairement 4 bits pour transmettre l'état de chacune des voies. Par conséquent l'IT16 (8 bits) ne peut transmettre que les états de 2 voies à la fois. Il faudra donc attendre que quinze trames soient transmises pour que l'ensemble des états des 30 voies soient complètement signalés. En pratique 16 trames sont utilisées, et forment une "supertrame". La durée d'une supertrame est donc $125\mu s \times 16 = 2ms$.

Calculons le débit obtenu : dans une fenêtre de durée $T = 125\mu s$ seconde est contenue un échantillon de chacune des 30 voies de 8 bits chacune, plus deux intervalles de temps de 8 bits contenant les IT0 et IT16. soit un total de $m = 32 \times 8$ bits dans $T = 125\mu s$. Le débit est donc :

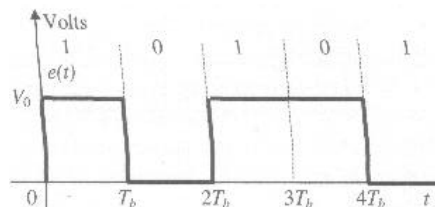
$$D_{MIC-E1} = \frac{m}{T} = 20480000 \text{ bits/s} = 2,048 \text{ Mbits}$$

LA SYNCHRONISATION**Echantillonnage bit :**

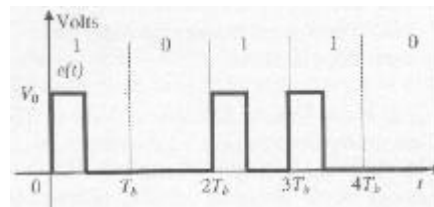
Un codage binaire est représenté par une simple suite de "0" et de "1". Cela paraît simple en apparence, mais il existe plusieurs façons de représenter une suite binaire.

Signaux NRZ: (non remise à zéro)

Si T est la durée d'un bit, le signal reste dans le même état pendant toute la durée T . Par suite, si plusieurs bits consécutifs sont dans le même état (par exemple n bits à 1) le signal transmis sera constant et égal à 1 pendant n période T d'horloge. L'avantage d'un tel système est la simplicité, l'inconvénient est le risque de perte de synchronisation, car rien n'indique s'il s'agit d'un seul bit de valeur 1 ou de plusieurs bits de valeur 1. La seule façon pour interpréter correctement ce signal est de disposer d'une horloge locale réglée sur la fréquence d'arrivée des bits.

**Signaux RZ: (Remise à Zéro)**

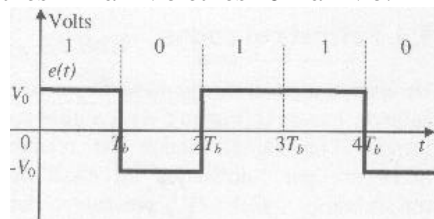
Dans ce type de représentation, le signal pour transmettre un "1" logique ne reste pas à V_0 durant la totalité de la durée T (durée d'un bit), mais seulement durant $T/2$ ensuite il est remis à "0". Cela facilite la synchronisation puisque même dans une longue séquence de 1 on retrouve le rythme de l'horloge.

**Signaux unipolaires :**

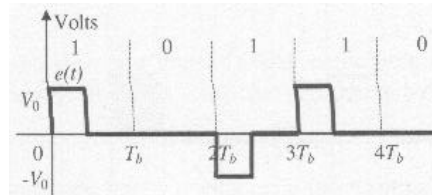
C'est la représentation la plus simple. Les niveaux hauts ("1") sont codés par une tension haute V_0 , alors que les niveaux bas sont représentés par une tension nulle. Dans le cas d'une longue suite de 1 la valeur moyenne du signal n'est plus nulle ce qui peut produire des erreurs.

Signaux bipolaires :

Pour maintenir une valeur moyenne nulle, il y a deux possibilités : Dans le cas de signaux bipolaires NRZ, on met les "1" à $+V_0$ et les "0" à $-V_0$.

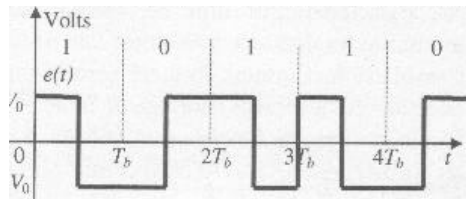


Pour les signaux bipolaires RZ, le premier bit à "1" est à la tension $+V_0$ (polarité positive), tandis que le deuxième bit à "1" suivant est représenté par la tension $-V_0$ (polarité négative).



Codage manchester :

Une variante consiste à transmettre les 1 par des fronts montants et les "0" par des fronts descendants.



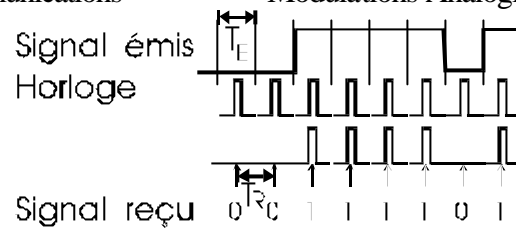
Signaux HDBn :

Les codes BHD(bipolaires haute densité ou high density bipolar codes) constituent une variante du code bipolaire de type RZ destinée à limiter le nombre de zéros successifs. Dans le cas de code bipolaire, une longue suite de zéros se traduit par une tension quasi nulle sans aucune transition, ce qui peut provoquer une perte de synchronisation du récepteur. Le code HDBn autorise au plus n intervalles de temps consécutifs sans impulsion. Le code HDB3 est très utilisé dans les réseaux téléphoniques numériques notamment dans la PDH.

Dans le code HDB3, l'émetteur remplace toutes les séquences de 4 zéros successifs par une autre séquence de la forme B00V, dans laquelle le quatrième zéro "V" est forcé à avoir la même polarité que le bit à "1" immédiatement précédent. Il y a donc violation de l'alternance des polarités. (bit de viol). B respecte l'alternance des polarités, c'est à dire qu'il aura le signe opposé à la dernière impulsion sauf dans le cas où il y a un nombre impair d'impulsion depuis la dernière violation de polarité. Dans ce cas B sera mis à zéro ce qui maintient une valeur moyenne nulle.

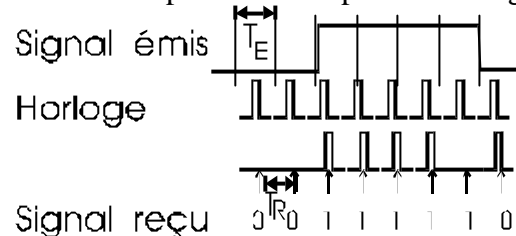
Décalage des horloges :

Considérons un signal simple de type NRZ. L'avantage d'un tel système est sa simplicité, l'inconvénient est le risque de perte de synchronisation, car comme nous l'avons vu rien n'indique s'il s'agit d'un seul bit de valeur 1 ou de plusieurs bits de valeur 1. La seule façon pour interpréter correctement ce signal est de disposer d'une horloge locale réglée sur la fréquence d'arrivée des bits. La séquence binaire reçue sera en fait le résultat de l'échantillonnage du signal physique reçu par une horloge locale, c'est à dire le produit de l'horloge par le signal reçu.

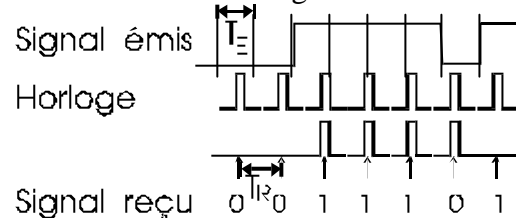
**Décalage des horloges :**

En réalité chaque élément de la chaîne de transmission a sa propre horloge. Ainsi aucun n'a rigoureusement la même fréquence que l'horloge de l'émetteur. Si ce décalage entre le rythme de l'horloge émettrice et l'horloge réceptrice est trop grand, il risque d'y avoir un bit doublé ou un bit oublié.

Prenons d'abord l'exemple d'une horloge de réception de fréquence f_r , et d'une horloge d'émission f_e avec $f_r = 8/7 f_e$. Lorsque l'émetteur cadence 7 intervalles de temps, le récepteur en compte 8. Cela se traduira dans l'exemple ci dessous par un doublage du bit 3:



En revanche si $f_e = 8/7 f_r$, l'émetteur enverra 8 bit lorsque le récepteur en décodera 7, ce qui provoquera l'oubli d'un bit comme le montre la figure ci dessous.



- De manière générale, si $f_r > f_e$ (on pose $f_r = 1/T_r$ et $f_e = 1/T_e$) on répète 1 bit sur N chaque fois que $f_r = N/(N+1) f_e$, soit $N/(N+1) = f_r/f_e$ d'où $N = f_r/(f_r - f_e) = f_r/\Delta f$. Il y a donc une erreur tous les $T = N \cdot T_r$ soit $T = 1/\Delta f$.

- Si $f_e > f_r$, on rate un bit parmi N émis chaque fois que $f_e = N/(N+1) f_r$, c'est à dire chaque fois que $N = f_e/(f_r - f_e) = f_e/\Delta f$. il y a donc une erreur tous les $T = N \cdot T_e = 1/\Delta f$, et la réception est sans erreur que sur une durée $T = 1/\Delta f$.

- Si on appelle p la précision sur les horloges $p = \Delta f/f$ on a $\Delta f = pf$, le nombre de bits qu'il est possible d'envoyer sans erreur est alors $N = T/t_e = 1/(T_e \cdot \Delta f) = f_e/\Delta f = f/\Delta f = 1/p$.

Transmission asynchrone :

Les transmissions asynchrones se réalisent pour des systèmes de transmission possédant chacun une horloge différente. Dans ce cas l'inévitable écart entre les fréquences d'horloges d'émission et de réception produit des erreurs.

En transmission asynchrone, on renonce donc à transmettre des paquets de bits supérieurs à $N=1/p$. La transmission n'est donc pas continue mais par paquets de $N < 1/p$ bits séparés par des "silences" pendant lesquels la ligne est laissée à l'état haut. Lorsqu'un paquet est envoyé la ligne redescend à zéro. Ce front descendant (signal START) fait office de signal de synchronisation pour démarrer l'horloge de réception et la mettre en phase avec l'horloge de l'émission.

A la fin du paquet la ligne est remise à l'état haut pendant une durée égale au moins à un bit (signal stop).

Transmission synchrone :

Pour s'affranchir des erreurs de bits répétées ou ratées, l'émetteur et le récepteur doivent avoir rigoureusement la même horloge. Une méthode consiste à envoyer au récepteur l'horloge de l'émetteur par un fil indépendant de la ligne de transmission, ou bien à fournir à tous les éléments de la ligne de transmission les mêmes impulsions d'horloges provenant d'une tierce machine. On peut aussi multiplexer temporellement le signal d'horloge et le signal contenant les informations.(codage Manchester ou HDB3).

PDH (PLÉSIOCHRONE DIGITAL HIERARCHIE)

Nous venons de décrire le multiplex numérique primaire E1 dont le rôle est de multiplexer 30 voies à 60kbit/s. Par ce procédé on divise par 30 le nombre de câbles à poser. Cependant cela reste largement insuffisant notamment dans le cas de réseaux urbains, et cela est encore moins suffisant pour des réseaux nationaux ou transatlantiques. Il est nécessaire donc de monter davantage en débit en multiplexant à des niveaux supérieurs.

Un premier niveau de multiplexage appelé MIC-E2 ou TN2 multiplexe 4 signaux à 2,048Mbit/s pour obtenir un signal multiplex secondaire 8 Mbit/s. 4 de ces multiplex secondaires peuvent à leur tour être multiplexés pour donner des multiplex tertiaires MIC-E3 à 34Mbit/s. Puis 4 multiplex peuvent être surmultiplexés en MIC-E4 à 140 Mbit/s. Le dernier multiplex MIC-E5 multiplexe 4 signaux MIC-E4 à 140 bit/s pour donner du 565Mbit/s, et le transmet par fibre optique sur des distances d'au plus 4 multiples de 23km (~92km). Ces liaisons interurbaines assuraient la liaison entre les principales villes de province jusqu'à la fin de l'année 2000. Par définition le passage à l'ordre de multiplexage supérieur $n+1$ consiste à multiplexer ensemble 4 multiplex d'ordre n .

L'ensemble des multiplex de différents niveaux (ordre) s'appelle une hiérarchie. Cette hiérarchie de multiplex successifs est dite PDH (Plésiochrone Digital Hiérarchy). Le multiplexage s'y fait bit à bit et non pas mot par mot. Les systèmes de multiplexages de la hiérarchie plésiochrone traitent chaque affluent de manière totalement transparent comme des flux de bits dont ils ignorent la structure.

Trame secondaire TN2 - MIC- E2:

La trame secondaire TN2 correspond à la norme G742 de l'ITU-T. La trame secondaire dure 100,4µs et est longue de 848 bits, qui se décompose de la manière suivante : 4 groupes de 212 bits chacun. Il faut remarquer tout d'abord que 848 bits ne correspond pas à $4 \times 8 \times 30 = 960$ bits. Nous allons expliquer pourquoi:

<i>Groupe I</i>		<i>Groupe II</i>		<i>Groupe III</i>		<i>Groupe IV</i>	
12	200	4	208	4	208	4	208

Figure 9 : Trame de multiplex E2

Le premier groupe est structuré de la façon suivante:

-Un mot de 12 bit indique le début de la trame multiplex E2 (appelé fanion de départ) et a pour valeur DEP=1111010000AR ou A et R sont des bits d'alarmes.

-200 bits qui contiennent les 50 premiers bits de chacune des 4 voies (affluent 0, 1, 2, 3). Les voies sont entrelacées, c'est à dire qu'elles sont multiplexées bit à bit et non pas mot à mot. Si un bit numéro $4k$ provient de l'affluent 0, le bit $4k+1$ provient de l'affluent 1, le bit $4k+2$ provient de l'affluent 2, et le bit $4k+3$ de l'affluent 3.

Dans chacun des 3 groupes suivant on peut distinguer 2 zones:

-un ensemble de 4 bits d'indication de justification dont nous allons expliquer la signification ci dessous.

-un ensemble de 208 bits ce qui correspond à 4 groupes de 52 bits provenant des 4 affluents, c'est à dire le multiplexage bit à bit de 52 des bits des 4 voies affluentes.

Groupe	Intervalle	Nombre de bits	Utilisation
I	1	12	Fanion de début de trame
	2	200	4 fois 50 bits entrelacés des affluents 0 à 3 Bits 1 à 50 de chaque affluent
II	1	4	Bits C_{j1}
	2	208	4 fois 52 bits entrelacés des affluents 0 à 3 Bits 51 à 102 de chaque affluent
III	1	4	Bits C_{j2}
	2	208	4 fois 52 bits entrelacés des affluents 0 à 3 Bits 103 à 154 de chaque affluent
IV	1	4	Bits C_{j3}
	2	208	4 fois 52 bits entrelacés des affluents 0 à 3 Bits 155 à 206 de chaque affluent ou Bits 155 à 205 si justification

Principe de la justification positive nulle négative :

Pour expliquer l'utilité des groupes de 4 bits C_{j1} C_{j2} et C_{j3} il nous faut considérer la synchronisation des 4 affluents. Chaque affluent possède une horloge qui n'est ni rigoureusement identique à celle des 3 autres, ni égale au quart du flux résultant T_{N2} .

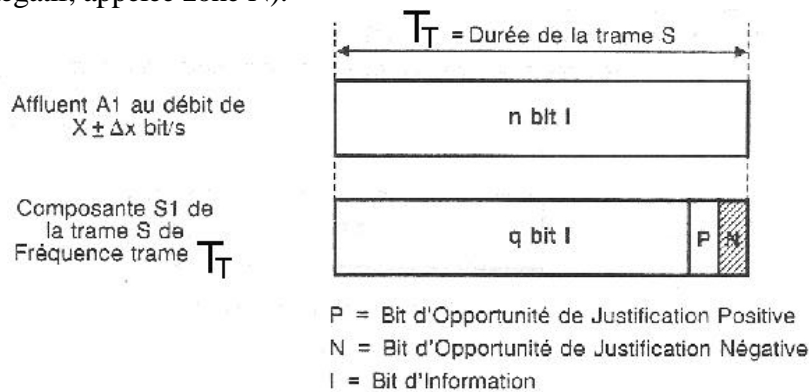
Considérons un signal affluent A1 au débit de X bits par seconde. Ce débit peut varier dans une plage $\pm \Delta x$ bit/s. On veut transporter ce signal dans une trame S dont la longueur est fixe L_T et la période est T_T .

Pour cela on peut prévoir dans la trame de transport S une place L_n allouée à chacun des affluents. Cette place doit être suffisamment longue pour transporter le nombre de bits maximum que l'on peut obtenir pendant la durée T_T dans le cas particulier du débit le plus rapide $X + \Delta X$ bit/s.

Dans le cas du débit maximum, si le débit est $X + \Delta X$ bits/s, le nombre de bits reçus pendant l'intervalle de temps T_T est $(X + \Delta X) \cdot T_T$. Une longueur $L_n > (X + \Delta X) \cdot T_T$ pourra donc contenir tous les bits reçus et à fortiori on aura suffisamment de place pour contenir le nombre de bits reçus dans le cas d'un débit affluent plus lent : $(X - \Delta X) \cdot T_T$.

L'opération permettant de transporter un signal de débit variable dans une trame S de longueur fixe L et de débit fixe L/T_T s'appelle la **justification**. Parmi les L bits de la trame S on distingue deux types de bits (les bits d'information à transmettre = bit I, et les bits de remplissage = bit R) et trois zones (zone pour les bits d'information, une zone pour les bits de

remplissage positif, appelée zone P sur la figure 10, et une zone pour les bits de remplissage négatif, appelée zone N).



Trois possibilités:

1- Justification nulle : l'affluent A1 a exactement le même débit L/T_T ($q+1$ bit en T_T), dans ce cas P est un bit d'information, N est rempli avec un bit quelconque dit bit de remplissage.

2-Justification positive : A1 est au débit $X+\Delta X$ ($q+2$ bit durant T_T). P et N sont remplis chacun avec un bit d'information de manière à compenser la variation $+\Delta X$.

3- Justification négative : A1 est au débit $X-\Delta X$, N et P sont des bits de remplissage.

Pour le TN2 la durée d'une trame secondaire est plus courte que celle d'une trame primaire ($100,4\mu s$ contre $125\mu s$). On a donc un période $T_2 = 100,4\mu s < T_1 = 125\mu s$, donc seule la justification négative existe. Pour préciser si le dernier bit est justifié ou non, on utilise les bits d'indication de justification.

Si les 3 bits $Cj1, Cj2, Cj3$, sont à 1,1,1 alors il y a présence d'un bit de justification pour la voie j. Le bit de justification est le dernier bit de chaque affluents du groupe IV. C'est à dire que seules les 4 derniers bits de la trame TN2 peuvent être des bits de justifications.

Calculons le débit d'une trame TN2 pour vérifier qu'elle peut bien transmettre sans perte 4 affluents à 2048Kbit/S:

•Avant le multiplexage du TN2, 4 affluents au débit de 2048 000bit/s sont présents pour une trame de période $125\mu s$. Le système TN2 doit donc mutliplexer 4×256 bits =1024 bits en $125\mu s$ au plus.

Soit $D_{TN2-1} = 1024/125\mu s = 8.192$ Mbit/s

•Parmi les 848 bits de la TN2 seuls $200+208+208+204+(4)$ bits sont des bits d'informations.

Sans justification $L=824$ bits

Avec justification $L^* = 820$ bits

La durée de la trame est $T_T = 100,4 \mu s$

Le débit est donc : $D_{TN2-2} = 8.207$ Mbit/S $> D_{TN2-1}$

avec justification : $D_{TN2-2justifié} = 8.167 \text{ Mbit/S} < D_{TN2-1}$

Notons que le TN2 transmet $L1=848$ bits alors que seulement $L2=824$ sont "utiles" pour le signal. L'efficacité correspondante est donc $L2/L1$.

Si le débit entrant est $4 \times D_{TN1} = 4 \times 2048 \text{ kbit/S}$ alors le débit réel sortant du TN2 en tenant compte de l'efficacité est :

$$D_{TN2} = 4 \cdot D_{TN1} \cdot L1/L2 \text{ soit un débit : } 8.430 \text{ Mbit/S}$$

Ce qui donne un nombre $n=131,7 \times 64 \text{ kbit/s}$. par souci d'arrondi on prend $n=132$ et donc $D_{TN2}=8448 \text{ k bit/S}$

Les ordres supérieurs des multiplex sont structurés de manières similaires.

SDH .(SYNCHRONOUS DIGITAL HIERARCHY). HIERARCHIE NUMERIQUE SYNCHRONE

L'inconvénient de la PDH est qu'il faut démultiplexer complètement l'ensemble des différents ordres de multiplexage pour extraire un signal. De plus, comme nous l'avons vu, la synchronisation de chaque émetteur avec chaque récepteur est une condition clef à une transmission sans erreur. Ceci est parfois délicat à obtenir dans des systèmes où les horloges donnant le rythme sont réparties sur les émetteurs et les récepteurs. Il y a alors autant d'horloges différentes qu'il y a de machines différentes dans le réseau. Pour palier ces problèmes une nouvelle hiérarchie de système de transmission a vu le jour : la "Hiérarchie, Numérique Synchrone" ou "synchronous Digital Hierarchy (SDH en Europe et SONET aux USA). L'un des aspects principaux est que tous les équipements de cette hiérarchie de "multiplexeur" possèdent la même horloge. Cela peut s'obtenir si une seule horloge délivre des impulsions donnant le rythme à tous les équipements de l'ensemble du réseau national via un réseau de fibres optiques.

Le premier niveau de cette hiérarchie est le STM1 (155,52 Mbit/s) (STM : Synchronous transfer module). Les niveaux supérieurs sont :

-STM16 à 2488,32 Mbits/S (dit 2,5Gbit/s)

-STM 64 à 9953,28 Mbits/s (dit 10 Gbit/S)

Ces informations transitent non plus sur des câbles, mais sur des fibres optiques.

Element de la hiérarchie synchrone:

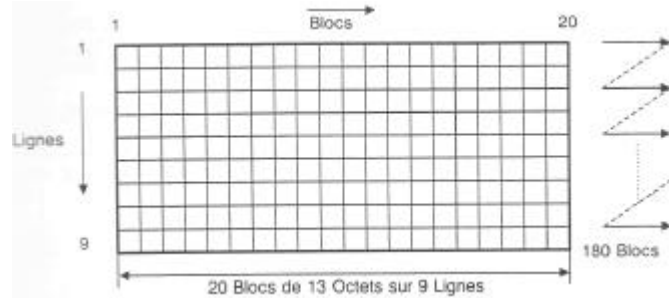
Afin d'expliquer la SDH il nous faut d'abord définir un certain nombre de termes:

Le conteneur : Cn (Contener):

Ce sont des blocs d'octets mis sous une certaine forme et dimensionnés pour assurer le transport d'un des différents débits affluents à la SDH. Il existe :

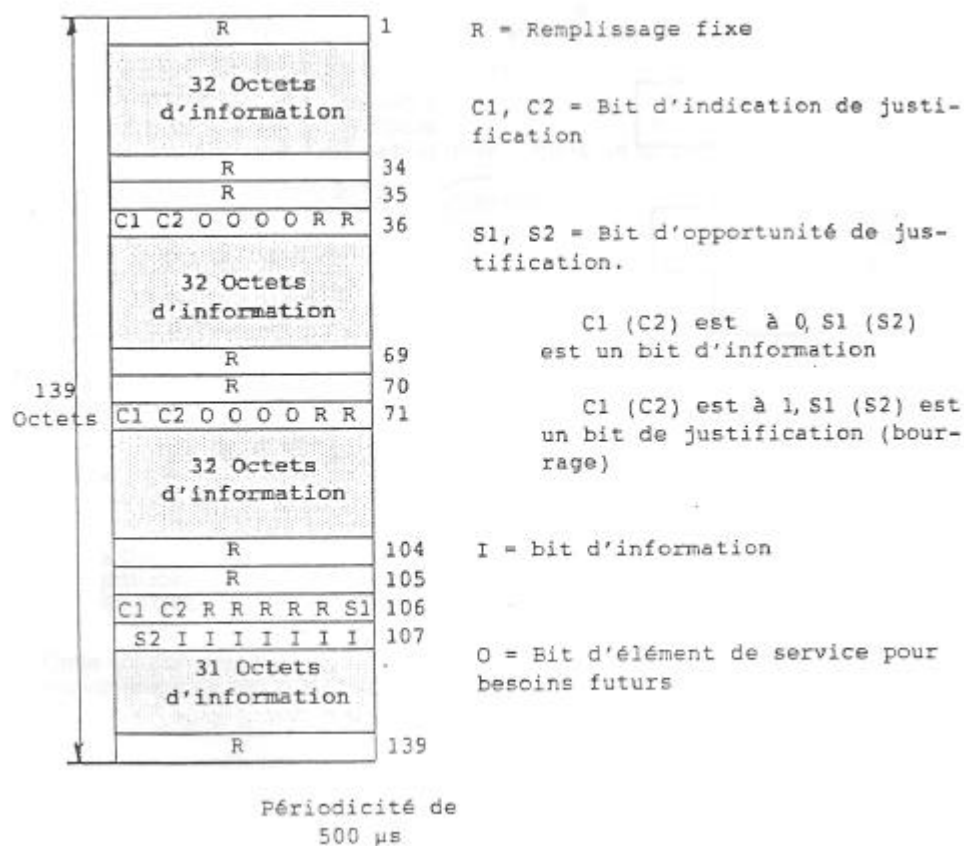
- C4 (permettant le transport de cellule ATM à 139264 Kbit/S)
- C3 (44736 Kbit/s ou le 34368 Kbit/s du TN3 de la PDH)
- C12 (trame MIC TN1 à 2048 Kbit/s)
- C11(1544Kbit/s).

C4 : Après récupération des données provenant d'un signal à 140 Mbit/s (ATM ou TN4 de la PDH) le C4 est élaboré en positionnant ces données dans un paquet de 180 blocs (9 x 20) chacun constitués de 13 octets. (9 x 20 X13 x 8 bits =18720 bits en 125 µs soit 149, 760 Mbit/s>139264 Kbit/s ce qui laisse des bits libres pour une éventuelle justification).



C3 : Il est élaboré à partir d'affluent à 34Mbit/s provenant du multiplex tertiaire TN3 de la PDH. Les données récupérées sont placées dans un conteneur de 756 octets (9 x 84 octets) soit un total de (9 x 84 x 8bits en 125 μ s = 6048 bits en 125 μ s = 48,384 Mbit/s)

C12 : La figure ci dessous illustre l'organisation du C12 pour un affluent à 2 Mbit/S :



Le conteneur virtuel : VCn (Virtual Contener):

Il est obtenu en ajoutant aux conteneurs précédemment définis un entête (PATH OVER HEAD = POH) utilisé pour la gestion du conteneur (routage, concaténation, justification...). Le VCn est l'entité gérée par le réseau SDH.

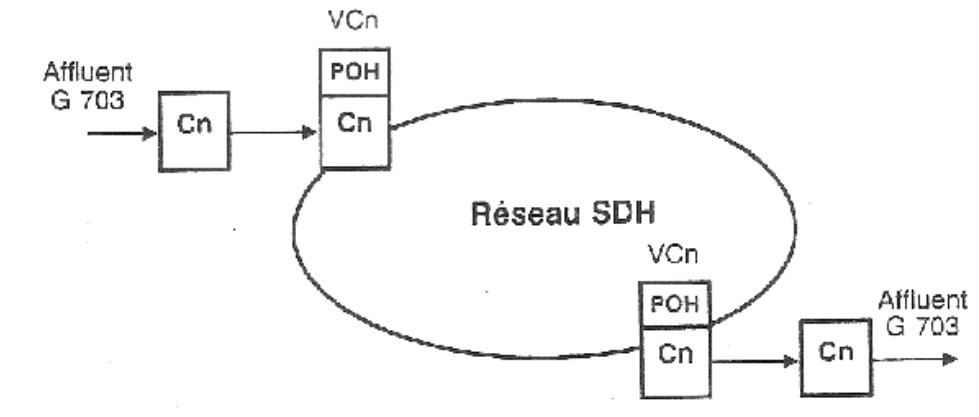


Fig : Conteneur et conteneur virtuel :

L'unité d'affluent : TUn (Tributary Unit)

L'unité d'affluent, TUn est composée du VCN et d'un pointeur PTR associé. La valeur de ce pointeur indique l'emplacement du VCN dans la trame de transport. Ce pointeur est associé au processus de justification du VC dans la trame de transport, c'est à dire que la position du VC dans la trame peut changer si il y a justification.

Le groupe d'unité d'affluent : TUGn (Tributary Unit Group)

Le groupe d'unité d'affluent n'est pas une nouvelle entité physique (informatique) mais représente une structure virtuelle de la trame reflétant le multiplexage des TUn. Cela permet de regrouper des TU pour les assembler en une entité (bloc) de dimension supérieure. Le TUG peut être considéré comme les règles de rangement des TU dans la trame de transport.

Par exemple :

- Le TUG 2 regroupe soit 3 TU12, soit 1 TU2.
- Le TUG 3 assemble 7 TUG 2 soit 1 TU3

L'unité administrative AU4 (Administrative unit) :

L'unité administrative, AU4 est composée du VC4 et du pointeur PTR associé. La valeur du pointeur indique le début du VC4 dans la trame de transport utilisée. Ce pointeur est associé au processus de justification du VC4.

Le groupe d'unité Administrative : AUG (Administrative unit Group) :

Le groupe d'unité administrative n'est pas une nouvelle entité physique mais représente une structure virtuelle de la trame. L'AUG correspond à la place que doit occuper l'AU4 dans la trame de transport.

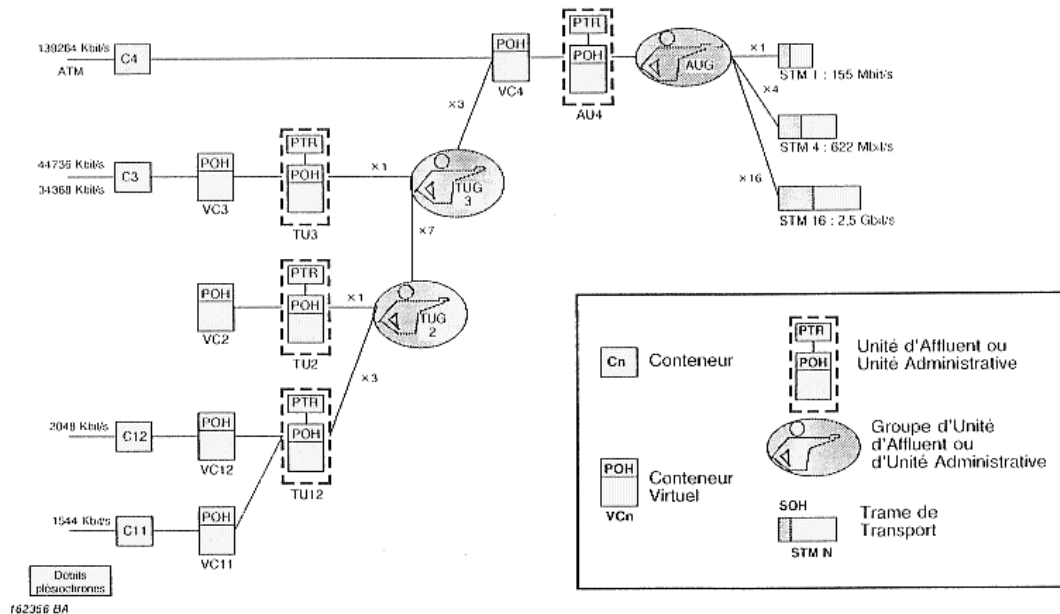


Fig. Structure de la hiérarchie numérique synchrone (SDH)

Les trames de transports STM_N (synchronous transport module):

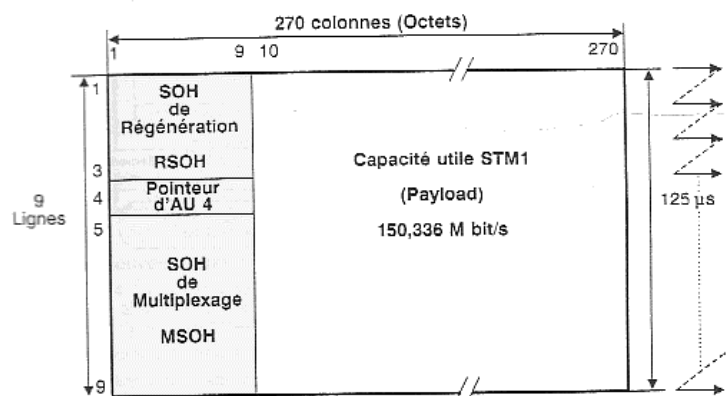
Le premier niveau de cette hiérarchie est le STM1 (155,52 Mbit/s). Les niveaux supérieurs sont :

- STM 4 à 622,08 Mbit/s
- STM16 à 2488,32 Mbits/S (dit 2,5Gbit/s)
- STM 64 à 9953,28 Mbits/s (dit 10 Gbit/S)

Structure de la Trame de base STM1 (insertion de l'AU4 dans la trame de base)

Comme dans le PDH, les informations sont réparties en trame, à la différence près que pour la SDH la trame ressemble plus à une page qu'à une suite de bits.

Pour le STM1 la page est longue de 2430 octets. Ceux ci sont répartis en 9 lignes et 270 colonnes. Parmi les 270 colonnes on distingue 2 zones: Les premières 9 colonnes font parties de l'en-tête, les 261 suivantes sont la zone où l'information proprement dite est rangée (payload).



La trame STM1 (155,520 Mbit/S = $270 \times 9 \times 8 / 125\mu s$) comprend trois zones :

-Le surdébit de section (SOH Section over head)

- Lignes 1 à 3 - colonnes 1 à 9 (SOH de régénération =RSOH).
- Lignes 5 à 9 - colonnes 1 à 9 (SOH de multiplexage=MSOH).

-Le pointeur d'AU4

- Ligne 4 - colonne 1 à 4

-La capacité utile (Payload)

- Ligne 1 à 9 des colonnes 10 à 270 (261 colonnes)

Pointeur d'AU4

Le Pointeur d'AU4 permet un alignement souple et dynamique du VC 4 dans le STM1, c'est à dire que le VC 4 peut " flotter " dans la trame. Le pointeur permet donc de tenir compte à la fois :

- du décalage du VC par rapport au début de la capacité utile.
- des différences de débit entre le VC 4 et la capacité utile de la trame de base.

a) Emplacement du Pointeur :

Le pointeur d'AU4 est contenu dans les octets H1, H2 et H3:

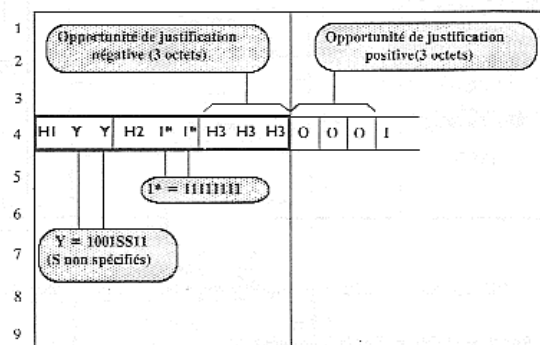


Fig : pointeur :

Seulement 10 bit des octets H1 et H2 sont utilisés pour désigner l'emplacement du premier octet du VC4. Ceci donne 1024 numéros possibles. Or il y a $261 \times 9 = 2349$ octets dans la capacité utile du STM 1 Il est donc impossible de désigner tous les octets. La solution consiste à grouper les octets par 3 et à attribuer un numéro à chaque groupe de 3, ce qui donne $2349 : 3 = 783$ groupes, numérotés de 0 à 782. En fait, pour le VC4, le pointeur désigne donc le premier d'un groupe de 3 octets.

La justification octet du VC4 porte par conséquent sur 3 octets à la fois

- 3 octets H3 pour la justification négative
- 3 octets 0 pour la justification positive

La SDH possède par rapport à la PDH des caractéristiques essentielles

- Tous les équipements SDH sont pilotés par le même signal d'horloge il y a quelque part une horloge atomique qui donne le rythme pour tout le pays et ce rythme est transmis de proche en proche à tous les équipements de transmission SDH du pays. C'est pour cette raison que ce mode de transmission est appelé synchrone. Pour les flux SDH, la justification n'est donc pas nécessaire.
- Les multiplexeurs d'ordre N savent multiplexer des affluents de n'importe quel ordre inférieur (au contraire de la PDH qui ne sait multiplexer que des flux de l'ordre immédiatement inférieur N). Il est donc possible de retrouver n'importe quel flux d'ordre N- 1 ou N-2 ou N-i, à la sortie d'un démultiplexeur SDH d'ordre N sans repasser par tous les niveaux intermédiaires de démultiplexage. Cette propriété rend possible la conception d'équipements pour injecter ou extraire des affluents de n'importe quel ordre inférieur sur un point intermédiaire du parcours de la transmission.
- Comme tous les affluents ne sont plus nécessairement issus de l'une des extrémités pour aller à l'autre extrémité de la transmission, une importante fonction de gestion se trouve nécessaire pour spécifier la destination de chacun des flux multiplexés.
- La SDH admet comme affluents aussi bien des flux plésiochrones (pour lesquels elle possède un mécanisme de justification) que des flux synchrones, c'est-à-dire rythmés par l'horloge unique du pays. Parmi les affluents plésiochrones, la SDH admet aussi bien les affluents de la PDH Européenne (Type E) que les affluents de la PDH Américaine (Type T).

.

.

8. INTRODUCTION AUX TELECOMMUNICATIONS OPTIQUES. COMMENT TRANSMETTRE DES SIGNAUX AVEC DE LA LUMIERE ?

COMMENT TRANSMETTRE DES SIGNAUX AVEC DE LA LUMIERE ?

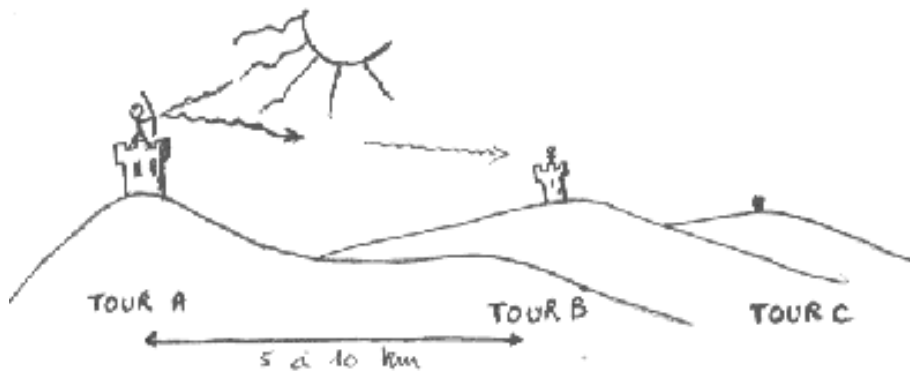


Figure 1 – Principe moyenâgeux de la transmission d'informations optiques.

Cette question n'est pas récente. Au moyen âge par exemple, le long de l'axe Rhin Rhône des signaux d'alerte étaient transmis de tours en tours sur des dizaines de kilomètres. Le principe était le suivant ; lorsqu'un danger était détecté à proximité d'un château fort, des signaux étaient envoyés au château fort voisin par l'intermédiaire d'un miroir réfléchissant les rayons du soleil ou bien par agitation d'un drapeau. Pour permettre l'observation visuelle directe des signaux, le château émetteur des signaux et le château récepteur devaient ne pas être trop distants l'un de l'autre (quelques kilomètres). Par transmission successive de château en château les signaux d'alerte pouvaient être communiquer à la ville voisine à une vitesse supérieure à la vitesse d'un cheval au galop et prévenir une invasion. Ce système apparemment simpliste, car soumis aux conditions climatiques qui pouvaient empêcher son fonctionnement (pluie, brouillard, contre-jour), a pourtant inspiré les premiers essais modernes de télécommunication optiques. En effet dans les années 70, des essais en

atmosphères libres visaient à transmettre des signaux optiques modulés provenant de lasers puissants sur des distances de quelques centaines de mètres. Par temps de brouillard les signaux n'atteignaient pas leur but, le spot laser à la réception pouvait avoir un diamètre de quelques mètres et sa position pouvait fluctuer sur quelques mètres également. Il apparaît donc que transmettre des signaux avec de la lumière nécessite une certaine technologie. Pour identifier les différents aspects et problèmes que peut rencontrer l'établissement d'une telle liaison, reprenons le schéma de principe du moyen âge et analysons-le. Volontairement nous partirons d'un schéma simpliste à partir duquel nous tenterons de reconstruire le monde des télécoms et d'inventer celui de demain.

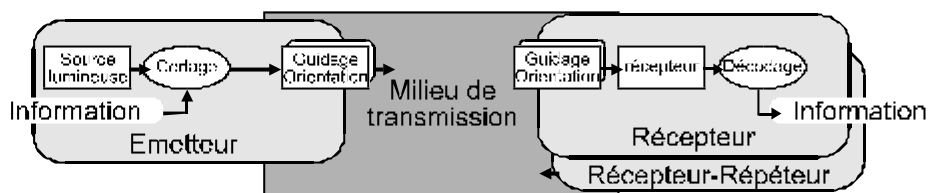


Figure 2 : Schéma synthétique d'une liaison à transmission optique

La
tour A
(Alice)
peut être
considér
é comme

l'émetteur, la tour B (Bob) comme le **récepteur** (Dans le cas de retransmission vers la tour C, Bob peut être considéré comme un **répéteur**). Entre les deux, un **milieu**, qui permet plus ou moins bien la transmission du signal. Quelques problèmes apparaissent déjà : comment s'assurer que le signal est bien **orienté** vers le bon destinataire ? (Est-ce que A a bien orienté ces reflets du soleil vers B et non pas dans une mauvaise direction ? Est-ce que le récepteur est en mesure de comprendre les signaux, (quel type de **codage** ?), (Est-ce que A et B se sont mis d'accord sur la signification des signes qu'ils font) Comment faire pour avoir une **source lumineuse** fiable (qui soit indépendante des aléas climatiques), et un récepteur capable de **détecter** même dans de mauvaises conditions (est-ce que Bob a de suffisamment bons yeux ?).

Ceci nous conduit au schéma synthétique suivant :

Nous distinguons donc des éléments distincts :

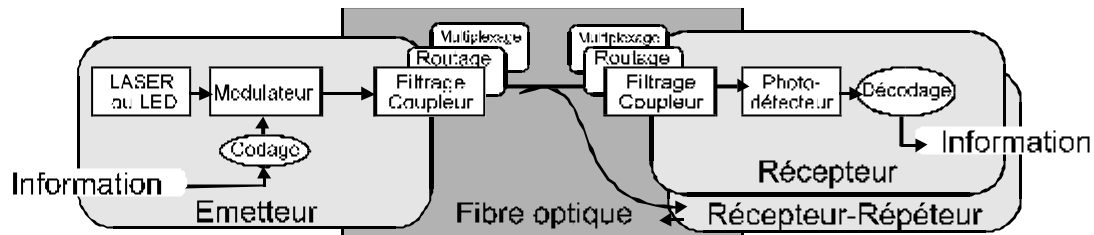
- La source lumineuse (au niveau de l'émetteur)
- Le détecteur sensible à la lumière (au niveau du récepteur)
- Le milieu de transmission
- La modulation de l'information dans la porteuse lumineuse (ce que nous avons appelé 'codage' au sens large)
- Le système d'orientation de l'information

Et nous identifions au moins trois systèmes complets

- Les émetteurs
- Les récepteurs
- Les répéteurs

Tous ces points vont être repris pour définir les chapitres de ce cours. Précisons-les davantage en traduisant le schéma synthétique figure 2, en un autre schéma plus détaillé et explicite incluant les différents équipements télécom modernes :

Figure 3 : Schéma de principe d'une ligne optique de transmission.



Nous les aborderons donc en suivant le plan ci après :

Première partie : Les éléments

CH. 8	Introduction
CH. 9	Optique Physique
CH. 10	Optique Géométrique
CH. 11	Fibres optiques
CH. 12	Les semi-conducteurs et leurs applications photosensibles
CH. 13	Les LEDS
CH. 14	Introduction à la physique des LASERS
CH. 15	Les photodétecteurs, Photodiode PIN et Avalanche.

9. OPTIQUE PHYSIQUE

Afin de pouvoir imaginer, comprendre et éventuellement concevoir des dispositifs optiques pour les télécoms, il est nécessaire de connaître ce qu'est la lumière ainsi que ces propriétés. La connaissance des propriétés de la lumière ainsi que celle de certains moyens physiques pour les altérer voire les contrôler permettra de réaliser des dispositifs.

Les propriétés physiques de la lumière sont fondées sur les équations de Maxwell. En effet la lumière étant une onde électromagnétique elle en vérifie les propriétés.

EQUATION DE MAXWELL :

Th. équivalent	Formes globales (Cours de 2 ^{ème} année) (Vrai seulement en Statique)	Formes locales : Eq. de Maxwell dans la matière (Vrai en statique et en dynamique)		Formes locales : Eq. de Maxwell dans le vide (pas de charges)
Th. De Gauss	$\text{Flux}(E) = \frac{Q_{\text{int}}}{\epsilon_0}$	$\text{div}.E = \frac{\rho}{\epsilon}$	$\nabla \cdot E = \frac{\rho}{\epsilon}$	$\nabla \cdot E = 0$
Potentiel nul le long d'une boucle fermée	Circulation de E $\oint_{\text{contour}} \vec{E} \cdot d\vec{l} = 0$	$\text{rot} \wedge E = -\frac{\partial B}{\partial t}$	$\nabla \wedge E = -\frac{\partial B}{\partial t}$	$\nabla \wedge E = -\frac{\partial B}{\partial t}$
Pas de monopole magnétique	$\text{Flux}(B) = 0$	$\text{div}.B = 0$	$\nabla \cdot B = 0$	$\nabla \cdot B = 0$
Th. D'Ampère	Circulation de B $\oint_{\text{contour}} \vec{B} \cdot d\vec{l} = \mu_0 J$	$\text{rot} \wedge B = \mu J + \epsilon \mu \frac{\partial E}{\partial t}$ avec $J = SE$	$\nabla \wedge B = \mu J + \epsilon \mu \frac{\partial E}{\partial t}$	$\nabla \wedge B = \epsilon \mu \frac{\partial E}{\partial t}$

Rappel mathématique sur les notations : Rotationnel gradient et divergence:

Opérateur NABLA : Pour simplifier les écritures on définit un opérateur (une fonction qui s'exerce sur les composantes d'un vecteur) :

$$\vec{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix}$$

Cet opérateur nous simplifie l'écriture du rotationnel de la divergence et du gradient :

Divergence :

$$\text{Div}(\vec{V}) = \vec{\nabla} \cdot \vec{V} = \begin{vmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{vmatrix} \cdot \begin{vmatrix} V_x \\ V_y \\ V_z \end{vmatrix} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}$$

Gradient :

$$\text{grad } V = \vec{\nabla} V = \begin{vmatrix} \frac{\partial V_x}{\partial x} \\ \frac{\partial V_y}{\partial y} \\ \frac{\partial V_z}{\partial z} \end{vmatrix}$$

Rotationnel :

$$\vec{\text{Rot}} V = \vec{\nabla} \wedge \vec{V} = \begin{vmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{vmatrix} \wedge \begin{vmatrix} V_x \\ V_y \\ V_z \end{vmatrix} = \begin{vmatrix} \frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \\ \frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \\ \frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \end{vmatrix}$$

Laplacien :

$$\vec{\nabla} \cdot (\vec{\nabla} E) = \vec{\nabla} \cdot \vec{\nabla} E = (\vec{\nabla} \cdot \vec{\nabla}) E = \nabla^2 E$$

$$\text{et l'opérateur Laplacien s'écrit : } \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

EQUATION DE PROPAGATION

Dans un premier temps nous allons étudier les équations dans le vide parce qu'elles ont l'avantage d'être plus simple en cela qu'elles n'ont pas de charges.

En combinant les équations on obtient une équation dite équation de

propagation dans le vide : $\nabla^2 E - \epsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2} = 0$ ou encore en développant $\nabla^2 E$:

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} - \epsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2} = 0$$

Par le même raisonnement on montre :

$$\frac{\partial^2 B}{\partial x^2} + \frac{\partial^2 B}{\partial y^2} + \frac{\partial^2 B}{\partial z^2} - \epsilon_0 \mu_0 \frac{\partial^2 B}{\partial t^2} = 0$$

Démonstration du calcul de l'équation de propagation:

Les équations à considérer sont les suivantes :

$$(1) \quad \nabla \wedge E = -\frac{\partial B}{\partial t} \quad \text{et} \quad \nabla \cdot E = 0 \quad (2)$$

$$(3) \quad \nabla \wedge B = \epsilon \mu \frac{\partial E}{\partial t} \quad \text{et} \quad \nabla \cdot B = 0 \quad (4)$$

on calcule le rotationnel de l'équation (1) :

$$\text{rot}(\text{rot}(E)) = \vec{\nabla} \wedge (\vec{\nabla} \wedge \vec{E}) = \vec{\nabla} \wedge \left(-\frac{\partial B}{\partial t} \right) = \text{rot}(-\partial B / \partial t) \text{ qui peut s'écrire :}$$

$$\overrightarrow{\text{grad}(\text{div} \vec{E})} - \text{laplacien}(\vec{E}) \text{ soit } \vec{\nabla}(\vec{\nabla} \cdot \vec{E}) - \nabla^2$$

mais selon (2) $\text{div}(E)=0$, ce dernier calcul se simplifie et devient :

$$-\nabla^2 \vec{E} = -\frac{\partial}{\partial t} (\vec{\nabla} \wedge \vec{B})$$

en utilisant (3) pour faire disparaître le terme B il vient :

$$-\nabla^2 \vec{E} = -\frac{\partial}{\partial t} \left(\epsilon \mu \frac{\partial E}{\partial t} \right) \text{ En final il vient donc : } 0 = \nabla^2 E - \epsilon \mu \frac{\partial^2 E}{\partial t^2}$$

D'une manière générale on montre que pour un potentiel V ou pour tout autres grandeurs physiques F l'équation de propagation s'écrit :

$$\frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} + \frac{\partial^2 F}{\partial z^2} - \epsilon \mu \frac{\partial^2 F}{\partial t^2} = 0 \quad (\text{Eq.1})$$

Si on se limite au seul axe X, l'équation s'écrit :

$$\frac{\partial^2 F}{\partial x^2} - \epsilon \mu \frac{\partial^2 F}{\partial t^2} = 0 \quad (\text{Eq.2})$$

Remarque : En notant [F] l'unité de la grandeur F, l'unité de l'expression $\mathcal{F}^2/\mathcal{X}^2$ est $[F]^2 m^{-2}$, et l'expression $\mathcal{F}^2/\mathcal{T}^2$ est $[F]^2 s^{-2}$. D'après l'équation (2) on peut écrire l'équation aux dimensions :

$$[F]^2 m^{-2} - [\epsilon \mu] [F]^2 s^{-2} = 0$$

où $[\epsilon \mu]$ est l'unité du produit $\epsilon \mu$. En simplifiant il vient $[\epsilon \mu] = m^{-2} s^{+2}$ soit : $\left[\frac{1}{\sqrt{\epsilon \mu}} \right] = m s^{-1}$

C'est donc une vitesse. Mais quelle est cette vitesse ? Faisons le calcul de l'application numérique dans le cas du vide avec :

$$\epsilon = \epsilon_0 = 8,854.10^{-12} \text{ F.m}^{-1} \text{ (Farad par mètre)}$$

$$\mu = \mu_0 = 4\pi \cdot 10^{-7} \text{ H.m}^{-1} \text{ (Henri par mètre)}$$

On trouve $\frac{1}{\sqrt{\epsilon \mu}} = 2.99792 \cdot 10^8 \text{ ms}^{-1}$ c'est à dire la vitesse C de la lumière. Ceci signifie que

l'équation de propagation définit une vitesse à laquelle les ondes électromagnétiques se propagent. Dans le vide, cette vitesse est C la vitesse de la lumière.

$$C = \frac{1}{\sqrt{\epsilon_0 \mu_0}}$$

Dans un milieu i la constante diélectrique ϵ_i et la permittivité diélectrique μ_i définissent une vitesse de propagation spécifique C_i au matériau en question :

$$C_i = \frac{1}{\sqrt{\epsilon_i \mu_i}}$$

ETUDE DE QUELQUES SOLUTIONS DE L'EQUATION DE PROPAGATION :

Si une fonction F est solution de l'équation de propagation, c'est qu'elle décrit une forme d'onde. Dans un premier temps nous nous limiterons à un problème avec un seul axe ce qui correspond à l'équation (2). Essayons maintenant de trouver une solution à l'équation (2) en considérant une onde de forme F se propageant suivant l'axe x .

On peut ainsi montrer que toute fonction f de la forme $f(x-ct)$ et $g(x+ct)$ est une onde vérifiant l'équation de propagation.

Considérons les fonctions $f(x-ct)$ et $g(x+ct)$.

En posant $u=x-ct$ et $v=x+ct$ il vient $x=\frac{u+v}{2}$ (i) et $t=\frac{v-u}{2c}$ (ii)

En dérivant $dx=\frac{du+dv}{2}$ (iii) et $dt=\frac{dv-du}{2c}$ (iv) et $df=\frac{\partial f}{\partial x}dx+\frac{\partial f}{\partial t}dt$ (v)

En remplaçant dx et dt par leur expression donnée par (iii) et (iv) il vient :

$$df=\frac{\partial f}{\partial x}\left(\frac{du+dv}{2}\right)+\frac{\partial f}{\partial t}\left(\frac{dv-du}{2c}\right)=\frac{1}{2}\left(\frac{\partial f}{\partial x}-\frac{1}{c}\frac{\partial f}{\partial t}\right)du+\frac{1}{2}\left(\frac{\partial f}{\partial x}+\frac{1}{c}\frac{\partial f}{\partial t}\right)dv \quad (\text{vi})$$

df peut aussi s'écrire en calculant sa dérivée par rapport au dérivée partielle du et dv :

$$df=\frac{\partial f}{\partial u}du+\frac{\partial f}{\partial v}dv \quad (\text{vii}) \text{ en identifiant terme à terme on peut en déduire que :}$$

$$\frac{\partial f}{\partial u}=\frac{1}{2}\left(\frac{\partial f}{\partial x}-\frac{1}{c}\frac{\partial f}{\partial t}\right) \quad (\text{viii}) \text{ et } \frac{\partial f}{\partial v}=\frac{1}{2}\left(\frac{\partial f}{\partial x}+\frac{1}{c}\frac{\partial f}{\partial t}\right) \quad (\text{ix})$$

En observant attentivement l'équation (2) on peut s'apercevoir qu'elle peut s'écrire :

$$\frac{\partial}{\partial x}\left[\frac{\partial F}{\partial x}+\frac{1}{c}\frac{\partial F}{\partial t}\right]-\frac{1}{c}\frac{\partial}{\partial t}\left[\frac{\partial F}{\partial x}+\frac{1}{c}\frac{\partial F}{\partial t}\right] \text{ dans chaque crochet on reconnaît l'opérateur } \frac{\partial}{\partial u}.$$

On peut essayer de mettre en facteur les opérateurs et on trouve alors :

$$\frac{\partial}{\partial x}\left[\frac{\partial F}{\partial x}+\frac{1}{c}\frac{\partial F}{\partial t}\right]-\frac{1}{c}\frac{\partial}{\partial t}\left[\frac{\partial F}{\partial x}+\frac{1}{c}\frac{\partial F}{\partial t}\right] \text{ ce qui en utilisant (viii) et (ix) permet de transformer (2) en : } \frac{\partial}{\partial u}\left[\frac{\partial F}{\partial v}\right]=0$$

ce qui peut s'intégrer facilement en deux étapes : on calcule d'abord la primitive par rapport à u , il reste

$$\frac{\partial F}{\partial v}=g(v) \quad g(v) \text{ est fonction de la seule variable } v, \text{ c'est une constante par rapport à } u.$$

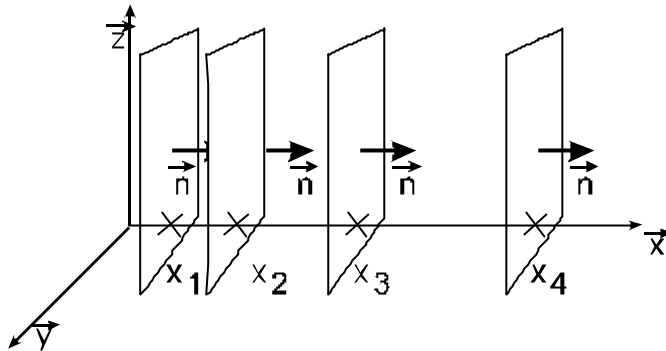
Intégrons par rapport à v

$$F=\int g(v)dv+\text{constant}(v)=g(v)+f(u) \quad \text{donc } F(u,v)=f(x-ct)+g(x+ct) \text{ est une}$$

fonction qui vérifie l'équation de propagation. C'est donc une onde.

Ondes planes sur un axe:

On dit qu'une onde est plane si F à la même valeur en tout point d'un plan perpendiculaire à une direction fixe définie par un vecteur unitaire \vec{k} qui est la direction de propagation.



Les fonctions f et g définies ci dessus, sont des fonctions de x et t . A t donné, f et g sont constantes dans tous plans $x=cste$ puisqu'elles ne dépendent ni de y ni de z . Dans l'espace, l'ensemble des points $x=cste$ sont des plans dont la normale est orientée suivant x . f et g sont donc des ondes planes normales à x .

Remarque : Si on considère une fonction f telle que $f(x,t) = f(x_1 - c.t_1) = f(x_2 - c.t_2)$ alors $x_1 - ct_1 = x_2 - ct_2$ d'où $c = \frac{x_2 - x_1}{t_2 - t_1}$. Si ce n'était déjà fait on pourrait remarquer que $C = \frac{1}{\sqrt{\epsilon\mu}} = \frac{x_2 - x_1}{t_2 - t_1}$ est une vitesse.

Conclusion : La dépendance en $(x-ct)$ traduit une onde plane se propageant suivant l'axe Ox à la vitesse C .

STRUCTURE PROGRESSIVE DE L'ONDE PLANE ELECTROMAGNETIQUE :

Ce que nous venons de conclure pour l'onde plane en général s'applique pour les ondes électromagnétiques. Mais n'oublions pas que les ondes électromagnétiques nécessite la présence d'un champ électrique E et d'un champ magnétique B . Cherchons maintenant à préciser la structure de l'onde électromagnétique plane et notamment comment sont orientées les vecteurs E et B .

Considérons un vecteur champ électrique E tel que :

$$\vec{E} = \begin{cases} E_x = E_x(x-ct) \\ E_y = E_y(x-ct) \\ E_z = E_z(x-ct) \end{cases}$$

E_x, E_y, E_z sont fonction des seuls x et t puisque nous avons supposés que ce sont des ondes planes se propageant suivant x .

Nous pouvons tenir le même raisonnement pour B :

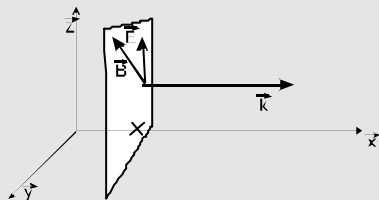
$$\vec{B} = \begin{cases} B_x = B_x(x-ct) \\ B_y = B_y(x-ct) \\ B_z = B_z(x-ct) \end{cases}$$

D'après les équations de Maxwell $\nabla \cdot \vec{B} = 0$. C'est à dire que $\frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} + \frac{\partial B_z}{\partial z} = 0$. Or B_y et B_z ne dépendent ni de y ni de z puisque B est une onde plane, donc $\frac{\partial B_y}{\partial y} = 0$ et $\frac{\partial B_z}{\partial z} = 0$. Du calcul de la divergence ci dessus nous en déduisons que $\frac{\partial B_x}{\partial x} = 0$ d'où B_x est une constante. Cette constante correspond à un champ magnétostatique et non pas à un champ magnétodynamique, et donc le pour le calcul de l'onde nous n'en tenons pas compte.

Le même raisonnement peut se conduire pour E puisque dans un milieu sans charge nous avons $\nabla \cdot \vec{E} = 0$.

$$\vec{E} = \begin{cases} E_y = E_y(x-ct) \\ E_z = E_z(x-ct) \end{cases} \text{ et } \vec{B} = \begin{cases} B_y = B_y(x-ct) \\ B_z = B_z(x-ct) \end{cases}$$

Ce qui signifie que \vec{E} et \vec{B} n'ont pas de composante suivant la direction de propagation. Les vecteurs \vec{E} et \vec{B} sont des vecteurs dans un plan perpendiculaire à la direction de propagation \vec{k} .



De l'équation $\nabla \wedge \vec{E} = -\frac{\partial \vec{B}}{\partial t}$ nous pouvons en déduire un dernier point sur l'orientation de \vec{E} par rapport à \vec{B} . Notons que

$$\vec{\nabla} \wedge \vec{E} = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ E_x & E_y & E_z \end{vmatrix} = \begin{vmatrix} 0 & \frac{\partial E_z}{\partial x} \\ \frac{\partial E_z}{\partial y} & 0 \end{vmatrix} \text{ et } \frac{\partial \vec{B}}{\partial t} = \begin{vmatrix} 0 & \frac{\partial B_y}{\partial t} \\ \frac{\partial B_z}{\partial t} & 0 \end{vmatrix} \text{ en remarquant que } \frac{\partial}{\partial x} = \frac{\partial}{\partial u} \times \frac{\partial u}{\partial x} = \frac{\partial}{\partial u} \text{ puisque } \frac{\partial}{\partial t} = \frac{\partial}{\partial u} \times \frac{\partial u}{\partial t} = -c \frac{\partial}{\partial u}$$

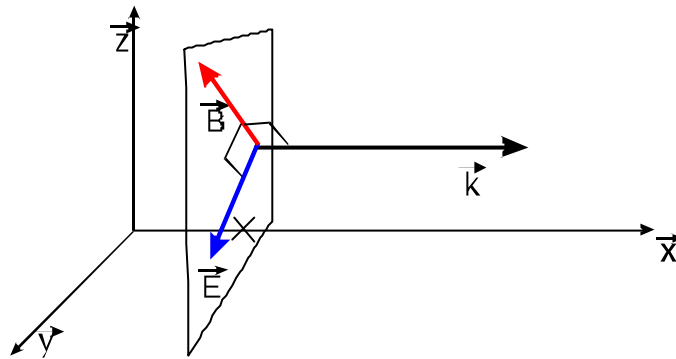
u=x-ct on peut en déduire que $\frac{\partial E_z}{\partial u} = -c \frac{\partial B_y}{\partial u}$ et $\frac{\partial E_y}{\partial u} = -c \frac{\partial B_z}{\partial u}$

Aux constantes près qui ne nous intéressent pas, $\frac{E_z}{E_y} = -\frac{cB_y}{cB_z}$ signifie que \vec{E} et \vec{B} sont orthogonaux.

Si $\vec{k} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ est la direction de propagation, $\vec{E} = \begin{pmatrix} 0 \\ E_y \\ E_z \end{pmatrix}$ est la structure du champ électrique,

alors $\vec{B} = -\frac{1}{C} \vec{k} \wedge \vec{E}$ puisque $\frac{1}{C} \vec{k} \wedge \vec{E} = \frac{1}{C} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \wedge \begin{pmatrix} 0 \\ E_y \\ E_z \end{pmatrix} = \frac{1}{C} \begin{pmatrix} 0 \\ E_z \\ -E_y \end{pmatrix}$

Une onde électromagnétique de vecteur \vec{E} et \vec{B} se propageant suivant \vec{k} a la particularité d'avoir \vec{E} et \vec{B} orthogonaux entre eux dans un plan normal à la direction de propagation \vec{k} .



ONDE ELECTROMAGNETIQUE PLANE MONOCHROMATIQUE :

On a montré que les ondes pouvaient s'écrire comme une fonction de $x-ct$ ou de $x+ct$. Mais peut-on avoir une expression plus détaillée ? En fait on peut montrer qu'un cosinus de $(x-ct)$ satisfait l'équation de propagation :

$$\|\vec{E}\| = E_0 \cos(k(x-ct)) \quad k \text{ est une constante.}$$

On peut montrer facilement que l'expression de $\|\vec{E}\|$ donnée ci-dessus est bien solution de l'équation de propagation :

$$\frac{\partial^2 \vec{E}}{\partial x^2} = -k^2 \vec{E} \quad \text{et} \quad -\frac{1}{C^2} \frac{\partial^2 \vec{E}}{\partial t^2} = +\frac{1}{C^2} k^2 C^2 \vec{E} \quad \text{d'où} \quad \frac{\partial^2 \vec{E}}{\partial x^2} - \frac{1}{C^2} \frac{\partial^2 \vec{E}}{\partial t^2} = 0$$

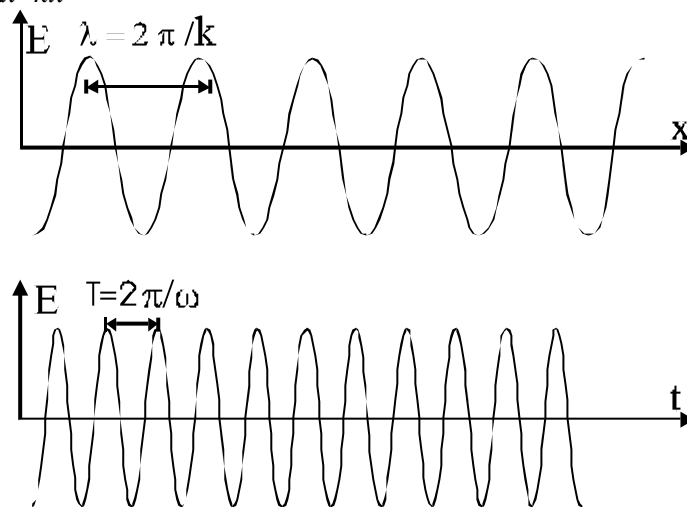
Donc $\|\vec{E}\| = E_0 \cos(k(x-ct))$ est bien solution de l'équation de propagation

D'autre part en développant $k(x-ct) = kx - kct$. Remarquons que kct est une phase. On peut écrire que $kc = \omega$ ce qui permet de faire apparaître une pulsation devant le terme temporel. Il n'est possible de faire apparaître ω que si l'onde électromagnétique a une seule fréquence, on parle alors d'onde monochromatique.

Conclusion : Une forme algébrique d'une onde plane monochromatique (une seule fréquence) se propageant suivant un seul axe x est : $E = E_0 \cos(kx - \omega t)$ ce qui donne une double périodicité à la fois dans le temps et dans l'espace :

Périodicité temporelle	$T = \frac{2\pi}{\omega}$	
Périodicité spatiale	$\lambda = \frac{2\pi}{k}$ avec $k = \frac{\omega}{c}$	$\lambda = cT = \frac{c}{n} = \frac{2\pi c}{\omega}$

La grandeur λ s'appelle la longueur d'onde c'est la distance parcouru par l'onde pendant une période temporelle. \vec{k} est appelé le vecteur d'onde. Si l'onde de direction \vec{k} se propageant suivant \vec{x} on a $\vec{k} \cdot \vec{x} = kx$



ECRITURE COMPLEXE DE L'ONDE PLANE MONOCHROMATIQUE

L'onde plane monochromatique électromagnétique peut aussi s'écrire sous forme complexe ce qui bien souvent permet de simplifier les calculs.

Considérons l'écriture suivante

$$\vec{E} = \begin{cases} E_x = 0 \\ E_y = E_0 \cos(\omega t - kx) \\ E_z = E_0 \cos(\omega t - kx) \end{cases} = \begin{cases} 0 \\ E_0 \cdot \Re(e^{j(\omega t - kx)}) \\ E_0 \cdot \Re(e^{j(\omega t - kx)}) \end{cases}$$

En fait on peut adopter la notation complexe d'une onde en remarquant que la forme \cos que nous utilisons habituellement n'est que la partie réelle. $\vec{E} = \Re(\vec{e}) = \Re(\vec{e}_0 e^{j(\omega t - kx)})$

Dans la suite de ce cours nous adopterons cette écriture complexe :

$$\vec{e} = \vec{e}_0 e^{j(\omega t - kx)}$$

Un des intérêts de cette méthode viens du fait qu'en représentation complexe

les opérateurs de dérivations et les opérateurs différentielles des ondes

planes monochromatiques ont des formes simples: $\frac{\partial \vec{e}}{\partial t} = j\omega \vec{e}$ et $\frac{\partial \vec{e}}{\partial x} = -jk \vec{e}$

DEPHASAGE :

Un autre intérêt de cette notation est de pouvoir comparer des ondes à des instants et à des endroits différents et de faire des calculs plus rapidement sur ces phases. En effet si ωt est une phase, le terme kx en est une aussi. Ainsi on peut exprimer facilement le déphasage entre deux instants différents. Ainsi une même onde est déphasée de $\Delta j = \omega \Delta t$ entre deux instants séparés de Δt .

De même le déphasage de l'onde entre deux points séparés de Δx est $\Delta j = k \Delta x$. En remplaçant k par son expression on obtient :

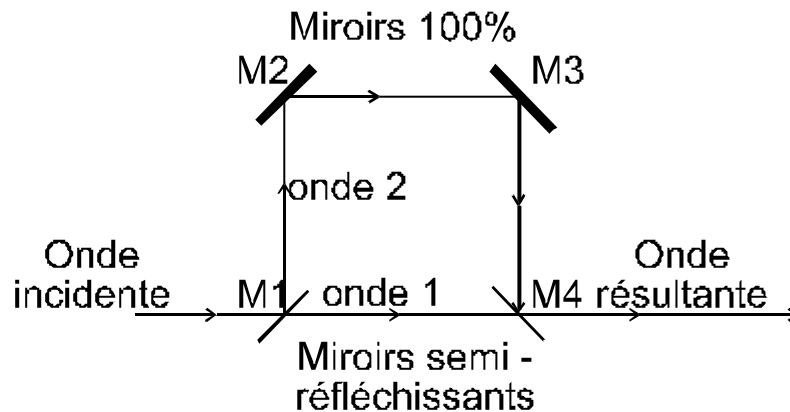
$$\Delta j = \frac{\omega}{c} \Delta x = 2\pi n \frac{\Delta x}{c} = 2\pi \frac{\Delta x}{\lambda}. \text{ Ce résultat est un élément de calcul des déphasages}$$

essentiel dans la suite du cours.

INTERFERENCES

Définition : On appelle interférence le phénomène physique qui résulte de l'addition de deux ou plusieurs ondes qui ont une relation de phase entre elles.

Exemple :



Un onde incidente est séparée en deux au niveau du miroir semi réfléchissant M1. Il en résulte deux ondes appelées "onde 1" et "onde 2" d'égale amplitude. L'onde 2 parcourt le trajet $2l+L$ défini par les miroirs M1, M2, M3 et M4, tandis que l'onde 1 parcourt la distance L à travers les miroirs M1 et M4 directement.

Avant M1 les ondes 1 et 2 s'écrivent :

$e_1 = \frac{E_0}{2} e^{j\omega t}$ et $e_2 = \frac{E_0}{2} e^{j\omega t}$ si on prend le miroir M1 comme origine des x . En M4, l'onde 1 à subit un déphasage $\Delta j_1 = 2\pi \frac{L}{\lambda}$ alors que l'onde 2 à subit un déphasage $\Delta j_2 = 2\pi \frac{2l+L}{\lambda}$.

D'où : $e_1 = \frac{E_0}{2} e^{j\omega t - 2p \frac{L}{T}}$ et $e_2 = \frac{E_0}{2} e^{j\omega t - 2p \frac{2l+L}{T}}$. En M4 les 2 ondes s'additionnent :

$$e_1 + e_2 = \frac{E_0}{2} e^{j\left(\omega t - 2p \frac{L}{T}\right)} + \frac{E_0}{2} e^{j\left(\omega t - 2p \frac{2l+L}{T}\right)} = \frac{E_0}{2} e^{j\left(\omega t - 2p \frac{L}{T}\right)} \left[1 + e^{-j2p \frac{2l}{T}} \right]$$

Le dernier terme peut se simplifier :

$$\left[1 + e^{-j2p \frac{2l}{T}} \right] = e^{-j2p \frac{l}{T}} \left[e^{j2p \frac{l}{T}} + e^{-j2p \frac{l}{T}} \right] = e^{-j2p \frac{l}{T}} 2 \cos\left(2p \frac{l}{T}\right)$$

$$e_T = e_1 + e_2 = E_0 \cos\left(2p \frac{l}{T}\right) e^{j\omega t - 2p \frac{L}{T} - 2p \frac{l}{T}} = E_0 \cos\left(2p \frac{l}{T}\right) e^{j\omega t + j} \text{ avec } j = -2p \frac{l}{T} - 2p \frac{L}{T}$$

On peut regrouper le terme $E_0 \cos\left(2p \frac{l}{T}\right)$ en une seule amplitude. Si la distance l est un multiple impair de la longueur d'onde λ , ($l = (2n+1)l$) le \cos s'annule, et l'amplitude de l'onde résultante est nulle; on parle d'interférence destructive. Les ondes se soustraient et s'annulent. En revanche si la distance l est un multiple pair de la longueur d'onde $l = (2n)l$, le \cos est égale à 1, et l'amplitude de l'onde résultante est maximale et égale à E_0 ; on parle d'interférence constructive. Les ondes s'additionnent. Evidemment toutes les phases intermédiaires sont possibles.

Si on est capable de faire varier la distance L , on est capable de contrôler la quantité de lumière transmise. C'est ce principe qui est utilisé pour réaliser des modulateurs optiques dans les systèmes de télécommunication modernes. Cependant le contrôle de la phase ne se fait pas par le contrôle d'une distance L , mais par un autre moyen qui nous oblige à nous intéresser à la propagation des signaux non plus dans le vide mais dans la matière. Nous reviendrons plus en détail sur les modulateurs dans un chapitre ultérieur.

LE CHEMIN OPTIQUE , PROPAGATION DE LA LUMIERE DANS LA MATIERE :

Les ondes que nous avons vu jusqu'à présent se propageaient dans le vide. Mais bien sur en télécoms optiques on veut aussi les faire se propager dans d'autres milieux comme le verre, pour les fibres optiques, le silicium, l'arséniure de Gallium et d'autres semi-conducteurs pour les leds, lasers, photodiodes, et amplificateur optiques. Qu'est ce qui change pour ces matériaux ? Souvenez vous des équations de Maxwell, dans un matériau la susceptibilité électrique ϵ change et il en est de même pour la permittivité μ . On remplace donc ϵ et μ :

$$\epsilon_0 \rightarrow \epsilon_{\text{matériau}} = \epsilon_r \epsilon_0$$

$$\mu \rightarrow \mu_{\text{matériau}} = \mu_r \mu_0$$

L'ensemble des matériaux vérifient $\epsilon_{matériau} > \epsilon_0$ et $\mu_{matériau} > \mu_0$ d'où

$$C = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \geq \frac{1}{\sqrt{\epsilon_{matériau} \mu_{matériau}}} = C_{matériau} . \text{ La vitesse de la lumière dans la matière est}$$

donc toujours plus faibles que dans le vide, le rapport entre les vitesses étant :

$$n = \frac{C}{C_{matériau}}$$

On appelle n , l'indice du matériau. Pour le verre $n=1,5$, pour le silicium et l'arséniure de Gallium $n \sim 3$ à $3,5$.

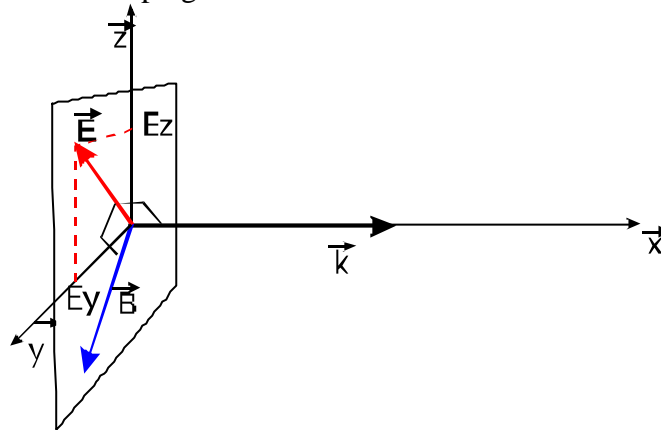
Une onde qui se propage dans un matériau d'indice n , est donc caractérisée par une vitesse C/n . Le déphasage associé est :

$$\Delta j = \frac{\omega}{C/n} \Delta x = 2\pi n \frac{n \Delta x}{C} = 2\pi \frac{n \Delta x}{\lambda} \quad \text{En tout ce passe comme si l'onde parcourait}$$

une distance non pas Δx mais $n\Delta x$. Par opposition au chemin géométrique Δx , on appelle chemin optique la grandeur $n\Delta x$.

POLARISATION D'UNE ONDE :

La structure d'une onde progressive est la suivante :



Le vecteur d'onde \vec{k} indique la direction de propagation de l'onde. \vec{E} et \vec{B} sont situés dans un plan auquel \vec{k} est normal. Comme \vec{B} est toujours perpendiculaire à \vec{E} , il suffit de s'intéresser à \vec{E} et \vec{k} pour connaître l'onde complètement. Dans la suite du cours on se concentrera uniquement sur \vec{E} et ce sera à vous de ne pas oublier que \vec{B} est là. Pour la direction de \vec{E} , on parle de direction de polarisation.

Le vecteur \vec{E} s'écrit :
$$\vec{E} = \begin{pmatrix} 0 \\ E_y = E_0 \cos(\omega t - j_1) \\ E_z = E_0 \cos(\omega t - j_2) \end{pmatrix}$$

Si $E_z=0$ on dit que l'onde est polarisée suivant l'axe Oy. On précise polarisation rectiligne.

10. OPTIQUE GEOMETRIQUE.

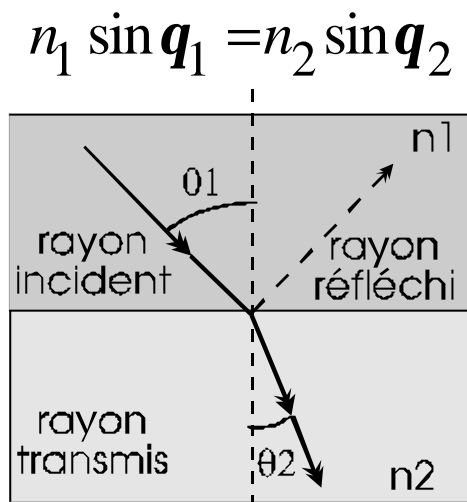
Loi de Descartes:

Dans un matériau 1, la vitesse de propagation de la lumière est V_1 . On définit l'indice n_1 :

$$n_1 = \frac{c}{V_1}$$

Considérons que ce matériau 1 est situé au dessus d'un autre matériau 2 d'indice n_2 . Que se passe t'il lorsqu'un rayon passe à travers ces 2 matériaux ?

Il existe une loi qui définit l'angle de sortie en fonction de l'angle d'entrée dans le matériau. C'est la loi de DESCARTES : Une onde incidente ϵ_1 dans un matériau 1 arrivant avec un angle θ_1 par rapport à la normale à la surface, peut être transmise au matériau 2 avec un angle θ_2 .



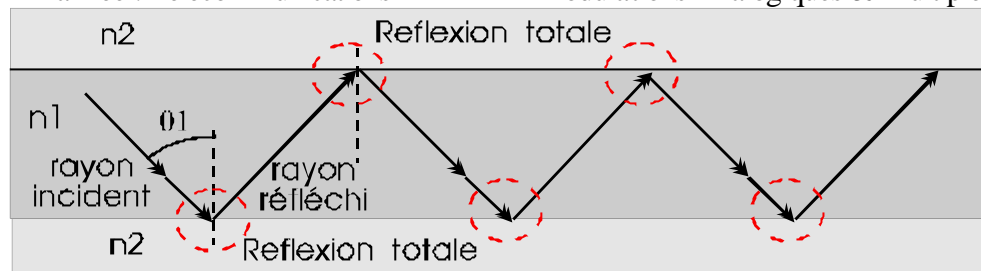
Remarque : La loi de Descartes peut aussi s'écrire :

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1$$

Si $n_1 > n_2$ alors $n_1/n_2 > 1$ et il est possible de trouver des angles θ_1 que $\frac{n_1}{n_2} \sin \theta_1 > 1$.

Or $\sin \theta_2 \leq 1$ par définition. Cette impossibilité mathématique traduit le fait qu'il n'y a pas de rayon transmis tout est réfléchi. On parle alors de réflexion totale.

Quel est l'intérêt d'un tel phénomène ? Le principe de la réflexion totale peut être appliqué pour réaliser des éléments qui guide la lumière. Il suffit pour cela de placer un matériau d'indice n_1 entre deux matériaux d'indice $n_2 < n_1$:



C'est exactement le principe des fibres optiques. Les deux interfaces agissent un peu comme des miroirs l'un en face de l'autre. L'objet ainsi réalisé s'appelle un guide d'onde

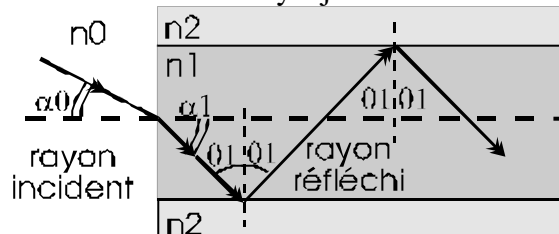
Angle limite :

L'angle limite θ_{lim} est l'angle θ_1 à partir duquel il y a réflexion totale. Si $\theta_1 > \theta_{lim}$ il y a réflexion totale mais si $\theta_1 < \theta_{lim}$ alors il n'y a plus réflexion totale mais transmission. D'après la relation ci dessus $\sin\theta_1$ est maximum lorsque $\sin\theta_2$ est maximum c'est à dire lorsque $\theta_2 = \pi/2$.

On a alors : $\sin q_{1\text{ limite}} = \frac{n_2}{n_1}$ d'où $q_{1\text{ limite}} = \text{Arcsin}\left(\frac{n_2}{n_1}\right)$.

Si la lumière arrive sur l'interface avec un angle θ_1 inférieur à $\theta_{1\text{ limite}}$ il y a transmissions dans le matériau 2. En revanche si θ_1 est supérieur à $\theta_{1\text{ limite}}$ alors il y a réflexions totales multiples et guidage. Une fois la lumière présente dans le guide elle se propage mais comment l'y introduire ?

Imaginons une fibre optique réalisée sur le principe décrit ci-dessus avec un matériau d'indice n_1 entouré de matériau d'indice n_2 . Comment y injecter de la lumière ?



Un faisceau de lumière rentre dans la fibre avec un angle α_0 par rapport à la normale à la surface. Puisque il y a un changement de milieu appliquons le principe de Descartes et calculons l'angle α_0 .

En appliquant le principe de Descartes l'angle dans le matériau d'indice n_1 est α_1 et vérifie $n_0 \sin \alpha_0 = n_1 \sin \alpha_1$ et $\alpha_1 = \frac{p}{2} - q_1$.

Pour l'angle limite on a : $\alpha_{1\text{ limite}} = \frac{p}{2} - q_{1\text{ limite}}$.

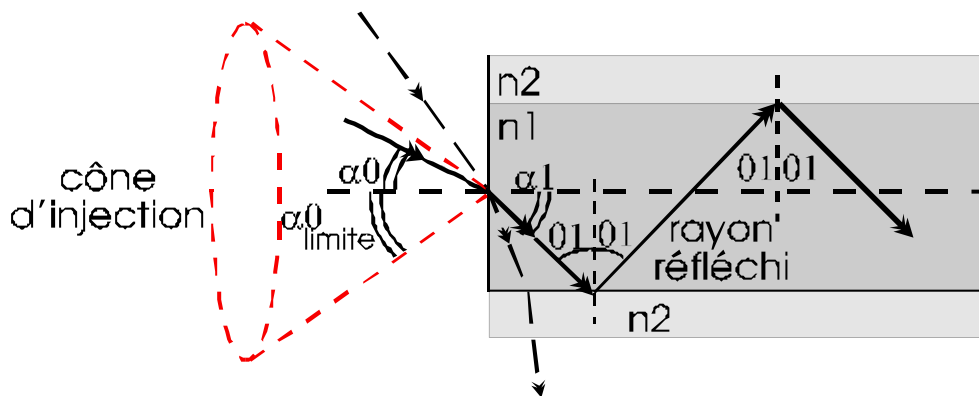
$$\sin \alpha_{0\text{ limite}} = \frac{n_1}{n_0} \sin \alpha_{1\text{ limite}} = \frac{n_1}{n_0} \sin \left(\frac{p}{2} - q_{1\text{ limite}} \right) = \frac{n_1}{n_0} \cos(q_{1\text{ limite}}) \text{ d'où}$$

$$\sin \alpha_{0\text{ limite}} = \frac{n_1}{n_0} \sqrt{1 - \sin^2(q_{1\text{ limite}})} = \frac{n_1}{n_0} \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2}$$

En première approximation n_0 l'indice de l'air est à peu près égal à l'indice du vide $n_0=1$:

$$\sin \alpha_{0 \text{ limite}} = n_1 \sqrt{1 - \left(\frac{n_2}{n_1}\right)^2} = \sqrt{n_1^2 - n_2^2}$$

Conclusion : pour injecter de la lumière dans une fibre, il faut que le faisceau lumineux arrive dans un cône d'angle $\alpha_{0 \text{ limite}}$ sinon la lumière est transmise dans le fibre avec un angle trop petit sur les surfaces séparant n_1 et n_2 et il n'y a pas réflexion totale.



Chaque fibre est caractérisé par un angle qui définit le cône d'injection on parle d'ouverture numérique.

Ouverture numérique N d'une fibre optique à saut d'indice :

$$N = n_0 \sin(\alpha_{0 \text{ limite}}) = \sqrt{n_1^2 - n_2^2}$$

L'ouverture numérique est donc égale au sinus du demi angle du cône d'acceptance. Plus N est grand (α_0 grand) plus on peut rentrer de la lumière.

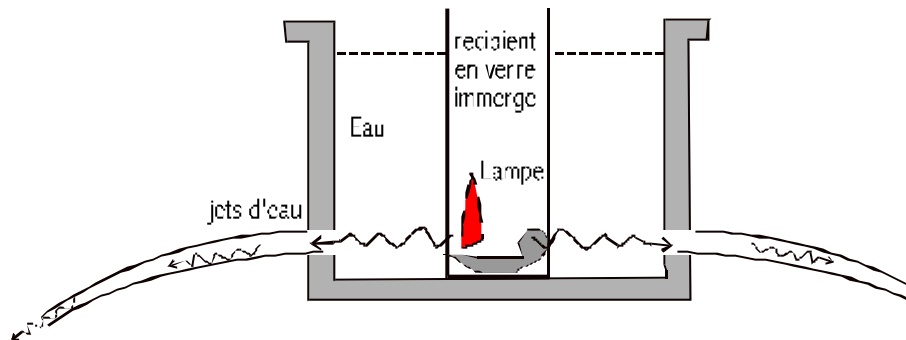
11. LES FIBRES OPTIQUES

HISTORIQUE

Les principes de guidage de la lumière sont connus depuis l'antiquité. Les romains lors de leur banquet exhibaient des jeux d'eau et de lumière : Les fontaines lumineuses : Ces fontaines consistaient en une lampe placée dans un compartiment étanche et transparent (verre) au milieu de leur réservoir. Sur les cotes du réservoir des trous d'où s'échappaient des jets d'eau, guidaient la lumière.

PRINCIPES

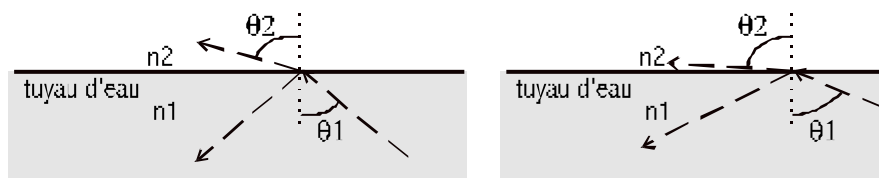
Mais comment est guidée la lumière ? Quel est le principe ? Essayons d'expliquer ce



phénomène simple, qui nous permettra de redécouvrir et de comprendre plus aisément les objets modernes guidant la lumière. Une fois que ce principe sera compris, il pourra être réappliqué afin de concevoir d'autres objets guidant la lumière.

Considérons tout d'abord un « tuyau » d'eau dans lequel se trouve de la lumière. L'eau a un indice n_1 , entourée d'air d'indice n_2 .

Que se passe-t-il lorsqu'un rayon lumineux n'est pas aligné avec la direction du tuyau, par exemple lorsqu'il fait un angle θ_1 avec la normale à la surface du tuyau ? Ce rayon



lumineux sort-il ? S'il sort comment la lumière peut-elle être guidée ?

Pour répondre à la question appliquons la loi de Descartes pour déterminer le trajet suivi par les rayons lumineux dans le tuyau :

A. Dupret, A. Fischer $n_1 \cdot \sin \theta_1 = n_2 \cdot \sin \theta_2$ 11-104

Si $n_1 > n_2$ ($n_1/n_2 > 1$) (la vitesse de la lumière dans l'eau est plus faible que dans l'air) alors $\sin \theta_2 = n_1/n_2 \cdot \sin \theta_1$ est plus grand que $\sin \theta_1$, c'est à dire que l'angle θ_2 est plus grand que l'angle θ_1 . Il peut exister une situation particulière où θ_2 est égal à $\pi/2$. Pour cette valeur de θ_2 , la lumière est totalement réfléchi et reste dans le matériau, sans pour autant que θ_1 soit parallèle à la direction du tuyau. En fait, on peut définir un angle limite $\theta_{\text{limite}} = \text{Arcsin}(n_2/n_1)$ au-delà duquel le rayon reste dans le matériau. Ce résultat est particulièrement intéressant parce qu'il signifie qu'on peut garder (guider) de la lumière dans le tuyau même si la direction de propagation de la lumière n'est pas exactement dans l'axe du tuyau. En fait tous les rayons lumineux "mal orientés", mais dont l'angle reste compris entre $\pm \theta_{\text{limite}}$ sont guidés dans le tuyau, tandis que les autres sont perdus.

Maintenant que nous avons compris ce principe nous pouvons imaginer remplacer l'eau par un plastique ou par du verre et réaliser ainsi de long barreau ou fil conduisant la lumière. En utilisant du verre nous pouvons réinventer la fibre optique. Pour cela il suffit de fabriquer un long tuyau de verre entouré d'un matériau d'indice plus faible.

LES FIBRES OPTIQUES MULTIMODES A SAUT D'INDICE

3.1 Fibre multimodes

Une fibre optique multimodes est une fibre en verre de section circulaire dont le cœur c'est-à-dire la partie centrale où se propage la lumière a un diamètre grand devant la longueur d'onde. On peut donc les étudier de façon simplifiée mais correcte par l'optique géométrique.

3.2 Gaine et Cœur

Le type le plus simple est la fibre optique à saut d'indice ou le **cœur** (la partie centrale de la fibre) d'indice de réfraction n_1 est entouré d'une **gaine** optique d'indice n_2 légèrement inférieur (Fig.3). Le diamètre du cœur est $2a=100\mu\text{m}$, pour un diamètre total (gaine) de $2b=150\mu\text{m}$. L'ensemble est entouré d'un revêtement de protection généralement en matière plastique.

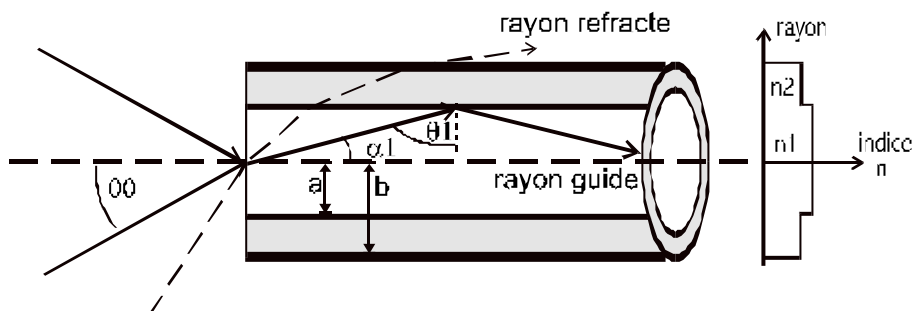


Figure 3 : Structure d'une fibre à saut d'indice

3.3 Ouverture numérique

Les indices optiques du cœur et de la gaine sont très voisins et de l'ordre de 1,45 pour les fibres en silice. La différence d'indice entre n_1 et n_2 est de l'ordre de $\delta n = 0.02$ à $\delta n = 0.01$.

Puisque nous avons défini un angle limite θ_{limite} dans la fibre en dessous duquel les faisceaux lumineux sont conservés, on peut aussi retrouver l'expression de l'angle limite θ_0 d'injection de lumière dans la fibre (Fig. 3). Cet angle vérifie :

$$\sin \theta_0 = n_1 \sin \alpha_{\text{limite}} = n_1 \sin \left(\frac{\pi}{2} - \theta_{\text{limite}} \right) = n_1 \cos \theta_{\text{limite}}$$

En remplaçant la valeur de θ_{limite} déterminée précédemment :

On définit l'ouverture numérique : $ON = n_0 \sin \theta_0$

Or par la relation de Descartes : $n_0 \sin \theta_0 = n_1 \sin \alpha_1$

En utilisant le résultat obtenu trois lignes plus haut et en remplaçant θ_{limite} par son expression il vient :

$$ON = \sin \theta_0 = n_1 \cos \left(\text{Arc sin} \left(\frac{n_2}{n_1} \right) \right) = n_1 \sqrt{1 - \left(\frac{n_2}{n_1} \right)^2} = \sqrt{n_1^2 - n_2^2}$$

L'ouverture numérique est un des paramètres les plus importants d'une fibre. Une forte ouverture numérique permet de coupler une grande quantité de lumière dans la fibre. Par contre, elle introduit également un étalement des impulsions transmises par effet de **dispersion modale**.

3.4 Dispersion modale

Comparons deux impulsions présentes à t_0 à l'extrémité d'une fibre de longueur L , et d'ouverture numérique ON . L'une des impulsions se propage suivant l'axe de symétrie de révolution de la fibre (angle nul), tandis que la deuxième arrive avec un angle égal à l'angle limite.

Ces deux impulsions vont donc avoir des trajets différents. Celle qui se propage suivant l'axe aura le trajet le plus court (longueur L , vitesse c/n_1). En revanche, l'impulsion caractérisée

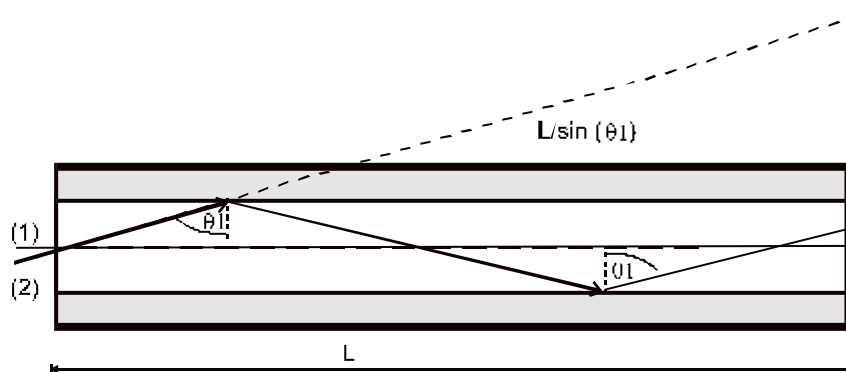


Fig. 4 Dispersion modale.

par un angle θ_{limite} , se propagera sur une longueur effective:

$$L_{\text{effective}} = \frac{L}{\sin(\theta_{\text{limite}})}$$

Le retard entre la première impulsion et la dernière est :

$$\Delta t = n_1 \frac{L_{effective} - L}{c} \quad \text{soit} \quad \Delta t = \frac{n_1}{c} L \left(\frac{1}{\sin(\theta_{limite})} - 1 \right)$$

en remplaçant θ_{limite} par son expression il vient :

$$\Delta t = \frac{L}{c} \left(\frac{n_1}{n_2} \right) (n_1 - n_2)$$

En réalité quand une impulsion est présente à l'entrée d'une fibre, tous les angles d'incidences compris entre 0 et θ_0 existent. En sortie de fibre, chaque trajet lumineux caractérisé par un angle différent subit un retard différent compris entre 0 et Δt . L'impulsion de départ se trouve donc élargie de Δt , indépendamment de la largeur de son impulsion initiale. C'est cet étalement de l'impulsion qui est désigné par dispersion modale. Chaque trajet différent étant un mode.

Les conséquences sont énormes en terme de bande passante. En effet, supposons que chaque impulsion corresponde à un bit d'un signal à transmettre. La durée entre deux bits doit être au supérieure ou égale à Δt , sinon chaque impulsion se voit de toute façon élargie de Δt au bout d'une longueur de fibre L. Le risque est alors que les bits successifs se chevauchent et créent des erreurs dans la transmission. Le débit de la ligne de transmission est donc limité à $1/\Delta t$ bits par secondes.

Le débit maximum est donc :

$$\Delta B_{dispersion\ modale} = \frac{1}{\Delta t} = \frac{c}{L} \frac{n_2}{n_1} \frac{1}{(n_1 - n_2)}$$

Il est à noter que ce débit diminue lorsque la longueur de la fibre augmente. Ce qui signifie qu'il est difficile par ce procédé de transmettre des hauts débits sur des grandes distances.

LES FIBRES OPTIQUES MULTIMODES A GRADIENT D'INDICE

C'est la raison pour laquelle une seconde génération de fibre à vue le jour; Les fibres à gradients d'indice ont été spécialement conçues pour minimiser cet effet de dispersion modale. Dans ce type de fibre, l'indice optique du cœur diminue de l'axe jusqu'à la gaine, suivant une loi parabolique comme indiqué sur la figure 5.

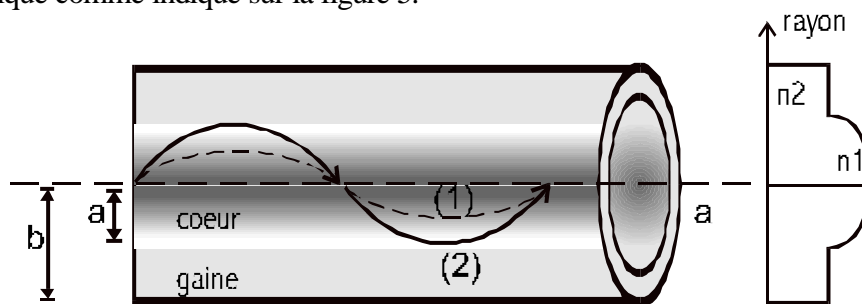


Fig.5 Fibre multimode à gradient d'indice.

L'indice du cœur à une distance r de l'axe est la loi parabolique:

$$n(r) = n_1 \sqrt{1 - 2\Delta \left(\frac{r}{a}\right)^2} \text{ avec } \Delta = \frac{n_1^2 - n_2^2}{2n_1^2} \text{ soit } \approx \frac{n_1 - n_2}{n_1}$$

Les rayons lumineux suivent des trajectoire d'allure sinusoïdale, et ceux ayant le trajet géométrique le plus long (2) passent par des milieux d'indice plus faible, donc de vitesses supérieures. Par ce procédé on égalise les temps de propagation.

LES FIBRES MONO-MODES

Ce sont des fibres dont la dimension du cœur est comprise entre 1 à 9 μ m. Elles sont conçues pour guider pratiquement sans perte la lumière à une longueur d'onde bien précise. Dans ce cas il n'y a plus de dispersion modale. Pour les autres longueurs d'ondes il est difficile d'injecter de la lumière et il n'y a presque pas de guidage. La dispersion chromatique est donc marquée.

12. LES SEMI-CONDUCTEURS ET LEURS APPLICATIONS PHOTOSENSIBLES.

INTRODUCTION :

Au cours des cinquante dernières années du dernier millénaire une nouvelle notion est apparue. Elle concerne les matériaux qui ne sont ni de bons conducteurs, ni de bons isolants : les semi-conducteurs. Leur utilisation massive en électronique et en optoélectronique en fait un objet d'étude incontournable.

Qu'est ce qu'un semi-conducteur ? Et comment l'utiliser ? C'est la question auquel le cours va tenter de répondre. Pour bien comprendre la différence entre les conducteurs et les isolants il est nécessaire de s'intéresser à la structure de la matière et notamment à l'électron.

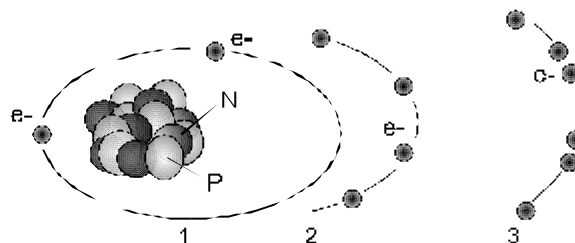


Figure 4 : Structure de l'atome. Un noyau ; N neutrons, P protons, P électrons répartis sur les orbitales. 1: première orbitale état fondamental, 2 : 2^{ème} couche d'électrons, 3 : 3^{ème} couche d'électrons...chaques couches correspond à un

STRUCTURE DE LA MATIERE

La matière est constituée d'atomes. Les atomes ont un noyau constitués de protons et de neutrons. Autour de ce noyau gravitent les électrons. Si on considère un atome avec N neutrons et P protons, il y a alors aussi P électrons afin de respecter la neutralité électrique. Les P électrons se répartissent sur les différentes orbitales autour du noyau en commençant par les orbitales les plus basses de façon que l'énergie de l'ensemble soit la plus basse possible.

NIVEAU D'ENERGIE

Il a été montré par la physique quantique que ces électrons ne peuvent occuper que des orbitales bien définies appelées états ou niveaux d'énergie. L'orbitale la plus basse obtenue pour un atome à - 273 degrés C est **l'état fondamental**. Les autres orbitales "plus hautes" correspondent à des niveaux d'énergie supérieurs, et un électron doit acquérir de l'énergie pour y accéder.

Attention : une orbitale est définie par 4 nombres quantiques (n, l, m, s). Ici nous ne parlerons que de n, donc nous pourrions utiliser l'expression couche électronique.

En général on trouve 2 électrons sur l'état fondamental 8 électrons sur la deuxième orbite et 18 électrons sur la troisième. La couche N possède $2N^2$ électrons. Si l'atome reçoit de l'énergie (température supérieure à -273 degrés) un électron peut passer sur une orbitale plus haute. Au lieu de parler de "hauteur" d'orbitale on parle d'état ou d'énergie, chaque état représentant un quantum d'énergie (énergie quantifiée). On observe que les atomes sont plus stables s'ils ont 8 électrons sur la dernière orbite. (Couche externe).

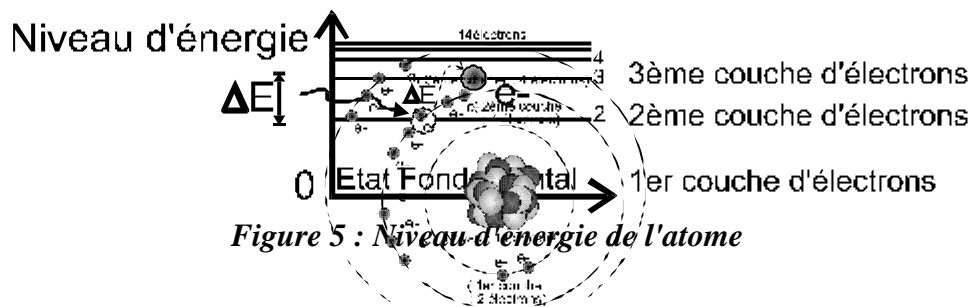


Figure 6 : Atome de silicium

Exemple du silicium.

L'atome de silicium est caractérisé par un noyau de 14 protons et donc 14 électrons.

Ils se répartissent de la manière suivante :

- 2 électrons sur la première couche.
- 8 électrons sur la deuxième couche.
- Reste 4 électrons sur la troisième couche.

Comme l'atome cherche à avoir 8 électrons sur la dernière couche, les structures dans lesquelles un atome de silicium partage 4 électrons avec 4 atomes voisins en s'organisant en cristal sont plus stables. La structure du cristal représenté dans un plan peut donc être la suivante :

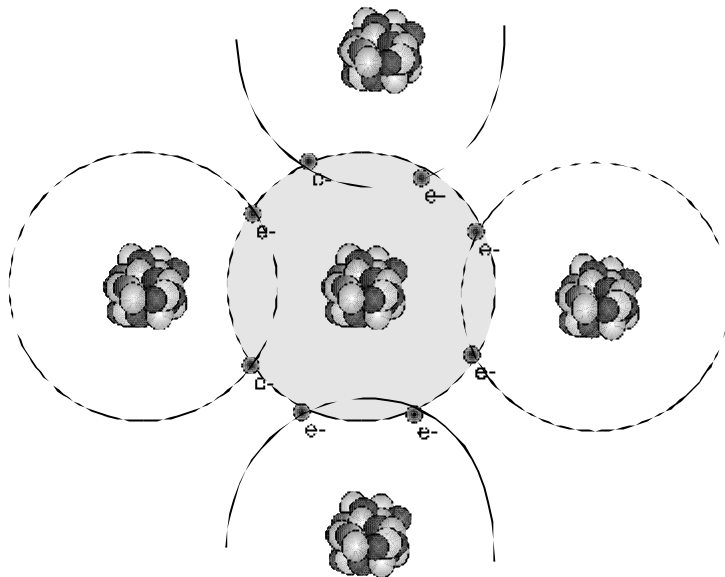


Figure 7 : Dans un cristal de silicium, un atome met en commun 4 électrons de sa dernière couche avec 4 de ses atomes voisins afin d'avoir 8 électrons sur sa dernière couche.

BANDE DE VALENCE - BANDE DE CONDUCTION

Ce type de liaison entre atomes s'appelle une liaison covalente. Les électrons qui assurent les **liaisons covalentes** sont appelés électrons de valences, ils assurent la cohésion du cristal de silicium. Ces électrons n'ont pas tous la même énergie, leurs niveaux d'énergie sont très proches les uns des autres mais non identiques. La juxtaposition de tous ces niveaux d'énergie voisins forme non plus un niveau discret unique mais une bande d'énergie dite **bande de valence**.

Bande de conduction

Lorsque les électrons sur les dernières couches acquièrent une énergie supplémentaire, ils peuvent quitter l'attraction de l'atome et passer dans d'autres atomes. Il s'agit alors d'électrons libres qui permettent le transport de charges électriques d'un atome à un autre (conduction). Le niveau d'énergie des électrons libres est donc supérieur à celui de la bande de valence. Comme pour les électrons de valence, le nombre de niveaux est tellement grand et les différences entre niveau sont telles que le niveau discret de l'atome seul s'épaissit en une

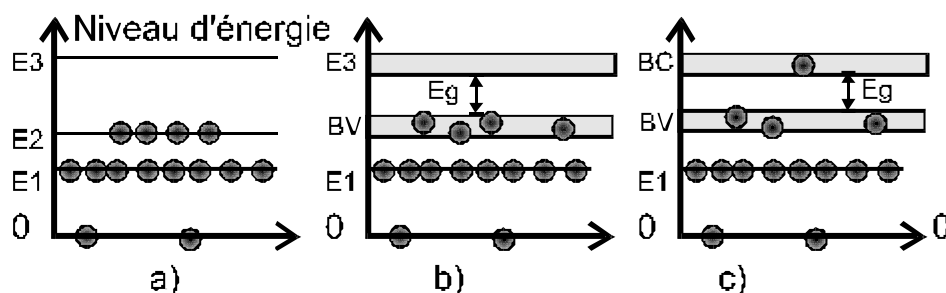


Figure 8 : niveau d'énergie : a) atome seul, b) atomes dans un cristal, pas de conduction. c) atomes dans un cristal, avec conduction. BV : bande de valence, BC bande de conduction. Eg énergie de Gap : bande interdite.

bande de conduction dans le cas de plusieurs atomes dans un cristal. Néanmoins, il reste une zone inaccessible (bande interdite) entre la bande de valence et la bande de conduction. Notons que la bande interdite est associée à la première réflexion de Bragg sur le réseau formé par les mailles du cristal.

ISOLANTS, SEMI-CONDUCTEURS, CONDUCTEURS.

Que se passe-t-il pour d'autres atomes qui ont plus d'électrons, et pour les atomes qui ont moins d'électrons ?

Si dans un cristal le nombre des électrons est tel que les bandes de d'énergie permises sont pleines [ou vides], alors ce cristal sera conducteur [ou isolant]. Si le nombre des électrons par atome est tel qu'il existe toujours des électrons dans la bande de conduction alors le cristal sera conducteur.

L'énergie du niveau rempli le plus élevé s'appelle le niveau de Fermi. Si le niveau de Fermi est au niveau de la bande de conduction, c'est qu'il y a toujours des électrons dans la bande de conduction, et donc le cristal est conducteur. Si le niveau de Fermi est dans la bande de valence, alors le cristal sera isolant ou semi-conducteur suivant la hauteur de la bande interdite. En effet, si la bande interdite est suffisamment faible ($\sim 1\text{eV}$) pour qu'un apport d'énergie (par exemple thermique) permette aux électrons d'atteindre la bande de conduction alors le cristal sera semi-conducteur. En revanche, si l'énergie de gap E_g est très importante ($\sim 10\text{eV}$), il faudra apporter beaucoup d'énergie aux électrons pour les voir accéder à la bande de conduction, il sera donc difficile d'avoir des électrons libres et le cristal sera isolant.

PHENOMENES DE TRANSPORTS, CREATION DE PAIRES ELECTRON-TROUS.

Lorsque les électrons sont dans la bande conduction, il existe plusieurs phénomènes de transports des charges. Tout d'abord l'excitation d'un électron le transformant en électron libre lui permet de se déplacer et d'atteindre d'autres atomes. L'atome qui perd ainsi un électron, n'est plus neutre électriquement et se retrouve donc chargé positivement. (Figure 10). Il existe alors deux charges électriques particulières, l'électron libre et l'atome chargé (ion)

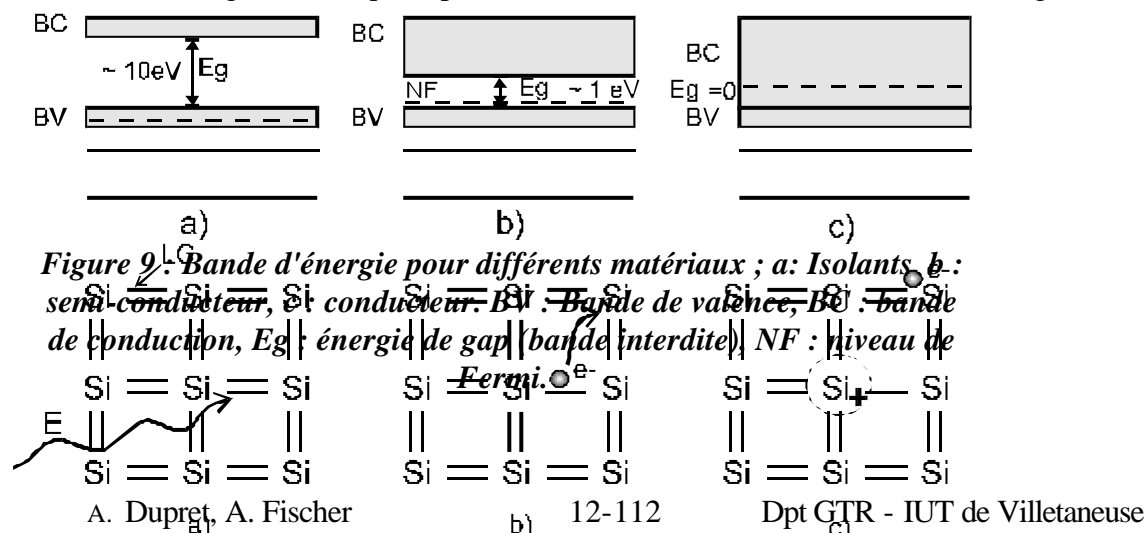


Figure 10 : création de paires d'électrons-trous : Un électron d'une liaison covalente (LC) recevant une l'énergie E supérieure à l'énergie de gap, devient libre et peut se déplacer librement dans le cristal. Il laisse l'atome avec une charge positive

positivement. On parle de **création de paire électron-trou**.

Déplacement des trous.

La charge positive ainsi créée est source d'un champ électrique qui peut attirer des électrons et récupérer un électron aux atomes voisins. L'atome voisin qui a cédé son électron se retrouve alors chargé positivement (ionisation). Tout se passe comme s'il y avait déplacement de la charge positive : on parle de déplacement de trous. Figure 11.

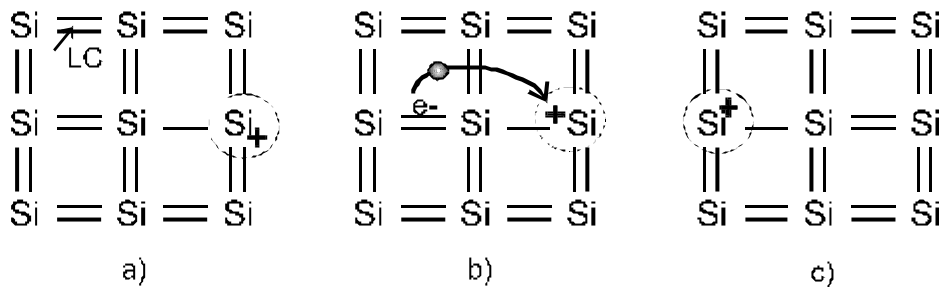


Figure 11 : déplacement d'un trou. Lorsqu'un électron se déplace de gauche à droite pour combler un trou, le trou se déplace de droite à gauche.

SEMI-CONDUCTEUR INTRINSEQUE, SEMI-CONDUCTEUR EXTRINSEQUE.

Ces transports de charges, qui existent à l'état naturel (3 atomes sur 10^{13} à 27°C), se font de manière désordonnée, et aucune direction de propagation des charges n'est favorisée, de plus la quantité de charges mises en jeu est très faible. Il est donc impossible d'utiliser ces propriétés pour aucune application.

Mais on peut tricher : puisque ce phénomène est rare à l'état naturel dans un cristal de silicium pur (semi-conducteur intrinsèque) on peut l'augmenter de la manière suivante : On rajoute dans le semi-conducteur des atomes (impuretés) qui possèdent seulement trois électrons sur leur couche périphérique (Bore - Indium). (Le semi-conducteur ainsi obtenu est dit : Semi-conducteur extrinsèque.

Il manque un électron pour réaliser les liaisons covalentes avec le silicium. Tout se passe comme si un des atomes de silicium avait cédé un électron à l'atome trivalent de bore créant un trou dans le cristal de silicium.

L'atome de bore qui capte un électron est appelé atome accepteur. Il a perdu sa neutralité pour devenir un ion négatif fixe. (Création d'un trou libre - semi-conducteurs de type P (positif)). A températures ordinaires tous les atomes accepteurs (dopants) sont ionisés. (Un atome accepteur pour 10^5 à 10^8 atomes de silicium).

En revanche, pour les semi-conducteurs de type N : On rajoute dans le cristal des atomes qui ont 5 électrons sur la couche externe, il y a alors un électron en trop pour la liaison covalente : Les atomes dopants sont des atomes d'arsenic ou phosphore (A_s , P).

Le cinquième électron est inutilisé et est faiblement lié à l'atome. L'énergie faible suffit à le rendre libre, on augmente ainsi le nombre des électrons libres.

Ainsi un semi-conducteur de **type N** possède beaucoup d'électrons libres (**e⁻ : porteurs majoritaires**) et est plus conducteur qu'un semi-conducteur intrinsèque. (Le nombre de trou est faible).

Pour un semi-conducteur de **type P**, la concentration en électrons est plus faible que pour un semi-conducteur intrinsèque, mais **la concentration en trous est majoritaire**. Il est donc difficile aux électrons de se déplacer dans un tel matériau.

Le dopage permet donc de contrôler la conductivité du cristal. Ce résultat est déjà très intéressant, mais il est encore plus intéressant de joindre des matériaux dopés différemment.

JONCTION, BARRIERE DE POTENTIEL

Imaginons maintenant que l'on "colle bout à bout" un semi-conducteur de type P avec un autre de type N. (Cf. Fig. 12) On appelle jonction P-N, un tel assemblage. Après déplacement des charges libres d'un côté et de l'autre, on obtient deux zones, l'une avec un excès de charges positives, l'autre avec un excès d'électrons libres, séparées par une barrière de potentiel. A l'équilibre il n'y a pas de migration de charges, et la jonction est bloquée.

Lorsqu'une tension est appliquée dans le sens direct à la jonction, il y a abaissement de la barrière de potentiel, les charges peuvent alors s'écouler. Dans le cas d'une tension appliquée en direct, le semi-conducteur de type P est relié à la borne plus de la tension alors que le semi-conducteur de type N est relié à la borne moins. Le type P est donc relié à une source de trou, alors que le type N est relié à une source d'électrons libres. Lorsqu'il y a recombinaison, un électron libre et un trou majoritaire sont consommés, de nouvelles charges

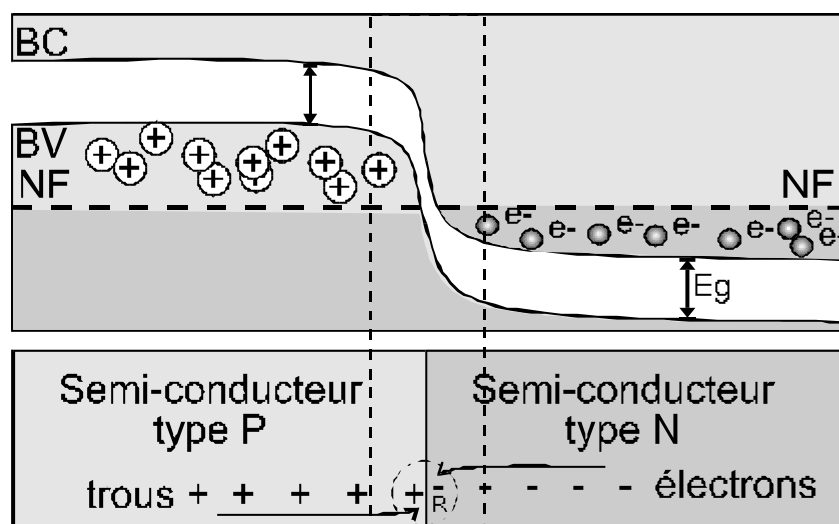


Figure 12 : Jonction PN non polarisée. Les trous de la zone P ne peuvent se recombinaison avec les électrons de la zone N.

peuvent donc quitter la source de tension pour venir se recombinaison dans la zone de charge d'espace.

En se recombinant, avec un trou majoritaire les électrons libres passent de la bande de conduction à la bande de valence en perdant de l'énergie.

EFFETS ELECTROLUMINESCENTS

Cette énergie perdue, est émise par le matériau au niveau de la zone de recombinaison

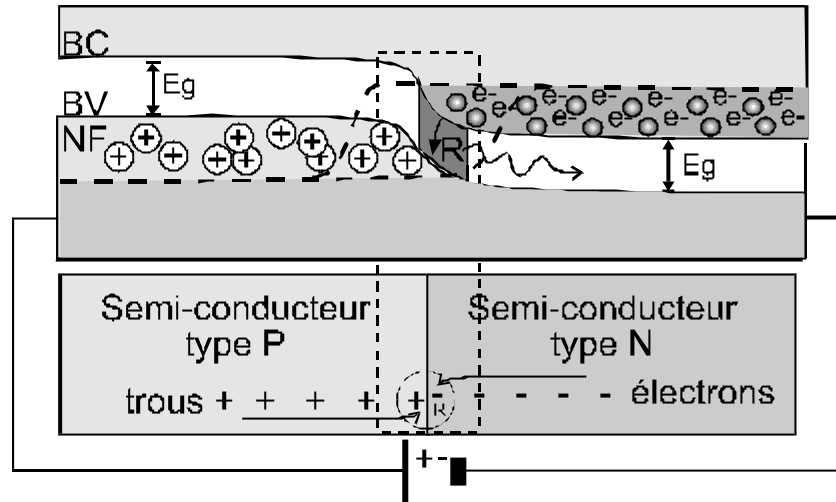


Figure 12 : Jonction en direct : les électrons libres en se recombinant passent de la bande de conduction BC à la bande de valence BV et libèrent

L'énergie ainsi émise peut avoir différentes formes (phonons (vibrations), photons (émission de lumière), thermiques). En optoélectronique il est intéressant de considérer une énergie émise sous forme de photons. Ainsi la recombinaison des charges au niveau de la zone de recombinaison produit une émission lumineuse. Lorsque des photons sont émis ils sont caractérisés par une fréquence ν , telle que :

$$E_g = h \cdot \nu$$

Cette fréquence ν définit la couleur de la lumière émise. On utilise couramment λ , la longueur d'onde d'émission définie par :

On distingue deux sortes d'émissions :

L'Emission spontanée :

$$\lambda = \frac{c}{\nu}$$

Lorsqu'un électron passe spontanément d'un niveau énergétique (par exemple la bande

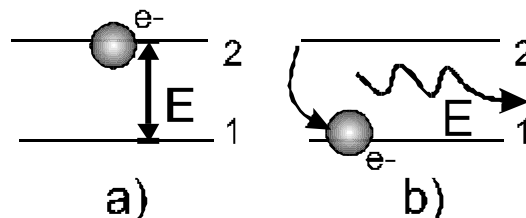


Figure 3 : Emission spontanée : a) En changeant de niveau, un électron libère de l'énergie et b) émet spontanément un photon.

de conduction) à un niveau moins énergétique (bande de valence) il libère de l'énergie et peut

émettre un photon. Cette émission est fortement influencée par les collisions entre atomes, elle présente de ce fait un caractère aléatoire marqué. C'est pourquoi elle est dite spontanée. Elle est prépondérante dans la plupart des sources lumineuses classiques : soleil, Lampes à incandescence, Lampes à décharges...

L'émission stimulée.

L'émission stimulée correspond à l'émission d'une onde de fréquence ν stimulée ou

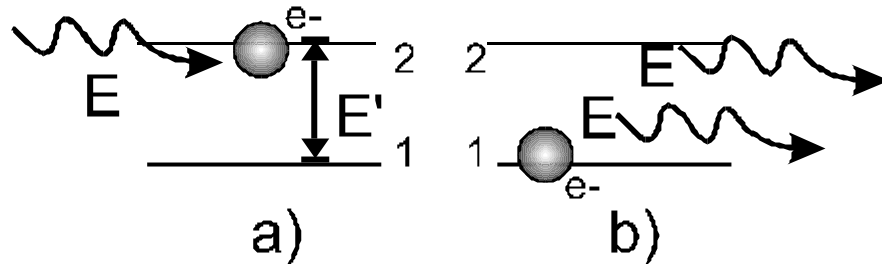


Figure 4 : émission stimulée ou induite. A) un photon incident d'énergie $E=hn$ "déclenche" le changement de niveau d'un électron, b) en changeant de niveau l'électron libère une énergie E identique à celle du photon incident, produisant ainsi un 2ème photon identique

induite par une onde incidente. Les ondes ainsi émises sont fortement reliées à l'onde incidente : L'émission induite est donc cohérente, ce qui la distingue fondamentalement de l'émission spontanée.

L'émission stimulée rend possible l'amplification optique puisque pour un photon incident, il y a deux photons en sortie. On a donc un gain optique.

Le phénomène inverse de l'émission est l'absorption :

L'Absorption : Lorsqu'un photon rencontre un atome, l'atome peut absorber l'énergie du photon, et un de ses électrons peut changer de niveau. Ce phénomène existe à l'état naturel et permet les échanges d'énergie notamment optique ou thermique.

Une jonction PN avec une énergie de gap E_g , peut donc générer des photons, bien conçue elle permet de réaliser une source lumineuse de longueur d'onde $\lambda=hc/E_g$.

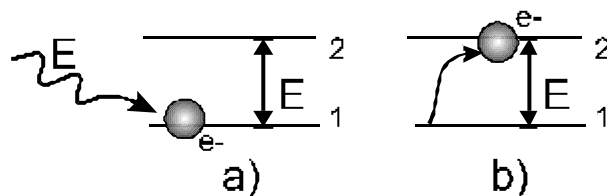


Figure 2 : Absorption : a) un électron reçoit un photon d'énergie E . b) l'électron peut passer du niveau 1 au 2 distant en énergie de E .

13. LES LEDS

Il existe différentes variétés de sources à semi-conducteurs. Le domaine qui regroupe les connaissances concernant ces composants s'appelle l'optoélectronique. Les plus connus sont les LEDs, les lasers, et les photodiodes.

La suite de ce cours sera composée des parties suivantes :

Les LED et DSL

Les diodes lasers

Les photodiodes

Les amplificateurs optiques à semi-conducteurs : (AOS)

Les plus simples d'entre eux, sont les diodes électroluminescentes ou DEL (LED en anglais), et sont l'objet de ce chapitre.

STRUCTURE DES LED

Nous venons de voir que Les jonctions PN peuvent servir de source lumineuse en profitant de l'émission spontanée. Le passage d'un courant direct dans une jonction, fait croître brusquement la concentration des porteurs près des frontières de la jonction PN. En diffusant vers les régions profondes du semi-conducteur les porteurs se recombinent en pénétrant en moyenne à une distance égale à la zone de charge d'espace de la jonction P/N. Si une partie considérable des actes de recombinaison s'effectue avec une émission de lumière, des conditions se réalisent pour la sortie de la lumière vers l'extérieur et donc la diode à semi-conducteurs peut être utilisée comme source de rayonnement lumineux. Une telle diode est appelée diode électroluminescente ou diode émettrice de lumière.

Dans les semi-conducteurs à bande directe, les recombinaisons conduisent à une émission de lumière, car les recombinaisons radiatives sont dominantes. En revanche dans les matériaux à recombinaison indirecte, l'efficacité d'émission de lumière est très faible, et la plupart des recombinaisons sont non radiatives, ce qui génère de la chaleur plutôt que de la lumière. Comme le silicium est un matériau à recombinaison indirecte, et donc que les recombinaisons radiatives sont faibles, ce matériau prépondérant en tous les domaines de l'électronique, se trouve disqualifié quand il s'agit d'émission de lumière. En optoélectronique on utilise plutôt de l'arséniure de gallium (saga).

La réalisation des LED à partir de jonction semi-conductrice pose 2 problèmes.

- 1- Pour minimiser l'absorption des photons par le matériau, il est essentiel que ces photons soient émis à proximité de la surface. Les photons émis lors des recombinaisons électrons-trous doivent l'être à proximité de la surface afin que la lumière ne soit pas reabsorbée par le matériau de la diode.
- 2- Étant donné l'indice élevé des matériaux semi-conducteurs ($n = 3,5$) une partie de la lumière n'est pas transmise, mais est réfléchi vers l'intérieur du matériau.

Lorsque que les photons traversent la surface du semi-conducteur pour être émis dans l'air, leurs angles θ_1 dans le matériau d'indice n_1 , et θ_2 dans l'air d'indice n_2 , vérifient la loi de Descartes :

$$n_1 \cdot \sin \theta_1 = n_2 \cdot \sin \theta_2$$

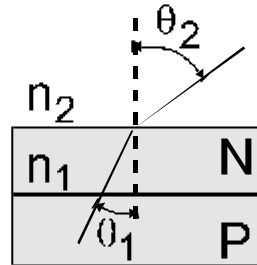


Figure1 : sortie de la lumière hors de la jonction.

L'angle de sortie est plus grand que l'angle dans le matériau. L'angle maximal de sortie θ_2 est $\pi/2$, au-delà, la lumière ne sort pas mais est réfléchi vers l'intérieur de la diode il lui correspond l'angle θ_1 :

$$\sin(\theta_1) = \frac{n_2}{n_1} \rightarrow \theta_1 = \text{Arc sin}\left(\frac{n_2}{n_1}\right)$$

Etant donné l'indice élevé $n_1 \sim 3$, l'angle θ_1 est très petit ($\theta_1 = 16^\circ$). Cela signifie que seuls les rayons dont l'angle dans le matériau est inférieur à 16° peuvent sortir. Les autres sont réfléchis. Comme les rayons sortants sont statistiquement distribués dans toutes les directions, seule une infime partie de la lumière est émise. Le reste est re-absorbé dans le semi-conducteur.

Les solutions à ces deux problèmes sont les suivants :

- 1- La zone P (surface) doit être la moins épaisse possible. Statistiquement, un photon parcourt $1\mu\text{m}$ avant d'être réabsorbé. L'épaisseur de la zone P doit donc être de l'ordre de quelques μm .
- 2- La surface de jonction doit être enrobée avec un matériau d'indice le plus grand possible afin d'augmenter l'angle maximal de sortie et de diminuer des pertes par réflexion.

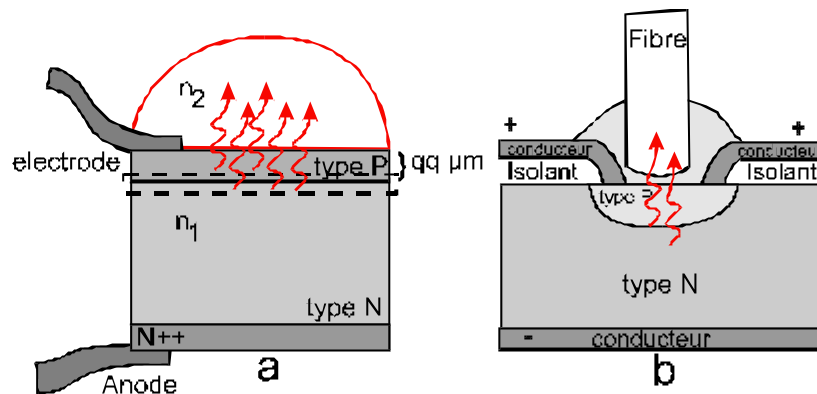
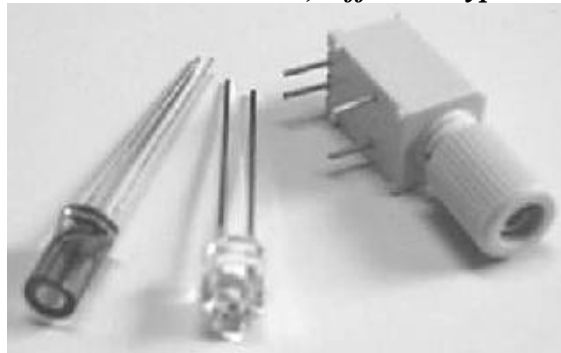


Figure 2 : Structures des LED : a/ LED encapsulée dans un milieu organique (plastique) d'indice élevé. B: LED fibrée noyée dans une résine d'indice élevé. En bas, différents types de LED.



Les structures ci dessus correspondent à des composants couramment utilisés. Néanmoins deux défauts majeurs réduisent leur efficacité :

- 1- Les défauts géométriques près de la surface font que les recombinaisons électrons-trous sont non-radiatives (pas d'émission de lumière).
- 2- Les électrons injectés dans la zone N et diffusant vers la zone P peuvent parcourir de longues distances avant de se recombiner. Les photons alors émis aussi éloignés de la surface sont majoritairement absorbés avant de l'atteindre.

HETEROJONCTION

Les LED à hétérojonction résolvent ces problèmes en modifiant la structure de la jonction afin de limiter l'étendue de la zone de contact PN.

Pour les hétérojonctions on utilise deux matériaux et deux dopants. Les matériaux sont l'Arséniure de Gallium (GaAs d'énergie de gap E_{g1}) et l'Arséniure d'Aluminium de Gallium (

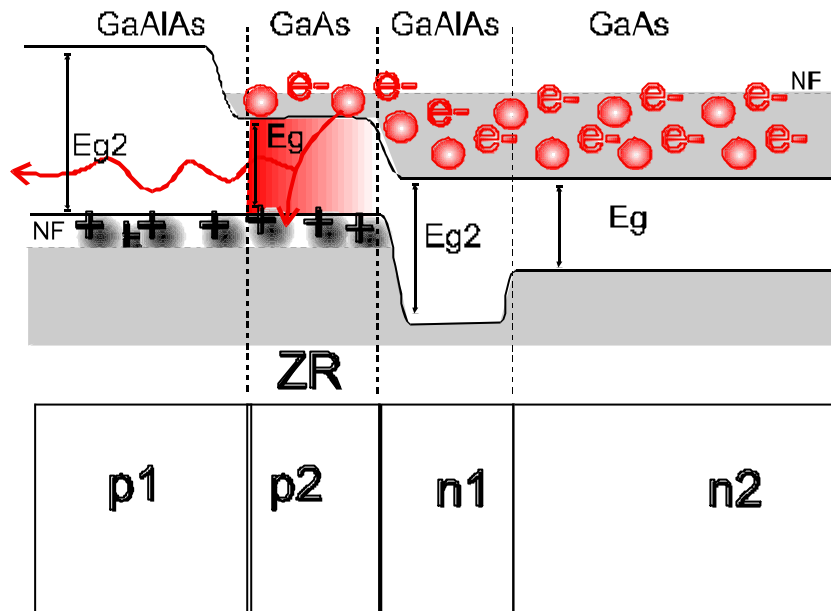


Figure 3 : Structure d'une hétérojonction.

$Ga_x Al_{1-x} As$ d'énergie de gap $E_{g2} > E_{g1}$). Les dopants différents sont (n1 GaAs, p1 GaAs) et (n2 GaAlAs et p2 GaAlAs).

Ainsi les zones p2 et n2 ont un gap trop important pour permettre l'émission à la bonne longueur d'onde, mais leur épaisseur peut être contrôlée très précisément et ainsi limité considérablement l'épaisseur de la zone de recombinaison (zone p1). On augmente ainsi considérablement l'efficacité de ces LED et donc leur puissance d'émission.

Pour mettre en œuvre ces composants il nous faut savoir, le type de liaison, la bande passante,

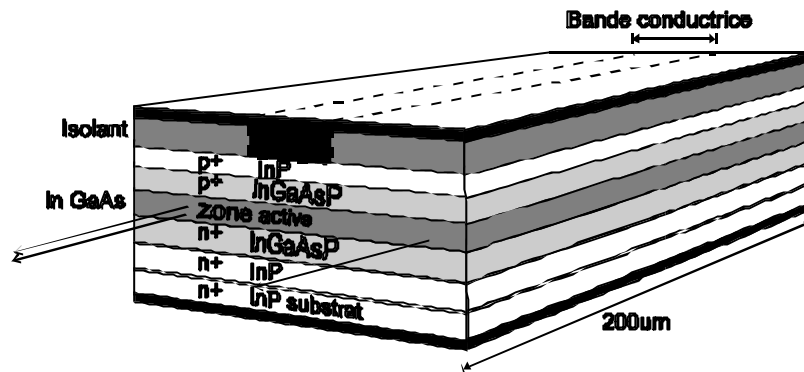


Figure 4 : Schéma d'une LED à émission par le côté. La région active est en $\text{In}_{0,47}\text{Ga}_{0,53}\text{As}$ ($E_g=0,8\text{eV}$) entre deux couches de InGaAsP ($E_g=1\text{eV}$). Les couches de confinement obligent la lumière à sortir par la tranche du semi-conducteur : effet de guidage

la vitesse de modulation, la puissance d'émission et les caractéristiques électriques de ces composants.

Dans les paragraphes suivants nous aborderons donc :

- Les caractéristiques électriques des LED
- Puissance d'émission des LED
- Temps de montée et bande passante
- Largeur spectrale des LED

Caractéristiques électriques :

En tant que diode de technologie bipolaire la caractéristique électrique d'une LED est similaire à celle de toute diode avec une relation courant-tension (I-u) logarithmique.

$$I(u) = I_o \left[\exp\left(\frac{eu}{k_b T}\right) - 1 \right]$$

avec k_b constante de Boltzman, e charge de l'électron, T température de la jonction, et I_o courant inverse.

Puissance d'émission

Si chaque électron-trou en se recombinant donnait un photon, un débit I/e d'électrons par seconde produirait un nombre $I_p = I/e$ de photons par seconde. Malheureusement seule une fraction η des recombinaisons est radiative. Le nombre par seconde de photons est donc :

$$I_P = h \frac{I}{e}$$

L'énergie de chaque photon étant :

$$E = h\nu = \frac{hw}{2\pi}$$

L'énergie émise :

$$E_P = h \cdot \frac{I}{e} \cdot h\nu$$

Puisque ce processus d'émission se fait sur une durée de temps τ_r . La puissance émise est alors :

$$P_{opt} = \frac{E_P}{t_r} = h \frac{1}{e} \cdot \frac{hw}{t_r} \cdot I$$

La puissance optique émise est donc proportionnelle au courant d'électrode.

En augmentant le courant au-delà du courant nominal on chauffe la jonction. Cet échauffement augmente les recombinaisons non-radiatives au détriment des recombinaisons radiatives et donc de la puissance optique émise. Tout se passe comme si la pente de la caractéristique diminuait. La pente s'infléchit vers une zone de saturation. En augmentant le courant d'électrode au-delà de la zone de saturation, la température peut dépasser la valeur critique T_c provoquant la destruction de la jonction.

La vitesse d'un composant est liée au temps nécessaire pour passer d'un état où la LED émet une certaine puissance, à un état où il n'y a plus d'émission. L'émission lumineuse s'arrête lorsque toutes les charges ont été consommées lors des recombinaisons radiatives. Mais les recombinaisons non-radiatives consomment également des électrons. En admettant que l'on puisse arrêter instantanément l'entrée des électrons dans la jonction, il faut encore

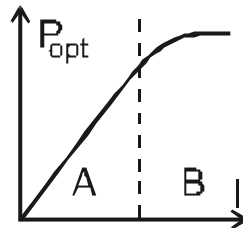


Figure 7 : Puissance optique d'une LED en fonction du courant d'électrode : après une partie linéaire (A) existe une zone de saturation (B). Au-delà de cette dernière zone il y a élévation de la température au risque de destruction de la jonction

attendre que tous les électrons à l'intérieur de la jonction se soient re-combinés et que les photons produits par recombinaisons radiatives en soient sortis pour que l'émission de lumière soit arrêtée. C'est donc principalement le temps de vie moyen des électrons se recombinant de façon radiative qui définit la réponse de la LED. Pour les LED, l'émission spontanée est prépondérante par rapport à l'émission stimulée, le temps de réponse de la diode est donc lié au temps de recombinaisons radiatives spontanées τ_{sp} des électrons. Statistiquement, ce temps de vie moyen de l'électron avant recombinaison par émission spontanée est de l'ordre de $\tau_{sp} \sim 1$ à 5 ns. Le temps de recombinaison radiative (temps statistique entre deux recombinaisons radiatives successives) est aussi lié au nombre de recombinaisons par seconde encore appelé taux de recombinaison radiative R_{sp} . Si l'intervalle de temps entre deux recombinaisons radiatives spontanées est τ_{sp} , le nombre moyen d'électrons se recombinant pendant 1 seconde, appelé taux de recombinaison radiative spontanée R_{sp} est :

$$R_{sp} = \frac{1}{\tau_{sp}} \quad \text{avec} \quad R_{sp} = \frac{e^2 n_r h \omega}{3 \pi \epsilon_0 m_0^2 c^3 h^2} |p|^2$$

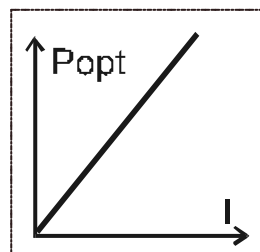


Figure 6 : Puissance optique émise en fonction du courant d'électrode

$$\text{et } \frac{2|p|^2}{m_0} \approx 22 \text{ eV}$$

n_r est l'indice de réfraction du semi-conducteur

m_0 la masse de l'électron libre

p est la matrice quantité de mouvement pour les échanges entre la bande de conduction et la bande de valence.

$$R_{sp} = 1,14 \cdot 10^{19} h \omega \quad \text{s}^{-1} \quad \text{avec } h\omega \text{ en (eV)}$$

$$t_{sp} = \frac{0,88}{h\omega(\text{eV})} \text{ en ns}$$

Ce qui définit une bande passante :

$$f = \frac{\omega}{2\pi} = \frac{1}{2\pi \cdot t_r}$$

Cette relation est valable lorsqu'il y a au moins autant de trous que d'électrons. Dans ce cas (cas le plus favorable) l'électron n'a pas à «attendre» qu'un trou soit «libre» et les recombinaisons radiatives ont lieu rapidement.

La variation d'électrons libres susceptibles de se recombiner dépend ;

- 1- Du courant d'électrode qui apporte de nouveaux électrons
- 2- Du nombre d'électrons consommés par les recombinaisons spontanées.

Un courant I apporte un nombre d'électrons de charge e égale à $\Delta N_1 = I/e$ par seconde et pour tout le volume V de la jonction. Par unité de volume et par unité de temps, le nombre d'électrons apportés est :

$$\Delta N_1 = \frac{I}{eV}$$

A chaque instant le pourcentage d'électrons libres se recombinant par émission radiative spontanée est R_p . Pour N électrons la fraction des électrons se recombinant par unité de temps et par unité de volume est :

$$\Delta N_2 = N \cdot R_{sp} = \frac{N}{t_{sp}}$$

La variation du nombre de porteurs libres est donc donnée par :

$$\frac{dN}{dt} = \Delta N_1 + \Delta N_2 = \frac{I(t)}{eV} - \frac{N}{t_{sp}}$$

On obtient donc une équation différentielle du premier ordre similaire à ce qu'on obtient avec un filtre RC. Si la LED est excitée avec un courant sinusoïdal :

$$I(t) = I_0 + I_m \cdot e^{j\omega t}$$

Une solution de l'équation différentielle est alors :

$$N(t) = N_0 + N_m \cdot e^{j\omega t}$$

avec
$$N_m(\omega) = \frac{t_{sp}}{1 + j\omega t_{sp}} \frac{I_m}{eV} \text{ et } N_0 = \frac{I_0}{eV} t_{sp}$$

L'amplitude N_m de la solution dépend de la fréquence d'excitation ω et fait apparaître une fréquence de coupure qui est la bande passante de la diode :

$$n_{limite} = \frac{1}{2p} \frac{1}{t_{sp}}$$

Expérimentalement les bandes passantes des diodes sont souvent dans la gamme 50-200MHz même s'il existe quelques composants jusqu'au GHz.

Largeur spectrale de la diode

La largeur spectrale est liée à la largeur des bandes de conduction et de valence et à l'énergie de gap. C'est l'énergie perdue par l'électron qui définit la fréquence d'émission. Cette énergie est approximativement égale à l'énergie de Gap, mais l'épaisseur de la bande

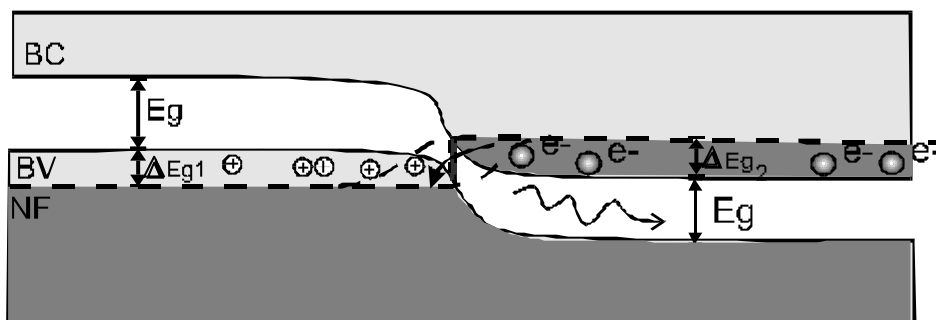


Figure 8 : La Largeur des bandes de conduction et de valence contribuent à la largeur spectrale.

occupée par les électrons libres (ΔE_{g2} cf. figure 8) ainsi que la largeur de bande à l'intérieur de laquelle les trous sont repartis interviennent. En se recombinant l'électron libère donc une énergie :

$$E_{gT} = E_g \pm \Delta E_{g1} \pm \Delta E_{g2}$$

définissant une énergie maximale et une énergie minimale :

$$E_{gTMax} = E_g + \Delta E_{g1} + \Delta E_{g2} \quad \text{et} \quad E_{gTmin} = E_g - \Delta E_{g1} - \Delta E_{g2}$$

Les variations ΔE_{g1} et ΔE_{g2} font que statistiquement lors des recombinaisons des photons de fréquences différentes sont émis. Ces photons ont une fréquence statistiquement comprise entre ν_{max} et ν_{min} :

$$n_{max} = \frac{E_{gTMax}}{h} = \frac{E_g + \Delta E_{g1} + \Delta E_{g2}}{h} \quad \text{et} \quad n_{min} = \frac{E_{gTmin}}{h} = \frac{E_g - \Delta E_{g1} - \Delta E_{g2}}{h}$$

La largeur spectrale ainsi définie $\Delta\nu = \nu_{\max} - \nu_{\min}$ est de l'ordre de quelques dizaines de manomètres (10 nm correspond à 1300 GHz à 1,55 μm – ou 1800 GHz à 1,33 μm – ou encore 4700 GHz à 0,8 μm) Les données constructeurs pour les largeurs spectrales des LED sont généralement de l'ordre de 20 à 50 nm.

Les DSL (diodes superluminescentes)

Ce sont des LED très performantes au point de vue de la puissance optique émise. En effet dans une LED optimisée incluant par exemple une double hétérostructure, le nombre de photons peut augmenter au point qu'un début d'émission stimulée soit rendu possible. En raison de l'effet de guidage produit par l'hétérojonction, la densité de photons peut être telle que le phénomène d'émission stimulée se produise. La puissance d'émission du composant est alors plus importante que pour une LED classique. Il y a alors un début d'amplification optique. Le composant agit alors comme une LED à grande irradiance. Grâce à sa structure qui favorise l'émission directionnelle de lumière (hétérostructure), il est possible d'injecter beaucoup plus de lumière dans une fibre que cela n'est possible avec une LED ordinaire. En télécom, on utilise ce type de structure appelée diode super-luminescente. DSL. Elles sont généralement fibrées.

-Les spectres optiques des DSL sont similaires aux LED bien que légèrement moins larges.

-Etant donnée la présence du phénomène d'émission stimulée qui vient se surajouter à l'émission stimulée, le taux de recombinaison des DSL (puissance optique d'émission) est supérieur à celui des LED. Par suite le temps de vie τ des électrons dans une DSL est inférieur à celui d'une LED, et les bandes passantes des DSL sont légèrement supérieures à celle des LED.

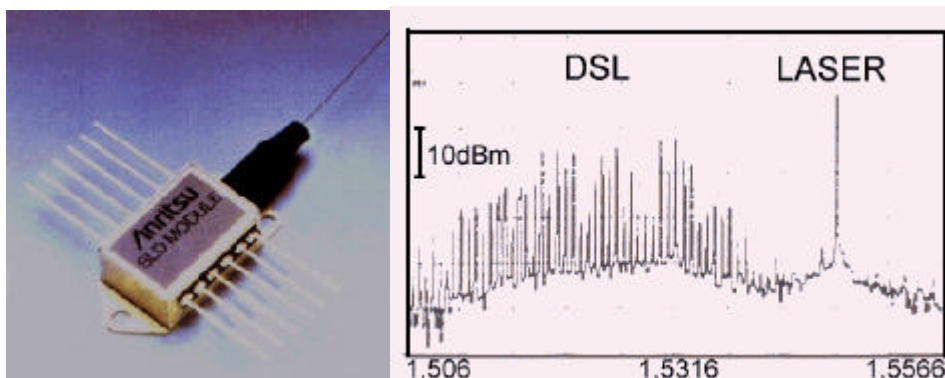


Figure 5 : à gauche : Diodes super luminescente fibrée "pigtailed SLD".

Construites, comme des LED elles ont une émission stimulée mais dépourvues de cavité optique elles ne lasent pas. A droite comparaisons

14. INTRODUCTION A LA PHYSIQUE DES LASERS

HISTORIQUE ET DEFINITION

Historique : Le mot "Laser" vient de l'acronyme anglais "Light Amplified by Stimulated Emission of Radiations ". Dès 1920 Einstein prévoit la possibilité de réaliser de l'émission stimulée première étape vers les sources lasers. Ce n'est qu'en 1960 que Mainman démontra le premier l'effet laser en utilisant un barreau de rubis.

Définition : Un laser est un oscillateur cohérent (spectre très étroit) dont la fréquence est située dans le spectre optique. Il émet donc de la lumière. On distingue deux éléments incontournables : **Un amplificateur** et un **organe de contre-réaction**.

AMPLIFICATION

Comment réaliser un milieu amplificateur pour les fréquences optiques ? Comment

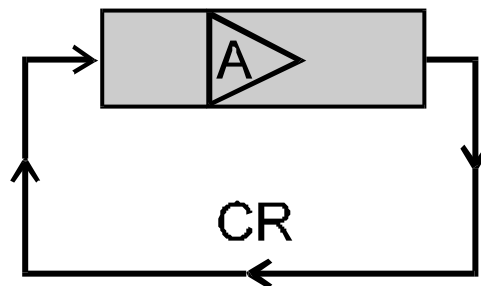


Figure1 : Un laser est un oscillateur composé d'un milieu amplificateur(A) et d'une boucle de contre-réaction(CR).

générer plusieurs photons à partir d'un seul ? La réponse est dans **l'émission stimulée**.

Inversion de population

Statistiquement à l'état naturel dans une structure épaisse (dimensions supérieures à quelques μm), un photon émis peut être absorbé par un autre électron. Globalement, à l'intérieur d'un matériau épais, il y a autant de photons émis que de photons absorbés. Pour cette raison, l'émission et l'absorption se "compensent".

Pour que l'amplification de la lumière puisse exister, il est nécessaire que, dans ce matériau, l'émission stimulée soit prépondérante et soit plus forte que l'absorption. Cela est possible si le nombre des électrons présents dans le niveau supérieur (2) est supérieur au nombre d'électrons dans le niveau inférieur (1). Cela s'appelle réaliser une «**inversion de population**». L'amplification de la lumière l'emportera sur l'absorption si la population des niveaux supérieurs est plus élevée que celle des niveaux inférieurs. Dans les conditions d'équilibre thermique, les populations des états d'énergie inférieurs sont toujours plus nombreuses que ceux des états d'énergie supérieurs, de sorte que la lumière est toujours absorbée. Pour obtenir une inversion de population, il faut faire sortir le système de son

équilibre thermique. Le milieu dans lequel on réalise l'inversion de population des niveaux s'appelle le milieu actif.

Dans un semi-conducteur l'inversion de population peut être obtenue par injection de porteurs minoritaires à travers une jonction PN dans des régions dégénérées du semi-conducteur. C'est le cas pour les diodes polarisées en direct.

La figure 2 montre une jonction PN. Le niveau de Fermi dans la région P est situé plus bas que le sommet de la bande de valence. Cela signifie qu'il y a toujours des trous dans la bande de valence dans cette région (excès de trou). De même, la position du niveau de Fermi

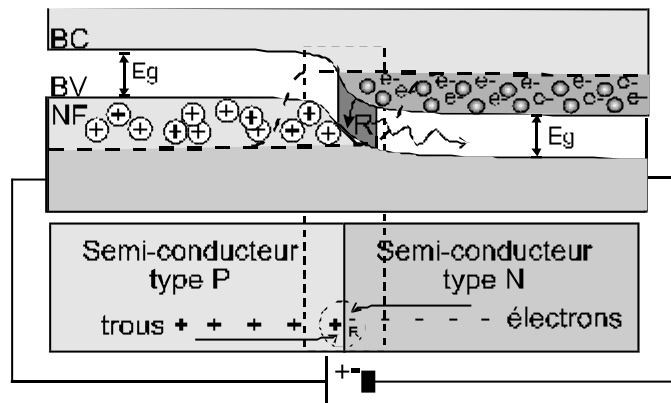


Figure 2 : Jonction PN polarisée en direct

dans la région N montre qu'il y a toujours des électrons libres dans la zone N. Si on applique à la jonction une tension V polarisée en direct qui réduit la barrière de potentiel, il apparaît dans cette jonction une inversion de population des bandes au niveau de la zone de recombinaison. Dans cette zone se trouve alors à la fois des trous et des électrons libres (niveau occupé de la bande de conduction) et les recombinaisons avec émission de lumière peuvent avoir lieu.

Ainsi, statistiquement il y a plus d'émission que d'absorption et la prépondérance de l'effet d'**émission stimulée** permet l'amplification optique.

CONTRE-REACTION OPTIQUE

L'amplification est considérablement augmentée si on impose aux ondes émises de traverser plusieurs fois le milieu actif. Alors un facteur d'amplification élevé peut compenser des pertes d'énergie dues à l'émission d'énergie vers l'extérieur et à l'absorption. En effet, la partie émise par le laser et utile pour l'utilisateur est une perte pour le laser. C'est pourquoi le milieu amplificateur est placé dans une boucle de contre-réaction. Comment est réalisée une contre-

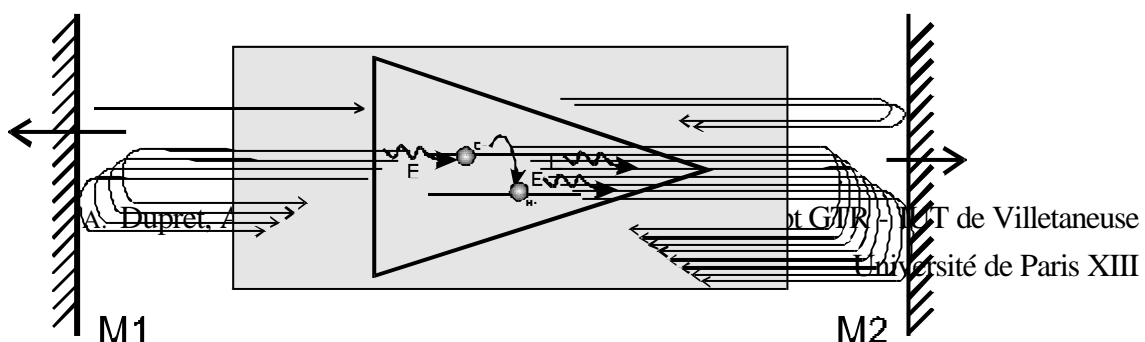


Figure 3 : Milieu amplificateur entre deux miroirs.

réaction optique (cavité optique) afin d'obtenir un résonateur optique ? Généralement il suffit de placer le milieu amplificateur entre deux miroirs.

La présence de la boucle de contre-réaction impose deux conditions au système : la **condition de gain** et la **condition de phase**.

Condition de gain :

Après N allers-retours dans le milieu amplificateur le nombre de photons augmente. Si l'augmentation se poursuit le système s'emballé. En revanche si l'amplification n'est pas suffisante pour compenser les pertes le système finit par s'amortir et s'arrêter. Pour obtenir un régime permanent il faut donc que le gain G compense exactement les pertes P.

$$\text{Gain} = \text{pertes}$$

Relation de phase

Comme pour tous les résonateurs, la boucle de contre-réaction de longueur L, introduit

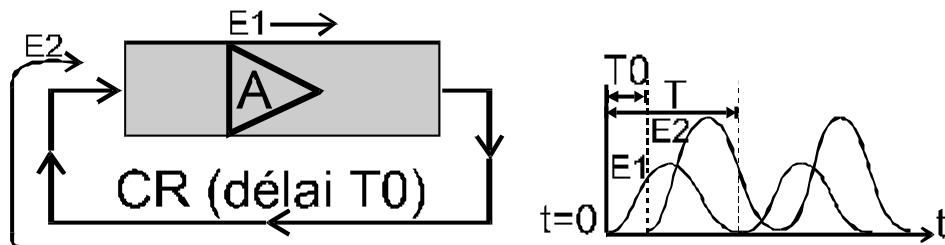


figure 4 : déphasage induit par la boucle de contre-réaction

un certain délai (temps de propagation T_0 du signal dans la boucle) ($T_0 = c.L$). Après un tour de boucle le signal de période T doit être en phase avec lui-même afin de conserver au système sa stabilité. Le déphasage du signal induit par le délai T_0 est $\Delta\phi = 2\pi.T_0/T$.

Une condition de stabilité est donc que le déphasage soit nul ou égale à un nombre entier de période ($T_0 = k.T$). La condition de phase peut s'écrire en fonction de la longueur d'onde :

$$l = c.T = \frac{1}{k}.c.T_0 = \frac{1}{k}.L$$

$$\Delta j = 0 \Rightarrow T_0 = k.T \quad (k \text{ entier})$$

Si la cavité est caractérisée par un indice de réfraction n , la condition s'écrit : $k.l = n.L$ (la longueur de la cavité définie la fréquence d'émission)

STRUCTURE D'UN LASER A SEMI-CONDUCTEURS

Les différents phénomènes précédemment cités, à savoir, l'émission stimulée, l'inversion

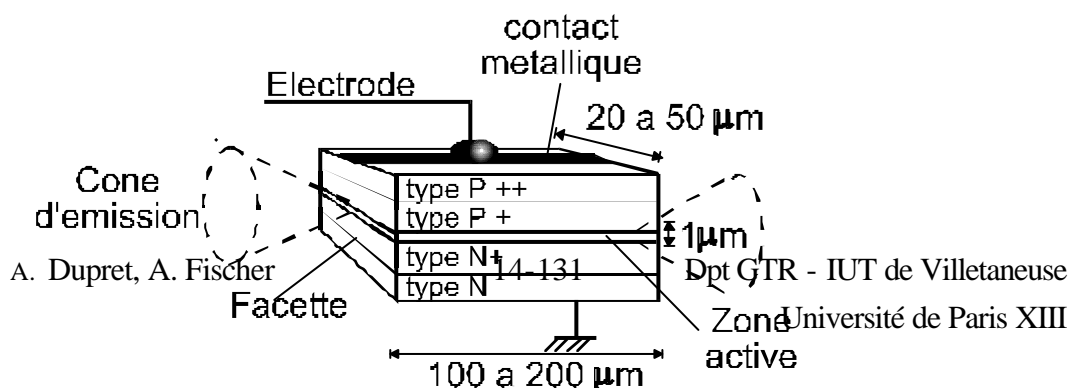


Figure 5 : structure d'un laser semi-conducteur.

de population, l'amplification optique, le guidage optique par double hétérostructure, rendent possible l'émission laser dans les structures à semi-conducteur. La structure des lasers semi-conducteurs est la suivante :

Une jonction PN à double hétérostructure, réalisée dans un matériau semi-conducteur (AsGa) est clivée à ses extrémités (cassure net suivant les plans du cristal, assurant une surface plate à la dimension des atomes près) afin de réaliser des surfaces réfléchissantes aux extrémités de la zone active définissant ainsi la cavité optique. Etant donné l'indice de réfraction $n \sim 3$ à $3,5$ du matériau semi-conducteur, seulement $R=30\%$ ($R=(n-1)/(n+1)$) de la lumière est réfléchi par les facettes vers l'intérieur de la structure. La mauvaise qualité optique de ce type de résonateur optique est compensée par le haut taux de conversion électron-photon.

Au sommet de la structure est placé un ruban semi-conducteur qui permet de limiter la zone active et d'augmenter la densité de courant tout en limitant le courant. L'utilisation d'une double hétérostructure permet de réduire l'épaisseur de la zone active à quelques μm . (épaisseur typique 1 à deux μm). Etant donné la faible dimension de la zone active, de même ordre de grandeur que la longueur d'onde, un phénomène de diffraction, donne au faisceau une forme elliptique.

CARACTERISTIQUES

Les caractéristiques des lasers abordés dans les chapitres suivants sont :

- Courant-tension
- Courant-puissance optique
- Spectre –fréquence
- Temps de réponse

. Courant-tension

La caractéristique courant-tension d'une diode laser est similaire à celle d'une LED. En tant que diode de technologie bipolaire la caractéristique électrique d'une LED est similaire à celle de toute diode avec une relation courant-tension ($I-u$) logarithmique.

$$I(u) = I_o \left[\exp\left(\frac{eu}{k_b T}\right) - 1 \right]$$

avec k_b constante de Boltzman, e charge de l'électron, T température de la jonction, et I_0 courant inverse.

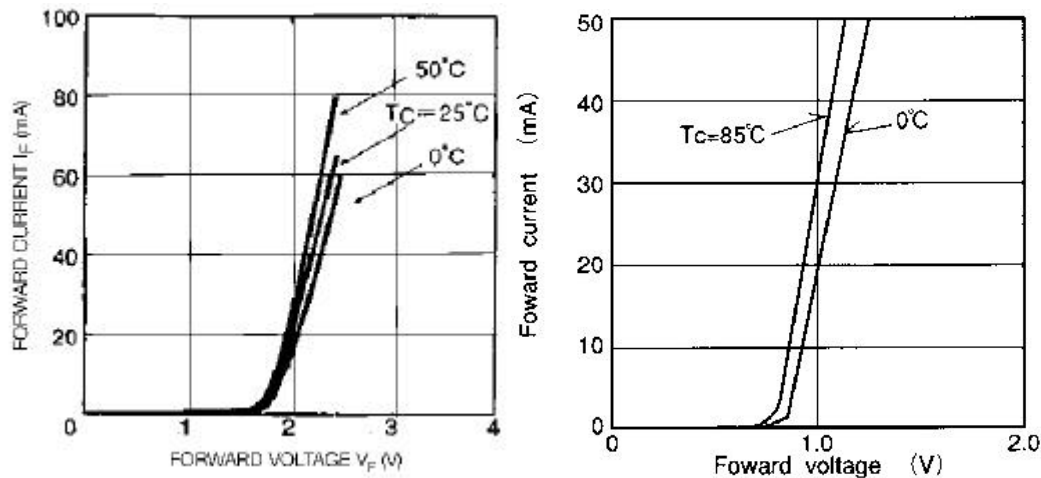


Figure 6 : Caractéristique électrique (U-I) d'un laser. A gauche laser Toshiba TOLD9225(S) tension de seuil 1,6V. A droite laser Mitsubishi ML976H11F, tension de seuil 0,8V.

Courant-puissance optique

Le mécanisme qui produit les photons à partir des charges est légèrement différent de celui d'une LED. Cela est dû à la prépondérance de l'émission stimulée. Tant que le nombre de photons n'est pas suffisant, il n'y a pas émission stimulée et le laser ne lase pas, seul un résidu de lumière spontanée est émis. C'est un effet non linéaire qui permet le démarrage du laser.

Pour calculer la puissance émise par la diode laser (DL), nous devons non seulement connaître les mécanismes sur les **électrons** mais aussi les mécanismes liés aux **photons**. Dans une troisième étape on calculera la puissance optique en fonction du courant.

Au niveau des électrons, les différents mécanismes faisant varier leur nombre sont outre le courant d'électrode, l'**émission spontanée** et l'**émission stimulée**.

Pour les photons il faut tenir compte du **gain** et des **pertes** par absorption ou par émission spontanée. Nous allons y venir dans les paragraphes suivants.

Qu'est ce qui fait varier le nombre d'électrons dans un laser ?

- Le courant apporte des électrons
 - L'émission spontanée consomme des électrons qui se recombinent pour donner des photons spontanés.
 - A la différence des LED, pour les lasers, l'émission stimulée est prépondérante, et elle consomme également des électrons qui se recombinent pour donner des photons stimulés.
- Le courant d'électrode I amène des électrons de charges e dans la zone de recombinaison de la jonction. Il y a donc

$$N_1' = \frac{I}{e}$$

charges dans le volume V de la jonction par unité de temps. Par unité de volume, il y a donc à chaque instant un apport de charges :

$$N_1 = \frac{I}{eV}$$

- Le dépeuplement dû à l'émission spontanée. Les électrons précédemment apportés par le courant d'électrode sont consommés en partie par l'émission spontanée. Si statistiquement le **taux** d'émission spontanée est $R_{sp} = 1/\tau_{sp}$ (τ_{sp} durée de vie des électrons), et que le nombre de charges est N , le nombre d'électrons se recombinant en émission spontanée par unité de temps est

$$N_2 = R_{sp} \cdot N = \frac{N}{t_{sp}}$$

- Le dépeuplement dû à l'émission stimulée.

Il est plus difficile de connaître le nombre d'électrons consommés par l'émission stimulée. En effet, parmi les électrons se recombinant par émission stimulée, certains participent à l'amplification et d'autres sont perdus par absorption ou par émission or de la structure. Etant donné le gain g de l'amplification optique, la présence de P photons dans la cavité, va déclencher l'émission stimulée de $g \cdot P$ photons. Chaque photon consommant un électron, la variation du nombre d'électrons par unité de temps est :

$$N_3 = g \cdot P$$

- Bilan électronique :

Si l'on fait le bilan pour chaque intervalle de temps dt , du nombre de charges consommées en tenant compte de **N1**, **N2**, **N3**, on obtient une variation **dN** :

$$\frac{dN}{dt} = \frac{I}{eV} - \frac{N}{t_{sp}} - gP$$

N_1 correspond à un apport de charges il est donc compté positivement, alors que N_2 et N_3 sont des diminutions de charges et sont donc comptés avec un signe moins.

Le résultat ainsi obtenu n'est pas directement explicite dans la mesure où N ne peut se calculer qu'en connaissant P et g , deux grandeurs qu'a priori nous ne connaissons pas. Nous sommes donc renvoyés à une étude similaire portant sur le nombre de photons.

Qu'est ce qui fait varier le nombre de photons dans un laser ?

A cette question, nous répondons par l'inventaire des pertes et création de photons :

- L'émission spontanée (création de photons)
- L'émission stimulée (création de photons)
- Les pertes par absorption (diminution de photons)
- Les pertes correspondant aux photons émis or de la cavité. (diminution de photons)

- Les recombinaisons de charges par émission spontanée consomment autant de charges qu'elles produisent de photons. Nous avons calculé précédemment le nombre **N2** de charges correspondantes lorsque le taux d'émission spontanée est $R_{sp}=1/t_{sp}$. Le nombre **P1** de photons produits par émission spontanée est donc :

$$P_1 = R_{sp} N = \frac{N}{t_{sp}}$$

- L'émission stimulée : Un raisonnement similaire nous conduit à assimiler la variation **P2** du nombre de photons stimulés avec **N3** :

$$P_2 = gP$$

A ce stade, nous ne connaissons toujours pas, ni le gain optique **g**, ni les pertes. Nous allons donc provisoirement interrompre notre raisonnement pour préciser ces notions.

Que signifie exactement le terme d'amplification d'une onde lumineuse ?

Lorsqu'une onde $e(t) = E.e^{j\omega t}$ d'amplitude E et de fréquence ω est amplifiée c'est son amplitude E qui augmente.

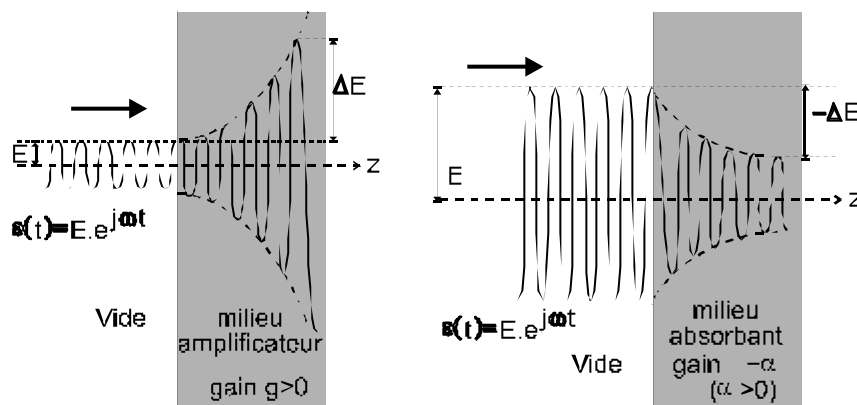


Figure 7 : modification de l'amplitude d'une onde électromagnétique suivant les milieux.

L'amplitude devient alors $E + \Delta E$ après que l'onde a progressé d'une distance Δz dans le milieu amplificateur. Si g est le gain de l'amplification optique on a :

$$\Delta E = g.E.\Delta z \quad \text{et donc} \quad \frac{dE}{dz} = gE$$

ce qui nous donne une expression de l'amplification du gain en fonction de l'épaisseur du milieu amplificateur :

$$E(z) = E_0 e^{gz}$$

L'amplitude de l'onde dépend de z , ce qui paraît naturel puisque, plus l'épaisseur du milieu amplificateur est grande, plus l'amplitude de l'onde est grande. C'est une amplification exponentielle. L'onde électromagnétique s'écrit alors : $\mathbf{e}(t) = E_0 \cdot e^{(gz)} \cdot e^{j\omega t}$.

Remarque : si au lieu d'une amplification, il y a atténuation par absorption on remplace g par $-a$ (a positif). L'amplitude de l'onde est alors amortie. $E = E_0 \exp(-az)$.

Nous savons donc maintenant comment se produit l'amplification d'une onde électromagnétique. Mais dans le cas des lasers, nous avons vu d'après la condition de gain, que cette amplification se stabilisait au niveau d'un gain g suffisant pour compenser les pertes G . Nous pouvons donc définir plus précisément le gain g en déterminant les pertes G dans la cavité d'un laser semi-conducteur de longueur L , composé d'un milieu amplificateur de gain g , d'indice n , avec une absorption a , et des réflexions r aux facettes.

Considérons une onde ϵ_1 présente en A ($z=0$) et voyageant vers B. En arrivant en B

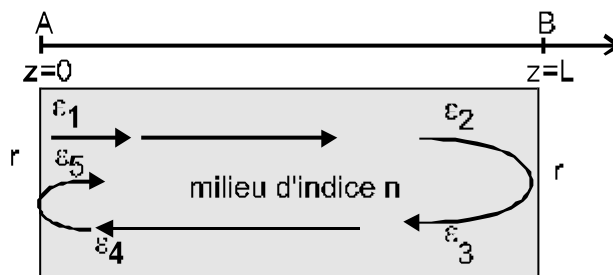


figure 8 : aller-retour d'une onde électromagnétique dans une cavité.

($z=L$), elle a été déphasée, amplifiée et absorbée, et peut s'écrire ϵ_2 :

$$\mathbf{e}_2 = \mathbf{e}_1 \cdot e^{(g-a)L} \cdot e^{j\mathbf{j}} \quad \text{avec} \quad \mathbf{j} = 2\mathbf{p} \cdot n \cdot \frac{L}{\lambda}$$

Une partie ϵ_3 de l'onde est réfléchiée par l'interface air-matériau.

$$\mathbf{e}_3 = r \cdot \mathbf{e}_2$$

L'onde se propage à nouveau dans le milieu, mais cette fois de B vers A. L'onde est à nouveau déphasée, amplifiée et absorbée pour donner ϵ_4 qui s'écrit facilement en fonction de ϵ_3 .

$$\mathbf{e}_4 = \mathbf{e}_3 \cdot e^{(g-a)L} \cdot e^{j\mathbf{j}}$$

Une partie de ϵ_4 est transmise hors du semi-conducteur, une autre partie ϵ_5 est réfléchiée vers l'intérieur du matériau.

$$\mathbf{e}_5 = r \cdot \mathbf{e}_4 = r^2 \cdot e^{2(g-a)L} \cdot \mathbf{e}_1 \cdot e^{j2\mathbf{j}}$$

On peut alors comparer l'onde après un aller-retour dans la cavité (ϵ_5), avec l'onde au départ (ϵ_1).

Si l'amplitude de ϵ_5 est supérieure à celle de ϵ_1 il y a amplification. Par contre, si l'amplitude de ϵ_5 est inférieure à celle de ϵ_1 il y a atténuation de l'onde, et le laser finira par s'arrêter. Si on imagine de multiple aller-retours, on conçoit aisément que le laser ne puisse être stable que si l'écriture complexe de l'onde ϵ_5 est égale en amplitude et en phase à ϵ_1 .

Nous avons vu au début de ce chapitre que le laser ne pouvait laser, que si les conditions de gains et de phases étaient respectées. Ces conditions sont des formes équivalentes.

La condition de gain nous impose :

$$\|e_1\| = \|e_5\| \quad \text{soit} \quad r^2 \cdot \exp(2(g - a)L) = 1$$

et nous pouvons en déduire :

$$g = a + \frac{1}{2L} \ln\left(\frac{1}{R}\right) \quad \text{avec} \quad R=r^2$$

Lorsque cette condition sur le gain est satisfaite (gain limite g_{limite}), l'instauration d'un régime d'émission stimulée est possible et le laser fonctionne. Le gain g_{limite} compense alors exactement les pertes totales a_T . N'oublions pas que c'est en augmentant le courant d'électrode du laser que l'on augmente le nombre d'électron

La condition de phase nous impose ;

$$\text{Arg}(e_1) = \text{Arg}(e_5) \quad \text{soit} \quad e^{j2j} = 1$$

$$\text{d'où} \quad 2p \left(2n \frac{L}{\lambda} \right) = 2kp \quad \text{avec } k \text{ entier}$$

Cette condition fixe la fréquence ou la longueur d'onde d'émission du laser puisque ;

$$k \cdot \lambda = 2nL$$

Au passage nous voyons que c'est la longueur optique de la cavité qui définit la fréquence du laser. En modifiant l'indice n ou la longueur L , il est possible de changer la fréquence du laser.

Nous pouvons maintenant revenir au bilan du nombre de photons par unité de temps et par unité de volume. Il nous reste à définir les pertes par absorption et les pertes par émission hors de la structure.

□ Les pertes par absorption et les pertes par émission hors de la structure

Connaissant les pertes totales α_T précédemment citées, nous pouvons distinguer les pertes par absorption en lien avec le coefficient d'absorption α , et les distingués des pertes par émission

hors de la structure liés aux transmissions de l'onde électromagnétique $\alpha_t = (ln(1/r^2))/2L$. Il faut ici remarquer, que l'onde émise par le laser et utile pour nous utilisateur, est en fait une perte pour la cavité du laser. Ces pertes ont pour unités des m^{-1} , puisqu'elles correspondent à un nombre moyen de photons absorbés par unité de longueur. La vitesse des photons dans le semi-conducteur étant C/n , l'intervalle de temps pendant lequel ils parcourent 1 unité de longueur est donc $1/(C/n)$. Si pour cette distance α_T photons sont perdus, c'est qu'en moyenne un photon est perdu toutes les :

$$\frac{1}{n.c.a_T}$$

On définit alors un temps de vie moyen statistique :

$$t_{ph} = \frac{1}{n.c.a_T}$$

On peut également définir un taux de perte :

$$R_{ph} = \frac{1}{t_{ph}} = nc.a_T$$

Pour un ensemble de P photons, le nombre de photons perdus à chaque instant est donc :

$$P_3 = \frac{P}{t_{ph}}$$

Nous pouvons donc maintenant faire le bilan de la variation du nombre de photons.

□ Bilan photonique :

En tenant compte des apports de photons par émission spontanée P1 et par émission stimulée P2, ainsi que des pertes P3 la variation totale du nombre de photons par unité de temps est :

$$\frac{\Delta P}{\Delta t} = P1 + P2 - P3$$

soit

$$\frac{dP}{dt} = \frac{N}{t_{sp}} + gP - \frac{P}{t_{ph}}$$

Ce bilan fait intervenir le nombre des électrons dans la cavité. Il y a donc un couplage entre le bilan photonique, et le bilan électronique.

□ Bilan photonique et électronique :

Nous venons de voir qu'il y avait un couplage entre les deux équations, et la connaissance du comportement du système passe par la résolution des deux équations différentielles couplées ;

$$\begin{cases} \frac{dN}{dt} = \frac{I}{eV} - \frac{N}{t_{sp}} - gP \\ \frac{dP}{dt} = gP - \frac{P}{t_{ph}} + \frac{N}{t_{sp}} \end{cases}$$

Résolution en régime permanent :

En régime permanent toutes les grandeurs sont constantes dans le temps, et leurs dérivées sont nulles. Le système précédent se simplifie et devient ;

$$\begin{cases} 0 = \frac{I}{eV} - \frac{N}{t_{sp}} - gP \\ 0 = gP + \frac{N}{t_{sp}} - \frac{P}{t_{ph}} \end{cases}$$

La résolution de la deuxième équation, en faisant l'hypothèse que les recombinaisons spontanées sont en nombre négligeable devant l'émission stimulée, nous donne la condition de seuil pour obtenir une émission laser :

$$g = \frac{1}{t_{ph}} = \Gamma$$

Le gain est égal aux pertes, conditions déjà énoncé précédemment.

La résolution de la première équation, en cherchant les conditions limites qui déclenche l'émission stimulée ($P=0$) nous donne une valeur de courant de seuil I_{seuil} au-dessus duquel le laser produit des photons stimulés principalement et en dessous duquel il n'émet que des photons spontanés (comme une LED).

$$I_{seuil} = \frac{NeV}{t_{sp}}$$

Si on admet qu'au-dessus du seuil N reste constant (N_{seuil}), on constate que le nombre de photons P , donc la puissance optique émise Phv , varie linéairement avec le courant d'électrode.

$$P = \frac{t_{ph}}{eV} (I - I_{seuil}) \text{ soit une puissance optique émise}$$

$$P = hn \frac{t_{ph}}{eV} (I - I_{seuil})$$

On peut résumer le fonctionnement du laser par le schéma ci-dessous :

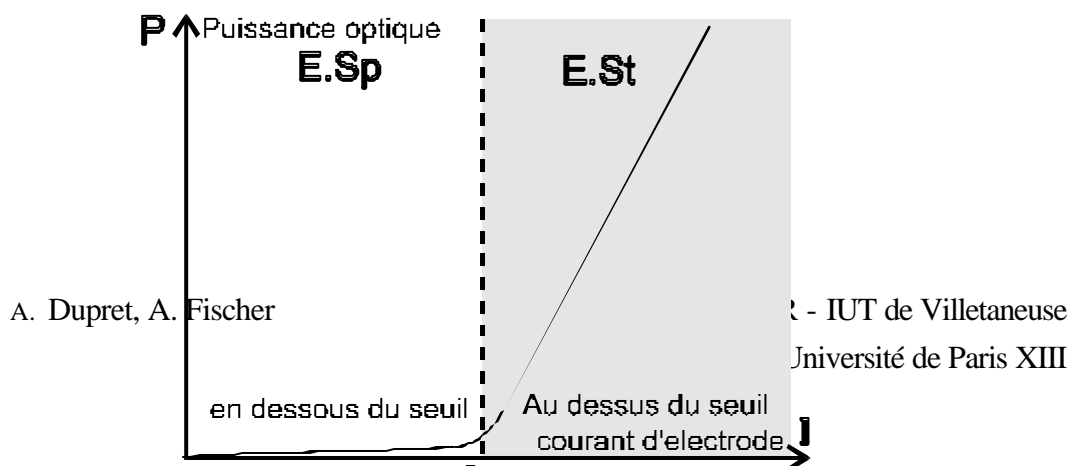


Figure 9: Caractéristique optique d'une diode laser***E.Sp* émission spontanée, *E.st* : émission stimulée**

En réalité lorsque le courant d'électrode atteint environ 200 à 300% de la valeur seuil, l'échauffement par effet joule des semi-conducteurs est tel qu'il y a modifications des propriétés du laser et un seuil de saturation amorçant la destruction de la cavité optique est visible au-delà de la partie linéaire. Les effets thermiques sont également présents au niveau du seuil, et peuvent avoir des effets destructeurs à long terme. Pour en prendre conscience, il faut comparer la puissance thermique dissipée par effet Joule, la puissance optique, et la puissance totale fournie électriquement.

$$P_{\text{électrique}} = P_{\text{optique}} + P_{\text{thermique}}$$

Partout où la puissance optique est faible la puissance totale est dissipée par effet Joule et peut provoquer la destruction de la diode. C'est notamment le cas en dessous du seuil ; La puissance optique est nulle et toute l'énergie électrique est transformée en chaleur. C'est la raison pour laquelle les DL ne sont jamais modulées au voisinage du seuil. Nous reviendrons sur ces aspects dans un chapitre ultérieur.

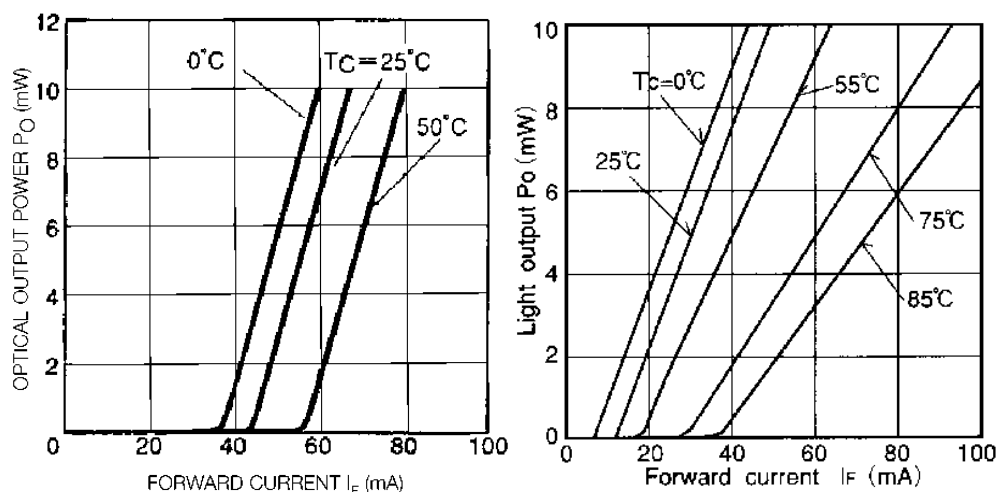


Figure 10 : Puissance optique d'une diode laser en fonction de son courant d'électrode. A gauche DL Toshiba TOLD9225 ($I_{\text{seuil}}=42\text{mA}$, $I_{\text{max}}=65\text{mA}$ à 25° , à droite DL Mitsubishi ML9116HF11. $I_{\text{seuil}}=15\text{mA}$, $I_{\text{max}}=50\text{mA}$ à 25° .

Les fonctions de guidage de l'onde lumineuse introduite par la double hétérostructure permet d'obtenir des rendements de puissance injectée dans une fibre optique supérieurs à 50%. L'injection peut se faire par collage de la fibre et en intercalant une lentille collimatrice entre la fibre et la DL.

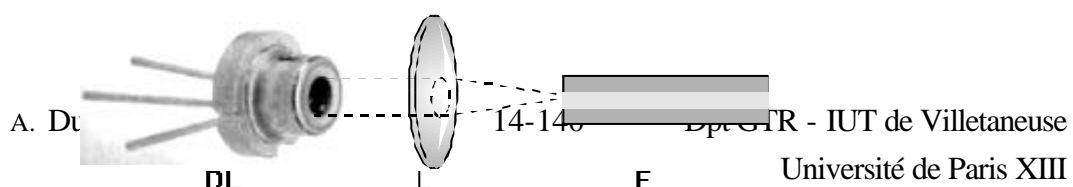


Figure 11 : Principe du collimatage d'un faisceau issu d'une diode pour couplage dans une fibre optique.

CARACTERISTIQUES FREQUENCIALES D'UN LASER SEMI-CONDUCTEUR.

Le but de ce chapitre est de déterminer d'une part la forme spectrale de la diode laser (spectre optique) et d'inventorier les différentes constantes de temps intervenant dans les mécanismes de modulation des diodes lasers afin de déterminer leur bande passante.

Laser Fabry-Pérot :

Les lasers qui ont été décrits précédemment sont de type Fabry-Pérot. En effet, la cavité optique est constituée du semi-conducteur lui-même terminé par les facettes clivées qui agissent comme des miroirs. La structure est donc similaire à celle de deux miroirs face à face, appelée Fabry-Pérot. Il y a alors un effet de filtrage en fréquence bien connu. La fonction de transfert du filtre pour deux facettes de coefficient de réflexion R et distantes de L dans un milieu d'indice n , est donné par :

$$T(s) = \frac{1}{1 + m \sin^2(2pnLs)} \text{ avec } m = \frac{4R}{(1-R)^2}, s = \frac{1}{I} \text{ et } f = \frac{p}{2} \sqrt{m}$$

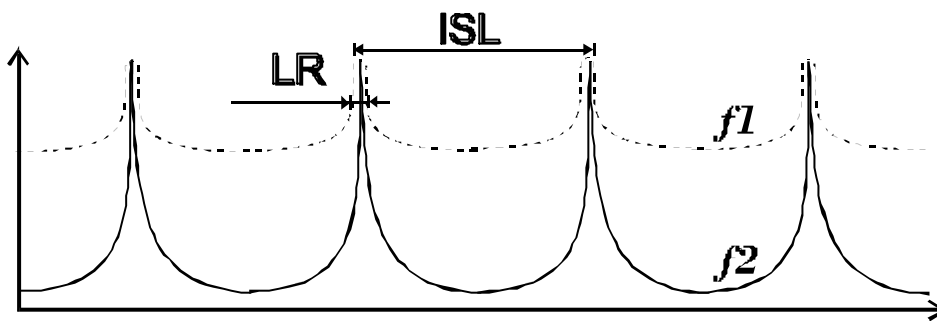


Figure 12 : Fonction de transfert d'un filtre Fabry-Pérot pour différentes finesesses $f1, f2$. ISL : interval spectral libre, LR : Largeur de raie

Cette fonction de transfert présente plusieurs pics de transmission séparés par un intervalle spectral libre (ISL ou FSR : free spectral range, $ISL = \Delta s = 1/2nL$). La largeur des pics de transmission est donnée par $ds = \Delta s / f$, c'est à dire que la largeur des pics $\delta \nu$ est f fois plus faible que l'ISL $\Delta \nu$, f étant la finesse du Fabry-Pérot fonction des coefficients de réflexion R des miroirs de la cavité ;

$$f = p \sqrt{m} = 2p \sqrt{\frac{R}{1-R}}$$

Typiquement dans un laser semi-conducteur, l'indice est $n \sim 3,5$, la longueur de la cavité est $L = 200 \mu m$, et l'indice de réflexion en énergie est $R = ((n-1)/(n+1))^2 = 0,3$, ce qui donne un intervalle spectral libre :

$$\Delta I = \Delta s \cdot I^2 = \frac{I^2}{2nL}$$

en prenant $I = \frac{c}{n}$ et en dérivant, il vient $\frac{dI}{dn} = c \frac{1}{n^2}$ d'où $\Delta n = c \frac{\Delta I}{I^2} = \frac{c}{2nL} = 200 \text{ GHz}$

et une finesse :

$$f = 10$$

Les pics de transmission espacés par un intervalle spectral libre sont appelés **Modes longitudinaux** de la cavité laser. Le laser peut donc émettre simultanément plusieurs fréquences lumineuses espacées de $\Delta n = 200 \text{ GHz}$. En réalité, le phénomène d'amplification optique n'a pas la même efficacité pour toutes les fréquences, ce qui revient à dire que le gain g dépend de la fréquence, et on peut alors définir une courbe de gain $g(n)$. Traçons sur une figure cette courbe de gain et la fonction de transfert de la cavité laser (Fabry-Pérot). Seuls les modes longitudinaux coïncidant avec le maximum de la courbe de gain, sont favorisés et participent à l'émission stimulée. Si le nombre de photons et d'électrons dans la cavité n'est pas suffisant pour fournir tous les modes, seuls celui ou ceux au voisinage du maximum de la courbe de gain participeront à l'émission laser.

En réalité le gain dépend du nombre de porteur et donc du courant d'électrode. Pour

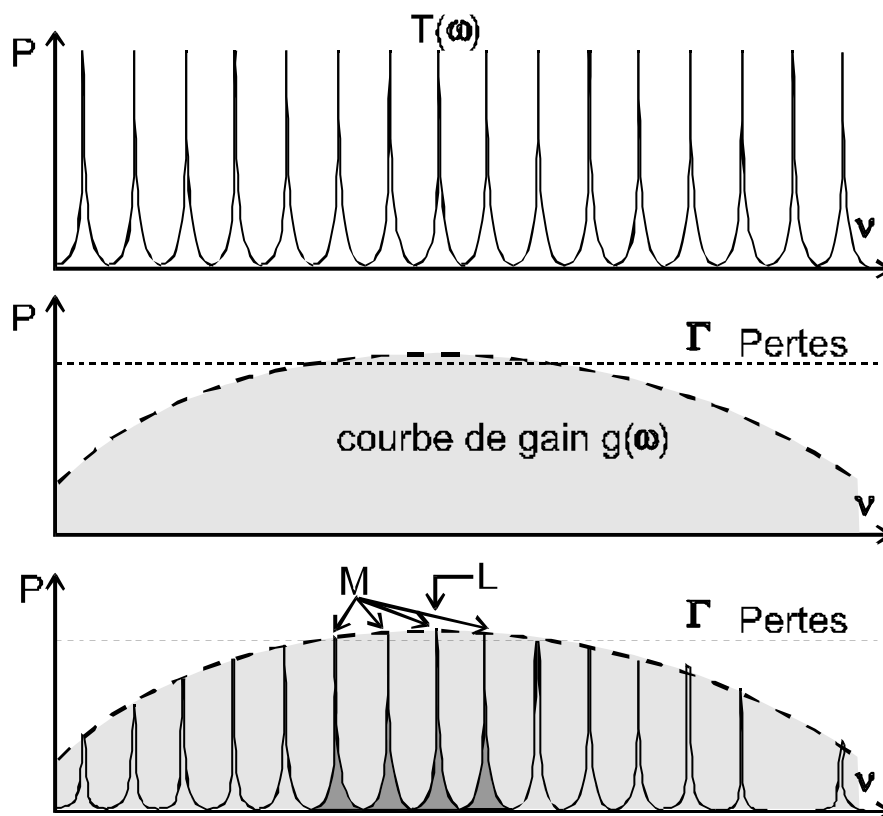


Figure 13 : sélection des modes longitudinaux par le gain. Figure en haut : fonction de transfert de la cavité laser, Au milieu courbe de gain et niveau des pertes. En bas sélection des modes de la cavité laser par la courbe de gain : seuls les modes ayant un gain supérieur aux pertes peuvent laser (M). Si la compétition des modes est suffisante, seul le mode ayant le plus grand gain (L) compose le spectre.

des gains faibles (faibles valeurs du courant d'électrodes par rapport au courant de seuil) plusieurs modes longitudinaux coexistent. Lorsque la valeur du courant augmente pour atteindre 5 à 10 fois le courant de seuil, le nombre de modes diminue et un seul mode peut subsister.

Lasers multimodes et laser monomodes

Les lasers ayant plusieurs mode longitudinaux simultanément sont dits laser multimodes.

Les lasers dont le spectre ne présente qu'un seul pic de transmission sont dits laser monomodes. Pour les télécommunications optiques et en particulier les communications longue distance les lasers utilisés sont des lasers monomodes. Nous reviendrons sur cette particularité au chapitre " fibres optiques ".

Largeur de raie

Dans le cas des cavités Fabry-Pérot passive la largeur de raie des pics de transmission est f fois plus faible que l'ISL. (Pour une cavité de longueur $L=200\mu\text{m}$ et un indice $n=3,5$, $R=0,3$, $f=10$, on a un $\text{ISL}=200\text{GHz}$ et on s'attend à une largeur de raie $\delta\nu=20\text{GHz}$). En fait pour les lasers (cavité Fabry-Pérot active en raison de l'émission stimulée) la largeur de raie est beaucoup plus faible et peut atteindre $\delta\nu=100\text{MHz}$ à 10MHz . L'émission stimulée provoque une compétition de mode, et il y a "sélection" d'une fréquence. Les photons stimulés tendent à "resserrer" la largeur de raie alors que les photons spontanés tendent à l'élargir. Cet élargissement est proportionnel au nombre $1/\tau_{sp}$ de photons spontanés, et le resserrement est proportionnel au nombre $1/\tau_{ph}$ de photon stimulé. Globalement la largeur de raie d'une cavité Fabry-Pérot active est t_{sp}/t_{ph} plus fine que celle d'une cavité Fabry-Pérot passive, soit :

$$\Delta\nu = \Delta\nu_{sp} \frac{1}{f} \frac{t_{sp}}{t_{ph}}$$

Pour un laser de longueur $200\mu\text{m}$ ($\Delta\nu=200\text{GHz}$) et d'indice $n=3,5$, $R=0,3$, $f=10$, une durée de vie des photons stimulés $\tau_{ph}\approx 5-10$ ps, et une durée de vie des photons spontanés $\tau_{sp}\approx 5-10\text{ns}$, la largeur de raie est alors ; $\delta\nu=20\text{MHz}$. Ce qui est l'ordre de grandeur moyen pour une largeur de raie.

Nous avons vu dans un chapitre précédent que la fréquence d'émission est donnée par la relation :

$$kI = 2nL \quad k \text{ étant un entier}$$

Remarquons au passage que l'intervalle spectral entre deux solutions correspond à l'ISL définit plus haut.

La longueur géométrique L dépendent de la température T , puisqu'une variation de température peut provoquer une dilatation du matériau. De même l'indice n dépend de la température mais aussi de la densité de porteurs de charge N dans le matériau. Cela peut se comprendre aisément, en effet lorsque la température du matériau augmente, l'excitation thermique des électrons change également, et le nombre d'électrons dans la bande de conduction augmente. Les photons se propageant dans le matériau sont alors absorbés en plus grand nombre, s'ils sont re-émis un instant plus tard leur propagation a tout de même été ralenti. La vitesse de propagation étant modifiée, l'indice l'est également. Le même phénomène s'applique lorsque le nombre d'électrons change par modification du courant d'électrode.

Pour modéliser ce changement de température on peut écrire ;

$$I(T, I) = \frac{1}{k} 2n(T, I)L(T)$$

La variation de longueur d'onde est alors ;

$$dI(T, I) = \frac{1}{k} 2 \left[\frac{\partial n(T, I)}{\partial T} L(T) + n(T, I) \frac{\partial L(T)}{\partial T} \right] .dT + \frac{1}{k} 2 \left[\frac{\partial n(T, I)}{\partial I} L(T) \right] .dI$$

Expérimentalement on trouve une dépendance en température :

$$\left[\frac{\partial n}{\partial T} L + n \frac{\partial L}{\partial T} \right] \approx 0,1 \text{ nm}/^\circ\text{C}$$

Et une dépendance en courant :

$$\left[\frac{\partial n}{\partial I} L \right] \approx 0,05 \text{ nm}/\text{mA}$$

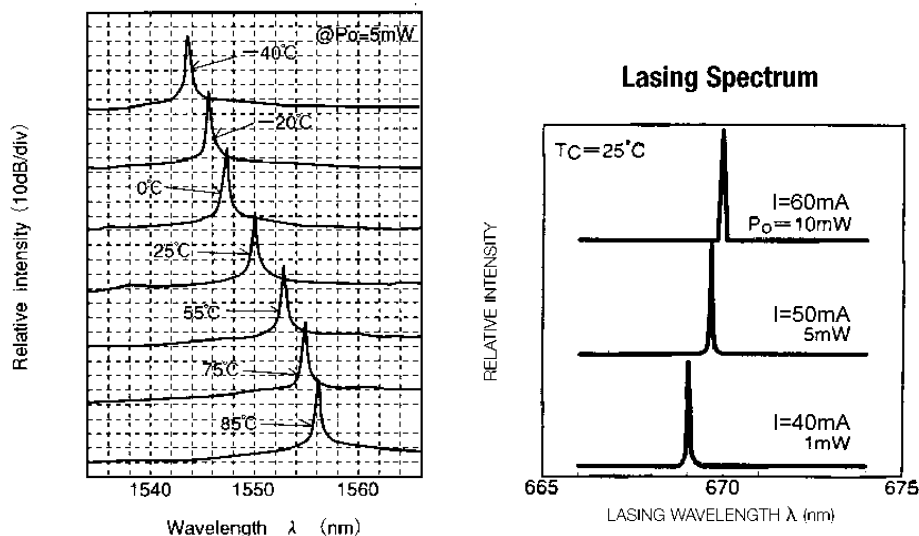


Figure 14 : A gauche, dépendance en température de la fréquence d'un laser (Mitsubishi ML9HL11), à droite dépendance en courant de la fréquence d'un laser (Toshiba TOLD9225ds)

Bande passante et vitesse de modulation

Constante de temps :

Les différentes constantes de temps intervenant dans les mécanismes des lasers sont :

Durée de vie des électrons : $\tau_{sp} \approx 5-10 \text{ ns}$

Durée de vie des photons : $\tau_{ph} \approx 5-10 \text{ ps}$

Oscillation de relaxation :

Afin de savoir comment se comporte le laser lorsqu'il est modulé par un signal électrique, il est nécessaire de reprendre les équations maîtresses et de les résoudre non plus en régime permanent, mais en régime dynamique.

$$\begin{cases} \frac{dN}{dt} = \frac{I}{eV} - \frac{N}{\tau_{sp}} - gP \\ \frac{dP}{dt} = gP - \frac{P}{\tau_{ph}} + \frac{N}{\tau_{sp}} \end{cases}$$

Si le laser est soumis à une modulation de la forme $I = I_0 + I_1 e^{j\omega t}$, on peut raisonnablement présupposer une solution pour le nombre de porteurs N et pour le nombre de photons P :

$$\begin{cases} N = N_0 + N_1 e^{j\omega t} \\ P = P_0 + P_1 e^{j\omega t} \end{cases}$$

où N_0 , P_0 sont les solutions en régime permanent déjà exposé au paragraphe 5.2.3. En

dérivant les équations ci dessus, on obtient :

$$\begin{cases} \frac{dN}{dt} = j\omega N_1 e^{j\omega t} \\ \frac{dP}{dt} = j\omega P_1 e^{j\omega t} \end{cases}$$

en remplaçant N, P, dN et dP dans les équations maîtresses il vient :

$$\begin{cases} \frac{dN}{dt} = j\omega N_1 e^{j\omega t} = \left(\frac{I}{eV} - \frac{N_0}{\tau_{sp}} - gP_0 \right) + \left[\frac{N_1 e^{j\omega t}}{\tau_{sp}} + gP_1 e^{j\omega t} \right] + I_1 e^{j\omega t} \\ \frac{dP}{dt} = j\omega P_1 e^{j\omega t} = \left(\left(g - \frac{1}{\tau_{ph}} \right) P_0 + \frac{N_0}{\tau_{sp}} \right) + \left[\left(g - \frac{1}{\tau_{ph}} \right) P_1 e^{j\omega t} + \frac{N_1 e^{j\omega t}}{\tau_{sp}} \right] \end{cases}^{(N_0, P_0)}$$

est solution en régime permanent, ce qui fait que les termes entre parenthèses peuvent être simplifiés, il reste alors les termes entre crochets :

$$\begin{cases} j\omega N_1 = - \left[\frac{N_1}{\tau_{sp}} + gP_1 \right] + I_1 \\ j\omega P_1 = \left[\left(g - \frac{1}{\tau_{ph}} \right) P_1 + \frac{N_1}{\tau_{sp}} \right] \end{cases}$$

En fonctionnement normal le laser est au-dessus du seuil, et $g \cdot \tau_{ph} = 1$. En, en tenant compte, et en remplaçant N_1 dans la première équation par son expression issue de la deuxième équation

il vient : $j\omega \left(j\omega \tau_{sp} P_1 \right) = -j\omega P_1 - \frac{1}{\tau_{ph}} P_1 + I_1$ soit

$$\frac{P_1}{I_1} = \frac{t_{sp} t_{ph}}{w^2 - jw \frac{1}{t_{sp}} - \frac{1}{t_{sp} t_{ph}}}$$

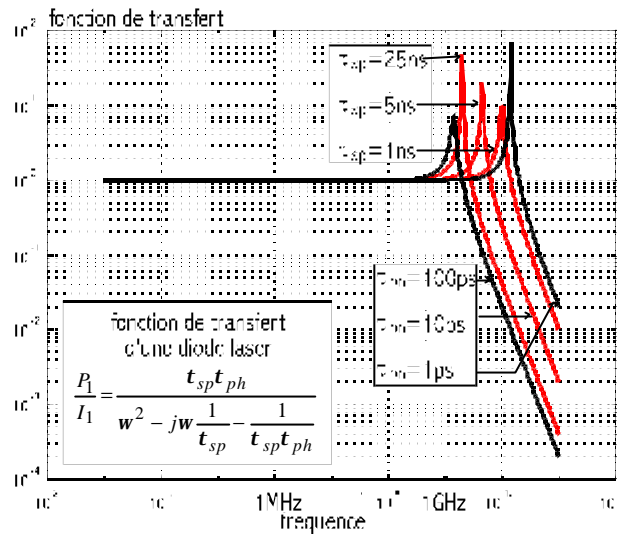


Figure 15 : Fonction de transfert typique d'une diode laser

En faisant abstraction du terme en $I_1 e^{j\omega t}$, on obtient l'équation suivante ;

$$w^2 - jw \frac{1}{t_{sp}} - \frac{1}{t_{sp} t_{ph}} = 0$$

On obtient donc une équation en w du second degré, dont les solutions indiquent que le laser a des fréquences propres d'oscillations. En résolvant

on obtient : $w_{oscillation} = \frac{1}{2} \left\{ \frac{j}{t_{sp}} \pm \sqrt{4 \frac{1}{t_{sp} t_{ph}} - \frac{1}{t_{sp}^2}} \right\}$

en posant : $b = \frac{1}{2t_{sp}}$ et $w_R = \sqrt{\frac{1}{t_{sp} t_{ph}}}$ on obtient :

$$w_{oscillation} = jb \pm \sqrt{w_R^2 - b^2}$$

La solution au système s'écrit donc $P = P_0 + P_1 e^{-bt} e^{j\Omega_{RO}t}$ ou la fréquence $\Omega_{RO} = \sqrt{w_R^2 - b^2}$ est appelée oscillation de relaxation, et β est le terme

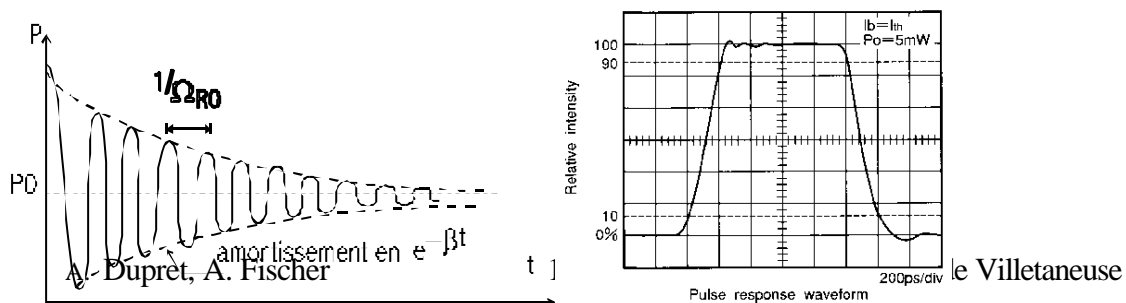


Figure 16 : A gauche : Oscillation de relaxation d'un laser avec amortissement. A droite réponse à une impulsion d'un laser mettant en évidence les oscillations de relaxation. Diode laser mitsubishi série

d'amortissement. Le laser tend naturellement à osciller à la fréquence Ω_{RO} . Le terme β amortit l'oscillation et la fait disparaître.

Une estimation de Ω_{RO} pour $\tau_{sp}=5\text{ns}$ et $\tau_{ph}=10\text{ps}$ donne $\Omega_{RO}=4,5\text{GHz}$, ce qui est une valeur couramment admise.

En régime permanent les oscillations de relaxation ne sont pas présentes, en revanche lors de transitoires (impulsions, créneaux) ce phénomène apparaît et peut parasiter le bon fonctionnement d'un laser notamment lors d'une modulation. Ce phénomène limite la bande passante des lasers bien qu'il puisse dans certains cas être mis à profit pour augmenter la bande passante de modulation.

Modulation large, Modulation petit signaux, Modulation rapide :

On distingue trois types de modulations existant :

1/Modulation large : Le signal électrique modulant varie entre deux valeurs dont l'une est en dessous du seuil, et l'autre au-dessus. Dans ce cas, un front montant permet au laser de passer d'un point de fonctionnement en dessous du seuil, à un point de fonctionnement au-dessus. Dans cette approche le laser est simplement éteint (en dessous du seuil) et allumé (au-dessus du seuil). En dessous du seuil, le mécanisme prépondérant est l'émission spontanée avec un temps de réponse τ_{sp} . Ce mode de modulation n'est donc pas rapide et le temps de réponse n'est pas meilleur que celui d'une LED. Ce type de modulation n'est pas utilisé en télécommunication, d'une part pour le faible temps de réponse (petit débit de transmission), mais aussi pour des raisons spectrales.

2/Modulation en petits signaux :

Le signal de modulation est dans ce cas la somme d'un signal continu qui polarise la diode laser au-dessus du seuil et d'un signal modulant de faible amplitude de sorte qu'à aucun moment le laser ne fonctionne au voisinage du seuil. Ce mode de modulation permet donc de franchir la limite imposée par l'émission spontanée (τ_{ph}). Cette technique fournit les plus faibles temps de réponse et les meilleures bandes passantes. (Des transmissions jusqu'à 50GHz ont ainsi été testées).

3 / Modulation rapide

Le dernier type de modulation est en fait une variante de la modulation en petits signaux avec des amplitudes de modulation supérieures et des fronts raides. Les fronts raides permettent l'excitation des fréquences de relaxation et la bande passante peut alors être étendue jusqu'à plusieurs GHz. Lors d'un front montant du signal de modulation, les oscillations de relaxation sont excitées et la réponse du laser s'accroît. Le terme d'amortissement doit être le plus faible possible afin que les oscillations s'amortissent rapidement. La vitesse de réponse du laser peut s'accroître encore par injection optique ou par contre réaction optique.

Lasers DFB et DBR – Lasers accordables

Il existe un autre moyen que les cavités de type Fabry-Pérot pour obtenir une amplification optique et une sélection en longueur d'onde. En effet les réseaux de type Bragg permettent également de réaliser des cavités lasers. Le réseau implanté au dessus de la zone active sélectionne les fréquences et son pas définit la longueur d'onde λ d'émission par la relation ;

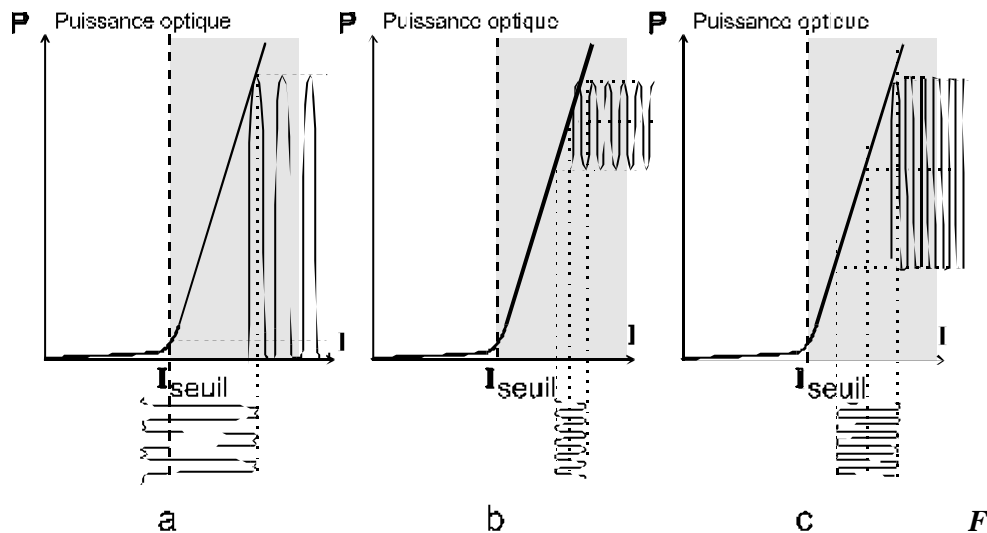


figure 17 : Les différents types de modulation d'une diode laser , a:, modulation large, b: modulation en petits signaux, , c: Modulation par codage d'impulsions.

$$\Lambda = 2n\lambda \quad \text{ou } n \text{ est l'indice et } \Lambda \text{ le pas du réseau}$$

Parmi ces structures les principales sont les lasers DBR et les lasers DFB. Tous sont basés sur l'utilisation des réseaux de Bragg. (On rappelle qu'un réseau de Bragg est une structure périodique de lignes parallèles espacées de quelques fractions de longueur d'onde).

Laser DBR

Principe : On place une structure périodique au-dessus d'une zone active :

Comme indiqué sur la figure, les réflexions sur la structure périodique peuvent remplacer

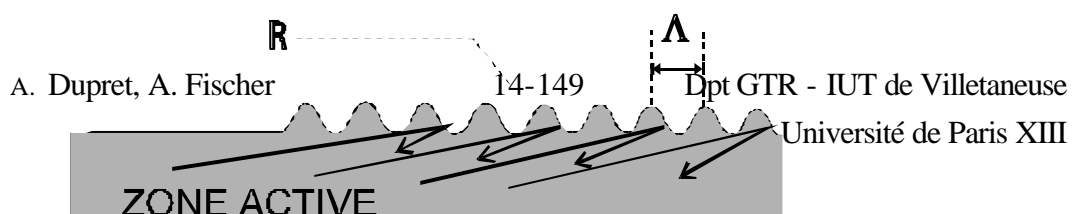


Figure 18 : principe de la réflexion rétrodiffusée

un miroir. Dans une cavité laser on peut donc remplacer les miroirs aux facettes par des réseaux.

Sur un substrat de silicium clivé ayant subi un traitement antireflet pour supprimer les réflexions aux facettes et donc, la structure Fabry-Pérot, on implante à chaque extrémités un réseau de Bragg.

On a alors un laser en trois parties : Un milieu amplificateur au centre et des réflecteurs de type Bragg aux extrémités, qui font office de miroir. Ce type de structure est appelé DBR

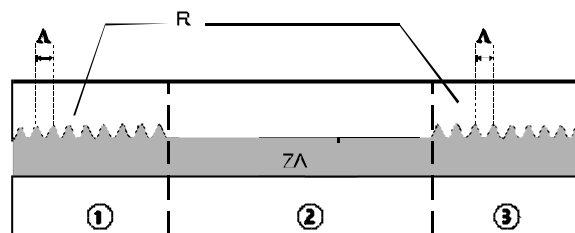


Figure 19 : principe d'un laser à réflecteurs de Bragg distribués

(Distributed Bragg Reflector). Le courant I_2 appliqué sur le milieu amplificateur contrôle la puissance d'émission.

Si des courants d'électrodes sont appliqués aux deux extrémités (zone 1 et 3), alors les indices locaux du matériau n_1 et n_3 vont varier avec ces courants I_1 et I_3 . Les courants I_1 et I_3 permettent donc de changer la densité de porteurs, donc l'indice du matériau, et donc la longueur d'onde d'émission. On a alors un moyen de contrôler la longueur d'onde du laser indépendamment de sa puissance, ce qui n'était pas le cas avec les lasers Fabry-Pérot.

Il est ainsi possible d'obtenir des lasers accordables en longueur d'onde avec des largeurs de raie très fines (quelques MHz). Les courants de seuils sont très faibles (quelques mA), et le spectre est toujours monomode.

Laser DFB

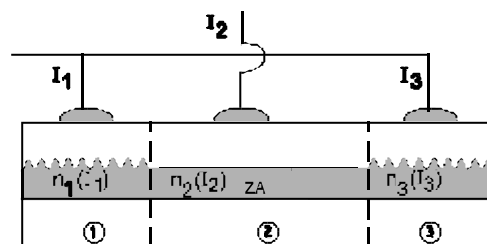


Figure 20 : Structure d'un laser DFB

La structure précédente peut être simplifiée. En effet dans les zones de contrôle en longueur d'onde il y a aussi un peu d'amplification. On peut donc répartir le réseau sur l'ensemble de la structure :

Dans ce cas, le réflecteur de Bragg est répartie non pas aux extrémités de la cavité (facettes) mais est distribué sur l'ensemble du laser. C'est pourquoi ces lasers sont appelés

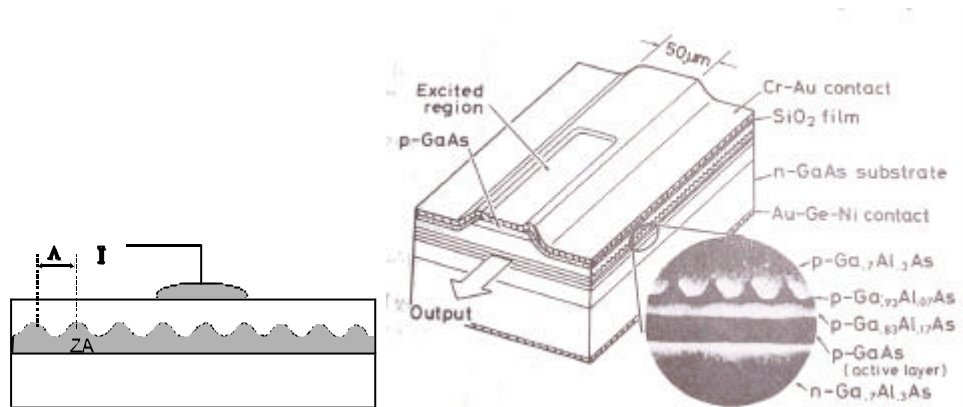


figure21 : structure d'un laser DFB

DFB.(Distributed FeedBack).

Laser DFB accordable

On peut distinguer deux zones dans la structure et les alimenter par deux courants différents :

On a alors deux indices légèrement différents $n_1(I_1)$ et $n_2(I_2)$ qui permettent de changer la longueur d'onde du laser. On obtient ainsi des lasers accordables sur 1 à 2 nm.(130-250GHz à $\lambda=1,55\mu\text{m}$)

Il existe aussi des DFB à trois électrodes plus stables, ayant une plus grande plage d'accord, et dont la puissance est constante quel que soit le courant d'électrode.

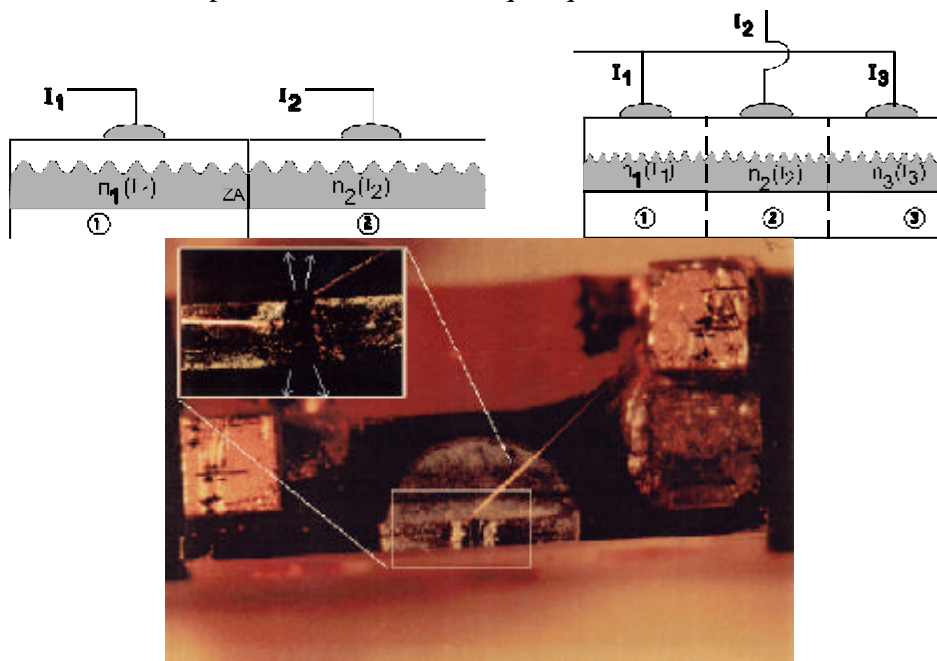


Figure 22 : Structures de lasers DFB accordables : en haut à gauche laser DFB à deux électrodes, en haut à droite laser DFB à trois électrodes. En bas photo d'un laser DFB trois électrodes. Les flèches indiquent la direction de l'émission.

PERTURBATIONS OPTIQUES DANS LES LASERS

Les perturbations optiques :

Les lasers semi-conducteurs sont extrêmement sensibles aux perturbations optiques extérieures. En effet leurs cavités optiques sont fermées par des miroirs (facettes clivées) ayant un mauvais coefficient de réflexion, ce qui facilite l'entrée de lumière venant de l'extérieur. Cette lumière extérieure injectée peut mettre à contribution l'émission spontanée pour modifier les caractéristiques du laser et notamment changer la fréquence et la puissance d'émission, voire induire des comportements dynamiques perturbatifs.

Les types de perturbation optiques sont de deux natures : La lumière injectée peut provenir d'un autre laser, ou bien il peut s'agir de la lumière émise par le laser lui-même et réfléchi par une interface optique quelconque (brisure ou raccord de fibre). Dans le premier cas on parle d'injection optique externe (optical injection), dans le second cas de réflexion (optical feedback).

Injection

Principe de l'injection optique : Un laser (esclave) est injecté lorsque le signal lumineux provenant d'un autre laser (maître) est injecté dans sa cavité optique.

Effets : Lorsqu'un laser à la longueur d'onde λ_1 est injecté par un laser maître de longueur d'onde λ_2 , il adopte la longueur d'onde d'émission λ_2 , la largeur de raie $\delta\lambda_2$ et le bruit du laser maître. Suivant la quantité de lumière injectée les oscillations de relaxation peuvent être excitées ou non.

Feedback

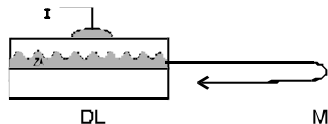


Figure 24 : Principe de la contre réaction optique (feedback)

Principe de la contre réaction optique : Lorsque la lumière émise par un laser est réfléchi par une surface quelconque et est renvoyée dans la cavité optique du laser, on parle de contre-réaction optique (feedback). La lumière réfléchi peut provenir d'un miroir, ou bien

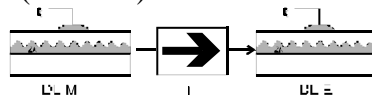


Figure 23 : Principe de l'injection optique d'un laser Maître DL M vers un laser esclave DL E. I : isolateur

de toute facette d'élément optique, (lentilles de collimation, isolateur optique, brisure dans une fibre, raccords de fibre...)

Effet : Cela a pour effet de créer une cavité optique externe (cavité Fabry-Pérot) avec ses propres caractéristiques, et elle a donc ses modes de cavité propre qui peuvent s'imposer au laser.

L'effet en régime permanent de l'injection de lumière est de provoquer un décalage en fréquence proportionnel à l'onde électromagnétique injectée.

Pour comprendre comment une perturbation optique peut modifier la fréquence d'un laser et son amplitude, il est nécessaire de s'intéresser au champ électrique complexe ϵ émis par le laser et qui contient à la fois les informations sur l'amplitude et sur la phase (donc la fréquence).

Pour modéliser le champ ϵ émis par le laser on peut se référer aux équations vues dans les chapitres précédents donnant le nombre de photons p . En effet, l'onde optique émise par les p photons n'est autre que le champ ϵ de l'onde électromagnétique.

Donc d'une manière similaire à ce qui a été fait pour modéliser p , nous pouvons dire que la variation $d\epsilon$ du champ par unité de temps est proportionnel au champ ϵ existant à l'instant considéré et au gain diminué des pertes, auquel on doit rajouter le terme définissant la fréquence de l'oscillation soit :

$$\frac{d\epsilon}{dt} = \left(G - \frac{1}{\tau_{sp}} + j\omega_0 \right) \epsilon$$

Une solution évidente est donc :

$$\epsilon(t) = E_0 e^{\left(G - \frac{1}{\tau_{sp}} \right) t} e^{j\omega_0 t}$$

On retrouve bien une onde de fréquence ω_0 et d'amplitude :

$$E_0 e^{\left(G - \frac{1}{\tau_{sp}} \right) t}$$

C'est à dire que l'amplitude s'atténue avec le temps si les pertes $1/\tau_{sp}$ sont supérieures au gain G , ou s'amplifie si le gain est supérieur aux pertes. En régime permanent le gain compense exactement les pertes, et l'amplitude de l'onde est constante.

Lorsqu'une perturbation est présente une partie de l'onde $\epsilon(t)$ revient dans la cavité après un parcours de longueur L introduisant un retard $T=2L/c$ (c vitesse de la lumière) et une atténuation k . Ce champ $\epsilon_{inj}(t)=k\epsilon(t-T)$ participe au bilan qui modifie $d\epsilon$. Le champ ϵ d'un laser avec une contre réaction optique peut donc être modélisé par une équation différentielle de la forme :

$$\frac{d\epsilon}{dt} = \left[\left(G - \frac{1}{\tau_{sp}} \right) + j\omega_0 \right] \epsilon + a k \epsilon(t-T)$$

ou a est un coefficient de proportionnalité ($a = \sqrt{1 + \alpha^2} / \tau_{ph}$) qui tient compte des pertes par absorptions $1/\tau_{ph}$ de l'onde injectée, et de la dispersion ($1+\alpha^2$ facteur d'élargissement spectral $\alpha \sim 3-5$). En séparant la partie réelle et la partie imaginaire de l'équation ci-dessus et

en considérant seulement le régime permanent, on montre que le laser change sa fréquence, et que sa nouvelle fréquence est ω vérifiant :

$$\omega_0 = \omega - \frac{E_{inj}}{E_0} \cdot \frac{\sqrt{1 + \alpha^2}}{t_{ph}} \sin(\omega T + \Delta f)$$

Où ω_0 est la fréquence du laser sans injection de lumière (par construction ω_0 est proportionnel au courant d'électrode du laser), E_0 l'amplitude de l'onde électromagnétique sans injection, E_{inj} l'onde injectée ($k = E_{inj}/E_0$), ω est la fréquence résultante, et α est le facteur d'élargissement spectral. On regroupe sous C le facteur de proportionnalité : $C = k\sqrt{1 + \alpha^2}/\tau_{ph}$.

Cette équation est représentée sur la figure 25 pour différentes valeurs de C afin d'en comprendre les implications.

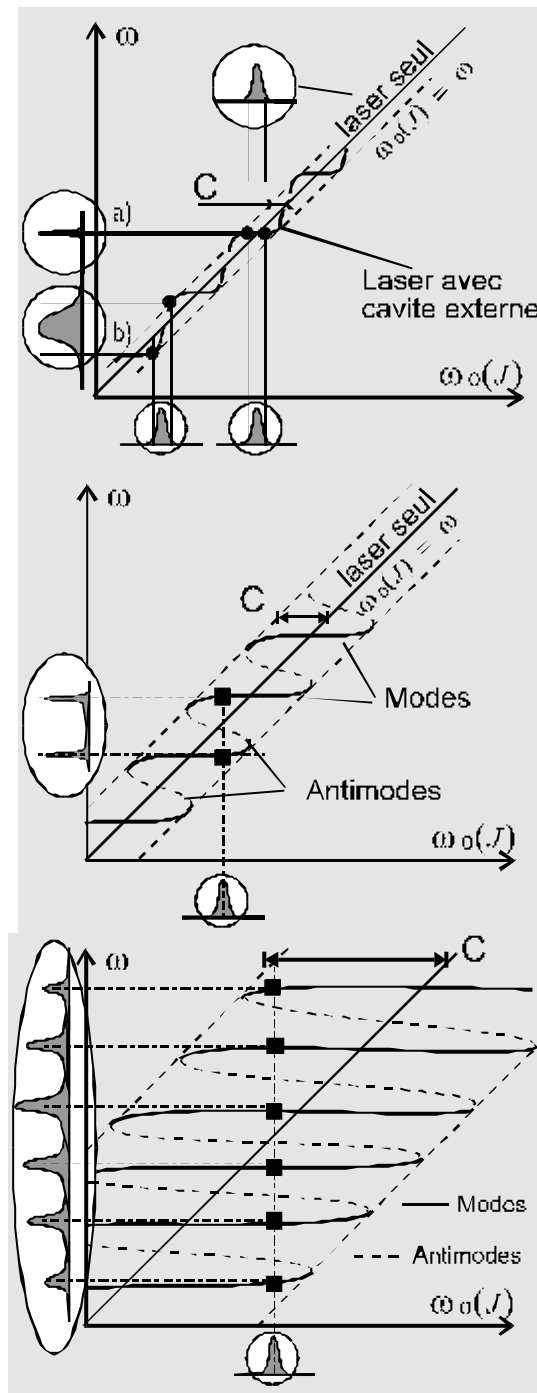
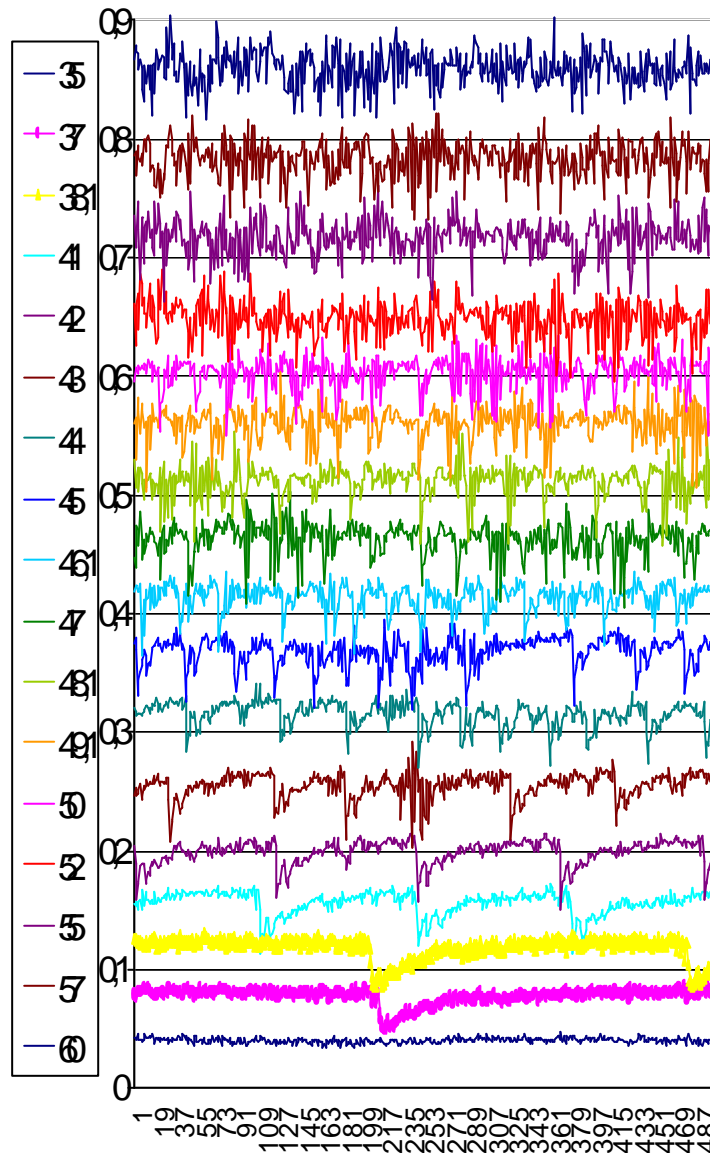


Figure 25 : Laser subissant une perturbation optique provenant d'une réflexion externe en haut pour un taux d'injection k de -45 - 40 dB $C=1$. L'effet de la pente sur un bruit présent sur le courant d'électrode avec un certain spectre peut se traduire par un élargissement du spectre du laser (pentes de la caractéristique supérieures à 1), ou au contraire un rétrécissement du spectre

LFF signifie Low Frequency Fluctuation. L'un des effets d'une contre-réaction optique est de diminuer les pertes. En effet l'onde émise par le laser hors de sa cavité est une perte pour sa structure puisqu'elle ne participe plus à l'émission stimulée. La présence d'une contre-

réacti
dimin
la cor
coura
dimin
possè
secon
dess
réfléc
l'émis
suffis
rapide
photo
brève
(5GH
mont



de cette onde
ain nécessaire à
me c'est par le
nt de seuil est
ente, un laser
seul $I_{\text{seuil-sol}}$, le
Cependant en
voir de lumière
nque provoque
n gain optique
lumineuse très
talement si les
impulsions très
laxation $\sim 20\text{ps}$
. La figure 26

Chapitre 6. LES PHOTODIODES.

Les photodiodes silicium sont des composants semi-conducteurs sensibles aux particules à hautes-énergies et aux photons. Les photodiodes absorbent des photons ou des particules chargées et génèrent dans un circuit externe, un flux de courant proportionnel à la puissance optique incidente. Les photodiodes peuvent être employées pour détecter la présence ou l'absence de quantités lumineuses très faibles et peuvent être calibrées pour des mesures extrêmement précises d'intensités en dessous du pW/cm^2 mais aussi des intensités au-dessus 100 mW/cm^2 . Les photodiodes silicium sont utilisées dans des applications aussi diverses que la spectroscopie, la photographie, l'instrumentation analytique, les détecteurs optiques de position, la caractérisation de surface, les télécommunications optiques, et l'instrumentation médicale.

PHOTODIODE DE SILICIUM EN TECHNOLOGIE PLANAIRE

Les photodiodes silicium en technologie planaire sont de simples jonctions PN. Une jonction PN peut être formée soit en diffusant des impuretés de type P (anode), tel que le Bore, dans des wafers de silicium de type N ou en diffusant des impuretés de type N, tel que le phosphore, dans des wafers de silicium dope type-p. La surface de diffusion définit la zone active de la photodiode. Pour former un contact ohmique une autre diffusion d'impureté appliquée sur le dos du wafer est nécessaire. Cette impureté est de type-N+ pour les wafers de type N et est de type P+ pour un wafer de type P. Les plots de contact sont déposés sur des zones prédéfinies de la face avant de la surface active, mais couvre complètement la face arrière du composant. Une couche antireflet est ensuite déposée sur la surface active afin de réduire les réflexions lumineuses pour des longueurs d'ondes spécifiques. Le dessus de la partie non-active de la surface est recouverte d'une couche épaisse d'oxyde de silicium. En contrôlant l'épaisseur du wafer, on peut contrôler la vitesse et la sensibilité de la photodiode. Il faut noter que les photodiodes doivent, être polarisées en inverse, c'est à dire qu'une tension négative doit être appliquée à l'anode et une tension positive sur la cathode.

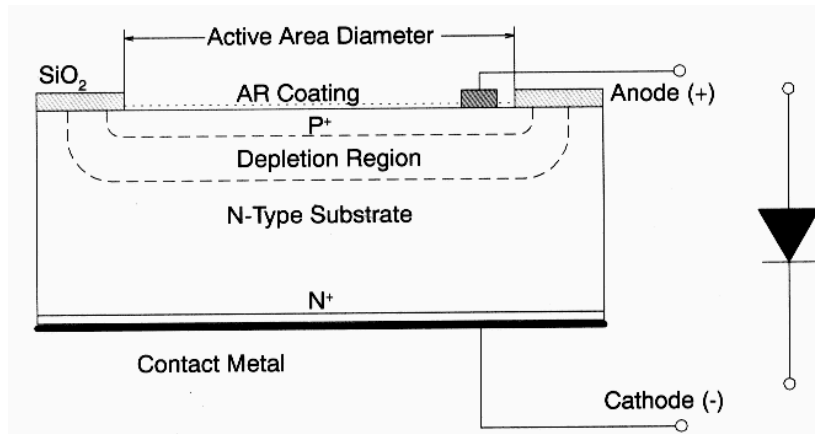


Figure 1: Photodiode silicium en technologie planaire.

PRINCIPE de fonctionnement

Le silicium est un semi-conducteur avec une énergie de gap de 1,12 eV à température ambiante. Le Gap est l'écart entre la bande de valence et la bande de conduction. A la température absolue de zéro, la bande de valence est complètement remplie et la bande de conduction est vide. Avec l'augmentation de température, les électrons acquièrent suffisamment d'énergie pour passer de la bande de valence à la bande de conduction par excitation thermique. Les particules de hautes énergies ou les photons dont l'énergie est supérieure à 1,12 eV (longueur d'onde plus courte que 1100 nm) peuvent aussi fournir suffisamment d'énergie à un électron pour que ce dernier quitte la bande de valence et atteigne la bande de conduction. Les électrons résultants, dans la bande de conduction, sont des électrons libres et conduisent le courant (électrons libres). A cause du gradient de concentration, la diffusion d'électrons des régions de type-N, vers les régions de type -P et la diffusion de trous des régions de type-P vers les régions de type-N produisent une différence de potentiel au travers de la jonction. La diffusion d'électrons et de trous entre les régions P et N à travers la jonction produit une zone sans porteurs libres. C'est la zone de déplétion. La tension à travers la région de déplétion produit un champ électrique dont le maximum est au niveau de la surface jonction, et il n'y a aucun champ en dehors de la région de déplétion. Toute application d'une différence de potentiel inverse s'ajoute à la tension et provoque une plus large zone de déplétion. L'absorption d'un photon crée une paire électron-trou qui est projetée dans la zone de déplétion après diffusion à travers la zone non déplétée. Le courant ainsi générée est proportionnel à la lumière incidente ou à la puissance irradiée.

L'absorption de la lumière est exponentielle avec la distance de pénétration et est proportionnelle au coefficient d'absorption. Le coefficient

d'absorption est grand pour des longueurs d'ondes courtes (région U.V.) et est plus faible pour les grandes longueurs d'ondes. Il s'ensuit que les photons de courtes de longueur d'ondes tels que l'U.V. sont absorbés dans les premiers micromètres en dessous de la surface tandis que le silicium devient progressivement transparent aux longueurs d'ondes de la lumière supérieure à 1200 nm. De plus, les photons de fréquence faibles (grandes longueur d'ondes) qui ont une énergie $E=h.v$ inférieure à l'énergie de Gap ne sont pas absorbés du tout.

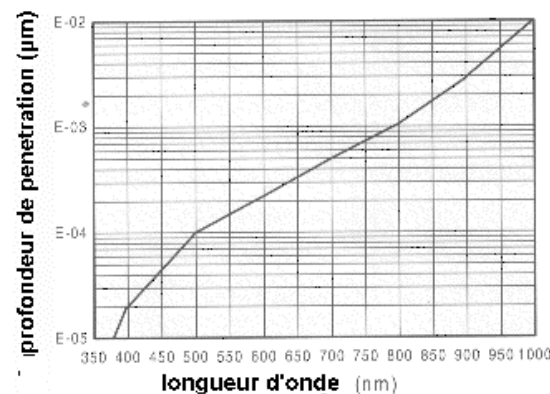


Figure 2: Profondeur de pénétration de la lumière dans un substrat de silicium en fonction de la longueur d'onde.

CARACTERISTIQUE ELECTRIQUE

Une photodiode silicium peut être représentée par une source de courant en parallèle avec une diode idéale. (Figure 3). La source de courant représente le courant généré par l'onde lumineuse incidente, et la diode représente la jonction PN. En outre, une capacité de jonction (C_j) et une résistance de court circuit (R_{SH}) peuvent être placées en parallèle avec les autres composants. La résistance de série (R_s) est reliée en série avec tous les composants dans ce modèle.

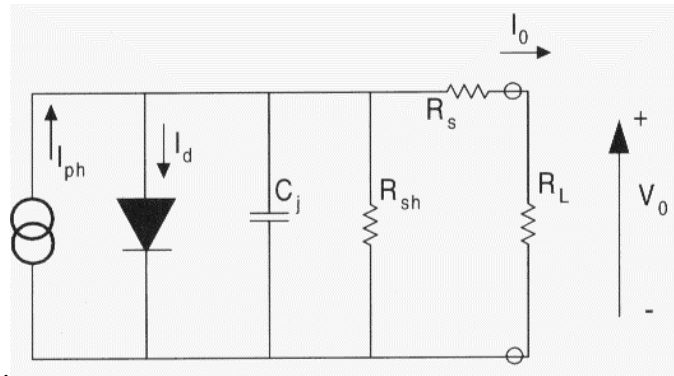


Figure 3 : modèle équivalent d'une photodiode.

Résistance de court circuit, R_{SH}

La résistance de court circuit donne la pente de la caractéristique courant – tension de la photodiode à l'origine, ($V=0$). Bien qu'une photodiode idéale doive avoir une résistance de court circuit infinie, les valeurs réelles varient de 10 à 1000 $M\Omega$. Expérimentalement cette valeur est obtenue en appliquant +10mV et en mesurant le courant et en en déduisant la résistance. La résistance de court circuit est employée pour déterminer le bruit de courant dans la photodiode sans tension d'offset. (Mode photovoltaïque). Lorsque les meilleures performances de photodiode sont désirées, les plus hautes valeurs de résistance de court circuit sont recherchées.

La résistance série, R_s

La résistance de série d'une photodiode provient de la résistance des contacts et de la résistance de la zone non déplétée du silicium (Figure 1). Elle est donnée par.

$$R_S = \frac{(W_S - W_d) \rho}{A} + R_C \quad (1)$$

Où W_s est l'épaisseur du substrat, W_d est la l'épaisseur de la région déplétée. A est la surface de diffusion de la jonction, ρ est la résistivité du substrat et R_c est la résistance de contact. La résistance de série est employée pour déterminer la linéarité de la photodiode en mode photovoltaïque (offset nul, $V=0$). Bien qu'une photodiode idéale doive n'avoir aucune résistance de série, les valeurs typiques varient entre 10 et 1000 Ω .

Capacité de JONCTION :

Les limites de la région de déplétion agissent comme les plaques parallèles d'un condensateur (Figure 1). La capacité de jonction est directement proportionnelle à la surface de diffusion et inversement proportionnelle l'épaisseur de la région de déplétion. En outre, les

grandes valeurs de résistivité du substrat correspondent à des capacités de jonction basses.

De plus, la capacité dépend de la tension de polarisation inverse :

$$C_j = \frac{\epsilon_{si} \epsilon_o A}{\sqrt{2 \epsilon_{si} \epsilon_o m \rho (V_A + V_{bi})}} \quad (2)$$

où $\epsilon_o = 8.854 \times 10^{-14}$ F/cm, est la permittivité du vide, $\epsilon_{sj} = 11.9$ est la constante diélectrique du silicium, $m = 1400$ cm²/V est la mobilité des électrons à 300K, ρ est la résistivité du silicium, V_{bi} est la tension de la jonction sans polarisation externe, et V_A est la tension de

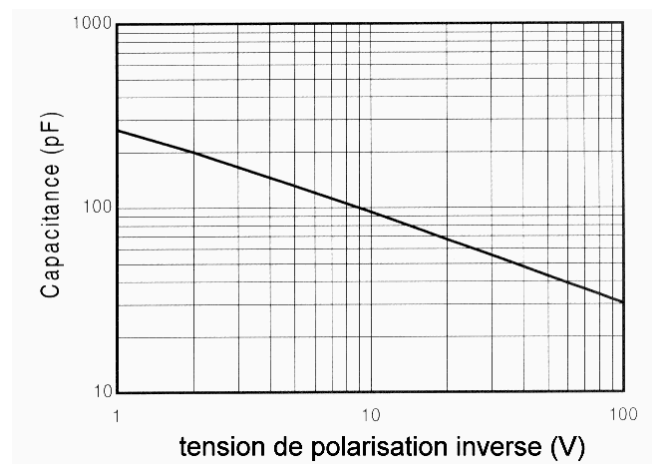


figure 4. capacité d'une photodiode en fonction de la tension inverse.

polarisation appliquée. La figure 4 montre la dépendance de la capacité en fonction de la tension de polarisation inverse appliquée. La capacité de jonction est utile pour déterminer la vitesse de réponse de la photodiode.

Temps de montée et de descente et réponse en fréquence $t_r/t_f/f_{3dB}$

Le Temps de montée/descente d'une photodiode est défini comme le temps pour nécessaire au signal pour passer de 10% à 90% ou de 90% à 10% de la valeur finale respectivement. Ce paramètre peut être exprimé aussi comme la réponse en fréquence, qui est la fréquence pour laquelle la sortie de la photodiode a diminuée de 3dB. Il est approximativement donné par :

$$t_r = \frac{0,35}{f_{3dB}} \quad (3)$$

Il y a trois facteurs définissant le temps de réponse d'une photodiode :

1. t_{drift} le temps de collecte des charges pour les porteurs de la région dépletée de la photodiode.
2. $t_{diffused}$ le temps de collecte des charges pour des porteurs dans la région non-dépletée de la photodiode.

3 . t_{RC} la constante de temps RC du circuit de la diode.

t_{trc} est déterminé par $t_{RC}' = 2,2 RC$, où R, est la somme de résistance série de la diode et de la résistance de charge ($R = R_s + R_L$), et C, est la somme de la jonction de la photodiode et de la capacité de fuite ($C_j + C_s$). Puisque la capacité de la jonction (C_j) dépend de la surface de diffusion de la photodiode et de la tension de polarisation inverse (Equation 2), on obtient des temps de montée plus rapides avec une surface de diffusion de la photodiode plus petite et des tensions de polarisation inverse plus grandes.

De plus la capacité de fuite peut être minimisée en utilisant des connections courtes et une configuration soignée du composant électronique. Le temps total de montée est déterminé par :

$$t_R = \sqrt{t_{derive}^2 + t_{diffusion}^2 + t_{RC}^2} \quad (4)$$

Généralement, en mode photovoltaïque, le temps de montée est dominé par le temps de diffusion dans le cas des surfaces de diffusion inférieures à 5mm², et par les constantes de temps RC dans le cas des surfaces plus larges. En travaillant en mode photoconductif (tension de polarisation inverse appliquée à la jonction.), Si la photodiode est complètement déplétée comme dans le cas des photodiodes pour fibre optiques, le factor dominant est le temps de diffusion. Cependant, dans le cas des photodiodes non complètement déplétées les trois facteurs contribuent à la réponse temporelle de la photodiode.

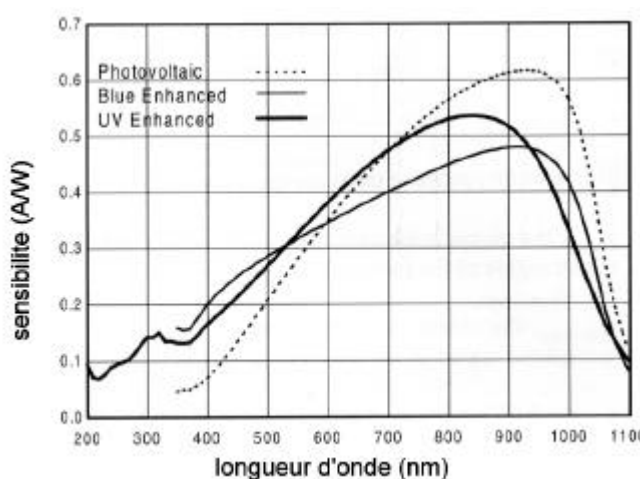
CARACTERISTIQUE OPTIQUE

Sensibilité RI

La sensibilité d'une photodiode silicium est la mesure de la sensibilité à la lumière, et est définie comme le rapport du photocourant I_p par la puissance optique incidente P qui l'a engendrée pour une longueur d'onde donnée :

$$R_I = \frac{I_p}{P} \quad (5)$$

En d'autres termes, c'est une mesure de l'efficacité de la conversion de la puissance lumineuse en courant électrique. Cela varie avec la longueur d'onde de la lumière incidente



A. Du

TR - IUT de Villetaneuse

Université de Paris XIII

Figure 5. Sensibilité spectrale typique de différentes photodiodes en technologie planaire.

La sensibilité augmente lumineusement avec la tension de polarisation inverse en raison de l'amélioration de l'efficacité de collecte des charges dans la photodiode. Il faut noter aussi les variations de la sensibilité due au changement de température comme indiqué sur les figures 6 et 7. Cela est dû à l'augmentation ou à la diminution de l'énergie de gap à cause de l'augmentation et la diminution de la température respectivement. La sensibilité spectrale peut varier énormément et dépend de la longueur d'onde. Cependant les variations relatives de la sensibilité peuvent être réduites à moins de 1% par sélection.

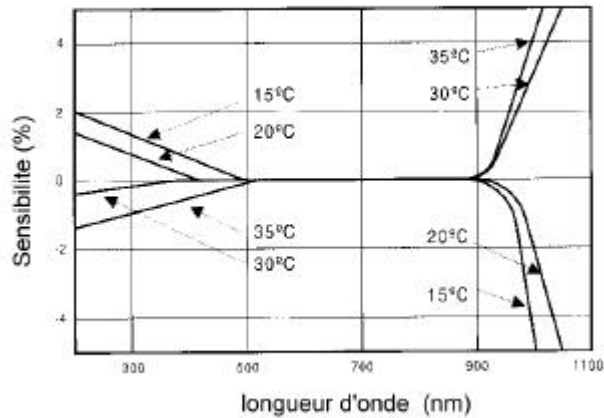


Figure 6 : Dépendance en température de la sensibilité.

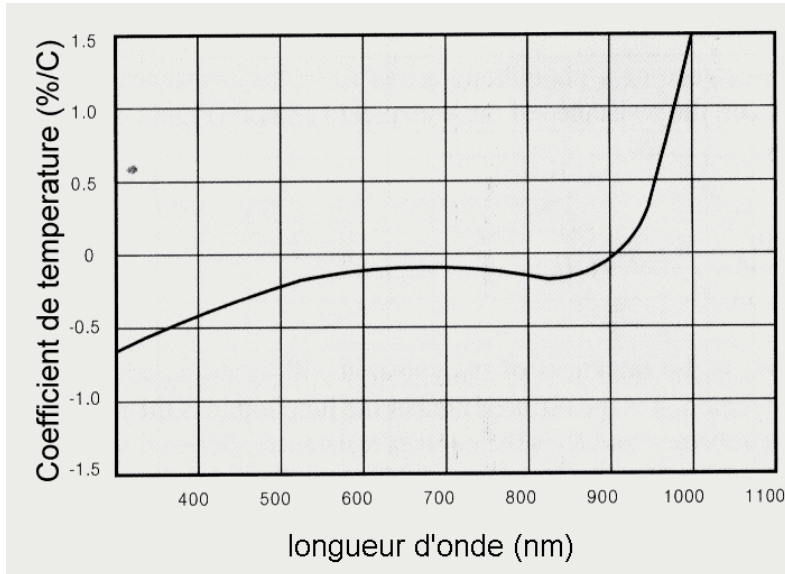


Figure 7 : Coefficient de température des photodiodes silicium

Efficacité Quantique EQ :

L'efficacité quantique est définie comme le pourcentage de la lumière incidente qui participe réellement au photocourant. Elle est reliée à la sensibilité par :

$$Q_E = \frac{R_{I \text{ observe}}}{R_{p \text{ ideal}}} = R_I \frac{hc}{Iq} = 1,24 \cdot 10^3 \frac{R_I}{I} \quad (6)$$

Où $h=6,63 \cdot 10^{-34}$ J-S est la constante de Planck, $c=3 \cdot 10^8$ m/s est la vitesse de la lumière, $q=1,6 \cdot 10^{-19}$ est la charge de l'électron, et $R\lambda$ est la sensibilité en A/W et λ est la longueur d'onde en nm.

Non – Uniformité :

La non-uniformité de la réponse est définie comme la variation de la sensibilité observée lorsqu'un point lumineux est balayé sur la surface active. La non-uniformité est inversement proportionnelle à la taille du point lumineux i.e. La non-uniformité augmente lorsque la tache lumineuse diminue.

Non-linéarité

Une photodiode silicium est considérée comme linéaire si le photo-courant généré augmente linéairement avec la puissance de la lumière incidente. La linéarité du photo-courant est déterminée en mesurant la variation locale de photo-courant lorsqu'il y a une petite variation de la puissance de la lumière incidente et en fonction du photo-courant totale ou en fonction de la puissance de la lumière incidente. La non-linéarité est le rapport de la variation de photo-courant ΔI provoquée par une variation de la puissance de la lumière ΔP , c'est à dire $\Delta I/\Delta P$. En d'autres termes, la linéarité correspond au changement de la sensibilité sur une

plage de la puissance lumineuse. Des non-linéarités inférieures à 1% peuvent être obtenues sur 6-9 décades pour des photodiodes en technologie planaire. La limite inférieure de la linéarité du photo-courant est limitée par le bruit en courant, et la limite supérieure par la résistance de série et la résistance de charge. Lorsque le photo-courant augmente, la non-linéarité s'établit, puis augmente graduellement avec l'augmentation du photo-courant, et finit par saturer, le photo-courant reste alors constant alors même que la puissance de la lumière incidente continue de croître. En général, la variation de photo-courant provoquée par une variation de la puissance lumineuse incidente est plus faibles pour les grandes valeurs de photo-courant que pour les faibles, si le photodétecteur est non-linéaire. La plage de linéarité peut être agrandie lumineusement si on augmente la tension de polarisation inverse à la photodiode.

CARACTERISTIQUES COURANT-TENSION

La caractéristique courant - tension d'une photodiode sans lumière incidente est similaire à celle d'une diode classique. Quand la photodiode est polarisée en direct il y a une augmentation exponentielle du courant. Quand une tension de polarisation inverse est appliquée, un faible courant de saturation inverse apparaît. Ce courant est lié au courant d'obscurité :

$$I_D = I_{sat} \left(e^{\frac{qV_A}{k_B T}} - 1 \right) \quad (7)$$

où I_D est le courant d'obscurité de la photodiode, I_{sat} est le courant de saturation inverse, q est la charge de l'électron, V_A est la tension appliquée, $k_B = 1,38 \cdot 10^{-23}$ J/K, est la constante de Boltzmann, et T est la température absolue ($273 \text{ K} = 0 \text{ C}$).

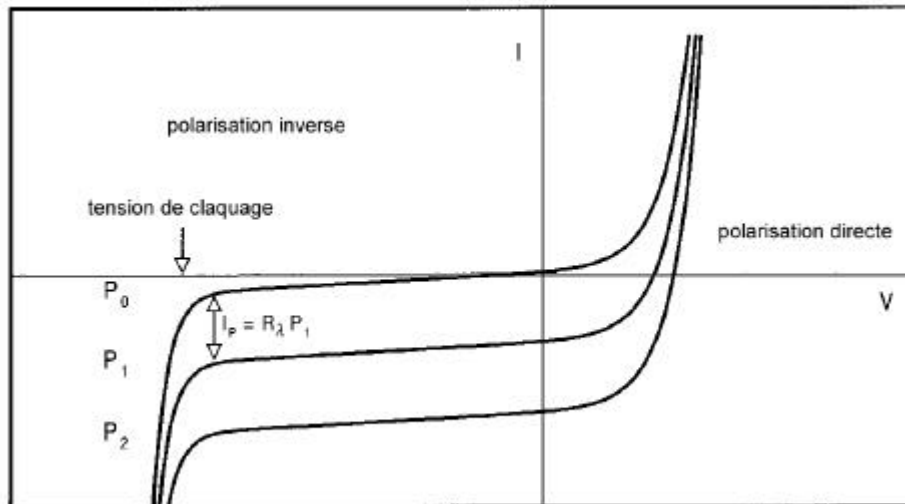


Figure 8. La caractéristique courant tension d'une photodiode en modes photoconducteur et photovoltaïque. P_0, P_1 et P_2 représente des puissances lumineuses différentes.

La relation 7 correspond à la figure 8. On peut distinguer trois états pour l'équation 7 :

- a) $V = 0$, Dans cet état le courant est courant de saturation inverse.
- b) $V > 0$, Dans cet état le courant augmente exponentiellement avec V . Cet état est aussi appelé mode de polarisation directe.
- c) $V < 0$ Quand une tension de polarisation négative est appliqué à la photodiode, le courant se comporte comme indiqué sur la Figure 8.

Lorsque la photodiode est illuminée avec une rayonnement optique, la caractéristique courant-tension est décalée de la quantité de photo-courant (I_p). Ainsi:

$$I_{Total} = I_{sat} \left(e^{\frac{qV_A}{k_B T}} - 1 \right) + I_P \quad (8)$$

où I_p est défini comme le photo-courant de l'équation 5.

Lorsque la tension de polarisation inverse est augmentée au-delà d'une valeur limite, il y a une augmentation brusque du photo-courant. Cette valeur limite est appelée **tension de claquage**. C'est la valeur de tension maximale qui doit être appliquée à la photodiode. (Connue aussi sous le nom de tension de polarisation inverse maximale). La tension de claquage varie d'une photodiode à une autre, et est généralement mesuré pour les petite surface active avec un courant de photodiode de $10 \mu A$.

LE BRUIT

Dans une photodiode deux sources de bruit peuvent être identifiées. Le bruit de grenaille, et le bruit Johnson.

Le bruit de grenaille:

Le bruit de grenaille est lié aux fluctuations statistiques à la fois du photo-courant et du courant d'obscurité. L'amplitude du bruit grenaille est donné par la racine carré (rms) du bruit de courant:

$$I_{sn} = \sqrt{2q(I_P + I_D)\Delta f} \quad (9)$$

où $q=1.6 \times 10^{-19}C$, est la charge de l'électron, I_P est le photo-courant, I_D est le courant d'obscurité du photo-détecteur, et Δf est la largeur spectrale du bruit. Le bruit de grenaille est la source de bruit dominante pour le mode photo-conductif mode (polarisation inverse).

Bruit thermique ou Bruit Johnson:

A chaque résistance de court-circuit d'un photodétecteur est associée un bruit Johnson. En effet l'augmentation de température générée par effet Joule, peut provoquer par génération thermique de nouveau porteurs. L'amplitude du bruit de courant ainsi généré est :

$$I_{jn} = \sqrt{\frac{4k_B T \Delta f}{R_{SH}}} \quad (10)$$

Où $k_B=1,38 \times 10^{-23}J/K$, est la constante de Boltzmann, T , est la température absolue en degré Kelvin ($273 K= 0 C$, Δf est la largeur spectral du bruit et R_{SH} est la résistance de court-circuit de la photodiode. Ce type de bruit est le bruit de courant dominant en mode photovoltaïque (non polarisée).

Note: Toute résistance est associée à un bruit Johnson, y compris la résistance de charge. Ce courant supplémentaire de bruit est grand et s'ajoute au bruit de courant Johnson provoqué par la résistance de court-circuit du photodétecteur.

Le bruit Total

Le bruit de courant total généré dans un photodétecteur est déterminé par :

$$I_{tn} = \sqrt{I_{sn}^2 + I_{jn}^2} \quad (11)$$

Puissance équivalente de bruit (Noise equivalent power) NEP:

Le NEP est la quantité de puissance lumineuse incidente sur un photodétecteur, nécessaire pour générer un photo-courant égal au courant de bruit. Le NEP est défini comme :

$$NEP = \frac{I_{tn}}{R_I} \quad (12)$$

Où R_L est la sensibilité en A/W et I_n est le bruit total du photodétecteur. Les valeurs de la NEP peuvent varier de 10^{-11} W/ $\sqrt{\text{Hz}}$ pour des photodiodes avec une grande surface active jusqu'à 10^{-15} W/ $\sqrt{\text{Hz}}$ pour des photodiodes avec de petites surfaces actives.

EFFETS DE TEMPERATURE

Les paramètres des photodiodes sont affectés par les variations de température. Cela inclut, les résistance de court-circuit, le courant d'obscurité, la tension de claquage, la sensibilité, et dans une moindre mesure d'autres paramètres tel que la capacité de jonction.

Résistance de court-circuit et Courant d'obscurité:

Dans une photodiode il y a deux principaux courants contribuant au courant d'obscurité et à la résistance de court-circuit. Le courant de diffusion, facteur dominant en mode photovoltaïque (non polarisé), détermine la résistance de court-circuit. Il varie avec le carré de la température. Cependant, en mode photo-conductif (polarisation inverse), le courant de dérive devient le courant dominant (courant d'obscurité) et varie linéairement avec la température. Ainsi, les changements de température affectent davantage les photodétecteurs en mode photovoltaïque qu'en mode photo-conductif. En mode photo-conductif le courant d'obscurité peut approximativement doubler pour chaque augmentation de 10°C de la température. En mode photovoltaïque, la résistance de court-circuit quand à elle peut approximativement doubler pour chaque diminution de 6°C de la température. Le changement exact dépend de paramètres supplémentaires tels que la tension de polarisation inverse, la sensibilité ainsi que l'épaisseur du substrat.

Le tension de claquage:

Pour les composants actifs de petites surfaces, la tension de claquage est par définition la tension pour laquelle le courant d'obscurité est de 10 μA . Puisque le courant d'obscurité augmente avec la température, la tension de claquage diminue, elle, avec l'augmentation de température.

La sensibilité.

Les effets de la température sur la sensibilité ont été discutés au paragraphe concernant la sensibilité.

POLARISATION:

Un signal de photodiode peut être mesuré soit sous forme de tension soit sous forme de courant. La mesure d'un courant offre une meilleure linéarité, offset, et performance en bande passante. Le photo-courant généré est proportionnel à la puissance lumineuse incidente mais il

nécessite d'être converti en tension par le biais d'une configuration trans - impédance. La photodiode peut être utilisée avec ou sans polarisation inverse suivant les contraintes l'application exigée. Ces deux modes sont nommés mode Photo-conductif (polarisé) et mode Photovoltaïque (non polarisé).

Mode PHOTO-CONDUCTIF (PC)

L'application d'une polarisation inverse (i.e. tension positive sur la cathode, tension négative à l'anode) peut améliorer énormément la vitesse de réponse et la linéarité d'un composant. Cela est dû à l'augmentation de la largeur de la région de déplétion et subséquemment à la diminution de la capacité de la jonction. En appliquant une tension inverse, cependant, on augmente les bruits de courants et le courant d'obscurité. Un exemple de photodiode en mode photo-conductif avec un niveau de puissance lumineuse bas et une grande vitesse de réponse est exposé Figure 9.

Dans cette configuration le détecteur est polarisé pour réduire la capacité de la jonction ainsi que pour réduire le bruit et le temps de montée (t_r). Un amplificateur à deux étages est utilisé dans cet exemple afin d'obtenir un grand gain et une largeur de bande large. Les deux étages incluent un amplificateur à trans-impédance pour la conversion courant - tension et un amplificateur non - inverseur pour l'amplification de la tension. Le gain et largeur de bande ($f_{3db\text{ MAX}}$) sont directement déterminés par R_F , grâce aux équations (13) et (14). Le gain du deuxième étage peut être approximé par $1 + R_1/R_2$. Une capacité de rétroaction (C_F) permet de limiter la réponse en fréquence et d'éviter les pics de gain.

$$f_{3dB}(Hz) = \sqrt{\frac{GPB}{2pR_F(C_j + C_F + C_A)}} \quad (13)$$

Où GBP est le Produit gain - bande de l'amplificateur (A_1) et C_A est la capacité d'entrée de l'amplificateur.

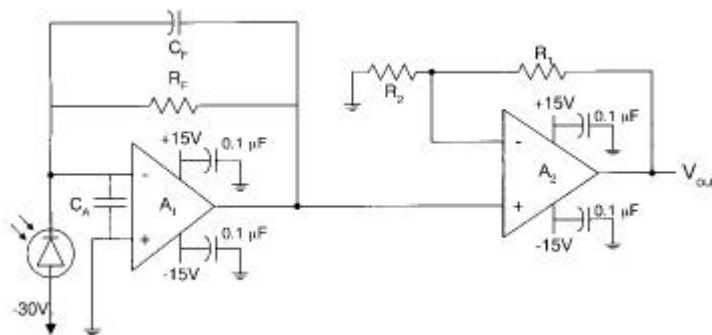


Figure 9. Mode Photoconductif : faible niveau lumineux / grande largeur de bande.

$$f_{3dB}(Hz) = \frac{V_{out}}{P} = R_F \left(1 + \frac{R_1}{R_2} \right) R_I \quad (14)$$

Dans des applications basses vitesse, un grand gain ($>10 \text{ M}\Omega$) peut être obtenu en introduisant une grande valeur de résistance de fuite R_F , dans ce cas un deuxième étage d'amplification n'est pas nécessaire.

Les composants typiques employés pour ces configurations sont :

Des amplificateurs CL-425, CLC 446 ou OPA- 637

R_f : 1 à $10 \text{ K}\Omega$ typique dépendant de c_j

R_1 : 10 à $50 \text{ k}\Omega$

R_2 : 0,5 à $10 \text{ k}\Omega$

C_f : 0,2 à 2pf

Cependant pour des mesures rapides, avec des niveaux lumineux importants, une approche différente est préférée. L'exemple le plus commun est la mesure de largeur impulsionnelle, d'impulsion laser courte de lasers à gaz, ou de diodes lasers, ou de tout autre source lumineuse à impulsion courte. Le signal de sortie des photodiodes peut être soit directement relié à un oscilloscope (Figure 10) ou bien il peut alimenter un amplificateur rapide. Lorsque l'oscilloscope est utilisé, la largeur de bande de l'oscilloscope peut être ajustée à la largeur d'impulsion de la source lumineuse pour obtenir le meilleur rapport signal à bruit. Dans ce genre d'application la tension de polarisation doit être grande. Deux diodes de protections tête bèches peuvent être placées à l'entrée de l'oscilloscope à travers l'entrée et la masse.

Afin d'éviter les oscillations du signal de sortie, le câble entre le détecteur et l'oscilloscope doit être court. ($< 20 \text{ cm}$) et être terminer par une charge de 50 ohms. (R_L). La

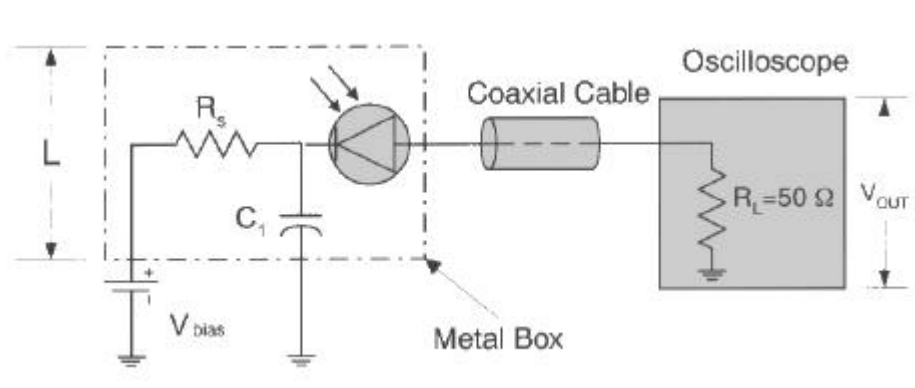


Figure 10. Exemple de circuit en mode photoconductif. Niveau lumineux intense, et signaux rapides.

photodiode doit être enfermée dans une boîte métallique, et si possible, avec des pistes

courtes entre le détecteur et la capacité, et entre le détecteur et le câble coaxial. La boîte métallique doit être reliée à la photodiode par une capacité (C_1) de longueur (L) inférieur à 2 cm, et tel que $R_L C_1 > 10\tau$ (τ est la largeur des impulsions en seconde). R_s est choisi de sorte que $R_s < V_{\text{polarisation}}/10 I_{\text{pdc}}$. Où I_{pdc} est la composante continue du photo-courant. La largeur de bande est définie comme $0,35/\tau$. Pour cette application, un minimum de 10V d'offset négatif est nécessaire. Il faut noter qu'un offset supérieur à la tension de polarisation inverse maximum ne doit pas être appliqué.

Mode photovoltaïque (PV)

Le mode photovoltaïque (non polarisée) est préférable quand la photodiode est utilisée dans un régime basses fréquences (jusqu'à 350 kHz) ainsi que pour niveaux lumineux extrêmement faible. En plus d'offrir une configuration simple, le photocourant dans ce régime présente une plus faible variation de la sensibilité avec la température. Un exemple de configuration pour niveau lumineux ultra faible, et faible vitesse est illustré Figure 11.

Dans cet exemple on trouve un amplificateur opérationnel FET ainsi qu'une grande résistance d'entrée R_F . Le détecteur est non polarisé pour éliminer toute source de bruit. La

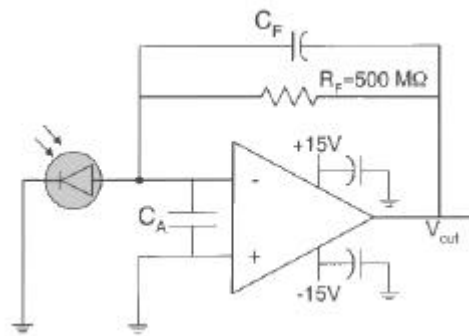


Figure 11. Exemple de circuit en mode photovoltaïque : Niveau lumineux Ultra faible et / basse vitesse.

sortie est déterminée par l'équation 15, et le bruit en courant de l'amplificateur opérationnel est défini par R_F dans l'équation(16).

$$V_{out} = I_P \cdot R_F \quad (15)$$

$$I_N \left(\frac{A_{rms}}{\sqrt{Hz}} \right) = \sqrt{\frac{4kT}{R_F}} \quad (16)$$

ou $k=1,38 \cdot 10^{-23} \text{J/K}$ et T est la température en K. Pour des questions de stabilité il est préférable de choisir C_F de sorte que :

$$\sqrt{\frac{GBP}{2pR_F(C_j + C_F + C_A)}} > \frac{1}{2pR_F C_{Ff}} \quad (17)$$

La bande passante effective avec compensation du pic de gain est :

$$f_{OP}(HZ) = \frac{1}{2pR_F C_F} \quad (18)$$

Pour une telle configuration on peut utiliser les composants suivants :

Amplificateur : OP-15, OP-16, OP-17

RF: 500MΩ.

Cet exemple ou toutes autres configurations pour des photodiodes simples peut être appliqué à tous capteurs monolithiques, ou barrette de capteurs. La sortie du premier étage d'amplification peut être relié à un circuit échantillonneur bloqueur et à un multiplexeur.

PHOTODIODE A AVALANCHE :

Il est possible d'amplifier de façon interne le photocourant. On réalise pour cela des structures particulières appelées Photodiode à avalanche. Dans ce genre de composant, une structure P-N classique avec une épaisseur de la région de déplétion suffisante pour absorber le signal optique, est polarisée avec une très grande tension inverse.

Cette tension inverse doit être très largement supérieure à la tension correspondante à l'énergie de Gap. Les porteurs qui traversent la région de déplétion ont acquis suffisamment d'énergie pour permettre l'excitation d'autres charges par impact. Le phénomène est illustré

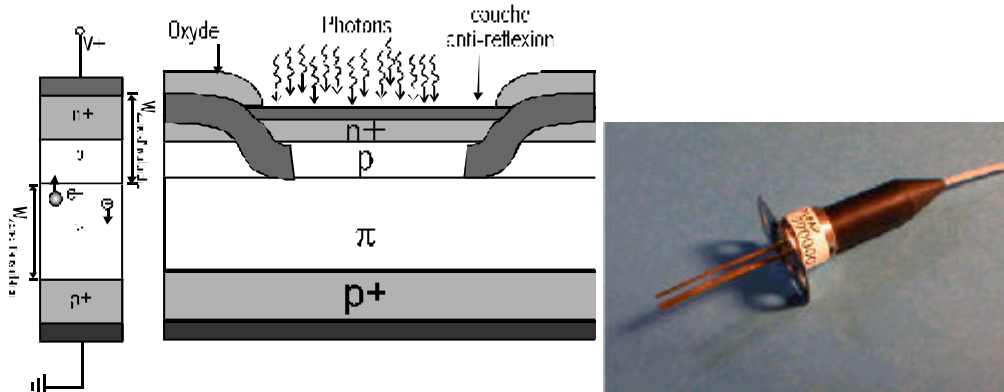


Figure 12 :A gauche Schéma d'une photodiode à avalanche.A droite photodiode à avalanche mitsubishi FU-318AP

figure 13.

Dans une photodiode à avalanche, la tension de polarisation inverse est telle que l'énergie des électrons présents en A et créés par absorption d'un photon possède

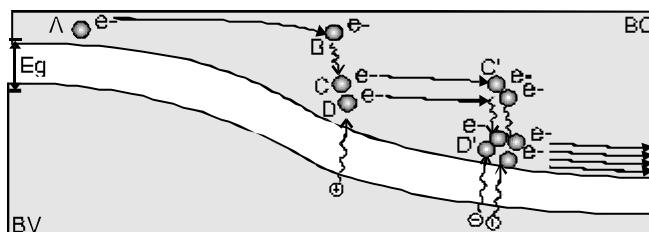


Figure 13 : principe de la photodiode avalanche. En A, électron haute énergie, qui diffuse à travers la jonction (B). De B en C, perte d'énergie de l'électron, et énergie transférée à un électron de la bande de valence qui lui permet d'atteindre la bande de conduction (D). Les deux électrons perdent de l'énergie de C' en D' et la communiquent à deux électrons de la bande de valence. Obtention de quatre électrons dans la bande de conduction.

suffisamment d'énergie pour les transmettre à d'autres électrons. En effet, un électron ayant transité à travers la jonction (B) peut perdre une partie de son énergie (passage en C) et la transmettre à un autre électron. Cet autre électron passe dans la bande de conduction (D). Les électrons ainsi créés continuent de diffuser à travers la jonction (C') et leur énergie peut

être telle qu'ils peuvent en perdre une partie (D') et la transmettre à d'autres électrons et produire davantage d'électrons libres. La multiplication des charges par ce procédé, permet d'atteindre des gains supérieurs à 100.

Le gain est fonction de la tension de polarisation comme indiqué figure 14. L'augmentation de la tension inverse permet évidemment d'augmenter le photo-courant. En revanche, avec une trop grande tension de polarisation inverse, le risque de déclenchement intempestif en l'absence de photoexcitation est réel. La valeur maximale sans déclenchement intempestif définit une valeur limite à la tension de polarisation inverse, appelée également **tension de claquage**.

Bruit d'une photodiode à avalanche :

Si chaque porteur de charge produit par photoabsorption produit M charges secondaires, le courant de bruit RMS du au bruit de grenaille est en reprenant l'équation 9 ;

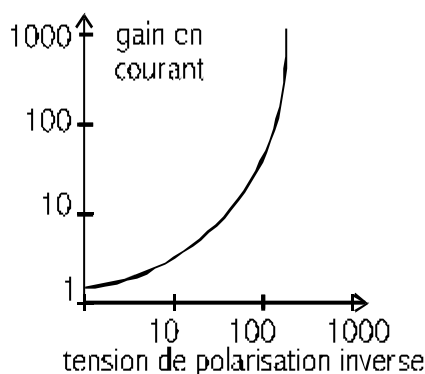


Figure 14 : Photocourant d'une photodiode à avalanche en fonction de la tension de polarisation inverse.

$$\Delta i_f = M \sqrt{2ie\Delta f} \quad (19)$$

ou i est le courant total, e la charge de l'électron, Δf la plage spectrale. Le processus de multiplication est un processus statistique, cependant toutes les charges présentes dans la jonction ne sont pas amplifiées, c'est le cas notamment des paires électrons-trous créés par ionisation. Si r est leur taux de génération spontanée, il faut introduire un coefficient correctif $F(M)$:

$$\Delta i_f = M \sqrt{2ei\Delta f F(M)} \quad \text{avec} \quad F(M) = M \left[1 - \left(1 - \frac{1}{r} \right) \left(\frac{M-1}{M} \right)^2 \right] \quad (20)$$

Bande passante d'une photodiode à avalanche :

L'intérêt des photodiodes à avalanche est leur grand gain, ce qui permet de détecter des signaux de très faible intensité. En revanche le prix à payer est en terme de bande passante, car ces composants ont une bande passante plus faible que celle des photodiodes pin classiques. La bande passante des photodiodes à avalanche est limitée par trois facteurs :

A: Le temps de transit à travers la zone absorbante de la jonction :

$$t_{transit} = \frac{W_{zone\ absorbante}}{v_{electrons}} \quad (21)$$

B: Le temps nécessaire à l'installation du phénomène d'avalanche :

$$\tau_{avalanche} \quad (22)$$

C: Le temps de transit des trous générés lors du phénomène d'avalanche nécessaire pour rejoindre la zone P en traversant la zone d'absorption et la zone d'avalanche $W_{avalanche}$:

$$t_{trous} = \frac{W_{zone\ absorbante} + W_{avalanche}}{v_{trous}} \quad (23)$$

Le retard temporel lié à $\tau_{avalanche}$ dépend du gain M. Si M électrons sont générés, il faut attendre qu'il soit tous sortis de la zone d'avalanche, soit approximativement M fois le temps de transit d'un seul électron. $\tau_{avalanche}$ dépend aussi de la différence entre le taux de génération des électrons et le taux de génération des trous (on comprend que si les trous sont générés moins vite que les électrons cela ralentit le processus), mais en première approximation on les considérera égaux. On considérera donc également que leur vitesse de transit est égale ($v_e = v_{trou}$).

$$\text{D'où } t_{avalanche} = \frac{MW_{zone\ d'avalanche}}{v_e} \quad (24)$$

Le temps de réponse total est donc :

$$t_{totale} = \frac{W_{zone\ absorbante} + MW_{zone\ d'avalanche}}{v_e} + \frac{W_{zone\ absorbante} + W_{zone\ d'avalanche}}{v_{trous}} I$$

Il est important de noter que pour les larges valeurs de gain, la grandeur M/τ est constante. Ainsi pour les photodiodes à avalanche avec un grand gain le produit gain bande reste constant.

PHOTOTRANSISTOR

Comme nous venons de le voir les photodiodes à avalanches sont caractérisées par un très grand gain, mais aussi de façon inhérente par un bruit important due à la nature du processus de multiplication. Il existe un autre type de composants permettant de détecter des signaux lumineux avec un grand gain, ce sont les phototransistors. Comparativement aux photodiodes à avalanches, les phototransistors sont moins bruités.

Les phototransistors bipolaires sont principalement constitués de deux jonctions pn couplées (npn) comme indiqué sur la figure 15a. Lorsqu'un transistor bipolaire est polarisé en mode direct (la jonction base-émetteur est polarisée en mode direct, et la jonction base-collecteur est polarisée en inverse) le profil des bandes est indiqué figure 15 b.

Dans un transistor bipolaire l'injection d'un faible courant de base i_b , provoque un petit changement de la tension de polarisation directe a travers la jonction émetteur-base, ce qui provoque alors une très grande injection de porteurs de charges dans la jonction polarisée en directe. Tout ce courant est principalement collecté au niveau du collecteur. Le gain en courant

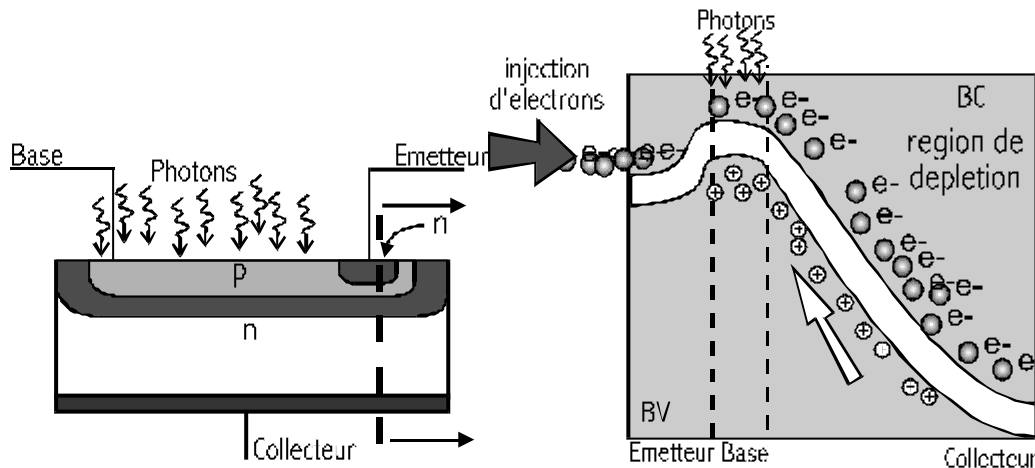


Figure 15 : A gauche (15a), structure d'un phototransistor. A droite (15b) position des bandes d'énergie.

définit comme le rapport du courant de collecteur sur le courant de base est très grand ($\sim 100-1000$).

Dans le cas d'un phototransistor le courant de base n'est pas produit par un courant externe (bien souvent les phototransistors n'ont pas de contacts sur leur base) mais par un signal optique. La lumière absorbée par la zone P du composant génère des paires électrons-trous. Les photo-transistors sont construits de telle sorte que la jonction émetteur-base soit très petite et que la jonction base-collecteur soit grande afin d'augmenter la surface photosensible.

Les phototransistors n'ont pas une très bonne réponse en fréquence (200-500KHz) en raison de la large capacité associée à la jonction base-collecteur.

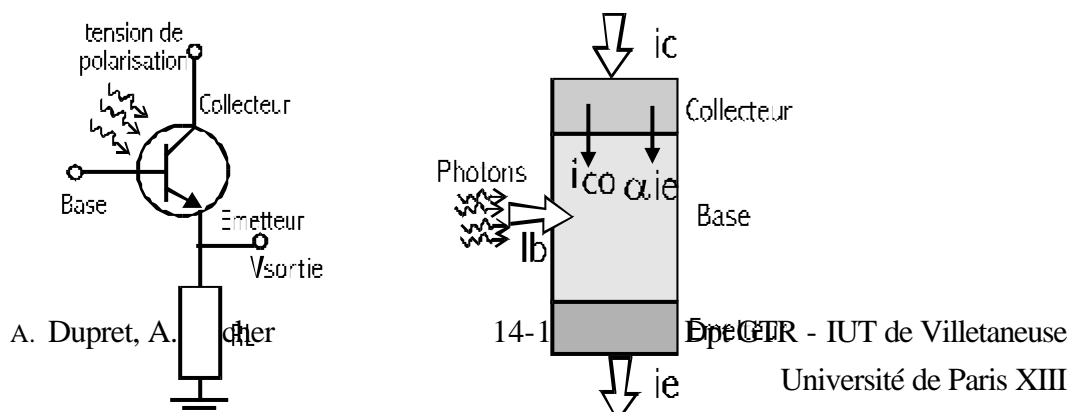


Figure 16: a gauche, schéma d'utilisation des phototransistors. A droite, schéma des courants.

Pour comprendre le fonctionnement du phototransistor, nous allons considérer les courants externes indiqués sur la figure 16 :

Tout se passe comme dans un transistor classique à l'exception du courant de base i_b généré par des photons. L'application de la loi des nœuds au niveau du phototransistor donne :

$$I_c + I_b = I_e \quad (25)$$

où i_e et i_c sont les courants d'émetteur et de collecteur respectivement. Le courant de collecteur à deux causes : (1) le courant de saturation inverse de diode normale i_{co} , (2) la partie du courant d'émetteur I_e qui traverse la jonction jusqu'au collecteur (le courant est porté par la diffusion des porteurs minoritaires à travers la base, mais tous les porteurs minoritaires quittant l'émetteur n'atteindront pas le collecteur). Ce dernier courant peut s'écrire αI_e ou α est un coefficient inférieur à l'unité (α est le gain de courant base commune) ; d'où un courant total ;

$$I_c = I_{co} + \alpha I_e \quad (26)$$

avec (25) et (26)
$$I_c = I_{co} + \alpha(I_c + I_b) \quad (27)$$

$$\text{d'où } I_c = \frac{\alpha I_b + I_{co}}{1 - \alpha} = \left(\frac{\alpha}{1 - \alpha} \right) I_b + I_{co} \frac{1}{1 - \alpha} = b \cdot I_b + \frac{b}{\alpha} I_{co} \quad (28)$$

où $b = \frac{\alpha}{1 - \alpha}$ est le gain en courant émetteur-commun du transistor. Les valeurs typiques de

β en courant d'obscurité sont de l'ordre de 100. Sans rayonnement optique $I_b=0$ et le courant $i_{co}' = i_{co}(\beta/\alpha)$ est le courant d'obscurité.

Lorsque la surface S du phototransistor est illuminée avec un faisceau d'irradiance I_o , le nombre de photons d'énergie $h\nu$ créés est $I_o/h\nu = I_o\lambda/hc$. Si chaque photon donne un électron de charge e , la charge totale créée par unité de temps est :

$$I_l = \frac{I_o S e \lambda}{hc} \quad (29)$$

Cependant tous les photons ne donnent pas naissance à un électron, et on en tient compte en introduisant un coefficient d'efficacité quantique η .

$$\text{Le courant de base effectif est donc : } I_b = I_l = \eta \frac{I_o S e \lambda}{hc} \quad (30)$$

Le courant de collecteur est donc ;

$$I_c = I_{co}' + b \cdot I_b = I_{co}' + b \frac{I_o S e \lambda}{hc} \quad (31)$$

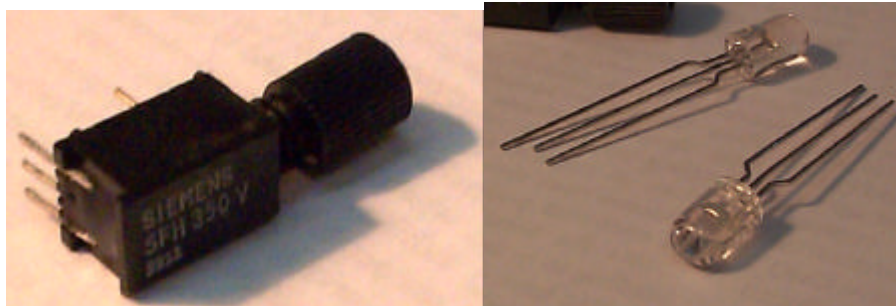


Figure 17 : Différents photo-transistors. A gauche SFH 350V en boîtier à connecteur pour fibre plastique, à droite SFH 350 (Siemens), notez la présence de trois broches; base, collecteur, émetteur.

15. HYPERFREQUENCE : INTRODUCTION **AU RADIOCOM.**

TRANSMISSION **DE** **L'INFORMATION** **LIGNES EN REGIME HARMONIQUE**

INTRODUCTION : RAPPELS SUR LA PROPAGATION DES SIGNAUX

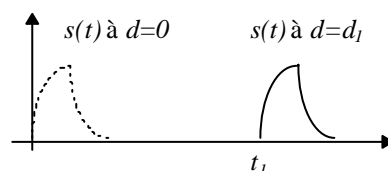
Pour transmettre une information, il est nécessaire de recourir à la modification d'une grandeur physique et, pour que ce changement soit perceptible à distance, il est nécessaire que cette modification se transmette de proche en proche (vibration de l'air dans le cas de signaux sonores, modification d'un signal lumineux...). Les ondes électromagnétiques (Hertziennes ou lumineuses) sont les plus couramment employées du fait de la vitesse élevée de la propagation de ces signaux. La transmission d'ondes électromagnétiques peut être effectuée sur différents supports. Le plus simple est le vide (ou l'air), on parle alors de transmission en espace libre. Les ondes peuvent être également transportées sur des supports tels que des câbles (signaux électriques) ou dans des guides d'ondes telles que les fibres optiques. Dans tous les cas, ces signaux obéissent à des phénomènes de propagation.

Propagation : rappels

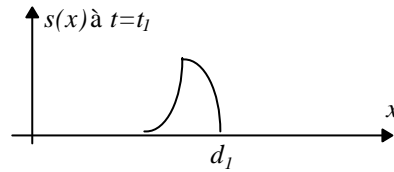
mise en évidence de la propagation des signaux

Considérons un signal créé par un dispositif d'émission et un milieu de transmission (un générateur suivi d'un câble coaxial infiniment long, par exemple) dans lequel le signal va se propager sans subir d'affaiblissement à une vitesse constante c_l . Au niveau du générateur, le signal créé à l'instant $t=t_0$ est une fonction du temps $s(t)$. A une distance d_1 du générateur, le signal parvient après un temps de retard $t_1=d_1/c_l$. A ce retard près, le signal observé en d_1 est caractérisé par la même fonction du temps, $s(t)$, que celui observé au niveau du générateur. A une distance d_2 du générateur, le signal parvient après un temps de retard $t_2=d_2/c_l$. De même que pour d_1 , le signal observé en d_2 est caractérisé, à un retard t_2 près, par la même fonction du temps $s(t)$ que celui observé au niveau du générateur.

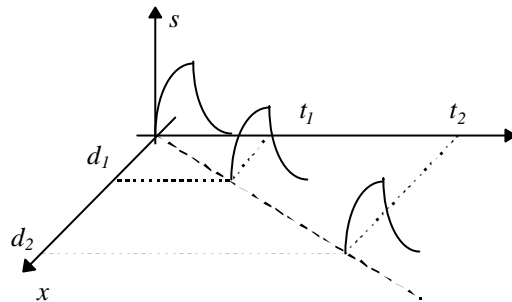
Représentons s en fonction du temps à une distance d_1 de la source. Nous avons alors un signal retardé de $t=t_1=d_1/c_l$ par rapport au signal observé en 0 :



Si nous représentons maintenant s en fonction de x à t fixé, nous avons un signal de la forme :



Une description plus générale de s est donc de considérer que s est une fonction de la distance et du temps. On a donc $s(x, t)$:



Propagation d'un signal selon la direction Ox.

Equation de d'Alembert

Considérons le signal $s(x, t)$ à $x+dx$ et à $t+dt$, de ce fait nous définissons la différentielle ds :

$$ds = \frac{\partial}{\partial x} s(x, t) dx + \frac{\partial}{\partial t} s(x, t) dt .$$

$s(x, t)$ est continue et dérivable en x et t donc : $\frac{\partial^2}{\partial x \partial t} s(x, t) = \frac{\partial^2}{\partial t \partial x} s(x, t) .$

De plus, nous avons une relation liant x et t : $x=c_0 t$. Donc :

$$\begin{aligned} \frac{\partial^2}{\partial x \partial t} s(x, t) &= \frac{\partial}{\partial x} \left(\frac{\partial}{\partial t} s(x, t) \right) \text{ soit : } \frac{\partial}{\partial x} \left(\frac{\partial}{\partial t} s(x, t) \right) = \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} s(x, t) \cdot \frac{dx}{dt} \right) \\ \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} s(x, t) \cdot \frac{dx}{dt} \right) &= \frac{\partial^2}{\partial x^2} s(x, t) \cdot \frac{dx}{dt} \text{ or } \frac{dx}{dt} = c_0 . \text{ Donc : } \frac{\partial^2}{\partial x \partial t} s(x, t) = c_0 \cdot \frac{\partial^2}{\partial x^2} s(x, t) . \end{aligned}$$

De même, on montre que : $\frac{\partial^2}{\partial t \partial x} s(x, t) = \frac{1}{c_0} \cdot \frac{\partial^2}{\partial t^2} s(x, t) .$

L'égalité : $\frac{\partial^2}{\partial x \partial t} s(x, t) = \frac{\partial^2}{\partial t \partial x} s(x, t)$ conduit donc à : $c_0 \cdot \frac{\partial^2}{\partial x^2} s(x, t) = \frac{1}{c_0} \cdot \frac{\partial^2}{\partial t^2} s(x, t) .$

Cette expression peut s'écrire :

$$\frac{\partial^2}{\partial x^2} s(x, t) - \frac{1}{c_0^2} \cdot \frac{\partial^2}{\partial t^2} s(x, t) = 0 .$$

On nomme d'Alembertien l'opérateur : $\square = \frac{\partial^2}{\partial x^2} - \frac{1}{c_0^2} \cdot \frac{\partial^2}{\partial t^2} .$

¹ Jean le Rond d'Alembert, mathématicien et philosophe français (1717–1783)

Tout signal $s(x, t)$ se propageant vérifie : $\square s = 0$.

Déterminons les solutions générales de : $\square s = 0$. Pour cela, nous pouvons remarquer que

l'équation : $\frac{\partial^2}{\partial x^2} s(x, t) - \frac{1}{c_0^2} \cdot \frac{\partial^2}{\partial t^2} s(x, t) = 0$, peut s'écrire :

$$\left(\frac{\partial}{\partial x} - \frac{1}{c_0} \cdot \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial x} + \frac{1}{c_0} \cdot \frac{\partial}{\partial t} \right) s(x, t) = 0.$$

Posons $p = t + x/c_0$ et $q = t - x/c_0$, en exprimant x et t en fonction de p et q , on a alors :

$$\begin{cases} \frac{\partial}{\partial x} = \frac{\partial p}{\partial x} \cdot \frac{\partial}{\partial p} + \frac{\partial q}{\partial x} \cdot \frac{\partial}{\partial q} = \frac{1}{c_0} \left(\frac{\partial}{\partial p} - \frac{\partial}{\partial q} \right) \\ \frac{\partial}{\partial t} = \frac{\partial p}{\partial t} \cdot \frac{\partial}{\partial p} + \frac{\partial q}{\partial t} \cdot \frac{\partial}{\partial q} = \frac{\partial}{\partial p} + \frac{\partial}{\partial q} \end{cases} \text{ ce qui conduit à : } \begin{cases} \frac{\partial}{\partial x} + \frac{1}{c_0} \cdot \frac{\partial}{\partial t} = 2 \cdot \frac{\partial}{\partial p} \\ \frac{\partial}{\partial x} - \frac{1}{c_0} \cdot \frac{\partial}{\partial t} = -2 \cdot \frac{\partial}{\partial q} \end{cases}$$

Puisque $\left(\frac{\partial}{\partial x} - \frac{1}{c_0} \cdot \frac{\partial}{\partial t} \right) \left(\frac{\partial}{\partial x} + \frac{1}{c_0} \cdot \frac{\partial}{\partial t} \right) s(x, t) = 0$, alors : $\frac{\partial^2}{\partial p \partial q} s = 0$. Cette expression

peut s'écrire :

$$\frac{\partial^2}{\partial p \partial q} s = \frac{\partial}{\partial p} \left(\frac{\partial s}{\partial q} \right) = 0 \text{ ce qui implique } s = g(q). \text{ D'autre part, } \frac{\partial^2}{\partial p \partial q} s = \frac{\partial^2}{\partial q \partial p} s = 0,$$

donc :

$$\frac{\partial^2}{\partial p \partial q} s = \frac{\partial}{\partial q} \left(\frac{\partial s}{\partial p} \right) = 0 \text{ ce qui implique } s = f(p).$$

En intégrant cette équation, on a alors $s = f(p) + g(q)$.

La solution générale de l'équation de d'Alembert est donc de la forme :

$$s(x, t) = f\left(t - \frac{x}{c_0}\right) + g\left(t + \frac{x}{c_0}\right)$$

$f(t - x/c_0)$ représente un signal se propageant dans le sens des x et $g(t + x/c_0)$ représente un signal se propageant en sens opposé.

La solution générale de l'équation de d'Alembert à une dimension peut donc s'interpréter comme la superposition de deux ondes progressives de vitesses opposées.

La représentation générale du D'Alembertien est :

$$\square = \nabla^2 - \frac{1}{c_0^2} \cdot \frac{\partial^2}{\partial t^2}$$

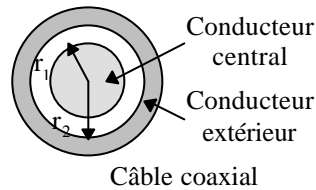
PROPAGATION GUIDÉE SUR DES LIGNES

Modélisation d'une ligne

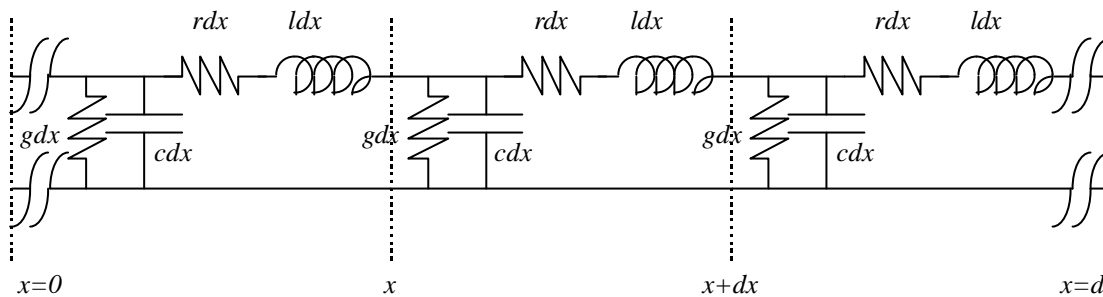
Une ligne, comme un câble coaxial ou une ligne bi-filaire, est caractérisée par une inductance linéique l (inductance par unité de longueur) et par une capacité linéique c (capacité par unité de longueur).

Par exemple, pour un câble coaxial :

$$c = \frac{2\pi \epsilon_0 \epsilon_r}{\ln(r_2/r_1)} \text{ et } l = \frac{\mu_0 \mu_r}{2\pi} \ln\left(\frac{r_2}{r_1}\right)$$



De plus, la résistance série r et l'admittance g reflètent les pertes d'énergie, respectivement dans les conducteurs et dans le diélectrique de la ligne. On parle alors de constantes réparties par opposition aux éléments discrets. De ce fait, la représentation de la ligne sur un tronçon dx est :

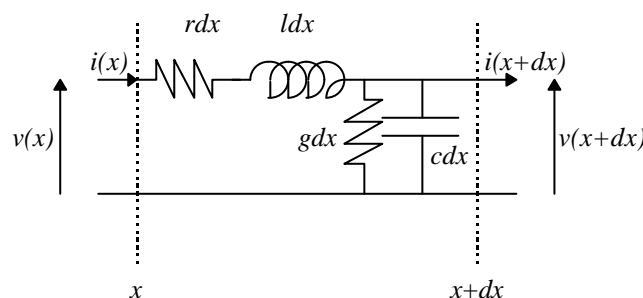


Lorsque la ligne est "courte", la chute de tension est négligeable sur un tronçon dx et, de façon corollaire, le courant entrant en $x=0$ est identique à celui en $x=d$. C'est le cas le plus fréquent aux basses fréquences (<30 MHz) en laboratoire. Il faut donc préciser le terme de "courte". Lorsqu'un courant variable circule dans la ligne; celui-ci engendre un champ électrique et un champ magnétique variable.

Equations de propagation

cas général :

Revenons à la ligne sur un tronçon dx entre les distances x et $x+dx$:



courants et tensions sur un élément de ligne.

Il est alors possible d'écrire les équations :

$$\begin{cases} \frac{\partial v}{\partial x} = -r \cdot i - l \cdot \frac{\partial i}{\partial t} \\ \frac{\partial i}{\partial x} = -g \cdot v - c \cdot \frac{\partial v}{\partial t} \end{cases}$$

En dérivant $\frac{\partial v}{\partial x}$ par rapport à x on obtient :

$$\begin{cases} \frac{\mathcal{V}_x^2}{\mathcal{V}_x^2} = -r \cdot \frac{\mathcal{I}_i}{\mathcal{V}_x} - l \cdot \frac{\mathcal{V}_i^2}{\mathcal{V}_x \mathcal{V}_t} \\ \frac{\mathcal{V}_i^2}{\mathcal{V}_x^2} = -g \cdot \frac{\mathcal{V}_v}{\mathcal{V}_x} - c \cdot \frac{\mathcal{V}_v^2}{\mathcal{V}_x \mathcal{V}_t} \end{cases}$$

Dans ce nouveau couple d'équations, en remplaçant $\frac{\mathcal{I}_i}{\mathcal{V}_x}$ (respectivement $\frac{\mathcal{V}_v}{\mathcal{V}_x}$) par

l'expression de la première équation, on obtient finalement l'équation dites "des télégraphistes" :

$$\begin{cases} \frac{\mathcal{V}_v^2}{\mathcal{V}_x^2} - l \cdot c \cdot \frac{\mathcal{V}_v^2}{\mathcal{V}_t^2} - (g \cdot l + r \cdot c) \cdot \frac{\mathcal{V}_v}{\mathcal{V}_t} - r \cdot g \cdot v = 0 \\ \frac{\mathcal{V}_i^2}{\mathcal{V}_x^2} - l \cdot c \cdot \frac{\mathcal{V}_i^2}{\mathcal{V}_t^2} - (g \cdot l + r \cdot c) \cdot \frac{\mathcal{V}_i}{\mathcal{V}_t} - r \cdot g \cdot i = 0 \end{cases}$$

Lorsque les pertes dues aux conducteurs et au diélectrique sont négligeables, on parle alors de lignes sans perte (L.S.P.), l'expression de cette équation est simplifiée :

$$\begin{cases} \frac{\mathcal{V}_v^2}{\mathcal{V}_x^2} - l \cdot c \cdot \frac{\mathcal{V}_v^2}{\mathcal{V}_t^2} = 0 \\ \frac{\mathcal{V}_i^2}{\mathcal{V}_x^2} - l \cdot c \cdot \frac{\mathcal{V}_i^2}{\mathcal{V}_t^2} = 0 \end{cases}$$

dans ce dernier cas, on reconnaît immédiatement un d'Alembertien, c'est-à-dire que nous sommes en présence de l'équation de signaux se propageant.

Régime harmonique :

Très souvent, les lignes sont utilisées pour transmettre des signaux modulés sur des porteuses sinusoïdales. Il est alors légitime de se placer dans le cas du régime permanent sinusoïdal et les expressions de l et c en fonction du temps sont remplacées par des impédances complexes. Les tensions et courant s'expriment alors sous la forme :

$$\begin{cases} \bar{v}(x,t) = \bar{V}(x) \times e^{j \cdot \omega \cdot t} \\ \bar{i}(x,t) = \bar{I}(x) \times e^{j \cdot \omega \cdot t} \end{cases} \text{ et les équations précédentes deviennent :}$$

$$\begin{cases} \frac{\mathcal{V}_x^2 \bar{V}}{\mathcal{V}_x^2} - l \cdot c \cdot \omega^2 \cdot \bar{V} - j(g \cdot l \cdot \omega + r \cdot c \cdot \omega) \cdot \bar{V} - r \cdot g \cdot \bar{V} = 0 \\ \frac{\mathcal{V}_x^2 \bar{I}}{\mathcal{V}_x^2} - l \cdot c \cdot \omega^2 \cdot \bar{I} - j(g \cdot l \cdot \omega + r \cdot c \cdot \omega) \cdot \bar{I} - r \cdot g \cdot \bar{I} = 0 \end{cases} \text{ ce qui peut s'écrire :}$$

$$\begin{cases} \frac{\mathcal{V}_x^2 \bar{V}}{\mathcal{V}_x^2} = (j \cdot l \cdot \omega + r) \cdot (g + j \cdot c \cdot \omega) \cdot \bar{V} \\ \frac{\mathcal{V}_x^2 \bar{I}}{\mathcal{V}_x^2} = (j \cdot l \cdot \omega + r) \cdot (g + j \cdot c \cdot \omega) \cdot \bar{I} \end{cases} \text{ et pour les L.S.P. : } \begin{cases} \frac{\mathcal{V}_x^2 \bar{V}}{\mathcal{V}_x^2} - l \cdot c \cdot \omega^2 \cdot \bar{V} = 0 \\ \frac{\mathcal{V}_x^2 \bar{I}}{\mathcal{V}_x^2} - l \cdot c \cdot \omega^2 \cdot \bar{I} = 0 \end{cases}$$

Posons : $\mathcal{g}^2 = (j \cdot l \cdot \omega + r) \cdot (g + j \cdot c \cdot \omega)$. γ est donc un nombre complexe qui s'écrit aussi :

$\mathcal{g} = ((j \cdot l \cdot \omega + r) \cdot (g + j \cdot c \cdot \omega))^{1/2} = \mathbf{a} + j\mathbf{b}$ ($\mathbf{a} > 0$). Nous avons alors :

$$\begin{cases} \frac{d^2 \bar{V}}{dx^2} - \mathbf{g}^2 \cdot \bar{V} = 0 \\ \frac{d^2 \bar{I}}{dx^2} - \mathbf{g}^2 \cdot \bar{I} = 0 \end{cases} \quad \text{Ces équations différentielles s'intègrent simplement et leur solution}$$

générale est de la forme :
$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{-\mathbf{g} \cdot x} + \bar{V}_r \times e^{+\mathbf{g} \cdot x} \\ \bar{I}(x) = \bar{I}_i \times e^{-\mathbf{g} \cdot x} + \bar{I}_r \times e^{+\mathbf{g} \cdot x} \end{cases}$$

La résolution de ces équation entraîne deux remarques importantes :

impédance caractéristique d'une ligne

$\begin{cases} \bar{V}_i, \bar{V}_r \\ \bar{I}_i, \bar{I}_r \end{cases}$ sont des constantes d'intégration liées entre elles par les équations précédentes

(en particulier, l'équation différentielle de $d^0=1$) : $\frac{\bar{V}_i}{\bar{I}_i} = -\frac{\bar{V}_r}{\bar{I}_r} = \left(\frac{r + j \cdot l\omega}{g + j \cdot c\omega} \right)^{1/2}$. Cette expression

est homogène à une impédance. On pose alors :

$$\bar{Z}_c = \left(\frac{r + j \cdot l\omega}{g + j \cdot c\omega} \right)^{1/2}, \text{ grandeur appelée impédance caractéristique de la ligne.}$$

Dans le cas des lignes sans pertes (L.S.P.), l'impédance caractéristique est :

$$\bar{Z}_c = \sqrt{\frac{l}{c}}.$$

Dans la majorité des cas Z_c peut être considéré comme réelle (et positive). Les valeurs classiques d'impédances caractéristiques sont : 50Ω (câbles coaxiaux ethernet, câbles coaxiaux de montée d'antennes UHF, etc...), 75Ω (câbles coaxiaux pour télévision), 300Ω (câbles bifilaires pour télévision).

équation de propagation

En remplaçant \mathbf{g} par $\mathbf{g} = \mathbf{a} + j \mathbf{b}$ et en exprimant $v(x, t)$ (respectivement $i(x, t)$), on obtient :

$$\begin{cases} \bar{v}(x, t) = \bar{V}_i \times e^{-\mathbf{a} \cdot x} \cdot e^{j(\mathbf{w} \cdot t - \mathbf{b} \cdot x)} + \bar{V}_r \times e^{+\mathbf{a} \cdot x} \cdot e^{j(\mathbf{w} \cdot t + \mathbf{b} \cdot x)} \\ \bar{i}(x, t) = \bar{I}_i \times e^{-\mathbf{a} \cdot x} \cdot e^{j(\mathbf{w} \cdot t - \mathbf{b} \cdot x)} + \bar{I}_r \times e^{+\mathbf{a} \cdot x} \cdot e^{j(\mathbf{w} \cdot t + \mathbf{b} \cdot x)} \end{cases}$$

$v(x, t)$ et $i(x, t)$ sont donc composées deux ondes, l'une est d'amplitude décroissante pour les x (positifs) croissants, l'autre est d'amplitude décroissante pour les x décroissants. La première est appelée onde incidente (d'où l'indice i) et est caractéristique d'un déplacement de la source (le générateur) vers l'extrémité de la ligne. L'autre est caractéristique d'une onde se propageant de l'extrémité de la ligne vers la source (le générateur). Cette onde est due à un phénomène de réflexion sur la ligne, exactement comme en mécanique, et est appelée onde réfléchie (d'où l'indice r).

Ces ondes sont caractérisées par :

- une période temporelle $T = \frac{2\pi}{\omega}$.

•une périodicité spatiale $\lambda = \frac{2\pi}{\beta}$ (dans le cas des L.S.P. : $\lambda = \frac{2\pi}{\beta} = \frac{2\pi}{\omega \sqrt{LC}} = \frac{v_p}{f}$).

•un affaiblissement, α , du aux pertes dans les conducteurs et dans le diélectrique.

Ces ondes se déplacent sur la ligne à une vitesse constante v_p appelée vitesse de phase déterminée par :

$$v_p = \frac{dx}{dt} = \frac{\omega}{\beta}. \text{ Dans le cas de L.S.P., } v_p \text{ est donné par : } v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}}.$$

On démontre que cette vitesse de phase est égale à la vitesse de propagation des ondes électromagnétiques dans le diélectrique constitué par la ligne. Nous comprenons alors que la ligne n'est plus "courte" si sa longueur n'est pas faible devant λ . Cette notion de "courte" et l'apparition des phénomènes de propagation sont donc liés à la fréquence de travail. On rencontre des phénomènes de propagation à quelques dizaines de MHz sur des longueurs de l'ordre du mètre mais également sur les lignes de transport électrique (EDF) de plusieurs kilomètres à 50 Hz.

Vitesse de groupe

En fait, le spectre des ondes transmises sur les lignes est rarement (essentiellement jamais) composé d'une seule raie (par analogie avec les ondes lumineuses, un signal dont le spectre ne comporte qu'une raie, est appelé monochromatique). On définit alors la vitesse de groupe qui est la vitesse à laquelle se déplace les "déformations" du signal correspondants à des signaux à d'autres fréquences que la fréquence pour laquelle l'étude a été menée (par exemple, les variations de fréquences autour de la porteuse d'un signal AM ou FM). La vitesse de groupe, v_g , est alors donnée par :

$$v_g = \frac{d\omega}{d\beta}.$$

On montre alors qu'un signal ne peut se transmettre sans déformation que si $v_g = v_p = C^{te}$.

Expression de l'impédance en un point de la ligne.

Impédance en un point x

Nous avons déterminé que :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{-\gamma \cdot x} + \bar{V}_r \times e^{+\gamma \cdot x} \\ \bar{I}(x) = \bar{I}_i \times e^{-\gamma \cdot x} + \bar{I}_r \times e^{+\gamma \cdot x} \end{cases} \text{ donc, en } x=0, \text{ nous avons : } \begin{cases} \bar{V}_0 = \bar{V}_i + \bar{V}_r \\ \bar{I}_0 = \bar{I}_i + \bar{I}_r \end{cases}.$$

Par ailleurs, nous avons : $\frac{\bar{V}_i}{\bar{I}_i} = -\frac{\bar{V}_r}{\bar{I}_r} = \bar{Z}_c$ (obtenue à partir de l'équation différentielle).

D'où :

$$\begin{cases} \bar{V}_i = \frac{\bar{V}_0 + \bar{Z}_c \bar{I}_0}{2}, \bar{V}_r = \frac{\bar{V}_0 - \bar{Z}_c \bar{I}_0}{2} \\ \bar{I}_i = \frac{\bar{I}_0}{2} + \frac{\bar{V}_0}{2 \cdot \bar{Z}_c}, \bar{I}_r = \frac{\bar{I}_0}{2} - \frac{\bar{V}_0}{2 \cdot \bar{Z}_c} \end{cases} \text{ . En remplaçant dans l'équation précédente :}$$

$$\begin{cases} \bar{V}(x) = \bar{V}_0 \cdot ch(\mathbf{g} \cdot x) - \bar{Z}_c \cdot \bar{I}_0 \cdot sh(\mathbf{g} \cdot x) \\ \bar{I}(x) = \bar{I}_0 \cdot ch(\mathbf{g} \cdot x) - \frac{\bar{V}_0}{\bar{Z}_c} \cdot sh(\mathbf{g} \cdot x) \end{cases} \text{ . D'où l'expression de l'impédance en un point } x :$$

$$\bar{Z}(x) = \frac{\bar{V}(x)}{\bar{I}(x)} = \bar{Z}_c \frac{\bar{Z}_0 - \bar{Z}_c th(\mathbf{b} x)}{\bar{Z}_c - \bar{Z}_0 th(\mathbf{b} x)}, \text{ ou, dans le cas de L.S.P. :}$$

$$\bar{Z}(x) = \frac{\bar{V}(x)}{\bar{I}(x)} = \bar{Z}_c \frac{\bar{Z}_0 - j \bar{Z}_c tg(\mathbf{b} x)}{\bar{Z}_c - j \bar{Z}_0 tg(\mathbf{b} x)} .$$

Impédance en un point $x=0$

A partir de maintenant, nous ne considérerons plus que le cas des lignes sans perte.

A partir de : $\bar{Z}(x) = \bar{Z}_c \frac{\bar{Z}_0 - j \bar{Z}_c tg(\mathbf{b} x)}{\bar{Z}_c - j \bar{Z}_0 tg(\mathbf{b} x)}$, on peut déterminer l'impédance ramenée en $x=0$ en fonction de l'impédance de charge Z_l , située à une distance d , de l'origine :

$$\bar{Z}_0 = \bar{Z}_c \frac{\bar{Z}_l + j \bar{Z}_c tg(\mathbf{b} d)}{\bar{Z}_c + j \bar{Z}_l tg(\mathbf{b} d)} .$$

Cas particuliers d'impédances Z_l et de longueur de ligne

$d=k \mathbf{1}/4$ (k entier)

On parle de ligne $\mathbf{1}/4$ lorsque $d=(2k+1) \mathbf{1}/4$. Dans ce cas, puisque $\mathbf{1} = \frac{2p}{b}$, $\mathbf{b} d = (2k+1) \frac{p}{2}$ et par conséquent, $tg(\mathbf{b} d) = tg((2k+1) p/2) \rightarrow \infty$. Ainsi, l'impédance ramenée en $x=0$ est : $\bar{Z}_0 = \frac{\bar{Z}_c^2}{\bar{Z}_l}$.

$d=k \mathbf{1}/2$ (k entier)

On parle de ligne $\mathbf{1}/2$ lorsque $d=k \mathbf{1}/2$. Dans ce cas, puisque $\mathbf{1} = \frac{2p}{b}$, $\mathbf{b} d = k p$, donc $tg(\mathbf{b} d) = tg(k p) = 0$. De ce fait, l'impédance ramenée en $x=0$ est : $\bar{Z}_0 = \bar{Z}_l$.

Cette propriété est particulièrement utile lorsqu'on ne peut accéder à une charge que l'on doit caractériser via une ligne.

$Z_l=0$

L'impédance ramenée en $x=0$ est : $\bar{Z}_0 = j \bar{Z}_c tg(\mathbf{b} d)$, d étant la longueur de la ligne.

L'impédance ramenée est purement imaginaire. Deux cas se présentent alors :

- $k p \leq d < \frac{p}{2} + k p$ alors $Z_c tg(\mathbf{b} d) > 0$, donc $\bar{Z}_0 = j \bar{Z}_c tg(\mathbf{b} d)$ est inductif. La ligne se comporte donc comme une inductance.
- $\frac{p}{2} + k p \leq d < p + k p$ alors $Z_c tg(\mathbf{b} d) < 0$, donc $\bar{Z}_0 = j \bar{Z}_c tg(\mathbf{b} d)$ est capacitif. La ligne se comporte donc comme une capacité.

Remarque : une ligne $\lambda/4$ terminée par un court-circuit est un isolateur en HF!

$Z_l=\infty$ (circuit ouvert)

L'impédance ramenée en $x=0$ est : $\bar{Z}_0 = \frac{\bar{Z}_c}{j tg(\mathbf{b} d)}$, d étant la longueur de la ligne.

L'impédance ramenée est purement imaginaire. Là encore, deux cas se présentent :

- $k\mathbf{p} \leq d < \frac{\mathbf{p}}{2} + k\mathbf{p}$ alors $\frac{\bar{Z}_c}{tg(\mathbf{b}d)} > 0$, donc $\bar{Z}_0 = \frac{\bar{Z}_c}{jtg(\mathbf{b}d)}$ est capacitif. La ligne se comporte donc comme une capacité.
- $\frac{\mathbf{p}}{2} + k\mathbf{p} \leq d < \mathbf{p} + k\mathbf{p}$ alors $\frac{\bar{Z}_c}{tg(\mathbf{b}d)} < 0$, donc $\bar{Z}_0 = \frac{\bar{Z}_c}{jtg(\mathbf{b}d)}$ est inductif. La ligne se comporte donc comme une inductance.

$$\underline{Z_l} = \underline{Z_c}$$

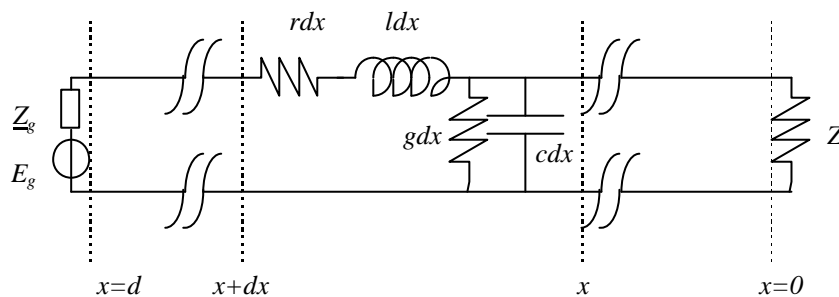
Quand l'impédance de charge est égale à l'impédance caractéristique de la ligne, $\bar{Z}_0 = \bar{Z}_c = \bar{Z}_l$.

Définition du coefficient de réflexion, des ondes progressives et des ondes stationnaires.

Nous avons vu que :
$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{-j\mathbf{b} \cdot x} + \bar{V}_r \times e^{+j\mathbf{b} \cdot x} \\ \bar{I}(x) = \bar{I}_i \times e^{-j\mathbf{b} \cdot x} + \bar{I}_r \times e^{+j\mathbf{b} \cdot x} \end{cases} \quad (\mathbf{g} \text{ a été remplacé par } j\mathbf{b})$$
 et

$$\begin{cases} \bar{V}_i = \frac{\bar{V}_0 + \bar{Z}_c \bar{I}_0}{2}, \bar{V}_r = \frac{\bar{V}_0 - \bar{Z}_c \bar{I}_0}{2} \\ \bar{I}_i = \frac{\bar{I}_0}{2} + \frac{\bar{V}_0}{2 \cdot \bar{Z}_c}, \bar{I}_r = \frac{\bar{I}_0}{2} - \frac{\bar{V}_0}{2 \cdot \bar{Z}_c} \end{cases} \quad \text{ou} \quad \begin{cases} \bar{V}(x) = \bar{V}_0 \cdot \cos(\mathbf{b} \cdot x) - \bar{Z}_c \cdot \bar{I}_0 \cdot \sin(\mathbf{b} \cdot x) \\ \bar{I}(x) = \bar{I}_0 \cdot \cos(\mathbf{b} \cdot x) - \frac{\bar{V}_0}{\bar{Z}_c} \cdot \sin(\mathbf{b} \cdot x) \end{cases}.$$

Il est utile d'établir, en fonction de l'impédance caractéristique \underline{Z}_c et de l'impédance de charge \underline{Z}_l , les caractéristiques des tensions et courants en chaque points de la ligne. Il est donc nécessaire d'effectuer un changement de repère et prendre comme origine l'extrémité de la ligne soit : $x=d$:



Changement des origines pour la ligne

Les expressions précédentes sur les tensions et les courants deviennent alors :

$$\begin{cases} \frac{\partial \bar{V}}{\partial x} = +r \cdot \bar{i} + j l \cdot \mathbf{w} \bar{I} \\ \frac{\partial \bar{I}}{\partial x} = -g \cdot \bar{v} - j c \cdot \mathbf{w} \bar{V} \end{cases} \quad \text{ce qui conduit à :} \quad \begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} + \bar{V}_r \times e^{-j\mathbf{b} \cdot x} \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} + \bar{I}_r \times e^{-j\mathbf{b} \cdot x} \end{cases}.$$

En exprimant courants et tensions uniquement à partir de l'onde incidente, les équations précédentes deviennent :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{V}_r}{\bar{V}_i} \times e^{-2j\mathbf{b} \cdot x} \right) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{V}_d - \bar{Z}_c \bar{I}_d}{\bar{V}_d + \bar{Z}_c \bar{I}_d} \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{I}_r}{\bar{I}_i} \times e^{-2j\mathbf{b} \cdot x} \right) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{Z}_c \bar{I}_d - \bar{V}_d}{\bar{Z}_c \bar{I}_d + \bar{V}_d} \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases} \quad \text{donc,}$$

en remarquant que $V_d/I_d = Z_l$, les relations suivantes apparaissent :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{V}_d/\bar{I}_d - \bar{Z}_c}{\bar{V}_d/\bar{I}_d + \bar{Z}_c} \times e^{-2j\mathbf{b} \cdot x} \right) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c} \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{Z}_c - \bar{V}_d/\bar{I}_d}{\bar{Z}_c + \bar{V}_d/\bar{I}_d} \times e^{-2j\mathbf{b} \cdot x} \right) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \frac{\bar{Z}_c - \bar{Z}_l}{\bar{Z}_c + \bar{Z}_l} \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases}$$

Il est alors d'usage de définir Γ_r , appelé coefficient de réflexion, avec : $\bar{\Gamma}_r = \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c}$.

Notons que $|\Gamma_r| \leq 1$. Nous utiliserons également la relation : $\Gamma_r = |\Gamma_r| e^{j\phi}$.

Deux cas peuvent alors se présenter : soit $\Gamma_r = 0$, soit $\Gamma_r \neq 0$.

Coefficient de réflexion nul : $\Gamma_r = 0$ ou adaptation d'impédance (Impedance matching).

Nous voyons alors que : $\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \end{cases}$. Cette équation entraîne deux remarques

essentielles :

- courant et tension sont en phase, il n'y a donc pas de puissance réactive. Toute la puissance est transmise à la charge.
- Il n'existe pas d'onde réfléchie.

Il s'agit donc d'un cas extrêmement favorable car en HF, la puissance coûte chère, il est donc impératif qu'elle transite intégralement de la source à la charge (du générateur à l'antenne, par exemple).

Il est d'usage de qualifier $V(x)$ d'onde progressive.

Coefficient de réflexion non-nul : $\Gamma_r \neq 0$ ou adaptation d'impédance (Impedance mismatch).

L'expression de la tension et du courant en chaque point de la ligne est donné par :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 - \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases}$$

Cette fois-ci, $\underline{V}(x)$ est différente de \underline{V}_i . Il y a donc une onde réfléchie et une onde incidente.

De plus, nous pouvons remarquer que $|\underline{V}(x)|$ et $|\underline{I}(x)|$ sont des fonctions périodique de x . Il est simple de montrer que la période du module du courant et de la tension est :

$$\frac{2\pi}{2\mathbf{b}} = \frac{\lambda}{2}.$$

Le fait que le module de $V(x)$, seule grandeur simplement accessible à la mesure, soit périodique conduit à la dénomination d'ondes stationnaires. S'il était possible de mesurer

courants et tensions en tout point de la ligne, on observerait des noeuds et des ventres, exactement comme lors de la propagation d'une déformation mécanique le long d'une corde.

Nous pouvons alors revenir à deux cas classiques : $Z_l=0$ et $Z_l=\infty$:

$Z_l=0$ (court-circuit)

Dans ce cas, $\bar{\Gamma}_r = \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c} = -1$. Les relations :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 - \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases} \text{ devient : } \begin{cases} \bar{V}(x) = \bar{V}_i \times \left(e^{+j\mathbf{b} \cdot x} - e^{-j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times \left(e^{+j\mathbf{b} \cdot x} + e^{-j\mathbf{b} \cdot x} \right) \end{cases}.$$

Le courant et la tension sont donc déphasés de $\pi/2$ en tout point de la ligne. $\underline{V}(0)$ est nulle (ce qui est heureux!!!) et $\underline{I}(0)$ est maximum. Le point de la ligne $x=0$ correspond à un noeud en tension et un ventre en courant.

$Z_l=\infty$ (circuit ouvert)

Dans ce cas, $\bar{\Gamma}_r = \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c} = +1$. Les relations :

$$\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 - \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases} \text{ deviennent : } \begin{cases} \bar{V}(x) = \bar{V}_i \times \left(e^{+j\mathbf{b} \cdot x} + e^{-j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times \left(e^{+j\mathbf{b} \cdot x} - e^{-j\mathbf{b} \cdot x} \right) \end{cases}.$$

Le courant et la tension sont donc déphasés de $\pi/2$ en tout point de la ligne. $\underline{V}(0)$ est maximum et $\underline{I}(0)$ est nul ce qui est cohérent avec ce que l'on attendait.

Rapport d'ondes stationnaires (ROS)

Les relations : $\begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} \left(1 + \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} \left(1 - \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x} \right) \end{cases}$ permettent de trouver une

amplitude de tension (respect. de courant) maximale que nous appellerons V_M et une amplitude de tension (respect. de courant) minimale, que nous appellerons V_m . Ceux-ci sont donnés par :

$$\begin{cases} V_M = V_i (1 + |\bar{\Gamma}_r|) & V_m = V_i (1 - |\bar{\Gamma}_r|) \\ I_M = I_i (1 + |\bar{\Gamma}_r|) & I_m = I_i (1 - |\bar{\Gamma}_r|) \end{cases}$$

Le rapport d'ondes stationnaires ou SWR (Standing Wave Ratio), également Taux d'Ondes Stationnaires (TOS), est la quantité r définie par $V_M/V_m = I_M/I_m$. Etant donné que les mesures de tensions (ou de courant) ne sont pas toujours simples à effectuer sur une ligne, il est également utile d'exprimer r en fonction du coefficient de réflexion :

$$r = \frac{1 + |\bar{\Gamma}_r|}{1 - |\bar{\Gamma}_r|}.$$

Notons que : $1 \leq r \leq \infty$.

Plus r se rapproche de 1, plus $|\bar{\Gamma}_r|$ est proche de 0 et donc plus le mode de fonctionnement de la ligne est voisin du fonctionnement en ondes progressive. A l'inverse,

lorsque r est grand, plus on se rapproche d'un mode de fonctionnement en ondes stationnaires pur (court-circuit ou circuit ouvert).

ABaque DE SMITH ET ADAPTATION

Généralités. Notion d'impédance réduite

Nous avons vu qu'il était essentiel, dans la majorité des cas, de parvenir à adapter la ligne et la charge. De plus, nous avons vu que :

$\bar{\Gamma}_r = \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c}$. Il est possible de réécrire cette équation de la manière suivante :

$$\bar{\Gamma}_r = \frac{\frac{\bar{Z}_l}{\bar{Z}_c} - 1}{\frac{\bar{Z}_l}{\bar{Z}_c} + 1}.$$

On appelle alors impédance réduite, z_l , la grandeur : $z_l = \bar{Z}_l / \bar{Z}_c$. Nous verrons par la suite l'utilité de cette normalisation par rapport à l'impédance caractéristique de la ligne. L'expression du coefficient de réflexion est alors ré-écrite sous la forme :

$$\bar{\Gamma}_r = \frac{z_l - 1}{z_l + 1}.$$

Nous pouvons également exprimer l'impédance réduite de la charge en fonction du coefficient de réflexion :

$$z_l = \frac{1 + \bar{\Gamma}_r}{1 - \bar{\Gamma}_r}.$$

Ces relations ont été données pour $x=l$ mais elles sont valables quel que soit le point de la ligne :

$$\bar{z}(x) = \frac{1 + \bar{\Gamma}(x)}{1 - \bar{\Gamma}(x)} \Leftrightarrow \bar{\Gamma}(x) = \frac{\bar{z}(x) - 1}{\bar{z}(x) + 1}$$

Nous disposons donc du moyen d'adapter la ligne après manipulation de ces équations. Il s'agit là d'un travail relativement ardu c'est pourquoi, il a été mis au point une technique graphique qui permet de déterminer $z(x)$ à partir de $\bar{\Gamma}$ et inversement. Malgré les progrès des calculateurs, cette technique reste très employée car elle permet une évaluation qualitative des mérites de l'adaptation réalisée.

Principe de l'abaque de Smith

Le problème fixé est : connaissant $\bar{\Gamma}$, comment déterminer z ?

Construction de l'abaque

Nous pouvons écrire : $\bar{\Gamma} = p + jq$ et $z = r + jy$. Donc connaissant p et q , il faut pouvoir déterminer r et y .

L'équation : $\bar{\Gamma}(x) = \frac{\bar{z}(x) - 1}{\bar{z}(x) + 1}$ devient alors : $p + jq = \frac{r + jy - 1}{r + jy + 1}$.

$$p + jq = \frac{r + jy - 1}{r + jy + 1} \Leftrightarrow \begin{cases} p(r+1) - yq = r - 1 \\ yp + (r+1)q = y \end{cases}$$

pour éliminer successivement x puis r du système :

Ainsi :

$$\begin{cases} p(r+1) - yq = r-1 \\ yp + (r+1)q = y \end{cases} \Leftrightarrow \begin{cases} p(p-1)(r+1) - (r-1)(p-1) = yq(p-1) \\ -(r+1)q^2 = y(p-1)q \end{cases}$$

$$\Leftrightarrow p(p-1)(r+1) - (r-1)(p-1) = -(r+1)q^2$$

$$\text{donc : } (r+1)p^2 - 2rp + (r+1)q^2 = -r+1$$

$$\text{d'où : } p^2 - 2\frac{r}{(r+1)}p + q^2 = \frac{-r+1}{(r+1)}$$

$$\text{enfin : } p^2 - 2\frac{r}{r+1}p + \left(\frac{r}{r+1}\right)^2 + q^2 = \frac{-r+1}{r+1} + \left(\frac{r}{r+1}\right)^2 = \left(\frac{1}{r+1}\right)^2$$

et :

$$\begin{cases} p(r+1) - yq = r-1 \\ yp + (r+1)q = y \end{cases} \Leftrightarrow \begin{cases} qp - yq^2 + q = q(1-p)r \\ yp(p-1) - y(p-1) + q(p-1) = qr(1-p) \end{cases}$$

$$\Leftrightarrow qp - yq^2 + q = yp^2 - 2yp + y + qp - q$$

$$\text{d'où : } -yq^2 + 2q - yp^2 + 2yp = y,$$

$$\text{donc : } -q^2 + 2\frac{q}{y} - p^2 + 2p = 1,$$

$$\text{d'où : } -q^2 + 2\frac{q}{y} - \frac{1}{y^2} - p^2 + 2p - 1 = 1 - 1 - \frac{1}{y^2},$$

Ainsi, ces équations peuvent se mettre sous la forme :

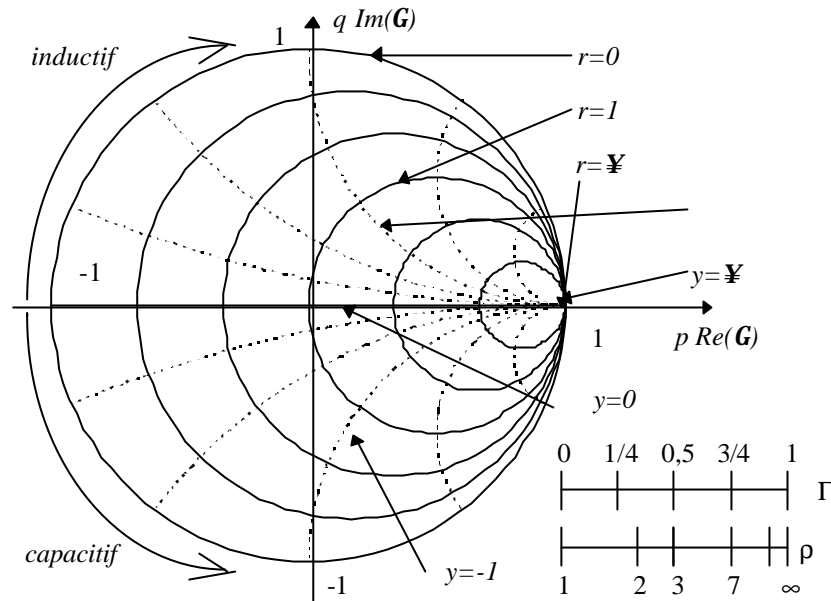
$$\left(p - \frac{r}{1+r}\right)^2 + q^2 = \left(\frac{1}{1+r}\right)^2 \quad \text{et} \quad (p-1)^2 + \left(q - \frac{1}{y}\right)^2 = \frac{1}{y^2}.$$

A partir de p et q , il est alors possible de déterminer r et y à partir de ces deux relations, mais il est plus judicieux de remarquer que pour la relation en r , celle-ci définit des familles de cercles de rayon $\frac{1}{1+r}$, centrés en $(\text{Re}(\Gamma) = \frac{r}{1+r}, \text{Im}(\Gamma) = 0)$. La relation en y définit, quant à elle une famille de cercles de rayon $\frac{1}{y}$ et de centre $(\text{Re}(\Gamma) = 1, \text{Im}(\Gamma) = \frac{1}{y})$.

Quatre cas particuliers sont intéressants :

- $r=0$: le cercle est alors un cercle de rayon 1, centré en $(0,0)$ ($|\Gamma|=1$ quelle que soit y).
On retrouve alors le fait qu'une ligne en court-circuit à un comportement purement réactif.
- $r=\infty$: le cercle est alors un cercle de rayon 0 (un point), centré en $(1,0)$.
- $y=\infty$: le cercle est alors un cercle de rayon 0 (un point), centré en $(1,0)$.
- $y=0$: le cercle est alors un cercle de rayon infini, centré en $(1,\infty)$.

Il est alors possible de représenter les 2 familles de cercles, comme le montre la figure ci-dessous :



Abaque de Smith.

Propriétés de l'abaque

A partir de cet abaque, connaissant \underline{z} , on peut déterminer Γ et réciproquement. En plus des cercles à $r=C^{te}$ et à $y=C^{te}$, les abaques de Smith comportent également des échelles sur leurs bords permettant de déterminer Γ , le module du coefficient de réflexion, ainsi que r , le ROS.

Cercle à $\rho=C^{te}$:

Lorsque $|\Gamma|=C^{te}$, les lieux de Γ forment un cercle de centre (0,0) et de rayon $|\Gamma|$ (seul $\text{Arg}(\Gamma)$ varie). Or le ROS est donné par :

$$r = \frac{1 + |\Gamma_r|}{1 - |\Gamma_r|}.$$

Donc les lieux à $\text{ROS}=C^{te}$ sont donnés par le cercle $|\Gamma|=C^{te}$.

Evolution de $\underline{G}(x)$ en fonction de $\underline{Z}(x)$

Nous avons vu que : $\bar{\Gamma}(x) = \frac{\bar{Z}(x) - \bar{Z}_c}{\bar{Z}(x) + \bar{Z}_c}$ et $\bar{Z}(x) = \bar{Z}_c \frac{\bar{Z}_l + j \bar{Z}_c \tan(\beta x)}{\bar{Z}_c + j \bar{Z}_l \tan(\beta x)}$. Donc :

$$\bar{\Gamma}(x) = \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_c + \bar{Z}_l} \times (\cos(2\beta x) - j \sin(2\beta x)).$$

Cette relation est fondamentale car elle permet d'établir :

- que le module coefficient de réflexion reste constant indépendamment de la longueur de la ligne,
- que les lieux de $\Gamma(x)$ forment un cercle de rayon $|\bar{\Gamma}| = \left| \frac{\bar{Z}_l - \bar{Z}_c}{\bar{Z}_l + \bar{Z}_c} \right|$ et parcouru dans le sens

inverse du sens inverse du sens trigonométrique lorsqu'on se déplace de la charge vers le générateur (mais dans le sens trigonométrique quand on se déplace du générateur vers la charge),

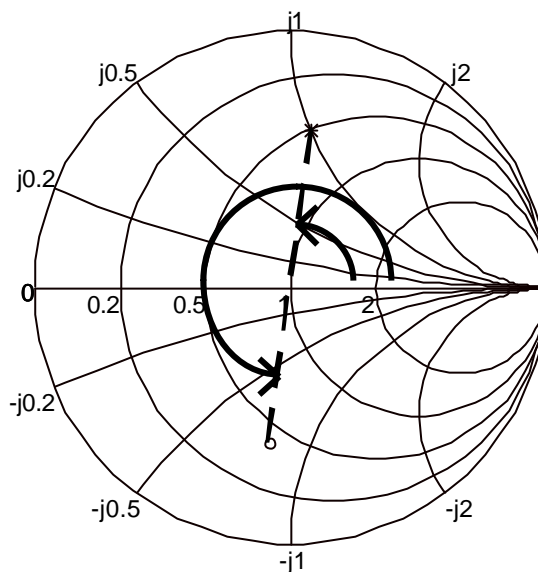
- que les lieux sont périodiques de période $\lambda/2$, autrement dit, un déplacement de $\lambda/2$ sur la ligne correspond à un déplacement d'un tour complet sur un cercle à $|\Gamma|=C^{te}$ de l'abaque de Smith. Il est donc normal de faire figurer des angles, exprimés sous forme de longueur d'onde, sur les bords extérieurs de l'abaque.

Utilisation de l'abaque

Voici quelques exemples de l'utilisation de l'abaque :

Détermination du coefficient de réflexion connaissant Z :

La première opération consiste à calculer l'impédance réduite : $z = Z/Z_c$. Ensuite, à l'aide des lieux en $r=C^{te}$ et $y=C^{te}$, on fait figurer sur l'abaque les impédances réduites : $z_1=0,5+i$ et $z_2=1/(0,5+i)$:



l'impédance z_1 est représentée par un "", z_2 par "o".*

Nous voyons que pour déterminer $\underline{\Gamma}$, il suffit de relever son module (pointillés) et sa phase (arc de cercle). Nous pouvons noter que $z_2=1/z_1$ que $|\Gamma_1|=|\Gamma_2|$ et $\psi_2=\psi_1+\pi$. En d'autre terme, les deux coefficients de réflexions sont placés de façon symétrique par rapport au point (0,0).

La figure précédente nous permet également de faire une remarque importante : étant donné que l'impédance réduite a été normalisée par rapport à Z_c , admittance réduite et impédance réduite sont diamétralement opposées.

Détermination de l'impédance réduite z connaissant coefficient de réflexion \underline{G} :

La méthode est pratiquement identique : il faut commencer par placer $\underline{\Gamma}$ ($|\Gamma|$ et ψ), puis de lire les valeurs de r et y .

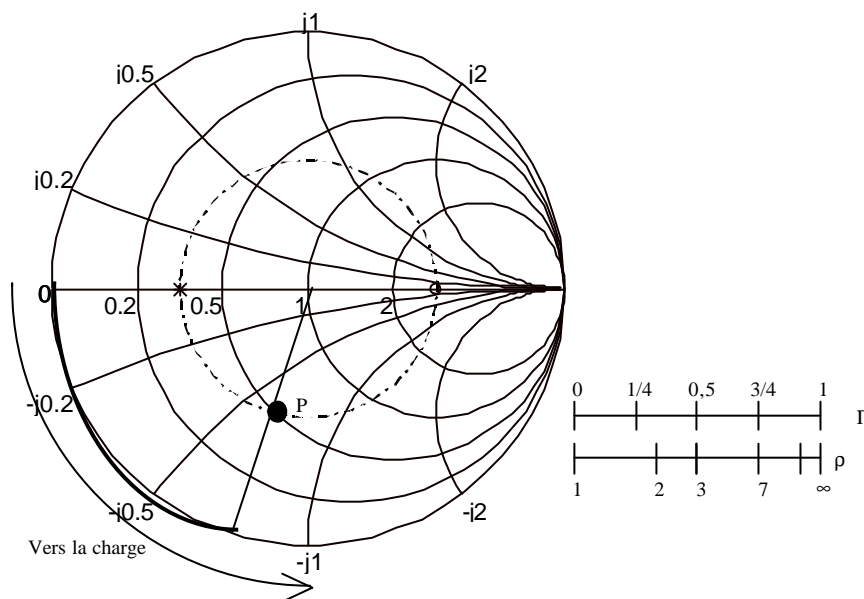
Détermination de l'impédance de charge d'une ligne à partir du Rapport d'Ondes

Stationnaires r :

Pour déterminer l'impédance de charge d'une ligne à partir du ROS, il faut commencer par tracer le cercle en ROS $r=C^{te}$. Prenons par exemple $r=3$. Après avoir déterminer le

module de Γ correspondant, on peut tracer le cercle des lieux donnant $r=3$. On place alors le point M (repéré par un 'o') sur l'axe des $r : 1 \rightarrow \infty$. Ce point correspond à un maximum de tension. Le point N (repéré par un '*'), diamétralement opposé par rapport au point $(r=1, y=0)$, correspond à un minimum de tension. En mesurant la distance x_M séparant la charge du point où le minimum de tension a été relevé, on peut déduire l'impédance de la charge en tournant sur le cercle vers la charge d'un angle x_M/λ . On peut alors placer le point P dont les coordonnées en (r, y) donnent l'impédance réduite de la charge. L'impédance réelle est alors obtenue en dénormalisant l'impédance réduite.

$$\text{Rappel : } \begin{cases} \bar{V}(x) = \bar{V}_i \times e^{+j\mathbf{b} \cdot x} (1 + \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x}) \\ \bar{I}(x) = \bar{I}_i \times e^{+j\mathbf{b} \cdot x} (1 - \bar{\Gamma}_r \times e^{-2j\mathbf{b} \cdot x}) \end{cases}$$



En pointillés, lieu des points vérifiant $\rho=3$. Les points M et N sont repérés par un 'o' et un '*'.

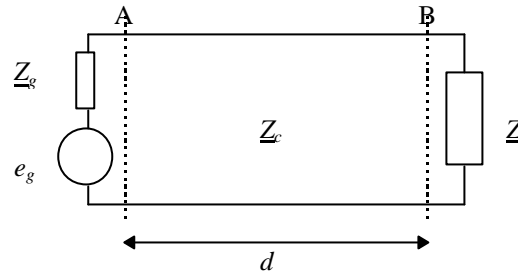
Principe et techniques d'adaptation d'une ligne

Enjeux

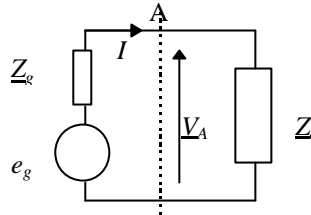
Les lignes constituent un média servant à transmettre un signal d'un générateur vers une charge, telle que l'entrée d'un amplificateur ou une antenne. Aux fréquences élevées, la puissance coûte chère, aussi est-il nécessaire que le générateur transmette le maximum de puissance et que la charge absorbe le maximum de puissance. Par ailleurs, nous avons vu que, dans le cas général, le phénomène de propagation se traduit par l'apparition d'une onde réfléchie. Pour ne pas risquer d'endommager le générateur, il est impératif d'éviter d'avoir de fort ROS.

Conditions d'adaptation

Le générateur, la ligne et la charge forment un ensemble représenté ci-dessous :



Considérons Z_e , l'impédance ramenée par une ligne chargée, au niveau du générateur. Nous avons alors le schéma équivalent ci-dessous :



Nous appellerons P la puissance active délivrée à Z_e par le générateur. P est donnée par :

$$P = \frac{1}{2} \Re(\bar{V}_A \bar{I}^*) \text{ or } \bar{V}_A = \bar{Z}_e \bar{I} = \bar{e}_g - \bar{Z}_g \bar{I}. \text{ En exprimant } P \text{ en fonction de } \underline{e}_g, Z_e \text{ et } Z_g :$$

$$P = \frac{1}{2} \Re \left[\bar{Z}_e \left(\frac{\bar{e}_g}{\bar{Z}_e + \bar{Z}_g} \right) \times \left(\frac{\bar{e}_g}{\bar{Z}_e + \bar{Z}_g} \right)^* \right].$$

Posons : $\bar{Z}_e = R_e + j X_e$ et $\bar{Z}_g = R_g + j X_g$. On obtient alors :

$$P = \frac{1}{2} \Re \left[\bar{Z}_e \left(\frac{\bar{e}_g}{R_e + j X_e + R_g + j X_g} \right) \times \left(\frac{\bar{e}_g}{R_e + j X_e + R_g + j X_g} \right)^* \right], \text{ soit :}$$

$$P = \frac{1}{2} \times \frac{R_e \times |\bar{e}_g|^2}{(R_e + R_g)^2 + (X_e + X_g)^2}.$$

Pour transmettre une puissance P maximale lorsque Z_g est fixée, il faut que $X_e = -X_g$ et $R_e = R_g$. En résumé : $\bar{Z}_e = \bar{Z}_g^*$.

Si nous nous plaçons maintenant en B, et en supposant les lignes sans pertes (Z_c est réelle), un raisonnement identique permet de montrer que la puissance active en x est donnée par :

$$P(x) = \frac{|V_i|^2}{2 Z_c} (1 - |\Gamma(x)|^2). \text{ Donc, au niveau de la charge :}$$

$$P_0 = \frac{|V_i|^2}{2 Z_c} (1 - |\Gamma|^2).$$

De ce fait, pour que la puissance soit maximale, il faut que $\Gamma = 0$. A contrario, lorsque $|\Gamma|$ est voisin de 1, la puissance active délivrée à la charge est voisine de 0. Cela revient à

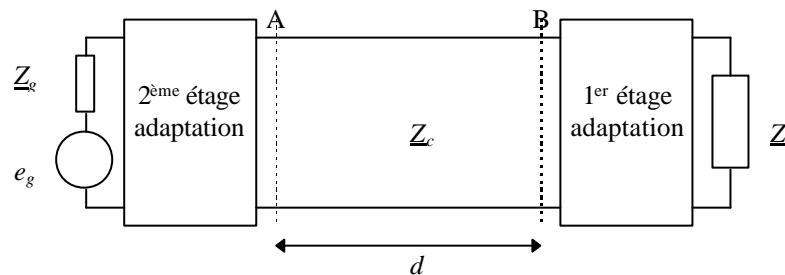
dire qu'il n'y a que de la puissance réactive, ce que nous avons vu précédemment. Or pour avoir $\Gamma=0$, il faut :

$$\bar{Z}_c = \bar{Z}_l.$$

Pour résumé, la transmission du maximum de puissance active implique les conditions :

$$\begin{cases} \bar{Z}_c = \bar{Z}_g^* \\ \bar{Z}_c = \bar{Z}_l \end{cases}.$$

Remplir simultanément ces deux conditions n'est que très rarement possible. De ce fait, afin d'éviter l'apparition d'ondes stationnaires sur la ligne, un étage d'adaptation est inséré entre la charge et la ligne, et, pour permettre au générateur de délivrer le maximum de puissance possible, un second étage d'adaptation est parfois nécessaire. En général, les amplificateurs et les générateurs sont conçus pour présenter une impédance de sortie de $50\ \Omega$ ou de $75\ \Omega$, compatible avec l'impédance caractéristique des lignes. En revanche, la charge ne peut pas toujours présenter une impédance normalisée (cas d'une antenne). Le schéma général du modèle d'une ligne est alors celui indiqué ci-dessous :

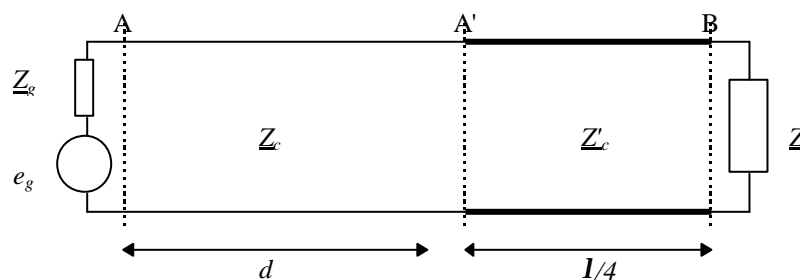


Ligne adaptée en entrée et en sortie (charge et générateur).

Techniques d'adaptation

Adaptation par ligne $\lambda/4$.

Une technique consiste à insérer une ligne $\lambda/4$, d'impédance Z'_c , entre la charge et la ligne :



Nous avons vu que l'impédance Z'_A ramenée par cette ligne $\lambda/4$ en A' est : $\bar{Z}'_A = \frac{Z'^2_c}{\bar{Z}_l}$.

Pour réaliser l'adaptation d'impédance en A', il faut alors que :

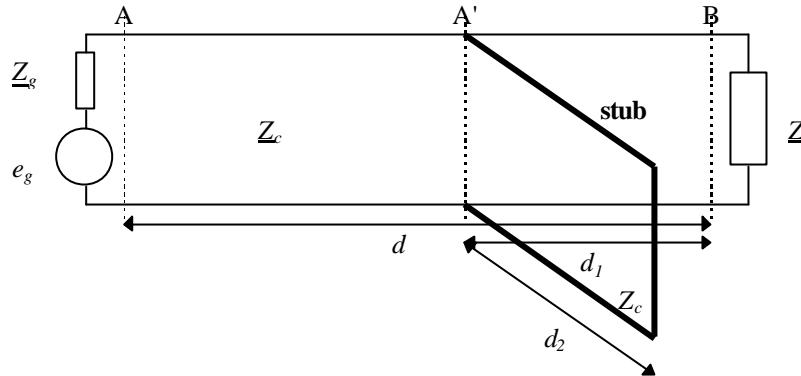
$$\bar{Z}'_A = Z_c = \frac{Z'^2_c}{\bar{Z}_l}, \text{ c'est-à-dire : } \bar{Z}'_c = \sqrt{Z_c \bar{Z}_l}$$

Notons que le tronçon $\lambda/4$ est le siège d'ondes stationnaires alors que le reste de la ligne fonctionne en régime d'ondes progressives.

Dans le cas où Z_L n'est pas réelle, il est nécessaire d'intercaler la ligne $\lambda/4$ à une d_1 distance de la charge de telle sorte que l'impédance ramenée à l'entrée de la ligne $\lambda/4$ soit réelle.

Adaptation par un stub.

Un stub est bout de ligne court-circuitée et placée en parallèle sur la ligne. Le modèle de la ligne adaptée à l'aide d'un stub est donné ci-dessous :



Ligne adaptée par un stub.

Le but du stub est de placer en A' une admittance \underline{Y}_s purement imaginaire qui ajoutée à l'admittance ramenée en A' par le tronçon de longueur d_1 soit égale à l'admittance caractéristique $1/Z_c$ de la ligne. L'impédance ramenée en A' par le stub est :

$$\bar{Z}_s = Z_c j \operatorname{tg}(\mathbf{b} d_2), \text{ donc l'admittance est : } \bar{Y}_s = \frac{-j}{Z_c \operatorname{tg}(\mathbf{b} d_2)}.$$

L'admittance ramenée par le tronçon de longueur d_1 en A' est $\underline{Y}_{A'}$ donnée par :

$$\bar{Y}_{A'} = \frac{1}{\bar{Z}_{A'}} = \frac{1}{Z_c} \frac{Z_c + j \bar{Z}_L \operatorname{tg}(\mathbf{b} d_1)}{\bar{Z}_L + j Z_c \operatorname{tg}(\mathbf{b} d_1)}.$$

Il faut alors trouver les longueurs des tronçons d_1 et d_2 telles que :

$$\bar{Y}_{A'} + \bar{Y}_s = \frac{1}{Z_c} \Leftrightarrow \frac{1}{Z_c} \times \frac{Z_c + j \bar{Z}_L \operatorname{tg}(\mathbf{b} d_1)}{\bar{Z}_L + j Z_c \operatorname{tg}(\mathbf{b} d_1)} + \frac{1}{j Z_c \operatorname{tg}(\mathbf{b} d_2)} = \frac{1}{Z_c}.$$

Le problème est plus simple si l'on utilise les impédances (ou admittances) réduites :

$$\bar{z}_{A'} = \frac{\bar{Z}_{A'}}{Z_c} \Leftrightarrow \bar{y}_{A'} = \bar{Y}_{A'} Z_c \text{ et } \bar{z}_s = \frac{\bar{Z}_s}{Z_c} \Leftrightarrow \bar{y}_s = \bar{Y}_s Z_c.$$

Remarquons encore une fois que les grandeurs réduites sont sans dimensions.

Les équations précédentes deviennent alors :

$$\bar{y}_{A'} = \frac{1}{\bar{z}_{A'}} = \frac{1 + j \bar{z}_L \operatorname{tg}(\mathbf{b} d_1)}{\bar{z}_L + j \operatorname{tg}(\mathbf{b} d_1)}.$$

$$\bar{y}_s = \frac{1}{\bar{z}_s} = \frac{1}{j \operatorname{tg}(\mathbf{b} d_2)}$$

$$\bar{y}_{A'} + \bar{y}_s = y_c = 1 \Leftrightarrow \frac{1 + j \bar{z}_L \operatorname{tg}(\mathbf{b} d_1)}{\bar{z}_L + j \operatorname{tg}(\mathbf{b} d_1)} + \frac{1}{j \operatorname{tg}(\mathbf{b} d_2)} = 1.$$

Pour cela, il faut que :

$$\begin{cases} \Re e \left(\frac{1 + j \bar{z}_l \operatorname{tg}(\mathbf{b} d_1)}{\bar{z}_l + j \operatorname{tg}(\mathbf{b} d_1)} \right) = 1 \\ \Im m \left(\frac{1 + j \bar{z}_l \operatorname{tg}(\mathbf{b} d_1)}{\bar{z}_l + j \operatorname{tg}(\mathbf{b} d_1)} \right) = -\frac{1}{j \operatorname{tg}(\mathbf{b} d_2)} \end{cases}, \text{ soit, en posant } \underline{z}_l = r + j x :$$

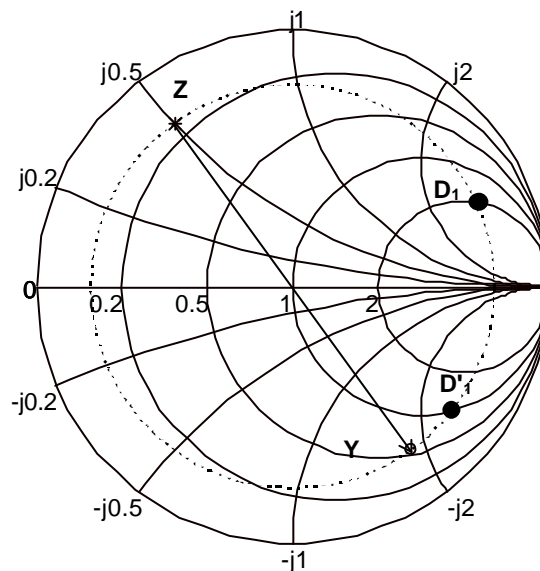
$$\begin{cases} \Re e \left(\frac{1 + j(r + jx) \operatorname{tg}(\mathbf{b} d_1)}{r + jx + j \operatorname{tg}(\mathbf{b} d_1)} \right) = 1 \\ \Im m \left(\frac{1 + j(r + jx) \operatorname{tg}(\mathbf{b} d_1)}{r + jx + j \operatorname{tg}(\mathbf{b} d_1)} \right) = -\frac{1}{j \operatorname{tg}(\mathbf{b} d_2)} \end{cases}$$

$$\Leftrightarrow \begin{cases} \Re e \left(\frac{[1 - x \operatorname{tg}(\mathbf{b} d_1) + j r \operatorname{tg}(\mathbf{b} d_1)](r - jx - j \operatorname{tg}(\mathbf{b} d_1))}{r^2 + (x + \operatorname{tg}(\mathbf{b} d_1))^2} \right) = 1 \\ \Im m \left(\frac{[1 - x \operatorname{tg}(\mathbf{b} d_1) + j r \operatorname{tg}(\mathbf{b} d_1)](r - jx - j \operatorname{tg}(\mathbf{b} d_1))}{r^2 + (x + \operatorname{tg}(\mathbf{b} d_1))^2} \right) = -\frac{1}{j \operatorname{tg}(\mathbf{b} d_2)} \end{cases}$$

Nous avons deux équations à deux inconnues. La première relation impose la longueur d_1 tandis que la seconde permet de trouver d_2 à partir de d_1 .

Plutôt que de résoudre le calcul de façon formelle, il est plus simple d'utiliser l'abaque de Smith et de déterminer d_1 et d_2 par construction graphique :

La première étape est de placer le point **Z** correspondant à l'impédance réduite de la charge et en déduire le lieu **Y** de l'admittance réduite, placée de façon symétrique à l'impédance réduite par rapport au point d'impédance (1,0). Il est d'ailleurs utile de remarquer que ce point correspond à une charge adaptée. Tout le problème se résumera donc à parvenir à ce point.

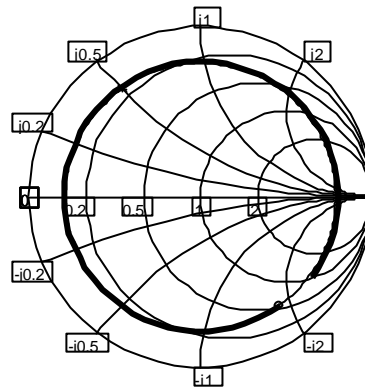
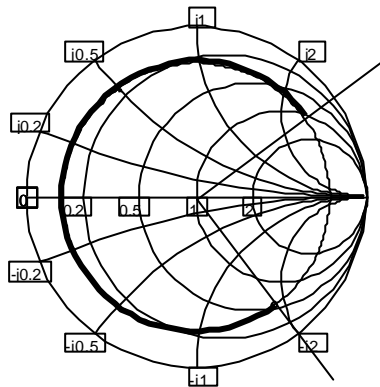


Abaque avec les différents points **Z**, **Y**, **D₁** et **D'₁**. Les lieux à $ROS = C^{te}$ figurent en pointillés.

Pour déterminer d_1 , il faut tout d'abord remarquer que l'impédance ramenée en A' par le tronçon d_1 fait parti du cercle à ROS=C^{te}. Par ailleurs, nous avons vu que l'admittance ramenée doit vérifier :

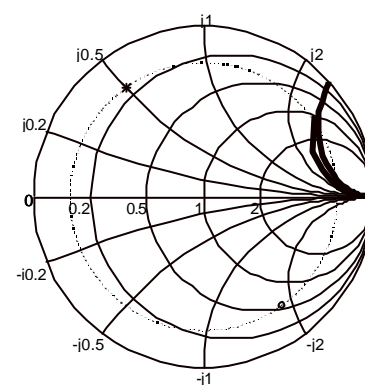
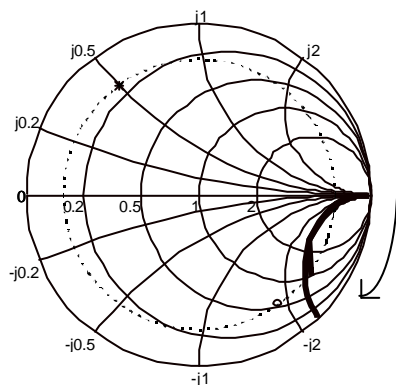
$$\Re\left(\frac{1 + j \bar{z}_l \operatorname{tg}(b d_1)}{\bar{z}_l + j \operatorname{tg}(b d_1)}\right) = 1.$$

La longueur du tronçon d_1 est donc telle qu'elle permet de passer du point **Y**, aux points **D**₁ et **D**'₁ situés sur le cercle $r=1$. Etant donné que le système vu précédemment est un système du second degré, il est normal de trouver deux points correspondants aux deux solutions (en fait, il en existe une infinité puisque que la ligne présente un comportement périodique de période $\lambda/2$). La longueur du tronçon d_1 est déduit de l'abaque en lisant la longueur qu'il faut parcourir pour passer de **Y** aux points **D**₁ et **D**'₁ en tournant vers le générateur (on s'est déplacé de la charge vers le générateur).



*Longueur de tronçon à insérer pour ramener une impédance correspondant aux points **D**₁ et **D**'₁.*

Pour déterminer d_2 , il suffit de déterminer l'admittance permettant de passer des points **D**₁ et **D**'₁ au point ($r=1, y=0$). Là encore, la distance d_2 est déterminée en lisant la distance en fraction de λ .



Les impédances des deux tronçons de stub figurent en gras.

Le choix de l'une ou l'autre des deux solutions ne dépend que de la facilité de la réalisation matérielle de l'adaptation. Un tronçon très court (quelques mm) sera impossible à réaliser (il est toutefois possible d'avoir des longueurs $k \lambda/2$ plus longues), alors qu'un tronçon

trop long entraînera un encombrement important et rajoutera des pertes. Notons que ces techniques d'adaptation sont transposables à toutes les fréquences. En particulier, aux très hautes fréquences (quelques Gigahertz), ce sont les pistes conduisant aux éléments amplificateurs qui sont utilisées pour effectuer l'adaptation par stub. L'adaptation d'impédance ne peut se faire qu'à une fréquence donnée. Or, les signaux ne sont jamais monochromatiques. Il est donc nécessaire de tracer le ROS en fonction de la fréquence sur l'abaque de Smith. Le but est alors que le ROS reste proche de 1, ($|\Gamma|$ proche de 0), ce qui conduit à la meilleure l'adaptation possible dans la plage de fréquences considérées.

Le stub constitue en fait une impédance purement imaginaire, il peut être remplacée par un élément discret telle qu'une self ou un condensateur.

16. ANTENNES **ET** **TRANSMISSIONS HERTZIENNES**

GENERALITES

La quantité d'informations que l'on peut transporter dépend de nombreux facteurs : la distance entre deux points d'une liaison, la sensibilité aux perturbations, la puissance que l'on peut allouer pour les émetteurs, etc.... Le type de communication que l'on souhaite réaliser peut s'accommoder d'une infrastructure plus ou moins lourde (câbles, antennes et pylônes, répéteurs...). Enfin, en fonction de l'application ou de la topologie du milieu, il peut être plus ou moins simple ou plus ou moins économique de tirer des lignes. Il peut donc y avoir transmission en espace libre ou sur des supports spécifiques (transmission guidée). Pour ces derniers, l'optique tend à se généraliser, alors que les transmissions en espace libre restent l'apanage des ondes électromagnétiques.

TRANSMISSIONS HERTZIENNES

Généralités

Les télécommunications hertziennes permettent des liaisons point fixe à point fixe (relais téléphoniques, relais de télévision, etc...) ou entre mobiles. La souplesse de l'infrastructure nécessaire permet de desservir des zones géographiques impropres aux communications filaires. En contrepartie, l'encombrement spectral limite le nombre des canaux de communication et ce d'autant plus que l'atmosphère impose ses propres contraintes. Les fréquences des systèmes de télécommunication sont donc attribuées par des organismes de normalisation tels que l'UIT-R et l'IFBR (International Frequency Registration Board).

Antennes

Afin de générer des ondes électromagnétiques, il est nécessaire de disposer d'éléments transformant les signaux électriques en onde et réciproquement selon que l'antenne est utilisée en émission ou en réception. Les éléments de réception et de transmission d'ondes hertziennes sont des antennes.

Principe

Une onde électromagnétique est composée d'un champ électrique $\vec{E}(t, x, y, z)$ et d'un champ magnétique $\vec{B}(t, x, y, z)$. En cours de physique, il a été démontré que cette onde se propage si les D'Alembertiens de $\vec{E}(t, x, y, z)$ et $\vec{B}(t, x, y, z)$ sont nuls. D'autre part, les champs $\vec{E}(t, x, y, z)$ et $\vec{B}(t, x, y, z)$ obéissent aux équations de Maxwell. Nous avons donc :

$$\left\{ \begin{array}{l} \nabla^2 \vec{B} - \frac{1}{c_0^2} \cdot \frac{\nabla^2}{\nabla^2} \vec{B} = \vec{0} \\ \nabla^2 \vec{E} - \frac{1}{c_0^2} \cdot \frac{\nabla^2}{\nabla^2} \vec{E} = \vec{0} \end{array} \right. \text{ et } \left\{ \begin{array}{l} \vec{\nabla} \cdot \vec{B} = 0, \vec{\nabla} \cdot (\vec{e} \vec{E}) = \vec{r} \\ \vec{\nabla} \wedge \vec{E} = -\frac{\nabla \vec{B}}{\nabla t}, \vec{\nabla} \wedge \vec{B} = \vec{m} \vec{j} + \vec{m} \frac{\nabla}{\nabla t} (\vec{e} \vec{E}) \end{array} \right.$$

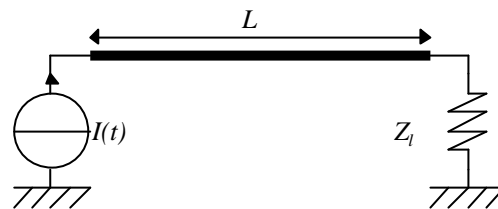
A partir de ces équations, il est possible de montrer que le champ électrique crée par une charge est :

$$\vec{E} = \frac{q}{4\pi \epsilon_0} \left[\frac{\vec{u}'}{r'^2} + \frac{r'}{c} \times \frac{d}{dt} \left(\frac{\vec{u}'}{r'^2} \right) + \frac{1}{c^2} \times \frac{d}{dt} \vec{u}' \right]$$

\vec{u} est le vecteur unitaire orienté du point d'observation de \vec{E} vers le point où se trouvait la charge à l'instant t .

Principe

Considérons un conducteur de longueur L , alimenté par une source de courant et terminé par une impédance de charge Z_l égale à l'impédance caractéristique de la ligne.



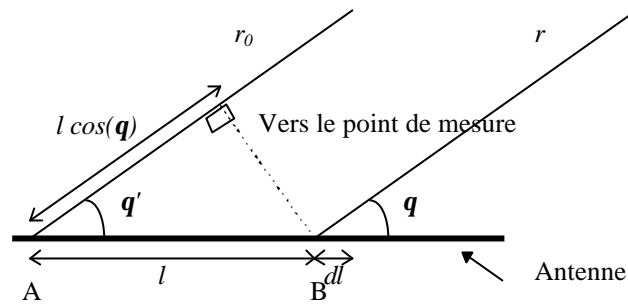
Le fil ainsi chargé constitue une ligne de transmission pour laquelle nous considérerons que les vitesses de phase et de groupe sont identiques et égales à c . Il y a donc propagation sans réflexion le long du conducteur. Le courant peut alors s'écrire :

$$I = I_0 e^{j(\omega t - \beta x)}.$$

Nous ferons plusieurs approximations afin de pouvoir obtenir un résultat analytique. La première approximation consiste à considérer qu' I_0 est constant le long de L . Cette approximation suppose qu'il n'y ait pas de rayonnement électromagnétique! En fait, le calcul numérique permet d'accepter cette approximation.

La seconde approximation consiste à considérer le champ créé en un point très éloigné de la source (c'est-à-dire à une distance r de l'antenne telle que r soit grande devant les dimension de l'antenne et de la longueur d'onde λ).

Pour calculer le rayonnement de l'antenne, nous pouvons donc supposer que les angles θ et θ' sont égaux. En revanche, les deux points A et B de l'antenne, séparés par une distance l , sont distants de r_0 et r du point où l'on mesure le champ. La différence de distance entre est $r_0 - r = l \cos(\theta)$. On obtient alors la figure suivante :



Le champ d'une antenne

En se souvenant des équations de Maxwell, et en particulier de relations telles que :

$$\operatorname{div} \vec{E} = \frac{\rho}{\epsilon_0}.$$

Le champ rayonné par l'élément dl parcouru par le courant I est donné par :

$$dE_q = \frac{-dl}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times \frac{dI}{dt}.$$

Or, en prenant le courant en A comme référence de phase, nous avons vu que le courant I est de la forme :

$$I = I_0 e^{j(\omega t - \beta l)}.$$

Par ailleurs, si le terme βl permet de tenir compte du déphasage dû à la distance l sur le courant, dans le calcul du champ, il faut également tenir compte du déphasage dû au temps de propagation jusqu'à la distance r . Nous avons donc :

$$dE_q = \frac{-dl}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times \frac{d}{dt} I_0 e^{j\left(\omega\left(t - \frac{r}{c}\right) - \beta l\right)}, \text{ soit, après dérivation :}$$

$$dE_q = \frac{-dl}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times j\omega I_0 e^{j\left(\omega\left(t - \frac{r_0 - l \cos(q)}{c}\right) - \beta l\right)}.$$

Le champ total E_q , à une distance r dans une direction q de l'antenne, est obtenu en intégrant par rapport à l , on obtient alors :

$$E_q = \frac{-1}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times j\omega I_0 e^{j\omega\left(t - \frac{r_0}{c}\right)} \times \int_0^L e^{-j\left(\beta l - \frac{\omega l \cos(q)}{c}\right)} dl.$$

Nous avons vu que : $\beta = \frac{\omega}{c}$, donc :

$$E_q = \frac{-1}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times j\omega I_0 e^{j(\omega t - \beta r_0)} \times \int_0^L e^{-j\beta l(1 - \cos(q))} dl, \text{ soit :}$$

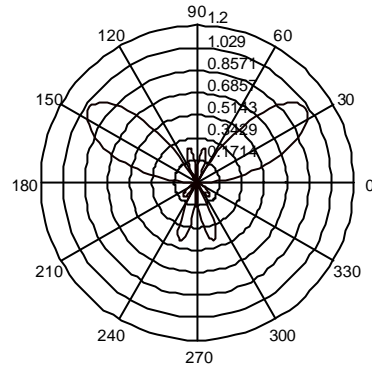
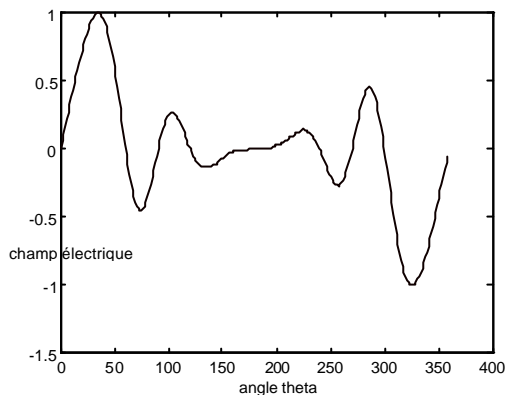
$$E_q = \frac{1}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times j\omega I_0 e^{j(\omega t - \beta r_0)} \times \frac{e^{-j\beta L(1 - \cos(q))} - 1}{j\beta(1 - \cos(q))}. \text{ Finalement :}$$

$$E_q = \frac{1}{4\pi\epsilon_0 c^2} \times \frac{\sin(q)}{r} \times j\omega I_0 e^{j(\omega t - br_0)} \times \frac{\sin\left(\frac{b}{2}L(1 - \cos(q))\right)}{j\frac{b}{2}(1 - \cos(q))} e^{-j\frac{b}{2}L(1 - \cos(q))}.$$

Comme on ne s'intéresse qu'au module de E_q :

$$E_q = m_0 I_0 \frac{\omega L}{4\pi} \times \frac{\sin(q)}{r} \times \frac{\sin\left(\frac{b}{2}L(1 - \cos(q))\right)}{L\frac{b}{2}(1 - \cos(q))}.$$

Le champ rayonné est donc de la forme :



$$G_a = 10 \log(g_a) = 10 \log\left(\frac{p_a}{p_{iso}}\right).$$

La surface constituée par g_a pour toutes les directions est appelée diagramme de rayonnement de l'antenne et la valeur maximum de g_a prend la dénomination de gain de l'antenne.

Une antenne à fort gain est une antenne directive.

Considérons une antenne, en réception, placée à une distance éloignée de la source EM et de façon être dans le lobe principal du gain d'antenne. Celle-ci reçoit une certaine puissance P_r . La surface plane S , placée perpendiculairement à la direction de propagation, recevant la même puissance P_r que l'antenne, est appelée « surface équivalente » de l'antenne. On démontre qu'il existe une relation entre le gain de l'antenne G et la surface équivalente S :

$$\Sigma = \frac{G I^2}{4 p}$$

Considérons maintenant une antenne en émission. La puissance due au rayonnement de l'antenne à une distance l ($l \gg \lambda$) est donnée par :

$$P_r = \frac{g_a P_E}{4 p l^2}$$

La puissance captée par une antenne en réception à une distance l ($l \gg \lambda$) est donc :

$$P_{captée} = P_r \Sigma$$

$$P_{captée} = \Sigma \frac{g_a P_E}{4 p l^2}$$

$$P_{captée} = \frac{G_a I^2}{4 p} \frac{g_a P_E}{4 p l^2}$$

$$P_{captée} = G_a g_a \left(\frac{I}{4 p l} \right)^2 P_E$$

Différents types d'antennes

L'antenne la plus simple est un simple brin conducteur de longueur $\lambda/4$. En fonction des applications, plusieurs types d'antennes ont été développés. Sommairement, leur conception vise à obtenir un diagramme de rayonnement adapté à l'application finale. Ainsi, des antennes fortement directives vont être utilisées pour la transmission entre des dispositifs fixes ou qu'il est possible de poursuivre (relais téléphoniques, satellites), les antennes faiblement directives vont être choisies pour permettre une diffusion omnidirectionnelle. Leurs formes peuvent également être liées aux fréquences d'emploi. L'impédance présentée par une antenne dépend de la fréquence, les valeurs données ci-après n'ont de sens qu'à la fréquence nominale de fonctionnement de l'antenne.

Tableau 3.42 Quelques antennes filiformes.

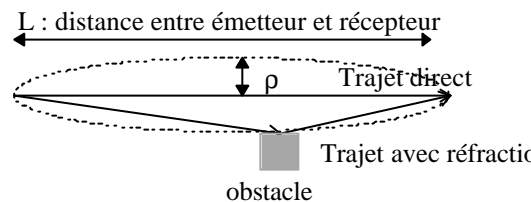
Type	Forme	Z_c	Diagramme de rayonnement	Gain $g(G)$	Utilisation
Dipôle (ou doublet de Hertz)				1,5 (1,8 dB)	Ondes longues moyennes et courtes
Dipôle quart d'onde		$\sim 73 \Omega$		1,64 (2,1 dB)	Ondes ultra-courtes
Dipôle $\lambda/4$ replié		$\sim 300 \Omega$			
Yagi		$< 300 \Omega$		8 ... 9 dB	Réception OUC (TV) (bande étroite)
Tourniquet		$\sim 70 \Omega$			Emission omnidirectionnelle
Dièdre		$\sim 130 \Omega$		~ 9 dB	Emission OUC
Papillon				5 dB	Emission OUC, TV
Cigare				16 dB	Faisceaux hertziens spéciaux
Hélice		90 ... 220 Ω			Poursuite et télécommande de satellites
Losange (rhombôidre)				15 ... 22 dB	Radiotélégraphie intercontinentale (ondes courtes)

Polarisation d'une antenne

Les vecteurs $\vec{E}(t, x, y, z)$ et $\vec{B}(t, x, y, z)$ sont orthogonaux au vecteur de vecteur d'onde (donc à la direction de propagation). L'orientation du vecteur $\vec{E}(t, x, y, z)$ par rapport au plan horizontal définit la polarisation de l'onde EM. En générale, la polarisation ondes TV est horizontale, les dipôles formant les antennes sont alors horizontales. Deux ondes aux polarisations orthogonales peuvent existées simultanément sans interférer l'une avec l'autre (cas des diffusions satellite).

Conditions de propagation

La propagation en espace libre introduit un certain nombre de contraintes et d'avantages. La rotondité de la terre limite la propagation en vision directe à une cinquantaine de kilomètres d'où l'emploi de tours. Par ailleurs la réfraction des ondes EM sur les différents obstacles du trajet de l'onde introduisent des trajets différents. L'onde résultante est la somme (vectorielle) de l'onde en trajectoire directe et des signaux parvenant par réfraction. De ce fait, si les signaux dus à la réfractons parviennent en opposition de phase avec le signal direct, le signal résultant peut être très sensiblement atténué.



Pour éviter ce genre d'inconvénient, il est nécessaire d'assurer un espace de dégagement, c'est-à-dire une hauteur minimum r donnée par :

$$r = \frac{1}{2} \sqrt{LI} \text{ , } L \text{ étant la longueur d'onde.}$$

Enfin, l'atmosphère terrestre n'est pas un milieu homogène. En premier lieu, les phénomènes météorologiques (pluie, vapeur d'eau) se traduisent par des phénomènes d'absorption pour certaines longueurs d'ondes. D'autre part, les couches supérieures de l'atmosphère (ionosphère) permettent la réflexion d'ondes EM de longueur d'ondes données. Contrairement aux ondes lumineuses, les transmissions hertziennes sont donc exploitables pour les télécommunications à longues distances.

Bilan d'une liaison

En fin de compte, tant pour les transmissions de signaux numériques que pour l'acheminement de signaux analogiques, le point crucial est le rapport signal sur bruit. En fonction de l'application, un certain niveau de rapport signal sur bruit doit être respecté. Les dispositifs d'émission, les milieux de propagation, les relais et enfin les dispositifs de réception sont tous entachés de bruits qui vont s'ajouter les uns aux autres. L'affaiblissement du signal dégrade également le rapport signal sur bruit. Le dimensionnement de la liaison dépend donc de ces paramètres.

Le bilan de liaison va donc consister à déterminer l'ensemble des atténuations et amplifications d'un signal. Le plus simple est alors d'exprimer les puissances en dBm ($P_{dBm} = 10 \log_{10}(P/1 \text{ mW})$) et les atténuations ou amplifications en dB. Exemple :

On souhaite réaliser une liaison de 12 km par fibres optiques avec un débit minimum exigé de 2 Mbits/s et un taux d'erreur admis de 10^{-8} . Le choix des composants est le suivant :

fibres optiques:

- fibre à saut d'indice: diamètre de cœur 200 μm , bande passante 100 MHz pour 100 m, atténuation 5 dB/km, livrée par rouleaux de 1 km.

- fibre à gradient d'indice: diamètre de coeur 50 μm , bande passante 100 MHz pour 1 km, atténuation 3 dB/km, livrée par rouleaux de 1 km.

émetteurs:

- LED: puissance d'émission 1 mW, bande passante 60 MHz, rendement de couplage -17 dB.

- LASER: puissance d'émission 10 mW, bande passante 600 Mhz, rendement de couplage -3 dB.

connectique:

- connecteur à l'émetteur: pertes de 1 dB.
- connecteur au récepteur: pertes de 1 dB.
- épissure par soudage: pertes de 0.3 dB.

Bilan en énergie de la liaison:

	LED		LASER	
émission	0 dBm		10 dBm	
couplage	-17 dB		-3 dB	
connecteur émission	- 1 dB		-1 dB	
connecteur détection	-1 dB		-1 dB	
épissures	11x(-0,3 dB)=- 3,3 dB		-3.3 dB	
Atténuation par les fibres	SI -60 dB	GI -36 dB	SI -60 dB	GI -36 dB
total	-82,3 dBm	-58,3 dBm	-58,3 dBm	-34,3 dBm