LA RADIO LOGICIELLE
ou
LE TRAITEMENT NUMERIQUE DU SIGNAL
expliqué aux analogiciens par un analogicien.

Robert BERRANGER F5NB

Quatrième partie : le calcul numérique (suite).

Article paru dans Radio-REF de mai 2007.

Dans la première partie, nous avons traité de l'échantillonnage. Dans la seconde, nous avons abordé le traitement numérique du signal, en particulier les oscillateurs (NCO) et les filtres. Dans la troisième, nous avons poursuivi avec le filtrage de décimation, la démodulation et les boucles de CAG. Dans celle-ci nous allons aborder la démodulation numérique et la transformée de Fourier.

La modulation numérique (digital modulation).

C'est une modulation analogique particulière car elle est discontinue, et composée d'une suite d'états stables [(1)].
Chaque état stable, appelé symbole, correspond à un nombre numérique et dure un temps défini et constant. Le nombre de symboles par seconde est mesuré en bauds.
Nous avons sur la figure 1 une représentation comparée des modulations analogiques et numériques.

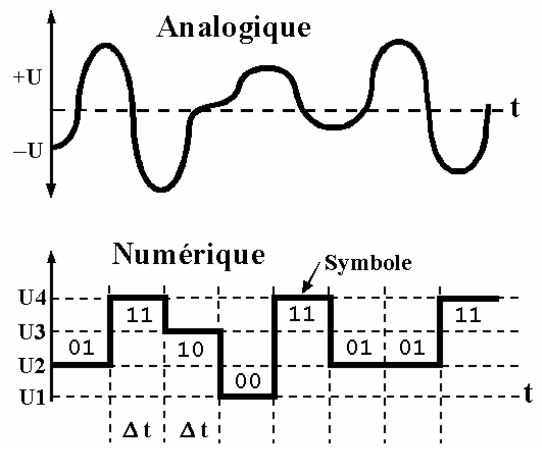


Figure 1

Ici, le symbole peut prendre 4 valeurs, donc transmettre un nombre à deux bits. Le débit en bits/sec sera alors le double du débit en bauds (égal lui, à 1s /Δt)[(2)].

Si l'on veut transmettre une modulation numérique sans déformation, il faut avoir une bande passante d'au moins cinq fois le rythme symbole[(3)]. Cela est très pénalisant, et en pratique, on s'arrange pour avoir une bande passante se rapprochant de la moitié de ce rythme. Alors, le signal démodulé n'aura plus de transitions brusques, comme montré sur la figure 2 (symboles à 2 états)

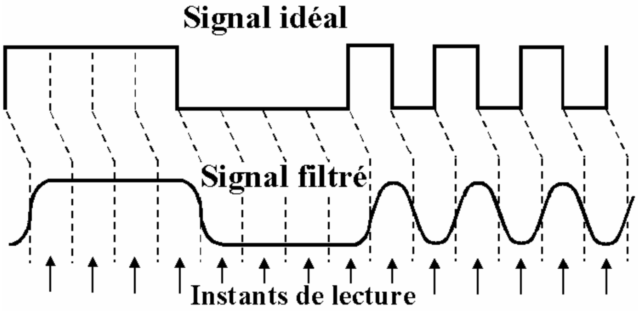


Figure 2

La conséquence principale affecte l'instant de lecture qui ne peut plus se faire sur la durée du symbole, mais à un endroit précis(4), comme montré sur la figure. Déterminer cet endroit constitue la « synchronisation bit ».

Synchro bit.

Le principe général consiste à transmettre au début de chaque trame un roulement constitué d'une suite de doubles bits (dibits) bistables (0, 1). A partir de ce roulement, le récepteur déterminera l'endroit idéal de lecture. Celui-ci sera valable pour toute la trame, espérant que la dérive différentielle entre les horloges émission et réception sera négligeable sur la durée de celle-ci.

Cas d'un signal numérique en bande de base.

Une première obligation consiste à avoir l'horloge de sortie numérique avec une fréquence multiple de celle de l'horloge symbole. Prenons le cas d'une bande passante égale au demi rythme symbole. Le théorème de Shannon et le filtre de canal (anti-repliement) demanderaient une horloge de sortie idéale égale à trois fois la fréquence la plus élevée, soit 1,5 fois le rythme symbole. Mais nous serons obligés de prendre une horloge égale à deux fois le rythme symbole pour avoir un rapport entier. Nous avons alors sur la figure 3 les cas limites de la position des échantillons par rapport au signal analogique (selon la phase relative entre les horloges émission et réception).

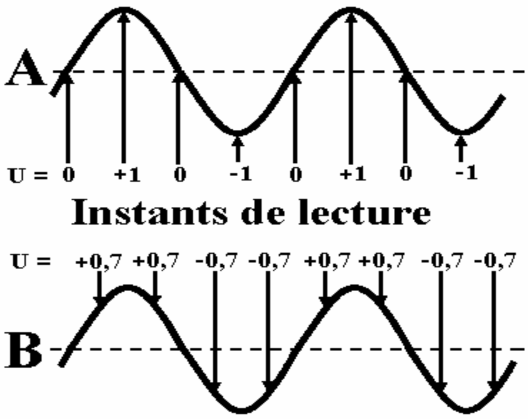


Figure 3

En A, nous avons une suite d'échantillons d'amplitudes égales à 0, +1, 0, -1, 0, +1, etc… En B nous avons la suite, +0,7, +0,7, -0,7, -0,7, etc… Il nous faudra trouver un algorithme valable pour les deux cas.

Nous pourrons par exemple faire la moyenne de deux échantillons successifs. Mais il y a deux possibilités en prenant, soit les échantillons n et n+1, soit les échantillons n-1 et n. On utilisera le roulement pour tester les deux possibilités[(5)]. On retiendra celle qui donnera le signal d'amplitude maximum (donc le plus cohérent).
Dans ce cas, la synchro bit est très simple à réaliser. Mais le signal subit une diminution du rapport S/B comprise entre 3 dB (valeur moyenne 0,7 dans le cas B) et 6 dB (valeur moyenne 0,5 dans le cas A). Pour améliorer cela, il faudra sur échantillonner le signal et faire des interpolations.

Interpolation.

L'interpolation consiste à calculer des valeurs intermédiaires régulièrement espacées entre les échantillons. Quand nous convertissons le numérique en analogique, c'est le filtre passe-bas anti-repliement (analogique) qui effectue cette opération dite « de reconstruction ». Nous avons alors une infinité de valeurs intermédiaires. En restant en numérique, nous ne pouvons que multiplier le nombre d'échantillons jusqu'à l'obtention d'une résolution acceptable.
L'interpolation la plus simple à faire est la multiplication par deux de la cadence d'échantillonnage. Il suffit d'ajouter entre deux échantillons, un échantillon à zéro, et de passer l'ensemble dans un filtre passe-bas qui nous donnera en sortie les valeurs intermédiaires (comme le filtre de reconstruction) [(6)]. Pour une dynamique de 40 dB (nous sommes après l'ALC), un FIR à 16 coefficients suffit. Avec 3 étages successifs, nous aurons une interpolation de 8 fois (2×2×2), ce qui nous fait 16 valeurs par symbole. La synchro bit consistera alors à trouver laquelle de ces 16 valeurs est à lire.

Cas d'un signal analogique en bande de base.

Cela peut être directement le signal de sortie d'un récepteur analogique que l'on va sur échantillonner avec un CAN. Avec un récepteur numérique, on peut passer en analogique avec un CNA suivi d'un filtre de reconstruction, puis d'un CAN avec sur échantillonnage.
Nous avons le système sur la figure 4.

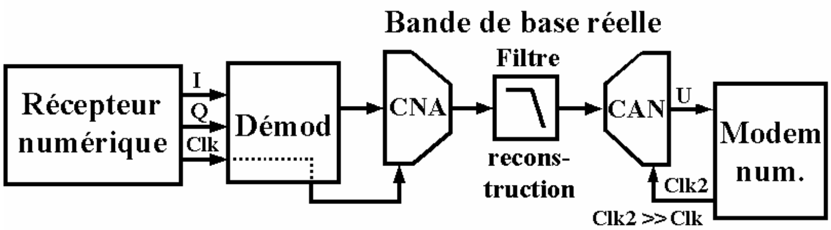


Figure 4

Noter qu'avec cette méthode, l'horloge de sortie du récepteur et celle de lecture des symboles n'ont plus l'obligation d'être dans un rapport entier, puisque le signal analogique a une infinité de points interpolés. Mais l'horloge du CAN et du modem devra toujours être un multiple du rythme symbole.

La synchro bit par corrélation.

Soit un signal avec 16 échantillons par symbole (après interpolation), et soit un roulement de 50 dibits {0,1}.
Le problème est de trouver le N° d'échantillon procurant le meilleur rapport S/B à la lecture.
La solution la plus simple est de faire une moyenne sur un grand nombre de dibits, par

exemple 32. Pour cela nous allons utiliser un buffer tournant de 32 valeurs par dibit (16×2). Nous allons créer un compteur par 16 qui constituera la référence de temps de l'horloge lecture. Nous allons accumuler les échantillons d'entrée sur 64 symboles successifs. A la fin, notre buffer contiendra 32 sommes de 32 couples d'échantillons. Du fait que le bruit n'est pas corrélé avec le signal utile, l'opération nous fait gagner une trentaine de dB sur le rapport S/B. Nous allons chercher dans le buffer les deux N° de changement de signe (après avoir calculé et soustrait la moyenne générale si le signal n'était pas signé). Ensuite, nous allons faire la moyenne des deux N°, puis ajouter 8 (modulo 16). Ce sera le N° de l'échantillon à utiliser pour la lecture du symbole.

Cette méthode donne une courbe identique au signal d'entrée, avec seulement une amélioration du rapport S/B. La figure 5 nous en montre une illustration (1 dibit).

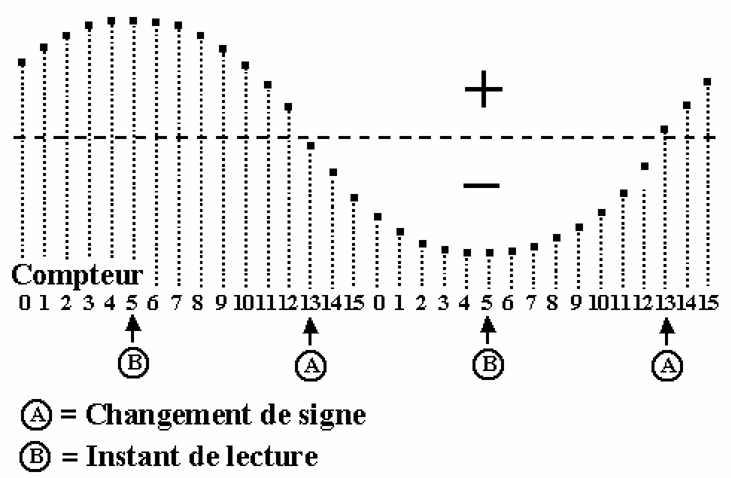


Figure 5

Il faudra que la synchro bit soit terminée avant la fin du roulement, de manière que l'on soit prêt pour la lecture du motif constituant la synchro trame. Mais nous sortons du sujet. Ceci sera expliqué plus en détail dans une série d'articles à paraître sur les modems.

Nous allons aborder maintenant la transformée de Fourier. Nous délaisserons les mathématiques pour nous intéresser au principe et à son application à la Radio Logicielle.

Les Transformées de Fourier

Descriptions temporelle et fréquentielle d'un signal variable.

La description temporelle concerne la variation de tension en fonction du temps. C'est ce que nous voyons quand nous appliquons le signal à un oscilloscope. La description fréquentielle concerne la variation d'amplitude en fonction du spectre de fréquence. C'est ce que nous voyons quand nous appliquons le signal à un analyseur de spectre, ou tout simplement en observant le S-mètre d'un récepteur dont nous faisons varier la fréquence.

Prenons un signal sinusoïdal et examinons ses deux descriptions sur la figure 6.

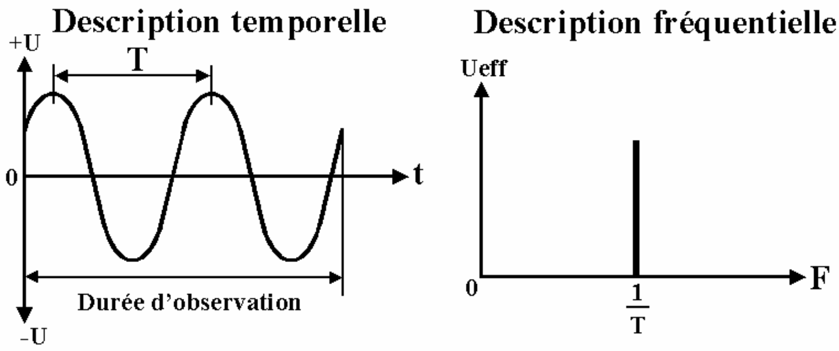


Figure 6

Dans la description fréquentielle, nous avons perdu un paramètre, qui est la durée d'observation du signal. Celle-ci sera au minimum égale à une période. Si elle est plus longue, cela n'aura pas d'importance si le signal est stationnaire, c'est à dire, invariable dans le temps. Sinon, l'amplitude de la raie ne sera plus égale à la valeur efficace, mais à la moyenne des valeurs efficaces par rapport au temps. En effet, l'obtention de l'amplitude est le résultat d'une intégration sur la durée d'observation du signal.

Décomposition en série de Fourier.

Fourier a démontré que n'importe quelle fonction **périodique** est la somme de fonctions **sinusoïdales** harmoniques de la période. Nous avons un exemple simple sur la figure 7.

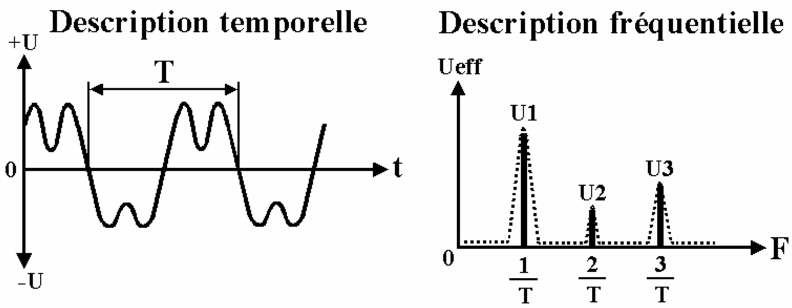


Figure 7

La décomposition du signal en série de Fourier peut se faire de deux manières :

1- La période T est connue.
   Il suffit d'appliquer le signal à une batterie de filtres centrés sur les fréquences 1/T, 2/T, 3/T, etc… et de mesurer les tensions à la sortie des filtres. Nous obtenons un ensemble de raies d'amplitudes correspondantes.
2- La période T est inconnue.
   Dans ce cas, on applique le signal à un filtre passe-bande dont la fréquence centrale varie progressivement de zéro à la fréquence maximum estimée. Alors, à la place des raies, on obtient une suite de courbes qui reproduisent toutes la bande passante du filtre (en pointillés sur la Fig. 7). C'est le principe de l'analyseur de spectre analogique.

On comprend que pour obtenir des valeurs exactes, il faut que le signal soit présent à l'entrée des filtres (durée d'observation) durant tout leur temps d'intégration, qui est d'autant plus long que leur bande passante est étroite et leur facteur de forme élevé.
En conséquence, l'analyseur de spectre ne fonctionne qu'avec des signaux stationnaires.

La transformée de Fourier.

Examinons la figure 8.

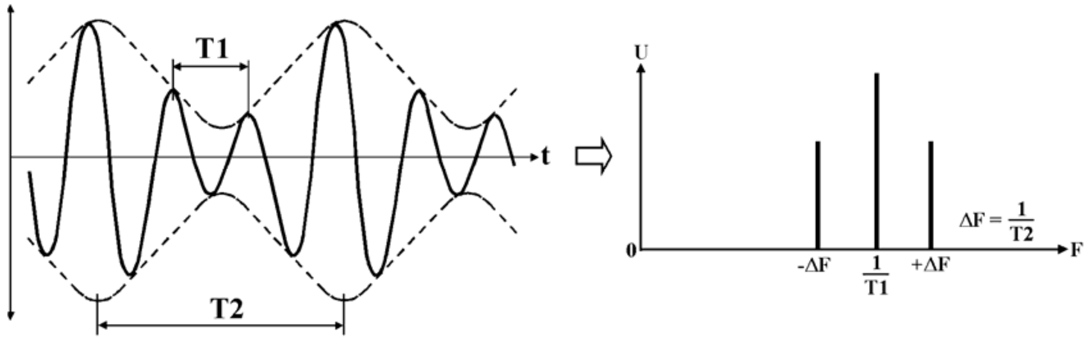


Figure 8

Nous avons un signal sinusoïdal (F = 1/T1) modulé en amplitude par un deuxième signal sinusoïdal (F = 1/T2). La description fréquentielle d'un tel signal fait apparaître de nouvelles propriétés. Nous avons maintenant un spectre centré sur la fréquence 1/T1 et formé de deux décompositions en série de Fourier symétriques, relatives à la période T2 du signal de modulation. Cela reste valable si le signal de modulation est de forme quelconque, mais toujours périodique.
Maintenant, allongeons la période du signal T2 (et par conséquent la durée d'observation). Alors, les raies de la décomposition en série de Fourier vont se rapprocher les unes des autres (ΔF = 1/T2). A la limite, avec une période infinie, les raies vont se toucher. Nous aurons un spectre continu. Si nous supprimons la porteuse à la fréquence 1/T1, pour ne garder que la modulation, le spectre va se retrouver centré sur la fréquence zéro, avec une moitié constituée de fréquences positives et l'autre moitié constituée de fréquences négatives[(7)]. Ce passage de la description amplitude/temps à une description amplitude/fréquence, avec un signal non périodique (puisque période infinie) est appelé « Transformée de Fourier ».
Vous me direz qu'on peut difficilement envisager une période infinie. Mais si le signal a une durée finie (partant de zéro et finissant à zéro), nous pourrons effectuer une transformée de Fourier sur la durée de ce signal. Nous avons ainsi sur la figure 9 la transformée de Fourier d'un signal rectangulaire unique.

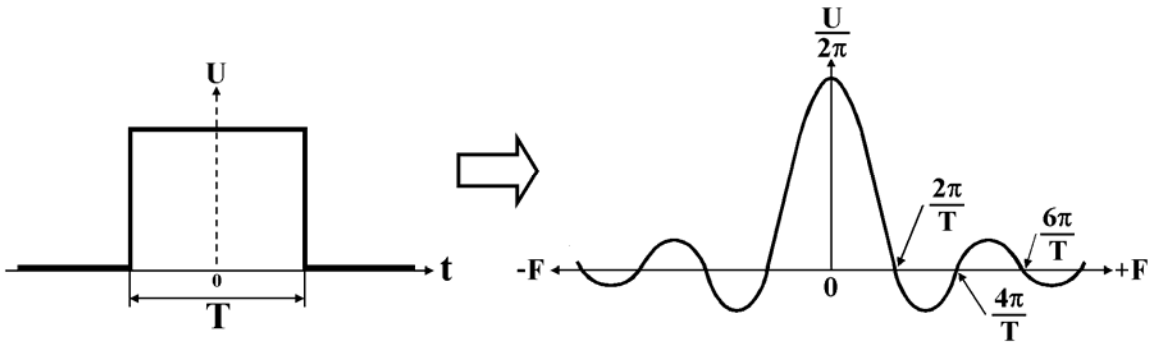


Figure 9

Noter que si le signal a une durée finie, sa transformée tend vers plus ou moins l'infini si la raideur de ses flancs tend vers zéro (temps nuls).
Q : *Et si nous voulions effectuer une transformée de Fourier sur une durée finie d'un signal de durée infinie ?*

R : C'est possible, et c'est ce qui est fait en pratique, mais cela entraîne deux conséquences :

1- Le spectre n'est plus continu, mais est constitué de raies espacées d'un ΔF égal à l'inverse de la durée d'observation du signal.
2- Le spectre est le résultat de la convolution[8] du signal analysé avec un rectangle de durée égale à celle d'observation, et leurs spectres s'ajoutent.

Examinons ceci sur la figure 10.

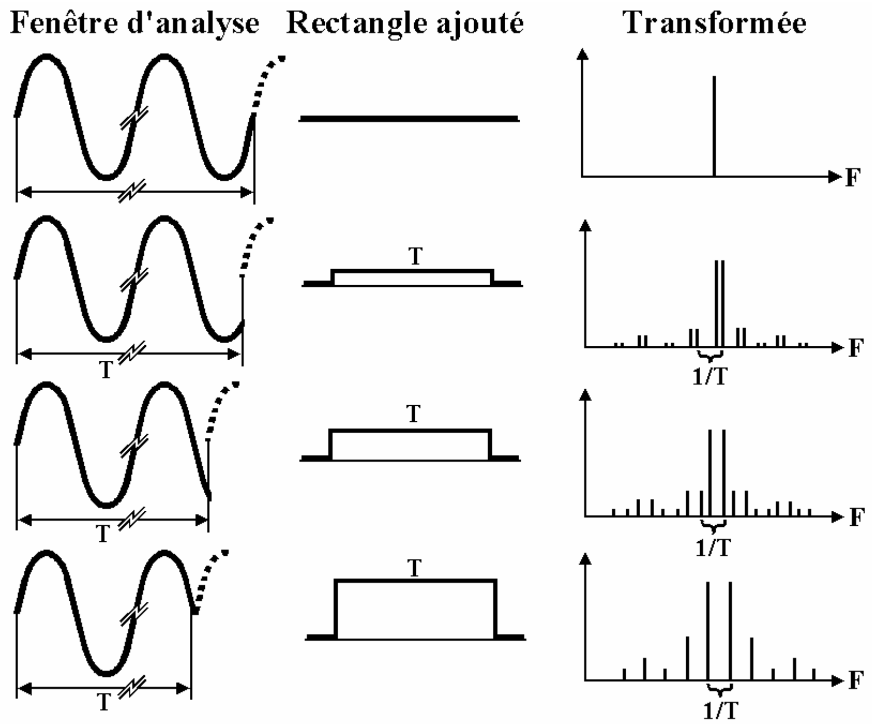


Figure 10

Pour une fréquence dont la durée d'observation contient un nombre entier de périodes, la raie correspondante tombe sur une raie de la transformée. Elle est alors à sa place exacte avec la bonne amplitude (le rectangle a une valeur nulle).
Si l'on décale progressivement la fréquence, la raie correspondante se dédouble en deux raies de niveau –3 dB qui s'écartent l'une de l'autre[9]. Quand la durée d'observation contient un nombre impair de demies périodes, l'écartement est maximum et correspond à un ΔF égal à l'inverse de la durée d'observation (ΔF de la transformée). Par ailleurs, le rectangle a sa valeur maximum, ce qui correspond à l'ajout d'un spectre en sin(x)/x de valeur maxi également.
N-B : *Sur la fig. 10, le sin(x)/x est resserré pour faciliter la lecture. En réalité il est plus étalé (le premier nul est à* $2\pi / T$*).*

Lorsque nous faisons la transformée d'un signal quelconque, composé d'un grand nombre de signaux de fréquences différentes, nous obtenons une multitude de raies qui n'ont pas de signification en elles-mêmes (ni aux bonnes fréquences, ni les bonnes amplitudes). Mais si nous les intégrons, nous obtenons pour chaque raie théorique la courbe d'un filtre passe bande du premier ordre, à la bonne fréquence, et avec la bonne amplitude. La faculté de séparer deux fréquences rapprochées est alors bien plus faible que le maximum théorique. C'est à dire que nous n'aurons pas ΔF = 1/T (T = durée d'observation), mais ΔF = K/T, avec K, fonction du pouvoir séparateur en décibels. Ce sera plus facile à comprendre en examinant la figure 11 qui montre la transformée de Fourier d'une addition de deux signaux sinusoïdaux.

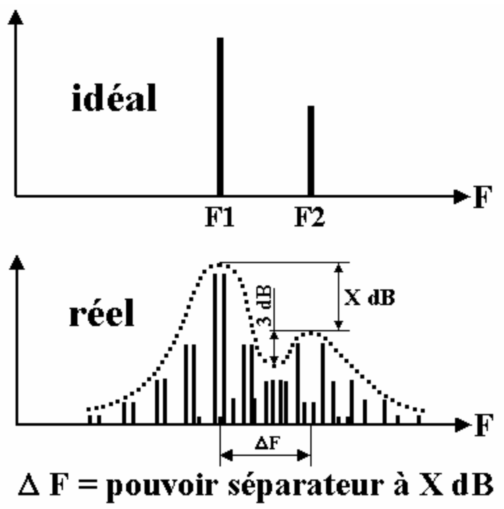


Figure 11

La courbe en pointillé est le résultat de l'intégration des raies sur un temps égal à $2\pi T$, ce qui veut dire qu'il a fallu dupliquer le signal sept fois au minimum (sept périodes).
Q : *Peut-on améliorer le pouvoir séparateur ?*
R : Oui, on peut augmenter la durée d'observation. Sinon, on peut procéder de la même manière que pour les coefficients d'un FIR, en remplaçant la fenêtre d'observation rectangulaire par une autre ayant un spectre moins étalé. Nous utiliserons pour cela les mêmes fonctions que pour les FIR. Nous ferons alors une convolution du signal avec la fenêtre avant d'en effectuer la transformée de Fourier.

Transformée de Fourier discontinue (DFT = Discrete Fourier Transform)

Jusqu'à présent, nous avons vu la transformée de Fourier d'un signal analogique. Qu'en est-il avec un signal échantillonné ?
Il existe une méthode mathématique, la DFT (Discrete Fourier Transform) qui effectue une transformée à partir des échantillons temporels. Celle-ci a quelques propriétés différentes de la transformée analogique. D'une part, elle est périodique, et d'autre part elle n'est composée que de raies harmoniques de la durée d'observation (1/T). Alors, il y aura autant de raies dans la période que d'échantillons temporels. Par ailleurs, nous avons le même principe de fenêtrage qu'en analogique, et il faudrait procéder à une intégration pour avoir des courbes ayant leurs maxima aux fréquences des raies réelles. Tout ceci est résumé sur la figure 12.

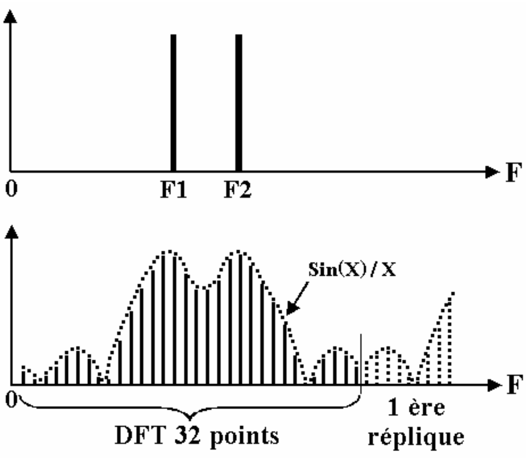


Figure 12

Nous avons ici une fenêtre rectangulaire, et nous voyons qu'il nous faudra faire une DFT sur un grand nombre de points (et d'échantillons) si nous voulons séparer des signaux proches d'amplitudes très différentes. Or, la DFT est gourmande en puissance de calcul, qui croît comme le carré du nombre de points. Heureusement, une partie de ces calculs est redondante, ce qui va nous permettre des simplifications.

La F.F.T. « Fast Fourier Transform »

Ce n'est qu'une méthode de calcul, qui ne change rien aux propriétés de la DFT. *Grosso modo*, le principe est basé sur la décimation. Prenons un exemple, avec une DFT sur 1000 points. Elle demande 1000^2 = 1000 000 de calculs. Avec une décimation par 2, les calculs se ramènent à faire deux DFT sur 500 points et à les additionner, soit $500^2 + 500^2 + 500 = 50500$ calculs, au lieu d'un million. On peut poursuivre la décimation, jusqu'à obtenir des DFT sur deux points, faciles à calculer (à condition que le nombre de points soit une puissance de deux). Avec cette méthode, on peut gagner plus de 50 fois sur la puissance de calcul nécessaire pour une FFT sur 512 points. Nous ne nous étendrons pas sur les algorithmes mis en œuvre car ce n'est pas le but de cet article, mais nous allons nous demander à quoi une FFT peut bien nous servir.

La transformée de Fourier, pourquoi faire ?

En radio, la transformée de Fourier, analogique ou FFT, ne sert, seule, que d'outil d'analyse. On peut par exemple vouloir le spectre des émissions présentes dans une bande radioamateur. C'est la fonction « réception panoramique » que l'on trouve dans certains postes du commerce.
En radio numérique, on peut utiliser la FFT pour faire du filtrage, en association avec une FFT inverse.

La transformée inverse de Fourier.

En fait, les transformées de Fourier sont doubles. Nous avons vu celle qui permet de passer du domaine temporel au domaine fréquentiel. Il y en a une autre qui permet de passer du domaine fréquentiel au domaine temporel. Elle est appelée « transformée inverse de Fourier », ou IFFT (Inverse Fast Fourier Transform) pour sa version discrète.
Nous pouvons alors imaginer effectuer successivement une transformée de Fourier suivie d'une transformée inverse. Bien sûr, nous ne retrouverons pas exactement notre signal de départ, à cause des imperfections liées au fenêtrage. Il faudra que celui-ci soit suffisamment long pour rendre ces imperfections négligeables.
Par ailleurs, nous ne pourrons plus nous contenter du module des transformées, comme en analyse, où seule la puissance nous intéresse. Pour reconstituer un signal fidèle, il nous faudra conserver la phase relative entre les composantes du spectre. En conséquence, nous devrons effectuer les transformées dans le plan complexe, en I et en Q.
Pour la suite, nous admettrons avoir affaire à des transformées complexes, même si ce n'est pas précisé.

Filtrage par transformées de Fourier.

Le principe est simple : Il suffit de faire une FFT, choisir les raies qui nous intéressent, puis faire une IFFT. Comme on a vu que les raies d'une FFT n'étaient pas réelles, mais un artefact, on imagine qu'il nous faudra un grand nombre de points pour obtenir un filtrage performant.

Prenons un exemple concret. Soit une FFT sur 4096 points avec une Horloge à 44,1 kHz. Nous voulons effectuer un filtrage passe-bande centré sur 11,025 kHz. Nous supposerons une fenêtre rectangulaire. Nous voulons trois bandes passantes : 43 Hz (4 points), 355 Hz (33 points) et 2,4 kHz (223 points). Nous avons sur la figure 13 l'allure des filtres obtenus.

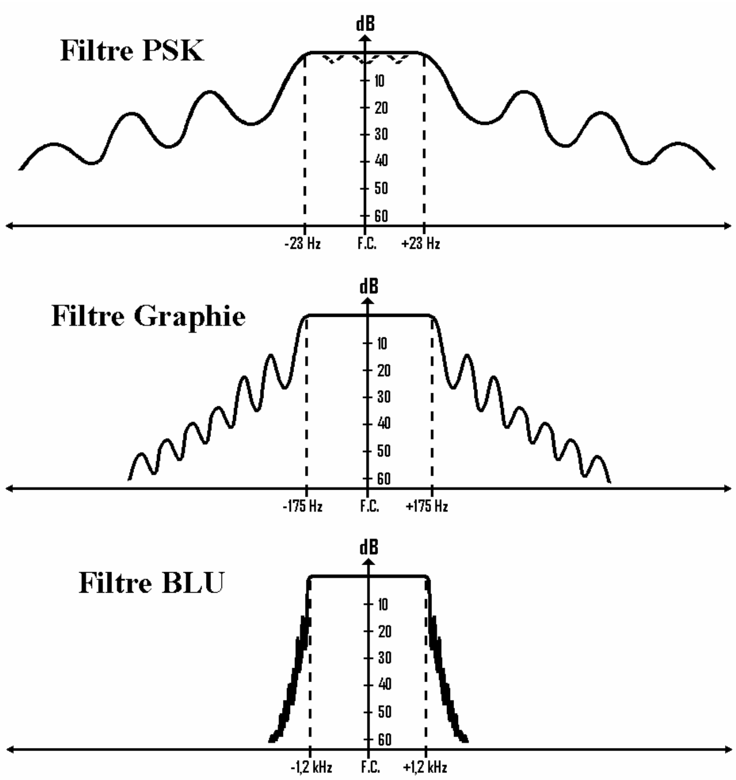


Figure 13

Nous constatons que la pente des filtres est une constante, quelle que soit la bande passante. Si nous avons un facteur de forme acceptable pour la bande de 2,4 kHz (BLU), il est déplorable pour la bande de 47 Hz (PSK). Dans ce mode, nous aurons une très faible réjection des brouilleurs.

Par ailleurs, ce filtrage nous demande des centaines de milliers de calculs, alors qu'avec un système composé de FIR et de décimations, nous nous en sortons avec une dizaine de milliers seulement. Et de cette façon, nous obtenons un facteur de forme constant (grâce aux décimations), donc une excellente tenue aux brouilleurs, même en PSK.

Q : *Alors, le filtrage par FFT, pas terrible ?*

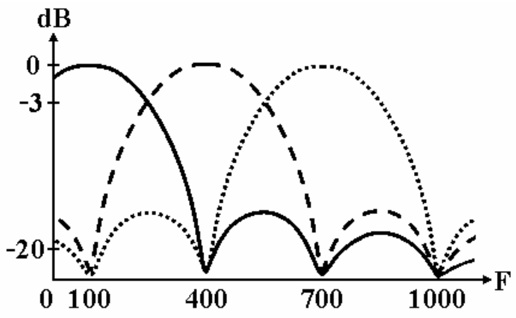
R : Peut-être pour un récepteur universel (toutes modulations, et brouilleurs de niveaux inconnus), où la dynamique est importante, mais il se rattrape dans certains cas particuliers que nous allons examiner.

La réception d'un ensemble de signaux modulés en analogique de provenances et de puissances inconnues, nous oblige à avoir pour le récepteur une réjection élevée des canaux adjacents, pouvant atteindre 60 dB en pratique. Mais avec des signaux provenant d'un même émetteur (radiodiffusion), ou ajustés pour arriver au récepteur avec la même amplitude (réseau de téléphonie sans fil), la dynamique se ramène au SINAD nécessaire pour une bonne démodulation. Ce SINAD peut être relativement faible avec des modulations numériques (10 dB pour du QPSK, 23 dB pour du 64-QAM). Alors, le filtrage par FFT devient performant, en particulier, dans le cas d'émission multi-porteuses qui est un moyen pour obtenir un étalement de spectre (il y en a d'autres). Ce principe consiste à remplacer une porteuse unique modulée très rapidement, par un grand nombre de porteuses modulées beaucoup plus lentement[10].

Si nous générons et filtrons chaque sous porteuse analogiquement, le facteur de forme des filtres nous obligera à espacer les porteuses en fréquence, d'où un encombrement spectral accru. Une solution existe avec :

L'OFDM

OFDM veut dire « Orthogonal Frequency Division Multiplexing ». Dans ce procédé, les porteuses adjacentes sont orthogonales (c'est-à-dire qu'elles ont entre elles certaines relations de phase qui permettent de les séparer lorsqu'elles se superposent), et leur espacement en fréquence est égal au rythme symbole. L'orthogonalité est obtenue grâce à une IFFT dont chaque point correspond à une porteuse. Nous avons sur la figure 14 le spectre d'un signal OFDM à trois porteuses (100, 400, 700 Hz) obtenues par IFFT.



Signal OFDM avec trois porteuses

Figure 14

En pratique, l'orthogonalité est dégradée par la chaîne de transmission (émetteur, propagation et récepteur), et l'on utilise divers moyens pour compenser cette dégradation, regroupés sous le vocable « égalisation ».

L'application la plus connue de l'OFDM est la TNT (Télévision Numérique Terrestre).

Dans le prochain article, nous verrons l'émetteur numérique (génération et modulation du signal de manière numérique et conversion N/A).

F5NB.

Notes.

1) *En principe. Nous verrons plus loin que ce n'est pas toujours le cas.*
2) *Ne pas confondre le nombre de bauds avec le nombre de bits par seconde, chaque symbole pouvant correspondre à un nombre de plusieurs bits.*
3) *C'est le cas, par exemple pour la télégraphie et le RTTY.*
4) *Prenons par exemple un signal ayant un SINAD de 10 dB (mini pour un taux d'erreurs acceptable). Avec le signal idéal de la fig. 2 (roulement 0-1), ce SINAD est constant pendant la durée du symbole. Avec le signal filtré, le SINAD n'est obtenu qu'à un seul instant précis. Avant et après cet instant, le SINAD diminue rapidement.*
5) *En faisant une corrélation sur tout le roulement pour améliorer le rapport S/B.*

6) Zéro est la moyenne d'un signal sinusoïdal, et je rappelle que nous devons faire une interpolation sinusoïdale pour se conformer au théorème de Shannon.

7) Le spectre négatif est une convention mathématique qui se réfère à un OL dont la fréquence peut être zéro. Physiquement, la transposition d'un signal HF unique (BLU) ne donne qu'un signal BF. Mais un ***deuxième*** *signal HF à la fréquence image, comme le canal adjacent en BLU, donne un spectre BF superposé au premier. Si ce sont des signaux audio, le résultat est inaudible. Heureusement, le spectre image est inversé en fréquence (miroir) et déphasé de 180° par rapport à l'OL, ce qui permet de les séparer par les méthodes du phasing et du weaver. Dans le cas d'un signal AM, les deux bandes latérales HF se replient l'une sur l'autre en BF. Si l'on décode de l'AM en mode BLU avec un récepteur analogique conventionnel, c'est parce que le filtre de canal à quartz élimine la bande latérale image.*

8) La convolution est une multiplication point par point des échantillons avec les valeurs correspondantes à leur rang dans la fenêtre. Avec une convolution de deux signaux, leurs spectres s'ajoutent.

9) La raie unique est en fait constituée de deux raies superposées en phase. Elles se dédoublent quand leur phase relative tourne (avec $\Delta F = \delta\varphi / \delta t$).

10) Pour permettre la réception par des mobiles.