INA LABO 1414

(Octobre 1993 / Janvier 1994)

Version 2.0 Janvier 2007

Introduction à la conversion audionumérique

Pierre-Antoine SIGNORET

Formation Supérieure aux Métiers du Son

Conservatoire National Supérieur de Musique de Paris

pa.signoret@free.fr

Remerciements

Ce document est un travail de synthèse basé sur la bibliographie donnée à la fin de ce volume. La publication dont je suis parti est le « tutorial » de ***Malcom Omar HAWKSFORD "An introduction to digital audio"****, publiée dans « Image of audio, Proceedings of the 10th international conference, 1991, September 7th - 9th ».*

*Les figures présentées ont été tirées en majeure partie de ces publications. Ce mémoire conclut un stage théorique de six mois et demi au sein de l'unité audionumérique de l'****Institut National de l'Audiovisuel*** *sous le tutorat de* ***Foued BERAHOU****.*

Je tiens à remercier particulièrement ***Bernard FOUQUET, Michel NOTTÉ, Claude MORETTI, Patrick LOUVET*** *et* ***David WIEM*** *pour leur patience, leurs conseils éclairés, leur bonne humeur ainsi que leur accueil.*

Les mesures physiques ont été faites grâce à la collaboration de tous au sein du ***Laboratoire d'Electronique et de Maintenance Audio (INA / LEMA)****. Un très grand merci à* ***Didier LECERT****, virtuose hors pair de « l'Audio Precision », sans les conseils et la compétence de qui, les mesures auraient été impossibles à mener à bien.*

Merci à mon alter ego « vidéo » ***Jean-Michel MOUSSU*** *pour son amitié.*

Je tiens aussi à remercier ***Thierry BARDON****, collègue et relecteur très attentif de ce document ; ses remarques et questions m'ont aidé à compléter celui-ci.*

Merci, enfin, à ***Gérard STRELETSKI,*** *musicologue et chef d'orchestre, pour sa curiosité technique et ses très précieux conseils de rédaction.*

Note de 2007

*Je me suis enfin décidé à relire mon mémoire de fin d'étude pour le mettre en ligne sur le site du «**Grenier à Son**» (http://www.legrenierason.com) suite aux questions régulières de mes étudiants. Les sujets soulevés dans ce mémoire sont, à mon étonnement, toujours d'actualité ; les convertisseurs ayant le mérite de poser de façon sensible certains des problèmes généraux et récurrents de l'audionumérique…*

*Merci donc à tous, et en particulier aux étudiants de la **Formation Supérieure aux Métiers du Son** et de l'**Option Son du Conservatoire National Supérieur de Musique de Paris**, aux étudiants du **Master Image et Son de l'Université de Bretagne Occidentale**, à ceux de L'**Ecole Nationale Supérieure des Arts Décoratifs*** ainsi qu'aux stagiaires du département de la formation professionnelle de l'***Institut National de l'Audiovisuel***...

SOMMAIRE

Comparatif analogique numérique

COMPARATIF ANALOGIQUE-NUMERIQUE

Analogique :

- Evolution asymptotique vers les limites physiques depuis le début du XXème siècle.
- Informations contenues dans des variations infinitésimales de paramètres continus (amplitude et temps).
- Temps analogique (vitesse constante). Temps continu.
- Systèmes linéaires sur des plages limitées bien définies (distorsion aux aguets).
- Dégradations résultantes de la somme de toutes les imperfections dues au trajet du signal dans les différents étages ⇒ nombre limite de passages et de générations.
- Problèmes et contraintes dus au support (enregistrement magnétique...).
- Sensibilité aux parasites, bruits et distorsions, dérive temporelle et thermique des composants.
- Reproductibilité difficile.

Numérique :

- Deux états binaires changent à des instants prédéterminés par une horloge stable ⇒ insensibilité aux parasites (seuils de remise en forme).
- Systèmes en temps différé (quelques périodes d'échantillonnage).
- Qualité indépendante du support physique (pas de pertes tant que les signaux restent identifiables).
- Pas de dérive des composants en filtrage, possibilité de changement virtuel de la structure des filtres, stabilité de leurs caractéristiques, phase linéaire possible, reproductibilité parfaite.
- Copies illimitées sans dégradations (vie "éternelle" des données si remises en forme régulières).
- Dynamique d'enregistrement faible par rapport à la dynamique restituée (bande passante élargie en conséquence) ⇒ pistes étroites, enregistrement à haute densité, vitesse tête-bande rapide, bandes métal E (*DASH* 76 cm/s, 48 pistes sur 1/2 pouce; *DAT* 3m/s, 2 pistes sur 1/8 de pouce par exemple).
- Diversité des supports : magnétiques, optique, magnéto-optique, informatiques (accès rapide à l'information).
- Communication par réseaux informatiques.
- Compression perceptive des données possible et économique.
- Opérations virtuelles, avantages du traitement informatique (montage, mixage, *DSP*...).

Défauts du numérique :

- Débits de l'ordre de quelques Mbit/sec ⇒ Procédés d'enregistrement à haute densité ⇒ fragilité des informations entraînant la nécessité de systèmes de correction d'erreur pour les supports à bandes magnétiques ou magnéto-optiques.
- Le stockage informatique requiert de grandes capacités mémoire et des temps d'accès rapides.
- Défauts perceptivement plus gênants qu'en analogique (distorsion plutôt que bruit).
- Qualité fixée par celle de la conversion analogique-numérique et la performance des calculateurs (dynamique / distorsion / bande passante).
- Prix et contraintes technologiques (*VLSI*).Les prix sont de plus en plus abordables.
- Perte totale de l'information au-delà du seuil de correction d'erreur.
- Problème de **stabilité d'horloge** (*jitter*) et **impératifs de (re)synchronisation** aux différentes étapes (temps de calcul…)
- Montage à la main délicat.
- Contraintes des supports (erreurs, longueur d'onde enregistrée, temps d'accès des disques durs, fragilité des supports optiques...).
- *Varispeed* délicat si supérieur à 10 ou 15 % (codage de voie…).
- Problèmes généraux du traitement numérique du signal (*DSP*).
- Complexité des enregistreurs, maintenance spécialisée.
- Compression parfois au détriment de la qualité…

Introduction à la conversion audionumérique

INTRODUCTION A LA CONVERSION AUDIONUMERIQUE

La conversion analogique-numérique se propose de coder l'information sonore sous forme de message binaire en vue du stockage et / ou du traitement informatique de celle-ci. Un décodage (conversion numérique-analogique) permettra ensuite de décoder le message binaire pour l'écouter.

Les codeurs / décodeurs se doivent donc d'être les plus transparents possibles.

Les systèmes audionumériques posent le problème psycho-acoustique de la plus grande sensibilité de l'oreille à la distorsion (non-linéarité d'un système / bruit coloré) qu'au bruit (aléatoire et blanc, par définition).

1.0.0 Echantillonnage, quantification et *dithering* :

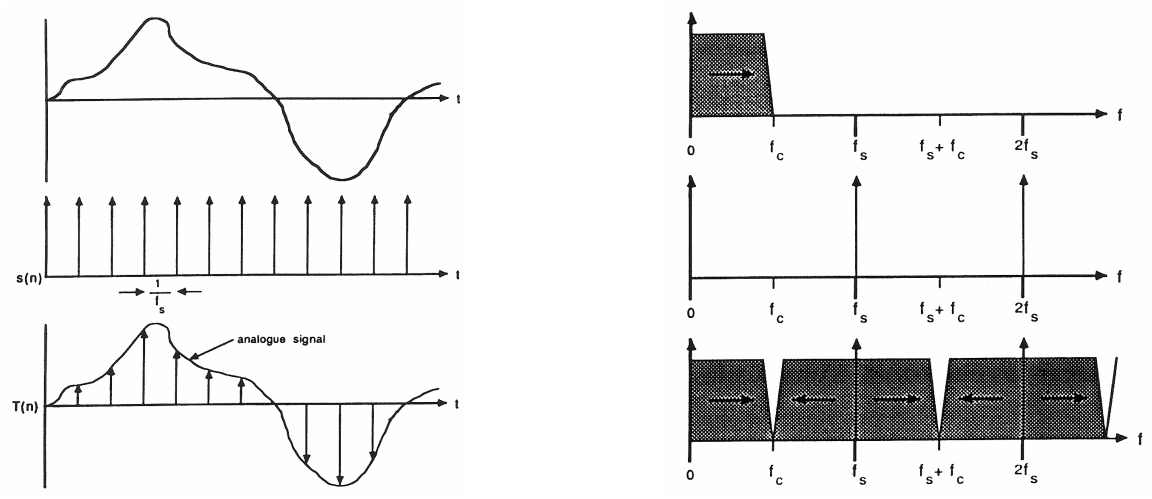
- Discrétisation d'un signal continu dans ses deux dimensions (amplitude et temps).
- Décorrélation de la distorsion de quantification du signal en vue d'obtenir un bruit résiduel uniforme.

1.1.0 Echantillonnage :

L'échantillonnage désigne l'étape de ***discrétisation temporelle***. Celle-ci est effectuée par multiplication du signal audio par un train périodique d'impulsions. Cette multiplication se traduit par une ***périodisation*** de ses composantes spectrales autour des multiples de la fréquence d'échantillonnage (*Fe*). Cette multiplication est équivalente à une convolution dans le domaine fréquentiel (sorte de "modulation d'amplitude périodique").

Pour garantir le non chevauchement des spectres périodisés et donc l'absence de distorsion due à ce phénomène **(repliement / *aliasing,*** cf. *figure n°1*), le signal à échantillonner doit avoir un spectre limité (d'où un filtrage **anti-repliement** nécessaire avant discrétisation) et la fréquence d'échantillonnage obéir au critère de ***Shannon / Nyquist / Kotelnikov*** : $Fe \geq 2Fmax$ (*Fmax* étant la fréquence la plus haute de la bande spectrale à coder). La fréquence ***2Fmax*** est parfois appelée **fréquence critique**.

Processus d'échantillonnage



Domaine temporel Domaine fréquentiel

Figure n°1

Mathématiquement ce processus peut se formaliser comme une multiplication du signal d'entrée par un peigne de ***Dirac*** qui, étant périodique (de période *Te = 1/Fe*), peut se décomposer en série de ***Fourier***.

Cette décomposition permet de mettre en évidence la modulation effectuée comme suit :

Signal d'entrée : *x (t)* Spectre d'entrée : *X(f)* Peigne de Dirac : $\delta(t-nTe)$

Signal échantillonné : *x*(t)* Spectre échantillonné : *X*(f)*

$$x^*(nTe) = \sum_{n=-\infty}^{+\infty} x(t)\delta(t-nTe), \delta \text{ étant la distibution de Dirac et n un entier relatif}$$

$$Avec : \sum_{n=-\infty}^{+\infty}\delta(t-nTe) = \frac{1}{Te}\sum_{n=-\infty}^{+\infty} e^{j2\pi nt/Te} \; on\; a : x^*(nTe) = \frac{1}{Te}\sum_{n=-\infty}^{+\infty} x(t)e^{j2\pi nt/Te} = \frac{1}{Te}\sum_{n=-\infty}^{+\infty} x(t)e^{j2\pi ntFe}$$

$$qui\; dans\; le\; domaine\; fréquentiel\; peuts'écrire : X^*(f) = \frac{1}{Te}\sum_{n=-\infty}^{+\infty} X(f-\frac{n}{Te}) = \frac{1}{Te}\sum_{n=-\infty}^{+\infty} X(f-nFe)$$

L'étape de discrétisation temporelle est prise en charge par le circuit **suiveur-bloqueur**.

1.1.1 Choix de la fréquence d'échantillonnage (en *PCM* linéaire) :

L'enregistrement d'un signal numérique suppose une bande passante beaucoup plus large que celle des systèmes analogiques. Historiquement, l'enregistrement audio a utilisé les capacités des systèmes vidéo et le choix des fréquences d'échantillonnage (en accord avec le critère de ***Shannon***) s'est donc fait en fonction des possibilités technologiques de ceux-ci et de leurs standards (bande passante de l'ordre de *3 MHz* à l'origine).

On obtient, en enregistrant *3* échantillons quantifiés sur *16* bits par ligne utile les fréquences suivantes :

PAL / SECAM : *625* lignes, *50 Hz*, *588* lignes utiles i. e. *294* par trame :

$$\Rightarrow 50 \times 294 \times 3 = 44.1\text{kHz}.$$

NTSC : *525* lignes, *59.94 Hz*, *490* lignes utiles i. e. *245* par trame :

$$\Rightarrow 59.94 \times 245 \times 3 = 44.056 kHz$$

TV USA N & B : *525* lignes, *60 Hz*, *490* lignes utiles i. e. *245* par trame :

$$\Rightarrow 60 \times 245 \times 3 = 44.1 kHz.$$

Standard de ***production*** **: *48 kHz***, marge pour le *varispeed,* rapport *3/2* avec le standard de diffusion.
Standard ***Compact Disc*** **: *44.1 kHz***, marge pour le filtrage anti-repliement.
Standard de ***diffusion*** **: *32 kHz***, contraintes des lignes de transmission.

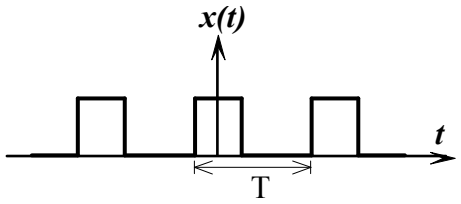
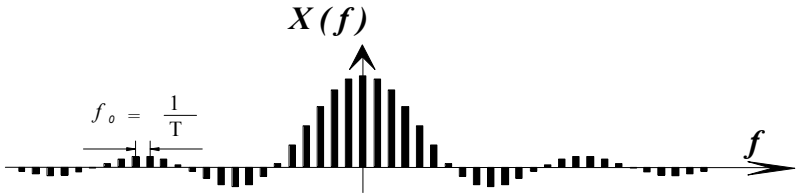
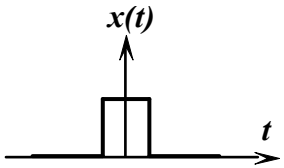
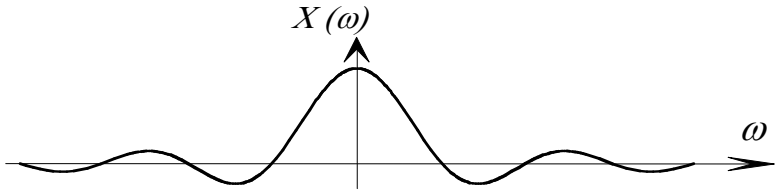
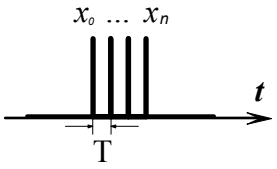
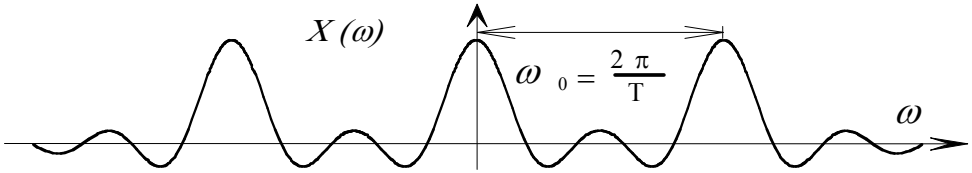
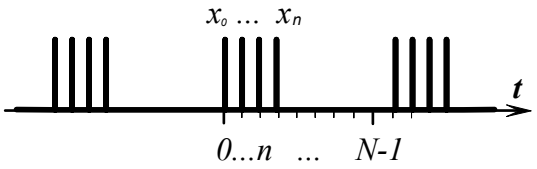
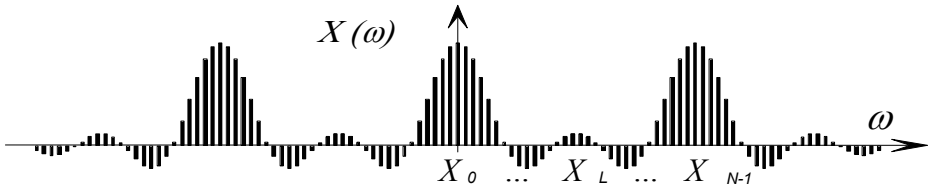
1.2.0 Filtre passe-bas anti-repliement :

Une limitation de la bande passante du signal avant échantillonnage est nécessaire pour éviter le repliement des spectres périodisés par la discrétisation temporelle. Le filtre passe-bas se caractérise par sa **fonction de transfert** dans le domaine fréquentiel (atténuation progressive donnée au delà d'une fréquence de coupure choisie) équivalente à sa **réponse impulsionnelle** dans le domaine temporel (la réponse impulsionnelle étant la transformée de ***Fourier*** inverse de la fonction de transfert).

Les propriétés générales des séries et de la transformée de ***Fourier*** sont présentées à la *figure n°2*.

PROPRIETES GENERALES DES SERIES ET DE LA TRANSFORMEE DE FOURIER

Figure n° 2

SIGNAL		SPECTRE
■ CONTINU ■ PERIODIQUE ■	*SERIE DE FOURIER*	■ DISCRET ■ APERIODIQUE ■
	⟷	
■ CONTINU ■ APERIODIQUE ■	*TRANSFORMATION DE FOURIER*	■ CONTINU ■ APERIODIQUE ■
	⟷	
■ DISCRET ■ APERIODIQUE ■	*TRANSFORMATION DE FOURIER*	■ CONTINU ■ PERIODIQUE ■
	⟷	
■ DISCRET ■ PERIODIQUE ■	*TRANSFORMATION DE FOURIER DISCRETE*	■ DISCRET ■ PERIODIQUE ■
	⟷	

Le problème posé est celui de la réalisation d'un filtre analogique à pente raide, phase linéaire et sans oscillations trop importantes.

Les limites du filtrage analogique sont (cf. *figures n°3a*, *3b* et *4*) :

- un temps de propagation de groupe non constant (phase non-linéaire / réponse impulsionnelle non symétrique ≡ distorsion temporelle).
- une atténuation progressive au delà de la fréquence de coupure fonction de l'ordre du filtre.
- une oscillation de leur réponse en fréquence fonction de l'atténuation.
- une instabilité due à la dérive des composants qui les constituent (temps et température) entraînant une reproductibilité difficile.
- une sensibilité aux parasites (rayonnements électromagnétiques, bruit thermique).

Ces imperfections se traduisent par de la distorsion (non-linéarité) lors des processus d'échantillonnage et de quantification (⇒ erreur d'acquisition).

Les filtres anti-repliement peuvent donc induire du bruit, une erreur d'amplitude due aux oscillations de leur réponse en fréquence (la causalité impliquant une troncature de la réponse impulsionnelle ⇒ phénomène de ***Gibbs*** et fenêtrages conventionnels du traitement du signal…), une distorsion par le repliement des composantes résiduelles et une erreur temporelle (retard de propagation de groupe - distorsion de phase).

Les modèles de filtres de ***Bessel***, ***Buterworth***, ***Chebyshev***, elliptiques, ..., ont été couramment employés pour le filtrage anti-repliement. Leurs caractéristiques ne permettent, néanmoins, qu'une approche du résultat recherché et risquent de condamner leur utilisation dans des systèmes de haute résolution.

L'horloge d'échantillonnage elle même peut induire une erreur si elle n'est pas suffisamment stable (cf. problèmes dus au *jitter*).

Remarque : Une limitation de bande passante dans le domaine fréquentiel se traduit par une dispersion temporelle du signal (limitation de la variation maximale / théorème de ***Bernstein***). Celle-ci est due à la convolution du signal (dans le domaine temporel) par la fonction $(\sin \pi f_c t)/(\pi f_c t)$ dans le cas d'un filtre passe-bas idéal de fréquence de coupure f_c. Un filtrage parfait n'est donc pas physiquement réalisable car il supposerait un temps infini (passé - présent - avenir)...

Système expérimental mis en œuvre pour observer l'effet de l'échantillonnage sur une impulsion

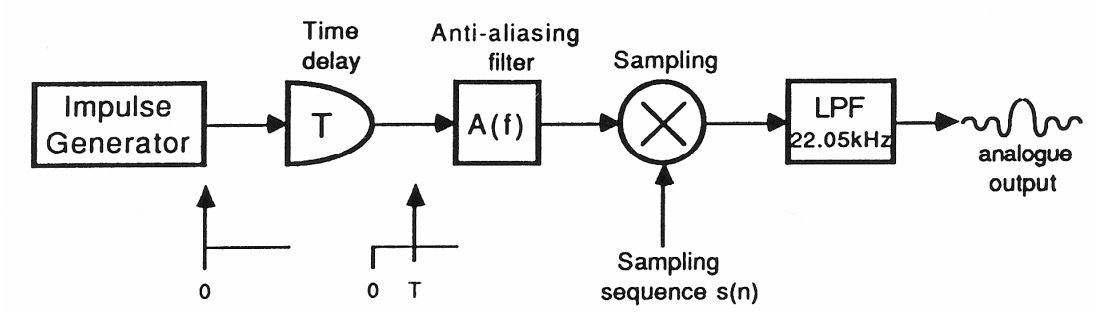
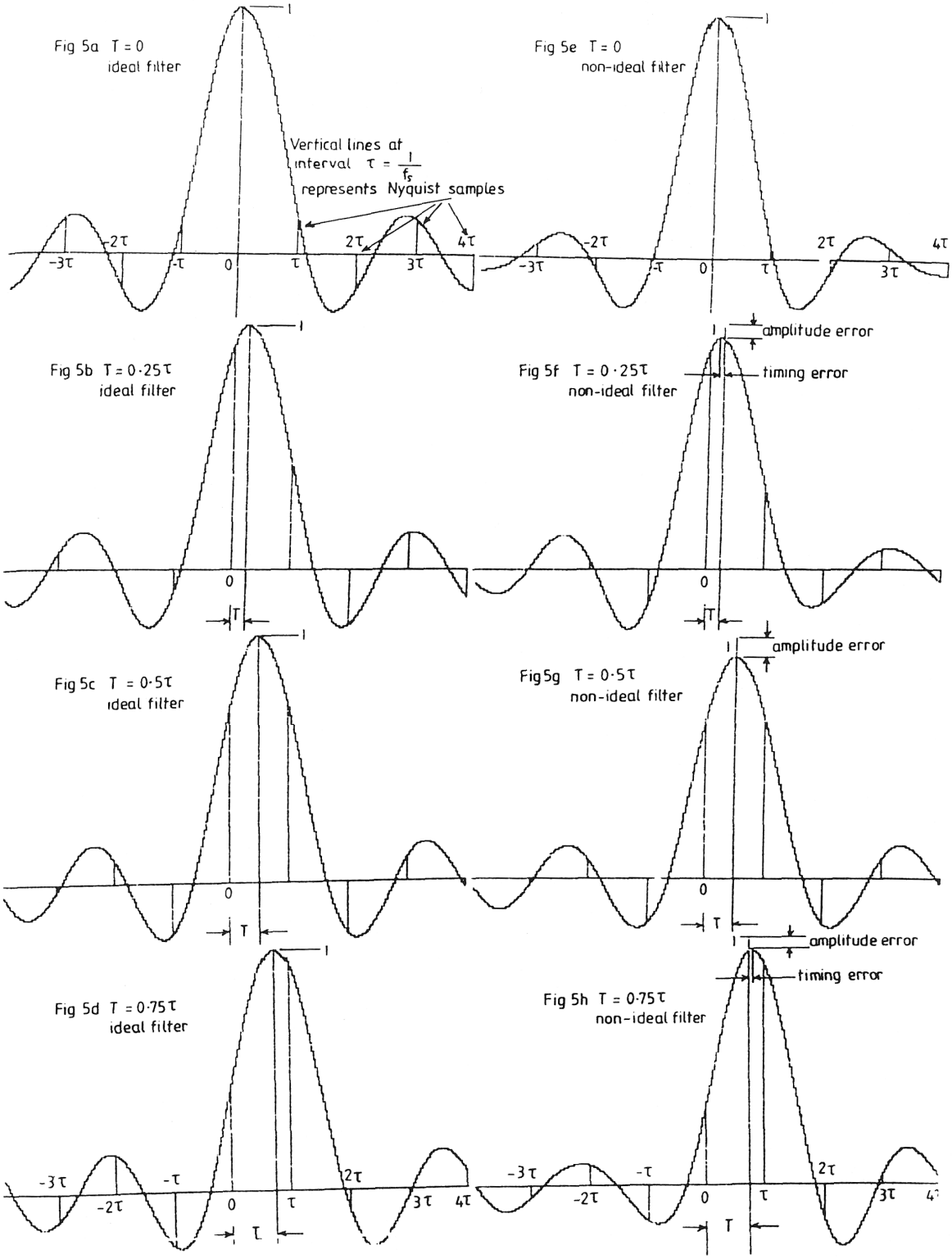


Figure n°3a

Réponse impulsionnelle du filtre anti-repliement



Filtre idéal ***Filtre réel***

Figure n°3b

1.2.1 Circuit suiveur-bloqueur (*track & hold circuit*)

Le circuit suiveur-bloqueur prend en charge l'étape de discrétisation temporelle et de maintien des échantillons le temps de la quantification. L'acquisition et le maintien induisent des erreurs de temps et d'amplitude.

• **Erreurs temporelles :** Le stockage de la valeur à convertir se fait par l'intermédiaire d'un réseau capacitif et pose donc des impératifs de délai d'ouverture, de temps de charge de la capacité et de maintien de celle-ci à la valeur à convertir. Schématiquement, on peut décomposer l'étape d'acquisition en délai d'ouverture ($\Delta t1$), temps de charge ($\Delta t2$) et "verrouillage" sur le signal d'entrée ($\Delta t3$). Le signal d'entrée est ensuite suivi puis maintenu le temps de la conversion. On remarque que les paramètres temporels d'acquisition dépendent de l'amplitude du signal d'entrée.

La phase de poursuite est effectuée à l'aide d'une boucle à verrouillage de phase (*PLL - Phase Locked Loop*) qui peut elle-même poser des problèmes de réglage et d'optimisation.

Le temps de charge total et la phase de maintien peuvent donc induire une erreur d'acquisition, différence entre la valeur stockée et la valeur du signal d'entrée à l'instant d'échantillonnage, qui peut se caractériser par le temps d'ouverture défini comme le temps entre l'ordre de démarrage du processus et celui de la conversion proprement dite. Celui-ci dépendant de la variation du signal d'entrée (cf. remarque sur le *jitter*), l'erreur introduite est donc la résultante du produit (dérivée temporelle du signal d'entrée) × (temps d'ouverture).

Dans les circuits suiveurs, la phase de poursuite n'est pas prise en compte dans le temps d'ouverture. On peut s'affranchir des erreurs causées par le temps d'ouverture par anticipation. Restent les erreurs dues au *jitter* et au *drooping*.

• **Erreurs d'amplitude :** On introduit une erreur d'amplitude lors de l'étape de maintien par le phénomène de *drooping*, dérive de la valeur par "fuite" du circuit de maintien (changement donné par unité de temps).

Remarque : Il existe aussi le phénomène de *feedthrough*, caractérisant le pourcentage de signal d'entrée que l'on retrouve dans la phase de maintien (sorte de diaphonie comprise entre *0.05 %* et *0.005 %* dans les technologiques actuelles c'est à dire quasi négligeable).

La corrélation entre ces erreurs d'acquisition et le signal entrant traduit une non-linéarité du système.

Les caractéristiques du circuit sont mises en évidence *figure n°4*.

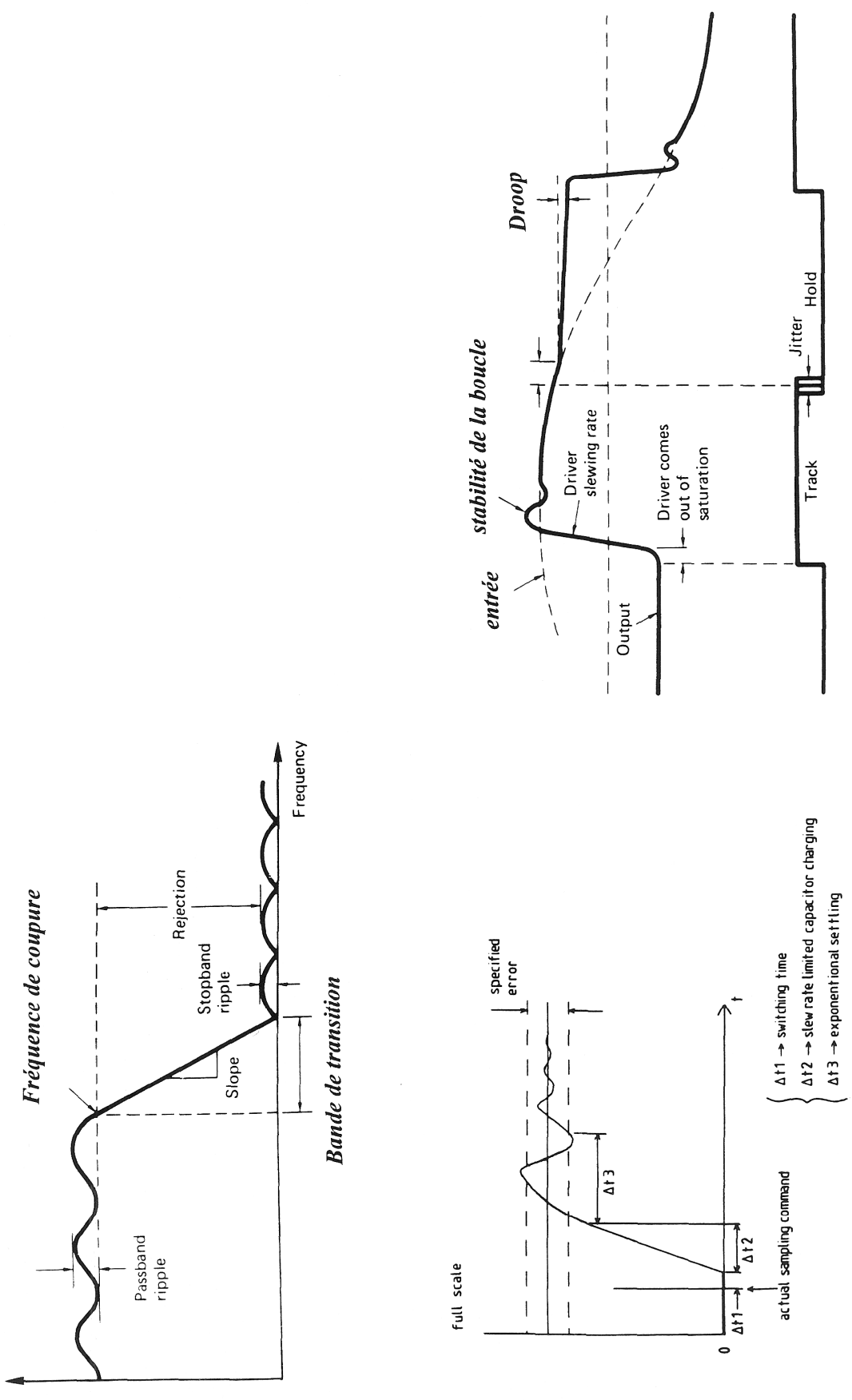

Droop
stabilité de la boucle
entrée
Driver slewing rate
Driver comes out of saturation
Output
Jitter
Hold
Track
Fréquence de coupure
Bande de transition
Frequency
Rejection
Stopband ripple
Slope
Passband ripple
specified error
full scale
Δt3
Δt2
Δt1
actual sampling command
0
t
Δt1 → switching time
Δt2 → slew rate limited capacitor charging
Δt3 → exponentional settling

Figure n°4

1.3.0 Quantification :

Le terme de quantification désigne l'étape de ***discrétisation en amplitude*** (échantillonnage vertical). L'amplitude crête à crête du signal est découpée en pas égaux dans les systèmes de ***quantification linéaire*** (pas de quantification *Q*, cf. *figures n°5*, *6a*, *6b*, *7*).

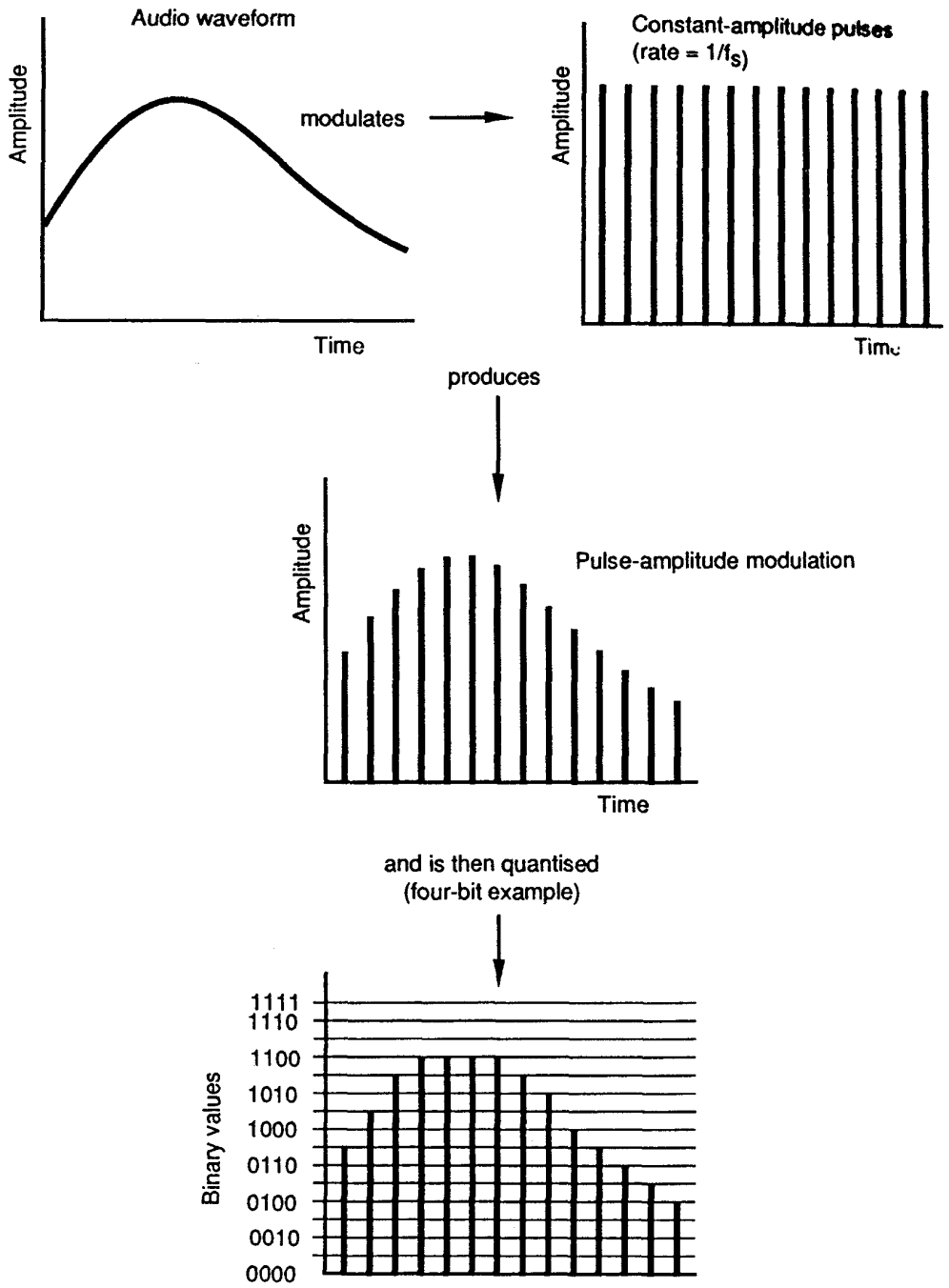


Figure n°5

La ***résolution*** de quantification est déterminée par le nombre de pas, à raison de 2^n pas pour un codage sur *n* bits ⇒ *16* bits : *65 536* pas, *20* bits : *1 048 576* pas, *24* Bits : *16 777 216* pas...

La quantification induit une distorsion (erreur corrélée au signal) par approximation des amplitudes comprises entre deux pas ($-Q/2 < erreur < Q/2$), le seuil de décision étant centré. Cette erreur est d'autant plus gênante que l'amplitude du signal est faible (ordre de grandeur des deux). On verra que le procédé de *ditherisation* permet d'uniformiser cette erreur par linéarisation statistique des marches de quantification (transformation de la distorsion en bruit résiduel).

La distorsion de quantification étant introduite après filtrage, le spectre de celle-ci (qui peut dépasser la bande de fréquence utile et s'atténue progressivement pour des raisons énergétiques) pourra se replier une fois périodisé par échantillonnage. L'échantillonnage modifie donc le spectre de distorsion par repliement sans en changer la puissance de façon caractéristique entre *0* et *Fe/2* (cf. *figure n°8*).

Si l'on considère, dans un premier temps, que l'erreur est **aléatoire** (densité de probabilité uniforme entre $-Q/2$ et $+Q/2$), la dynamique **au niveau maximum et en régime harmonique** est reliée à *n* par la relation :

$$\frac{S}{B}(endB) = 6.02n + 1.76$$

Amplitude maximale du signal : $A = \frac{2^n Q}{2}$ Amplitude efficace : $\frac{A}{\sqrt{2}}$

Valeur quadratique moyenne du signal : $\left[\frac{2^n Q}{2\sqrt{2}}\right]^2$

Puissance moyenne du bruit de quantification (variance) : $\sigma_b^2 = E[b^2] = \frac{1}{Q}\int_{-Q/2}^{+Q/2} b^2\, db = \frac{Q^2}{12}$

On a donc ***S/B (en dB) = 10 log 2^{2n} + 10 log 3/2 = n × 20 log 2 + 10 log 3/2***

(20 log 2 = 6.02 et 10 log 3/2 = 1.76).

On se méfiera de l'emploi de grandeurs statistiques et de densité de probabilité uniformes avant *dithering* (cf. § 1.5.0.) et sur-échantillonnage (cf. § 2.0.0).

Si le bruit est considéré uniforme, sa densité de puissance peut s'exprimer comme :

$$N_b(f) = \frac{Q^2}{12Fe} = \frac{2^{-2n}}{3Fe}$$

Remarques :

• L'étape de quantification est un échantillonnage au même titre que la discrétisation temporelle, la période spatiale d'échantillonnage en amplitude étant de Q. Nous sommes ici en présence d'une pseudo modulation de phase... L'échantillonnage et la quantification peuvent donc se résumer à une double modulation en amplitude et en phase du signal de départ.

• **Lien entre la stabilité de l'horloge (*jitter*) et la distorsion (erreur) de quantification :**

On peut, pour rendre compte du défi technologique, calculer la déviation τ de l'horloge d'échantillonnage qui introduirait une erreur de quantification d'un pas (Q) dans un système à n bits.

Le calcul peut s'effectuer en régime sinusoïdal dans les conditions les plus critiques, *20 kHz* pleine échelle (variation maximale en une période d'échantillonnage) près de son passage à zéro (cf. *figure n°9*).

La pente est de $A\omega$, A étant l'amplitude maximale i.e. $A=(Q \times 2^n)/2$ et ω la pulsation :

$$\omega=2\pi f=2\pi \times 20\,000 \text{ rad.s}^{-1}$$

$$\text{On a donc } \tau = Q/(A\omega) \quad \tau = (2^{1-n})/\omega \quad \text{et donc en 16 bits } \tau = 243 \text{ ps (soit } 243 \times 10^{-12} \text{ secondes)}$$

(signal sinusoïdal : $A \sin\omega t$, pente $A\omega \cos\omega t = A\omega$ en 0)

Le même calcul peut se faire pour évaluer la limite entre la distorsion due au *jitter* d'horloge et la distorsion de quantification. Dans ce cas l'erreur maximale admise est de $Q/2$ et le résultat donc deux fois moindre c'est à dire. ***121.5 ps*** pour le cas du ***16 bits***. On retrouve la valeur proposée par la recommandation *AES* concernant la synchronisation des équipements numériques (recommandation *DRAFT AES 11-19XX*, donnée en annexe).

On peut présenter les résultats obtenus pour différentes résolutions :

RESOLUTION	*ERREUR DE Q*	*ERREUR DE Q / 2*
16 BITS	*243 ps*	*121.5 ps*
18 BITS	*60.7 ps*	*30.4 ps*
20 BITS	*15.2 ps*	*7.6 ps*
22 BITS	*3.8 ps*	*1.9 ps*
24 BITS	*1 ps*	*0.5 ps*

Cette analyse suppose que des conversions de haute résolution ($\geq$ *16 bits*) puissent se faire par le même type de technologie ce qui, on le verra plus loin, n'est pas possible. Cela permet tout de même de donner un ordre d'idée des problèmes de stabilité d'horloge qui peuvent se poser.

L'erreur introduite par le *jitter* est donc plus importante aux hautes fréquences (variation du signal d'entrée) et induit une distorsion de modulation (l'effet dépendant du niveau et du spectre du signal d'entrée).

La variation d'horloge (hors convertisseurs) peut aussi introduire un effet de « pleurage et scintillement ».

Dans les systèmes *multibits* (*PCM* linéaires), l'importance de la distorsion de quantification et de la distorsion due au *jitter* dépend du couple niveau / spectre.

Les tolérances sont donc très serrées et la précision d'horloge devient une condition déterminante de la qualité de la conversion, le *jitter* étant synonyme d'erreur d'acquisition.

L'***AES*** propose de synchroniser l'ensemble d'une installation audionumérique sur un signal de référence (***Digital Audio Reference Signal***). On peut imaginer utiliser l'horloge au césium disponible par satellite ou radio[1]. Cette horloge a une précision de l'ordre de la picoseconde et peut donc convenir pour des systèmes audio évolués (vers le *24* bits...).

L'incertitude temporelle est analogue à une incertitude sur les niveaux de quantification. On verra que les problèmes de *jitter* sont moins critiques dans les systèmes sur-échantillonnés par moyennage de l'erreur d'horloge sur un grand nombre d'échantillons lors du retour à la fréquence d'échantillonnage standardisée. Les tolérances données dans le tableau sont dans ce cas à multiplier par le facteur de sur-échantillonnage.

La corrélation entre l'erreur de quantification et le signal d'entrée peut être mise en évidence en considérant qu'une quantification parfaite (sans erreurs / pas égaux en amplitude) peut se représenter comme une modulation de la largeur des marches de quantifications en fonction du signal d'entrée (modulation de la *Fe*). La fréquence d'échantillonnage étant fixée (largeur des pas), elle introduit une distorsion d'amplitude qui dépend du signal entrant (corrélation).

Cette présentation a l'avantage de permettre de faire le lien avec les problèmes de *jitter*.

Définition du bruit : Le bruit est un signal aléatoire se superposant au signal (décorrélé de celui-ci). Le problème de corrélation du bruit de quantification dans les systèmes *PCM* linéaires (cf. types de codages § 4.0.0) est à l'origine du procédé de *ditherisation*.

Le cumul des erreurs des différents étages de la conversion analogique-numérique (filtrage anti-repliement, circuit suiveur-bloqueur, quantificateur) ne permet pas le développement de techniques de haute résolution et limite ce genre de système (*PCM* linéaire) au *16* bits / *48 kHz*. Une technologie de conversion haut de gamme devra donc s'affranchir de ces étapes (vers le *24* bits en production).

[1] La seconde est définie comme la durée de 9 192 631 770 périodes de la radiation correspondant à la transition entre les deux niveaux hyperfins de l'état fondamental de l'atome de Césium 133, au repos et à une température de 0° Kelvin.(13ème conférence des Poids et Mesures, 1967).

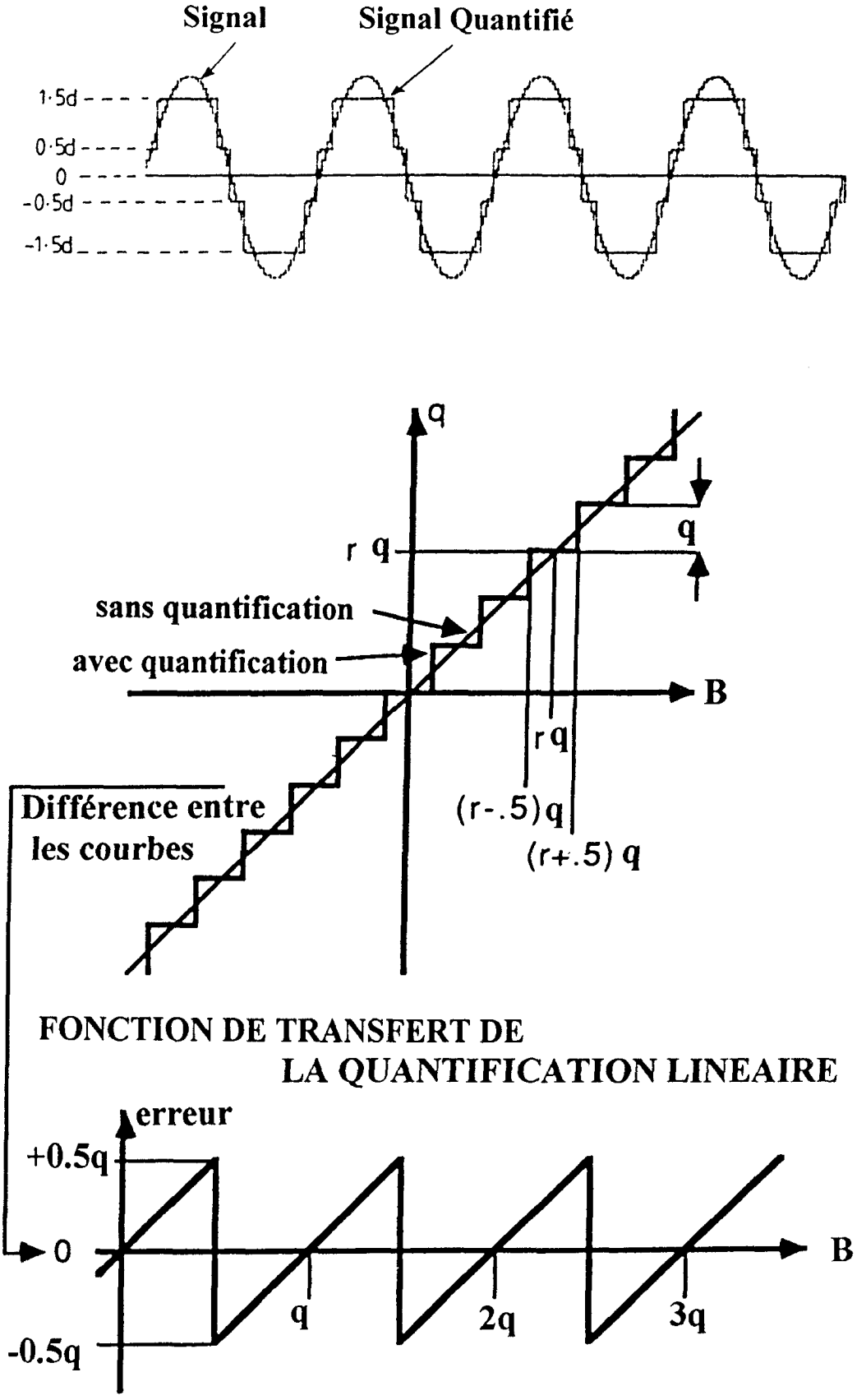
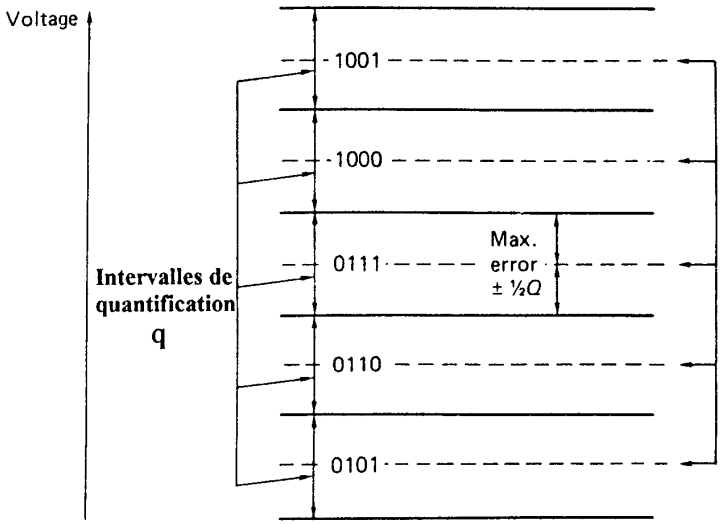

Signal
Signal Quantifié
1·5d
0·5d
0
-0·5d
-1·5d
q
r q
q
sans quantification
avec quantification
B
r q
(r-.5)q
(r+.5) q
Différence entre
les courbes
FONCTION DE TRANSFERT DE
LA QUANTIFICATION LINEAIRE
erreur
+0.5q
0
B
q
2q
3q
-0.5q

Figure n°6a



$$-\frac{q}{2} \leq Erreur \leq +\frac{q}{2}$$

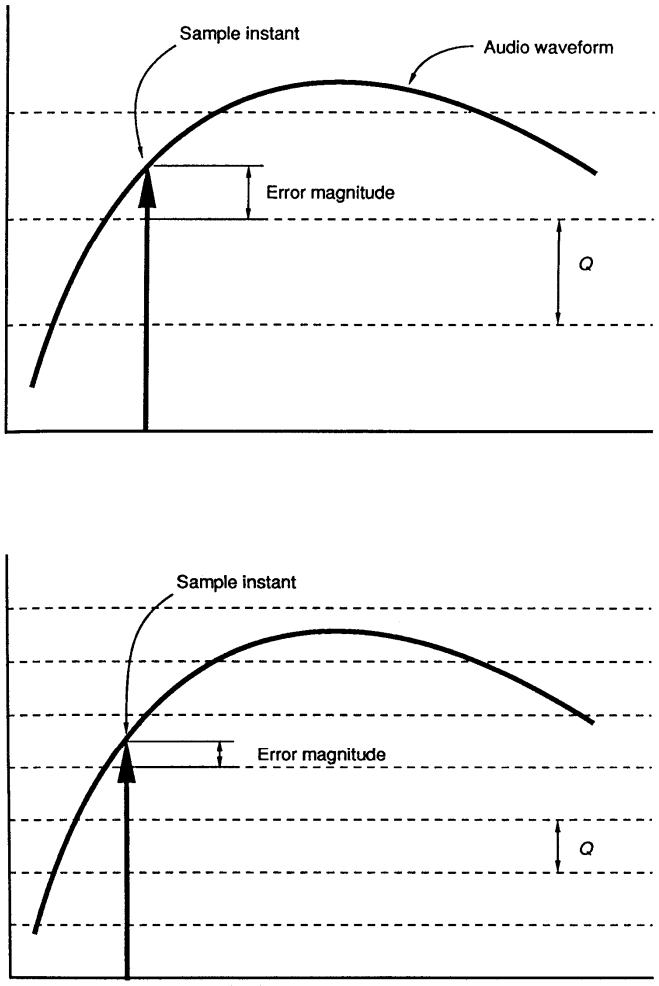


Figure n°6b

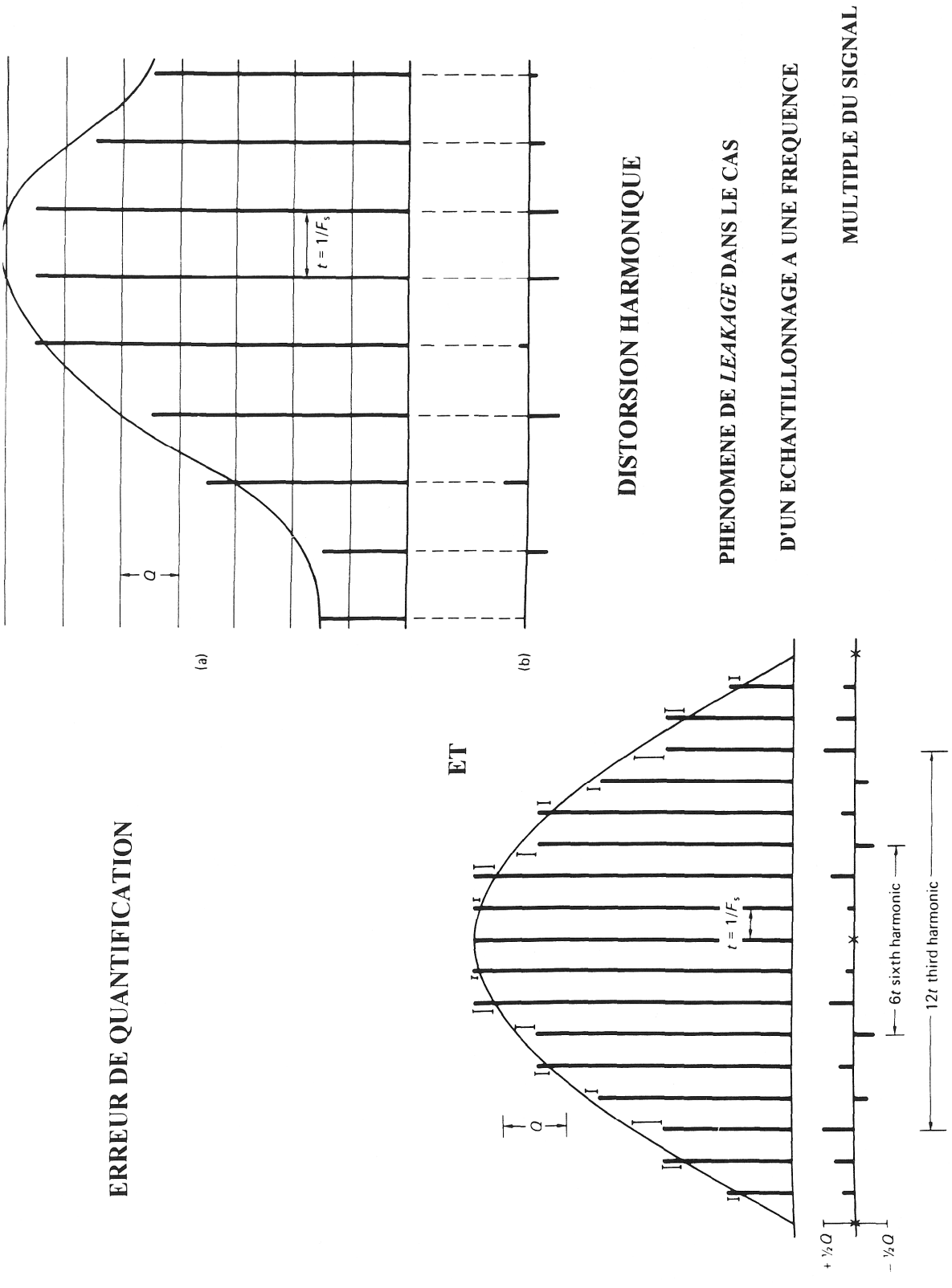

ERREUR DE QUANTIFICATION
(a)
(b)
Q
$t = 1/F_s$
ET
DISTORSION HARMONIQUE
PHENOMENE DE *LEAKAGE* DANS LE CAS
D'UN ECHANTILLONNAGE A UNE FREQUENCE
MULTIPLE DU SIGNAL
$t = 1/F_s$
Q
+ ½Q
- ½Q
6t sixth harmonic
12t third harmonic

Figure n°7

Phénomène de repliement de la distorsion de quantification

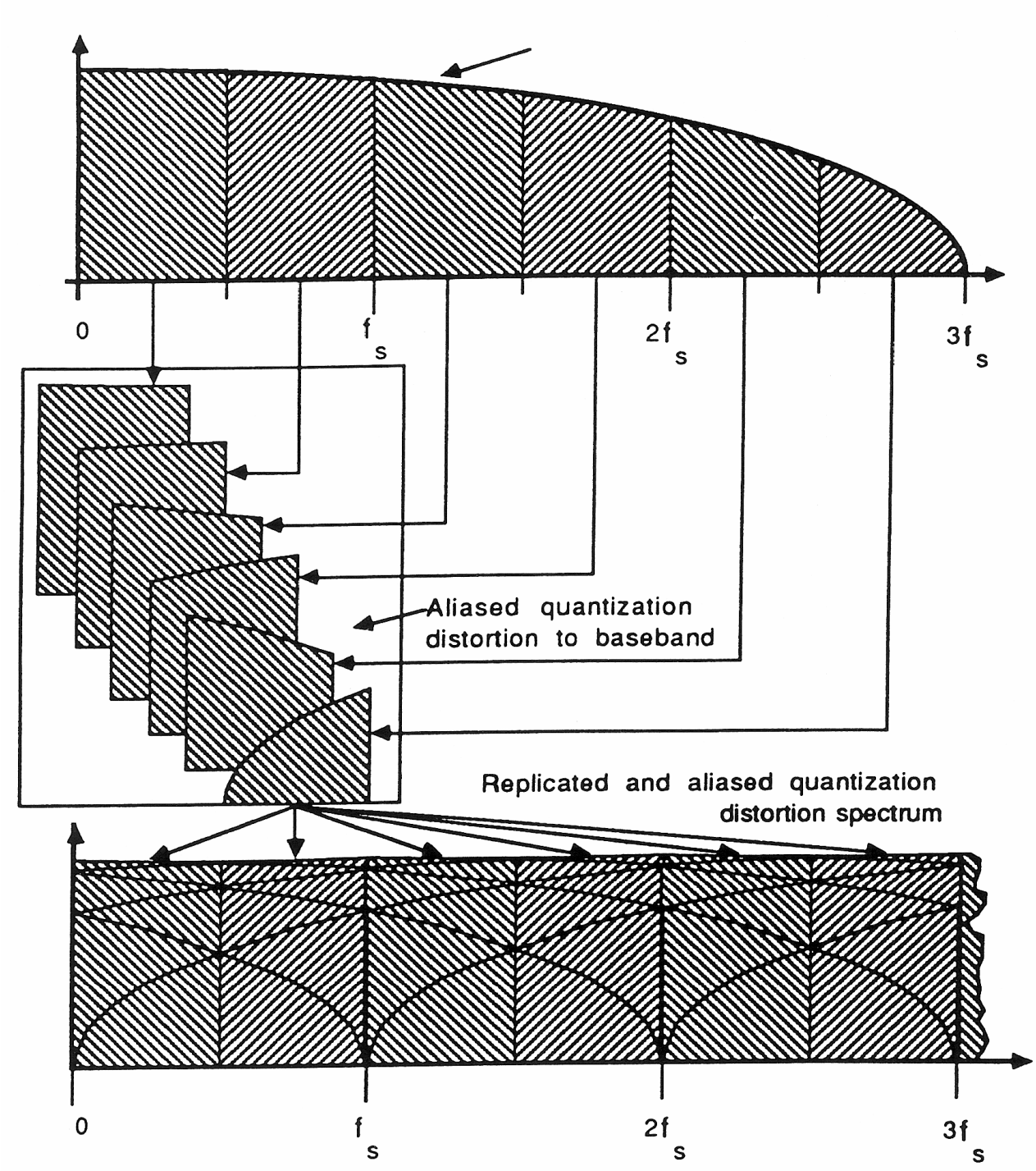
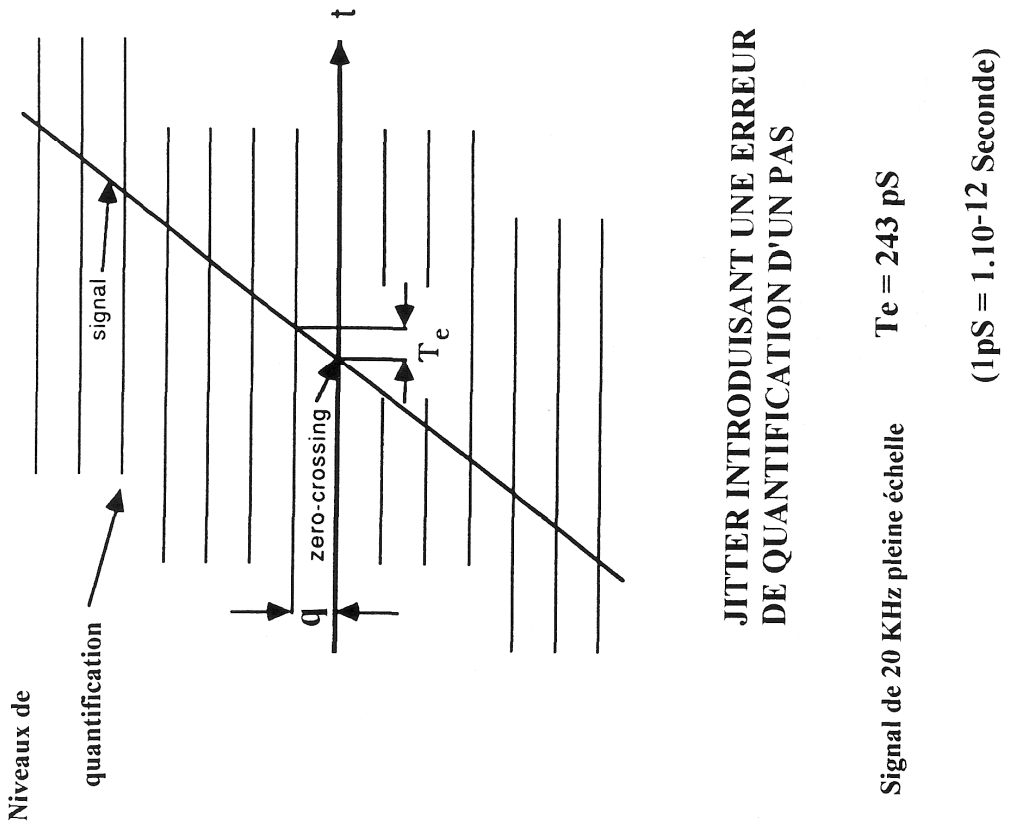


Figure n°8


Niveaux de
quantification
signal
zero-crossing
q
t
T_e
JITTER INTRODUISANT UNE ERREUR
DE QUANTIFICATION D'UN PAS
Signal de 20 KHz pleine échelle
Te = 243 pS
(1pS = 1.10-12 Seconde)

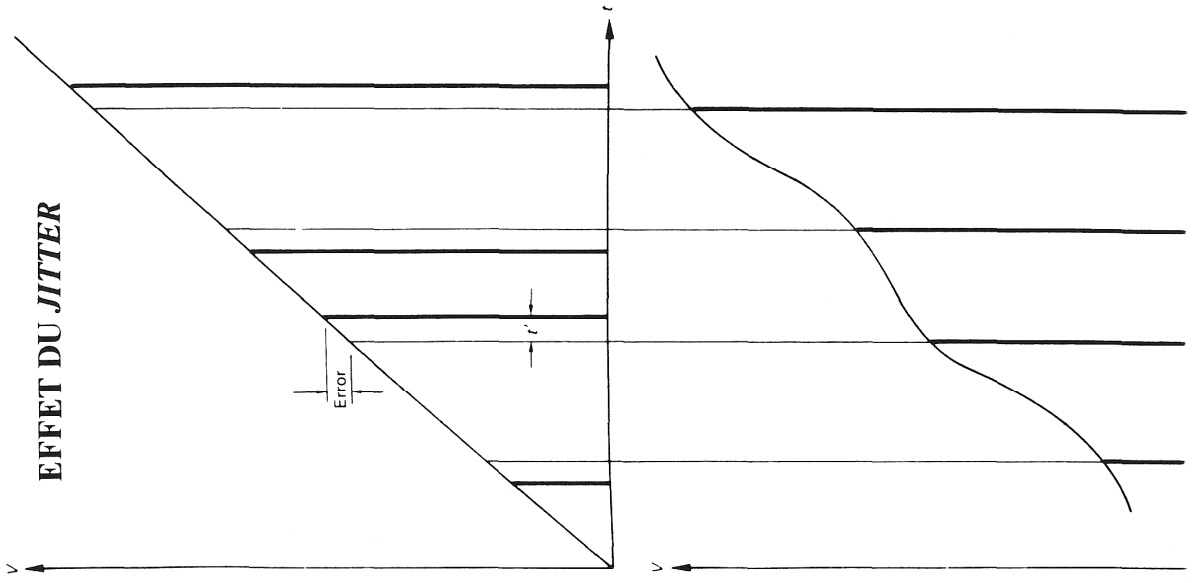

EFFET DU *JITTER*
Error
t'
t
v
v

Figure n°9

1.4.0 Reconstitution du signal:

La reconstitution du signal se compose d'une phase de maintien des échantillons à la tension correspondante à leur quantification (les informations numériques n'ayant plus de dimension énergétique) et d'un filtrage analogique.

Le filtre de reconstitution a les mêmes caractéristiques que le filtre anti-repliement (filtre passe-bas de lissage, interpolation par intégration analogique, sachant qu'une limitation de bande passante est équivalente à une limitation de variation).

Le maintien des échantillons (porte temporelle) se traduit par un filtrage passe-bas en *sinx/x* (transformée de ***Fourier*** de la porte temporelle). Si le maintien est fait sur toute la durée de la période d'échantillonnage, une atténuation de *-4 dB* à *Fe/2* est obtenue (cf. *figure n°10*).

Pour compenser ce filtrage on peut soit re-échantillonner le signal après convertisseur de manière à réduire la durée de la porte et rester dans la partie linéaire de la réponse en fréquence du filtre (compromis durée, énergie / plage linéaire du filtre), soit introduire une pondération en *x/sin x*.

Cette dernière solution est la plus employée car elle n'a pas d'effet sur la dynamique du système.

Dans les systèmes sur-échantillonnés, les opérations de filtrage sont "rejetées"dans le domaine numérique.

1.5.0 *Dithering* :

Le procédé de *ditherisation* permet de décorréler l'erreur de quantification du signal par ajout d'un bruit aléatoire de faible niveau (dont on peut connaître la densité de probabilité) avant numérisation. Le but étant de transformer la distorsion de quantification en bruit résiduel. En principe, le bruit ajouté avant conversion devra être ensuite soustrait dans le domaine numérique pour éviter de dégrader la dynamique disponible. En pratique, cette soustraction n'est pas toujours effectuée en raison du faible niveau de celui-ci.

Cette technique permet de coder les amplitudes du signal entre les niveaux de quantification sous forme de modulation de largeur d'impulsion (*PWM* : *Pulse Width Modulation*) dans les systèmes sur-échantillonnés. Un moyennage temporel pourra ensuite permettre de retrouver ces valeurs intermédiaires.

Le *dithering* étant un procédé statistique de linéarisation de la caractéristique de quantification, il demande un moyennage (cf. *figure n°11*). Son application optimale est donc réservée aux systèmes sur-échantillonnés qui permettent d'effectuer cette moyenne au cours de la décimation.

Dans les "systèmes de type " ***Nyquist*** ", son utilisation permet une décorrélation de la distorsion de quantification au prix d'une augmentation de la densité de puissance de bruit. Le bruit étant perceptivement moins gênant que la distorsion, on "noie" celle-ci.

Le *dither* à une efficacité maximale quand son amplitude efficace est de $Q/3$ et que sa densité de probabilité est triangulaire (*Triangular Probability Density Function -TPDF-,* cf. ***Lipshitz***, *« Dither Theory »*).

Blesser a montré qu'un *dither* de bande étroite centré sur la fréquence de ***Nyquist*** pouvait être utilisé.

Un générateur de séquences pseudo-aléatoires (auquel on impose une certaine densité de probabilité) converties par *DAC* (*Digital to Analog Converter*) peut servir de générateur de *dither*. L'escalier de transfert est dans ce cas linéarisé par flou statistique, cf. *figure n°12*.

La numérisation du signal + *dither* est donc statistiquement linéaire et il sera théoriquement possible de retrancher le *dither* en tenant compte du retard de conversion après numérisation.

Le procédé de *ditherisation* permet d'approcher la dynamique théorique calculée § 1.3.0 et trouve une application naturelle dans tous les cas de quantification ou de requantification (conversion, convertisseurs de fréquence d'échantillonnage, algorithmes de traitement numérique des signaux, problèmes de troncature et d'arrondis...).

Remarques :

• En réalité, l'erreur de quantification n'est pas strictement corrélée au signal* en raison des parasites inhérents aux systèmes électroniques (bruit thermique, rayonnements électromagnétiques) et des précisions d'horloge (*jitter* minimal). Ce bruit résultant (dont on ne connaît pas la densité de probabilité) est parfois utilisé comme *dither*. Dans ce cas l'étape de soustraction n'est évidemment plus possible.

* si c'était le cas, on pourrait imaginer pouvoir soustraire cette distorsion du signal utile après une première analyse (systèmes en temps différés, restauration sonore...)

• On peut parler indifféremment d'erreur ou de distorsion de quantification. Dans ce cas le terme de distorsion est pris au sens large (non-linéarité d'un phénomène). Cette distorsion n'est harmonique que dans les cas particuliers du *leakage* (distorsion harmonique paire cf. *figure n°7*) et de la sous-quantification extrême d'un signal sinusoïdal (distorsion harmonique impaire par codage d'un signal sinusoïdal sous forme de signal carré / écrêtage). Elle est comparable dans le cas général à une distorsion "de modulation" (pas de signal, pas de distorsion). La notion de corrélation reflète ces caractéristiques.

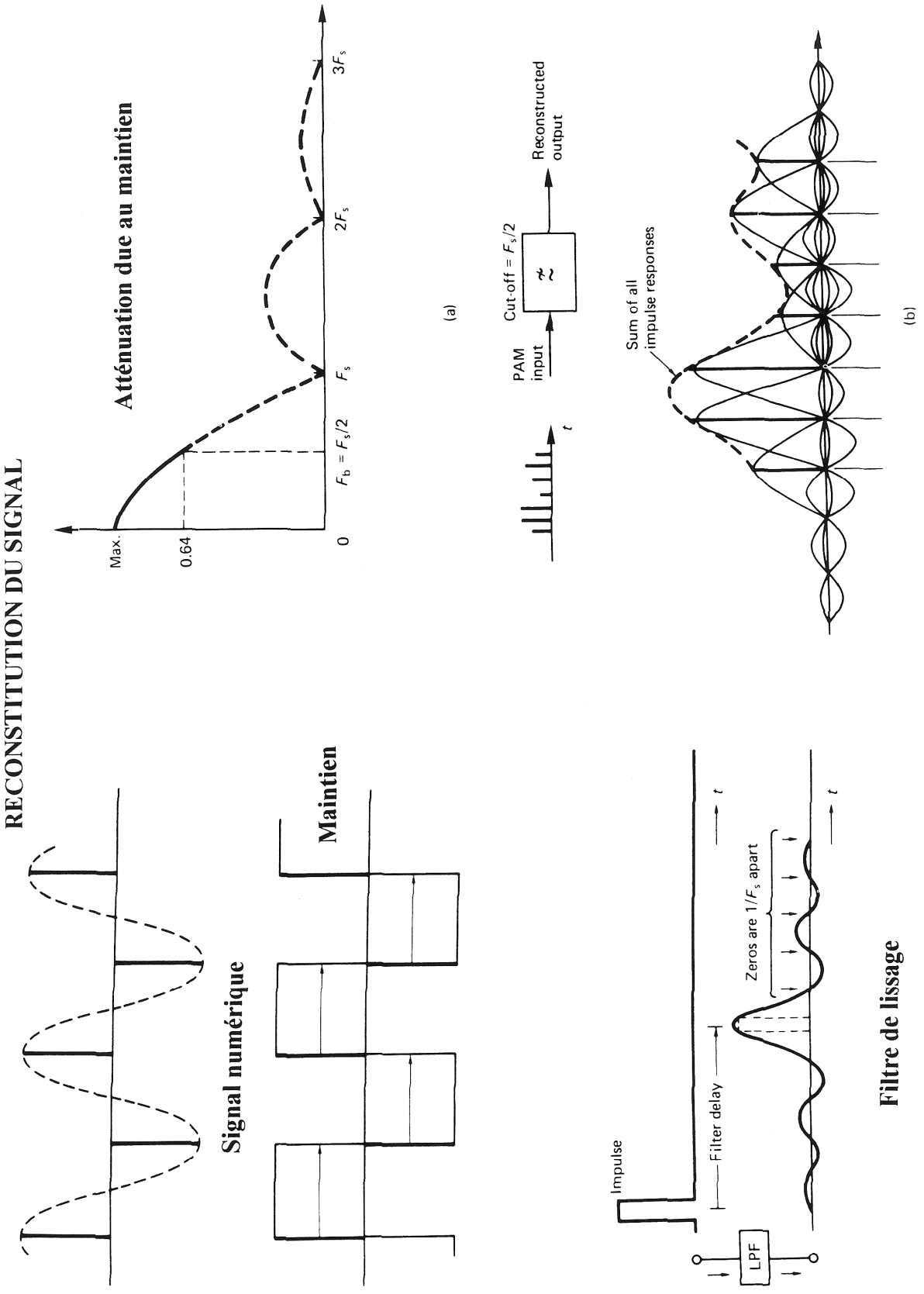
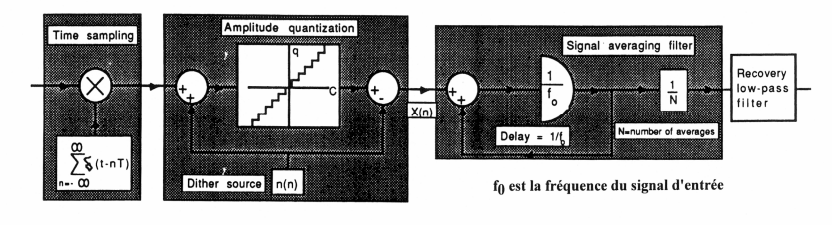
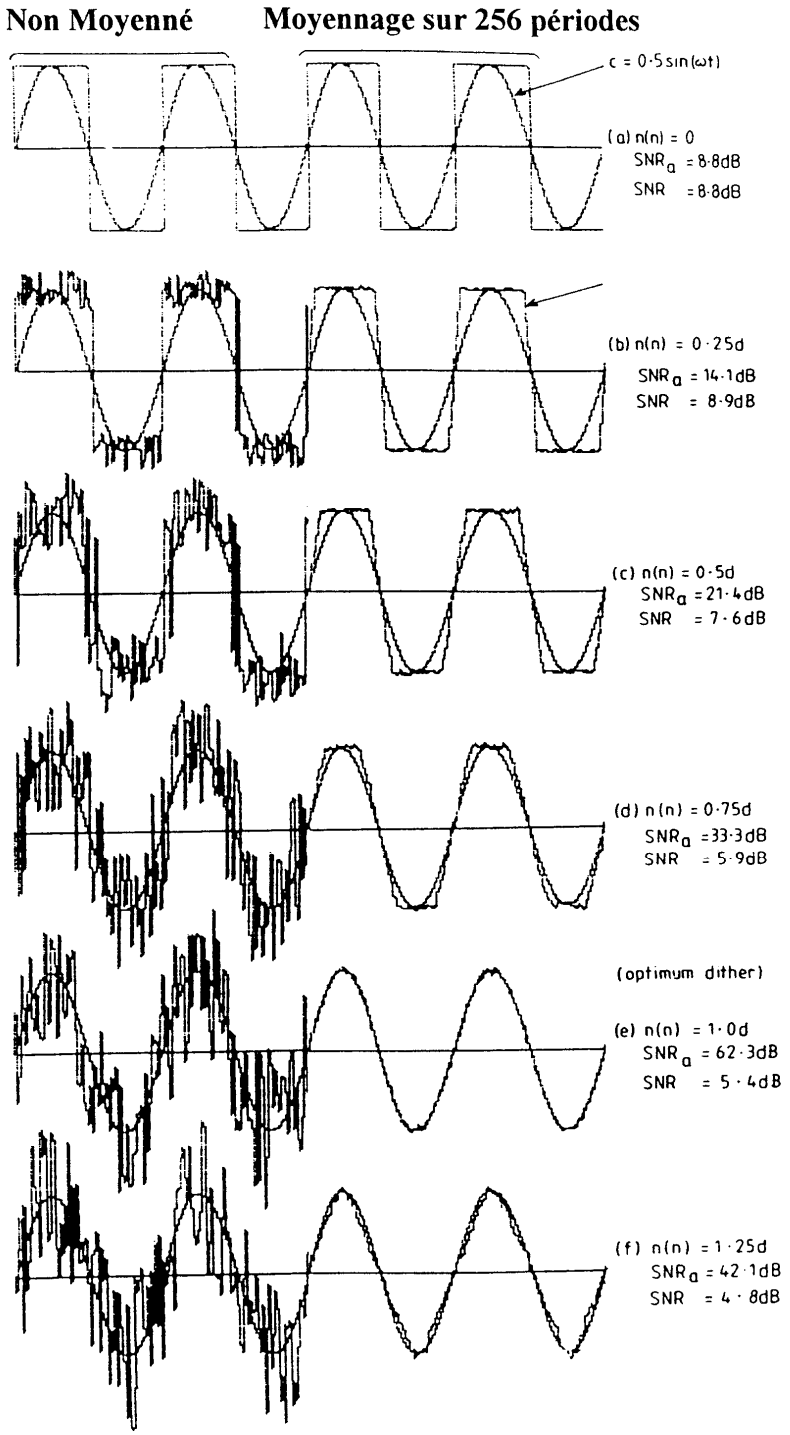

RECONSTITUTION DU SIGNAL
Signal numérique
Maintien
Atténuation due au maintien
Max.
0.64
0
$F_b = F_s/2$
F_s
$2F_s$
$3F_s$
(a)
PAM input
Cut-off = $F_s/2$
Reconstructed output
t
Sum of all impulse responses
(b)
Impulse
Filter delay
Zeros are $1/F_s$ apart
LPF
Filtre de lissage

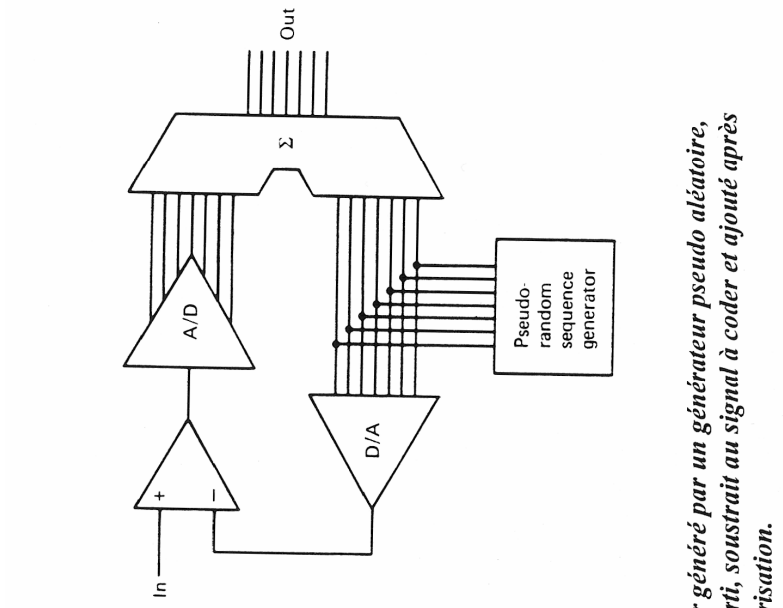
Figure n°10





Le dither à une amplitude variant de 0 à 1,25 Q

Figures n°11



Dither généré par un générateur pseudo aléatoire, converti, soustrait au signal à coder et ajouté après numérisation.

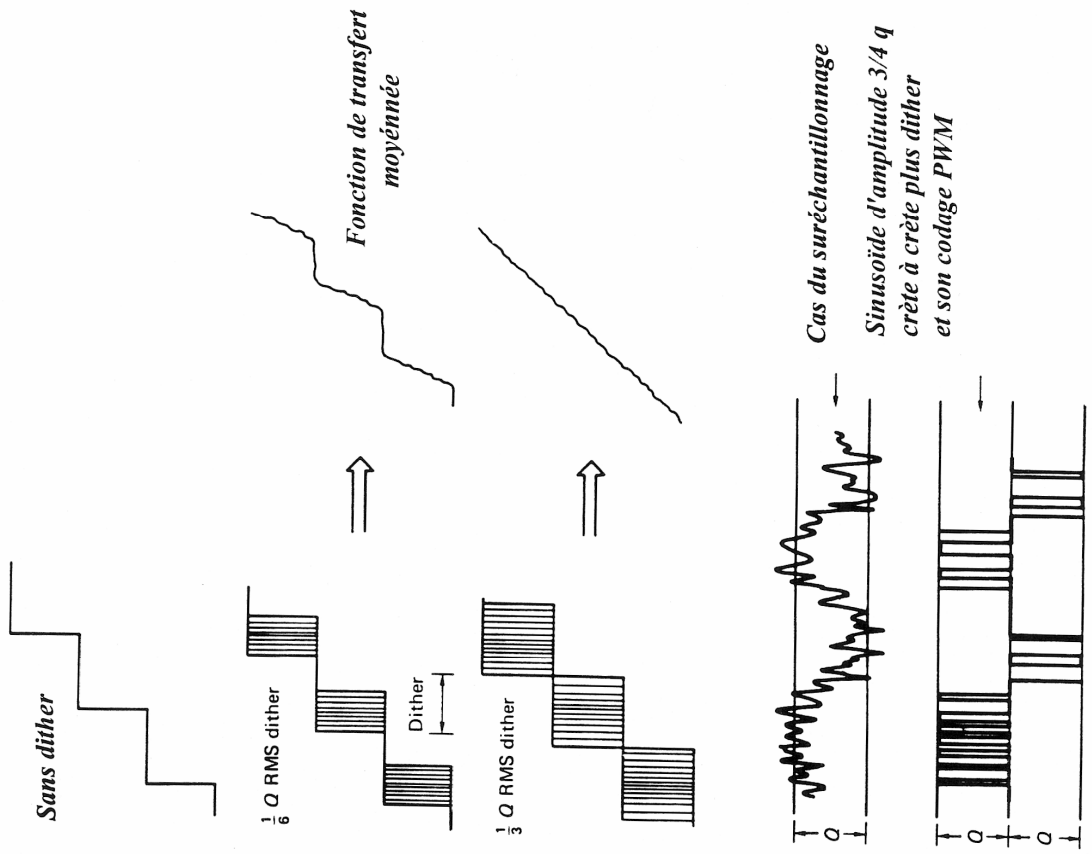


Figure n°12

2.0.0 Sur-échantillonnage :

Le sur-échantillonnage (échantillonnage à une fréquence supérieure à la fréquence critique) permet de s'affranchir de certains défauts de la conversion *multibits* (conversion parallèle classique) en vue d'atteindre les performances théoriques et d'envisager de meilleures résolutions.

Parmi ces défauts, on peut citer :

• Dans la chaîne de conversion :

- bruit et distorsion dus au filtre anti-repliement (cf. § 1.2.0).
- acquisition non-linéaire des circuits suiveurs bloqueurs aux fréquences de fonctionnement (erreurs d'amplitude et de temps).
- *jitter* entraînant des erreurs d'acquisition / restitution traduites sous forme de distorsion

• Convertisseur proprement dit (cf. *figure n°13*) :

- Temps de conversion...
- erreur de gain (rotation de la pente de la fonction de transfert) pouvant entraîner une erreur d'*offset*.
- erreur d'*offset* (ajout d'une composante continue) ⇒ écrêtage, perte de définition en montage / mixage.
- linéarité de la fonction de transfert (sans tenir compte des marches de quantification) ⇒ mêmes effets que la linéarité en analogique (distorsion harmonique, par inter-modulation...)*.
- précision absolue : si toutes les sources de courant dérivent de façon analogue, la linéarité est maintenue bien que l'on s'écarte des valeurs idéales.

* La linéarité est divisée en deux classes, linéarité différentielle (égalité des pas de quantification) et monotonie (mêmes variations en entrée et sortie). La non monotonie étant un cas particulier de non-linéarité différentielle (stabilité et précision des références de quantification).

On définit R, le **facteur de sur-échantillonnage** par $Fse = RFe$, Fse étant la fréquence de sur-échantillonnage. En général R est choisi comme puissance de deux (implémentation informatique) mais cette restriction n'est pas limitative.

Si $2 \leq R \leq 16$, le sur-échantillonnage est dit léger par comparaison au cas où $R > 16$.

La réduction de fréquence d'échantillonnage est la **décimation**, son augmentation l'**interpolation**.

La décimation est une perte irréversible d'information par réduction de bande passante, l'interpolation n'est, en aucun cas, un gain d'information malgré l'élargissement spectral possible par cette méthode (la bande passante utile restant la même).

Défauts classiques des convertisseurs

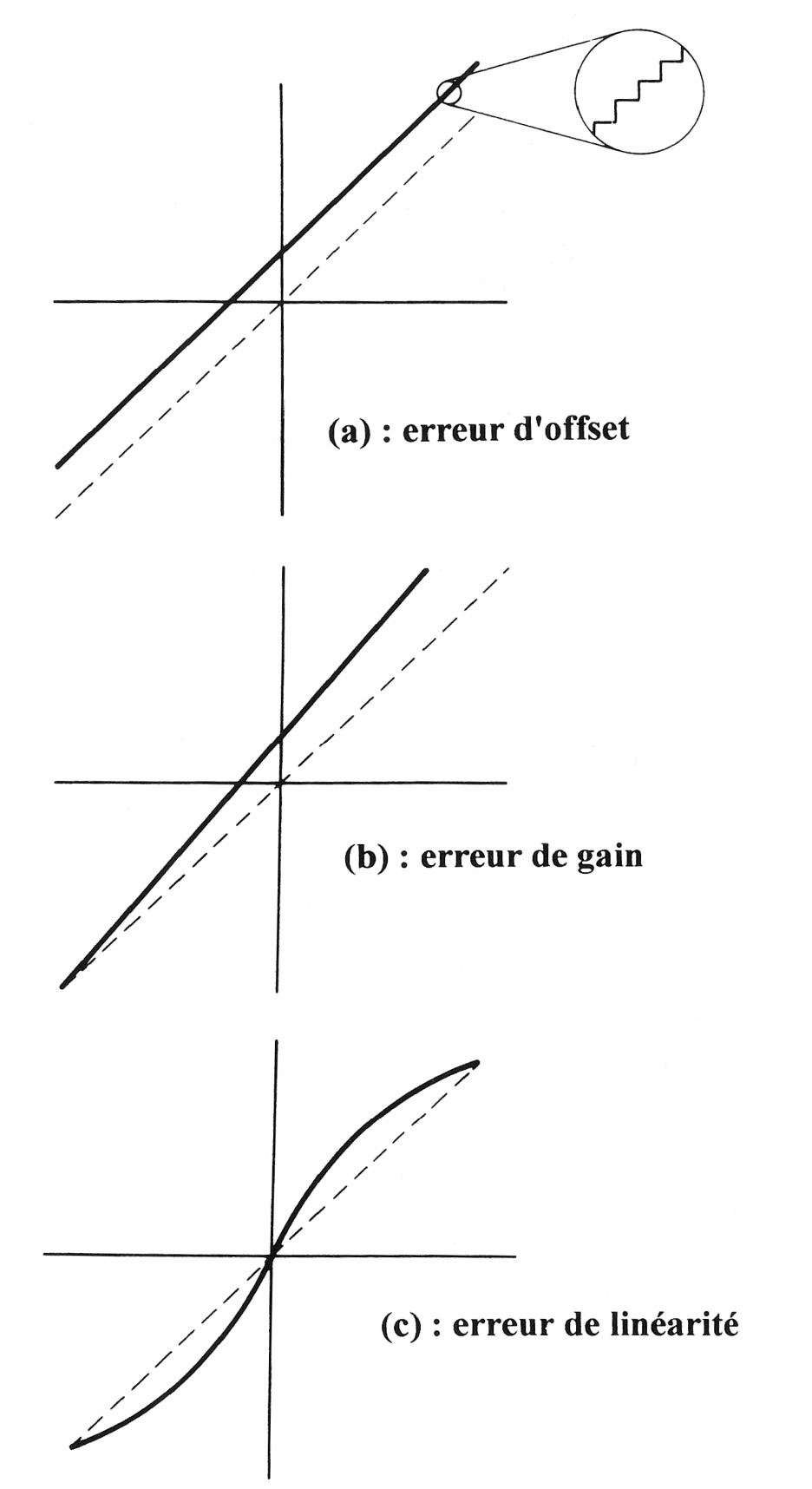


Figure n°13

2.1.0 Exemple d'un convertisseur analogique-numérique sur-échantillonné quatre fois :

On suppose que l'optimisation du *dither* entraîne une répartition uniforme du bruit résiduel entre *0* et *Fse/2 = 88.2 kHz* et que *Fse = 4 x Fe = 4 x 44 100 = 176.4 kHz*.

Le signal d'entrée doit donc être limité à *176.4 - 22.05 = 154.35 kHz* pour éviter le repliement dans la bande utile (cf. *figure n°14*).

Ce type de convertisseur nécessite donc un filtrage anti-repliement beaucoup moins efficace que dans le cas de l'échantillonnage à la fréquence critique (le repliement hors de la bande utile n'étant pas gênant).

On remarque que la répartition du bruit de quantification sur une bande passante quatre fois plus large entraîne une réduction de sa puissance en *1/R* (ici *1/4*) dans la bande *[0; Fe/2]*.

Ceci suppose l'uniformité de son spectre dont on s'est assuré par une *ditherisation* optimale.

Le sur-échantillonnage par *4* permet donc une réduction de la puissance de bruit de *6 dB (10 log 4,* cf. *figure n°15)*.

Cette augmentation de dynamique est équivalente à une résolution supérieure d'un bit (*n' = n+1*).

On peut généraliser la relation entre le rapport signal sur bruit d'un système sur-échantillonné et sa résolution comme :

$$\textbf{\textit{S/B (en dB) = 6,02n + 1,76 + 10log R}}$$

et l'on peut exprimer la densité de puissance de bruit sous la forme :

$$N_b(f) = \int_{-F\,max}^{+F\,max} \frac{Q^2}{12}\frac{1}{Fse} = \frac{Q^2}{12}\frac{2F\,max}{Fse} = \frac{Q^2}{12R}$$

On en déduit le nombre de bits équivalents n' comme **:** $n' = n + \frac{\log R}{\log 4}$

Sur-échantillonnage par 4

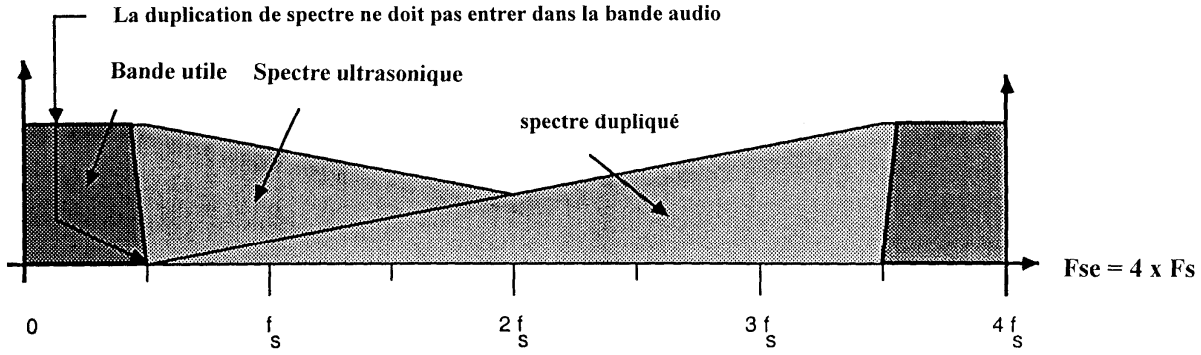


Figure n°14

• Limites :

- une augmentation de la fréquence d'échantillonnage entraîne des dérives du système **suiveur-bloqueur** (circuit ***track & hold***, cf. *figure n°4*).

- La chute naturelle du spectre de la distorsion de quantification au delà de *nFe* rend caduc le calcul effectué ci-dessus.

- Une réduction de la résolution entraîne, pour une qualité équivalente, un *R* très grand. Si l'on imagine un système codé sur 1 seul bit (n=1 et n'=16), on a alors $\boldsymbol{R= 4 \times 10^{15}}$. Il n'est donc pas envisageable de penser réduire la résolution à un bit par le sur-échantillonnage seul…

L'étape de décimation (sous-échantillonnage) devra être précédée d'un filtrage numérique de toutes les composantes spectrales entre *22.05* et *154.35 kHz* pour éviter un repliement de spectre lors du passage en *44.1 kHz*. (cf. *figure n°16* et la remarque sur le lien décimation / filtrage transversal).

Le sous-échantillonnage peut être effectué, si besoin est, par moyennage ou en ne considérant plus que quelques-uns des échantillons acquis. Les spectres périodiques modulés sont, dans ce cas, recentrés sur de nouvelles porteuses périodiques (ici *Fse/4* par exemple).

Le sur-échantillonnage permet donc de ***transférer l'étape de filtrage anti-repliement dans le domaine numérique et donc de minimiser les problèmes du filtrage analogique classique***.

La souplesse du filtrage numérique autorise la synthèse de filtres à oscillation limitée, pente raide, retard de propagation de groupe nul (phase linéaire), invariance thermique et temporelle de leurs caractéristiques, reproductibilité parfaite, programmation par soft, faible sensibilité aux parasites et aux dérives d'alimentation et, de plus, meilleur marché. Leur dynamique est fixée par le format de calcul et les troncatures éventuelles.

La sensibilité au *jitter* est diminuée par moyennage (sur *R* échantillons) de l'erreur d'acquisition lors de la décimation, et le *dither* optimisé.

2.2.0 Conversion numérique-analogique :

Le procédé d'interpolation permet d'augmenter la fréquence d'échantillonnage par calcul d'échantillons intermédiaires. Comme dans le cas précédent, l'avantage réside dans le transfert d'opérations de filtrage dans le domaine numérique.

La résolution ne pouvant dépasser celle de l'acquisition, l'interpolation reste un moyen d'approche des limites théoriques en contournant les difficultés du filtrage de reconstitution.

2.2.1 Principe (cf. *figure n°17*) :

- insertion d'échantillons nuls supplémentaires*.
- calcul de ces points intermédiaires par filtrage passe-bas numérique (réjection de bande périodique).
- la restitution peut ensuite se faire par filtrage passe-bas analogique d'ordre faible.

* Dans le cas d'un rapport *Fse/Fe* d'entiers, *y* échantillons nuls sont insérés tous les *x* échantillons acquis :

$$\Rightarrow Fse = Fe \times \frac{x+y}{x} = Fe \times \frac{A}{B} (\ A \text{ et } B \text{ entiers naturels }\ )$$

Le filtre passe-bas d'interpolation numérique (spectre périodique) est analogue au filtre pré-décimation de l'acquisition. On retrouve ici la symétrie des filtres d'anti-repliement et de restitution des systèmes de ***Nyquist***.

La précision du filtrage d'interpolation est synonyme de précision d'interpolation au sens mathématique (résolution et format du calculateur).

La limitation temporelle (temps de calcul total) est fixée par la fréquence de sortie (de sur-échantillonnage) du système. Le retard total de traitement (retard entrée / sortie de la chaîne audionumériques) est fonction des tolérances de l'oreille (temps de fusion, environ *15* ms soit *15 × 48 = 720* échantillons en *48 kHz*), des impératifs de synchronisation selon la destination du système (temps réel, consoles numériques, synchronisation image...) et de la capacité du *buffer*.

Si l'interpolation est optimisée (calcul sur un grand nombre d'échantillons et dynamique de calcul supérieure à la résolution d'acquisition), on peut alors considérer que les échantillons calculés sont exacts, ou en tout cas plus précis que les échantillons réellement acquis au sens mathématique.

La dynamique de calcul étant limitée par le processeur, le problème de troncature est le facteur limitant de cette interpolation. De plus, si le résultat obtenu doit être conformé au format d'acquisition, une nouvelle troncature (i.e. "requantification") sera faite. Le *dithering* trouve ici des applications naturelles.

L'uniformisation de la distorsion de quantification est assurée une fois de plus par ajout de *dither*. La distorsion de calcul (normalement inférieure à celle de l'acquisition) est sujette au même phénomène de repliement de spectre qu'au paragraphe 1.3.0.

Le bruit résiduel est, par conséquent, réparti sur la bande passante utile de façon uniforme. Le calcul de dynamique tenant compte du facteur de sur-échantillonnage fait (sous réserves) dans le cas de la conversion inverse est donc valable pour des systèmes *ditherisés* correctement.

On verra plus loin que l'erreur de quantification peut faire l'objet d'une mise en forme indépendante des principes du *dithering*. On parlera, dans ce cas, de *noise shaping* (rejet du bruit hors de la bande utile par filtrage). Néanmoins, l'optimisation de ces systèmes suppose l'utilisation de *dither*.

Du point de vue de la dynamique, le gain en *10 log R* dans les systèmes à sur-échantillonnage est équivalent à une résolution supérieure de *(10 log R)/6* bit (voir la *figure ci-dessous* pour le cas où R=4).

Un *CNA* peut donc avoir comme le *CAN* une résolution moindre à dynamique égale, dans le cas du sur-échantillonnage.

Il semble toutefois préférable d'utiliser ces techniques afin d'améliorer les performances des systèmes *16* bits qui, par sur-échantillonnage, peuvent gagner en dynamique (définition dans les bas niveaux).

Le sur-échantillonnage permet donc d'approcher les questions de filtrage dans le domaine numérique (**interpolation** et **décimation**) et de gagner en rapport signal sur bruit par élargissement spectral du bruit résiduel. Le filtrage devient donc une algorithmie précise et complexe ne dépendant plus des aléas des composants analogiques.

Il reste cependant impossible de dépasser les performances théoriques calculées au paragraphe 2.1.0.

On verra que l'optimisation des systèmes fortement sur-échantillonnés passe par l'emploi de codeurs à résolution réduite (quantification plus grossière) dont le bruit de quantification est mis en forme. Une décimation en conséquence pourra ensuite permettre de gagner en définition lorsque l'on se ramènera à une fréquence d'échantillonnage standardisée (*48 kHz* par exemple).

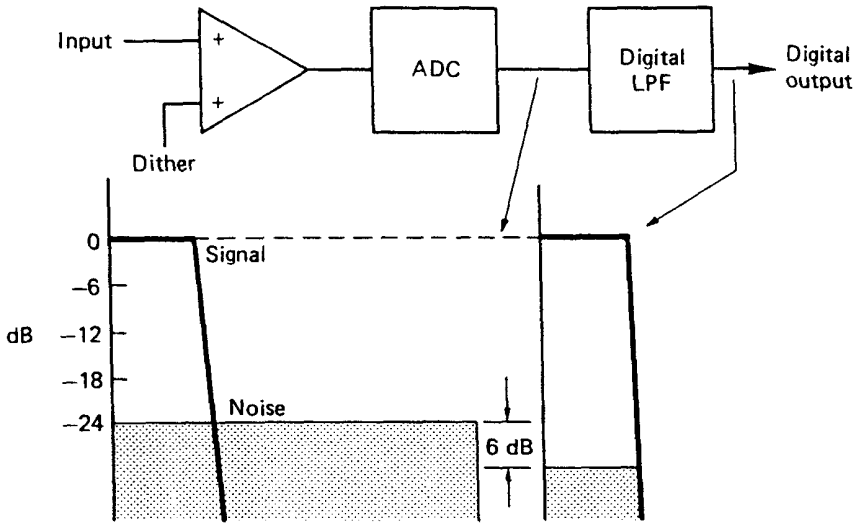


Figure n°15

Décimation dans le domaine fréquentiel

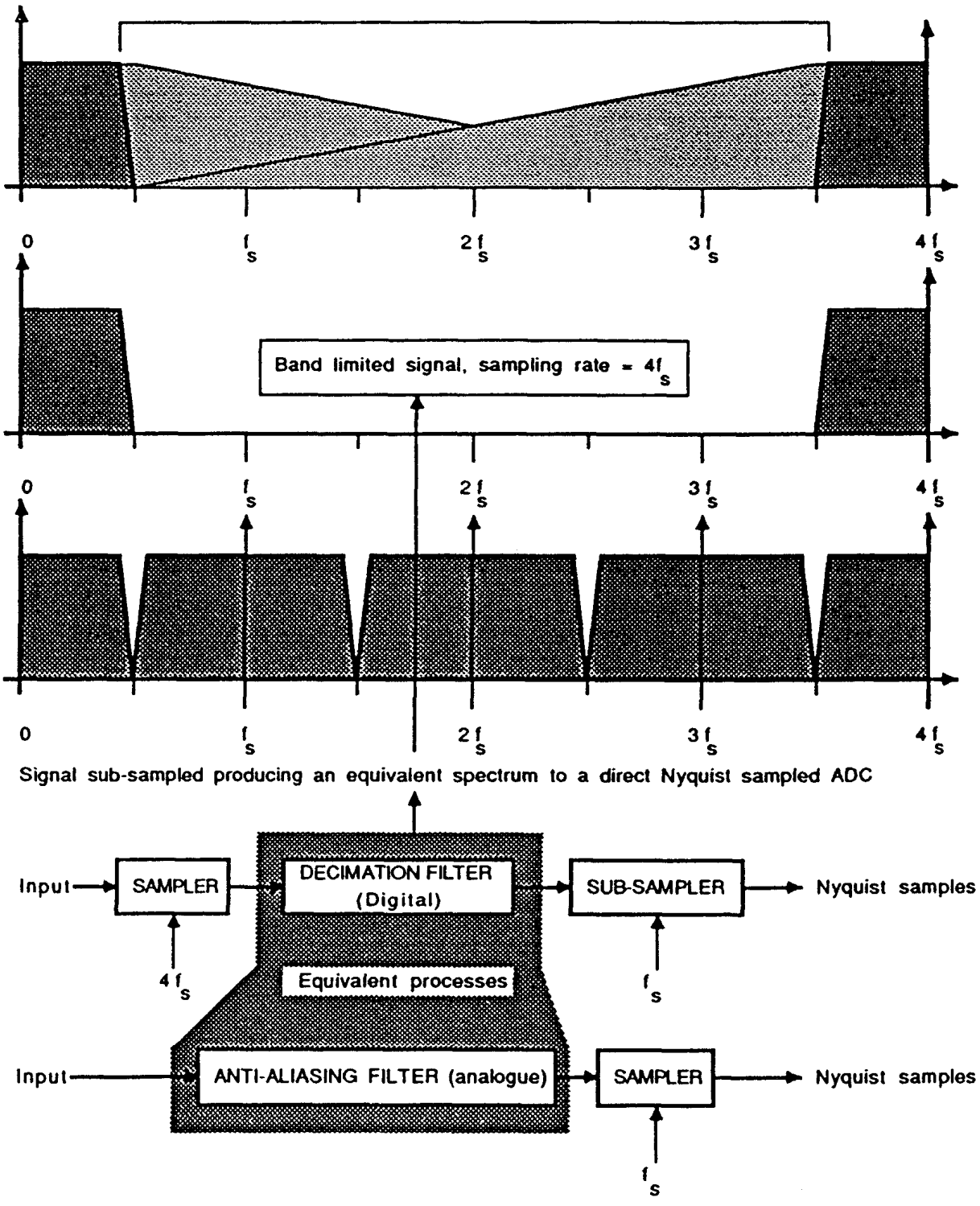


Figure n°16

Interpolation

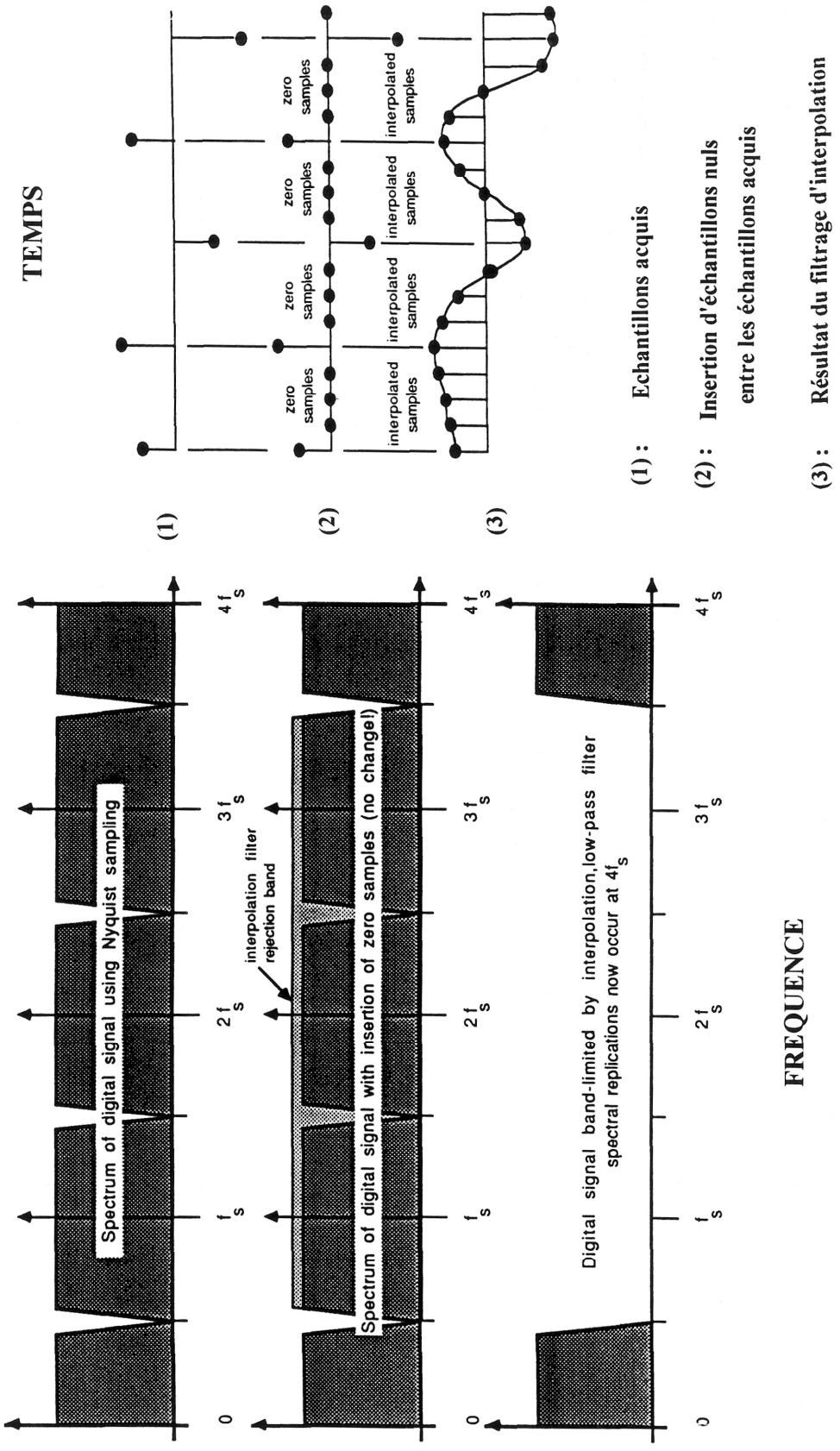


Figure n°17

2.3.0 Remarques sur le filtrage numérique (cf. *figures n°18* et *19*) :

Les filtres numériques se divisent en deux catégories ; les filtres à réponse impulsionnelle finie (*FIR*, filtrage transversal) et ceux à réponse impulsionnelle infinie (*IIR*, filtrage récursif).

La propriété de causalité des filtres réalisables physiquement implique, dans certains cas, une troncature de leur réponse impulsionnelle (transversal ou récursif tronqué).

Un filtre à phase linéaire (retard de propagation de groupe nul) supposant une réponse impulsionnelle symétrique, un décalage temporel sur la moitié de la durée de celle-ci devra être prévu (temps de traitement).

Seules les structures transversales (plus gourmandes en calcul) permettent d'obtenir une réponse en phase linéaire.

La précision mathématique d'un filtrage numérique est directement fonction du nombre de coefficients utilisés, de leur précision, et de la forme de la fenêtre de troncature (format et algorithmie du calculateur).

Dans les cas où le temps de *processing* est un facteur limitant (consoles numériques...), l'utilisation de structures récursives (plus rapides) peut s'imposer, malgré les problèmes de phase qu'elles impliquent.

Les opérations de filtrage numérique se résumant à une série d'addition et de multiplications, il est important de remarquer que l'addition de deux mots de n bits donne un résultat sur $n+1$ bits et que leur multiplication tient sur $2n$ bits. Le format de calcul, par les troncatures qu'il introduit, détermine la précision du traitement (taille de l'accumulateur et algorithme d'arrondi).

2.4.0 Remarque sur la conversion de fréquence d'échantillonnage :

Pour effectuer une conversion de fréquence d'échantillonnage, la méthode idéale passe par le sur-échantillonnage à la plus petite fréquence commune multiple et la décimation appropriée. Elle n'est néanmoins pas toujours techniquement envisageable.

Pour un passage $F1 \leftrightarrow F2$: $F1 \times C = F2 \times D = PPCM$ (sur-échantillonnage par interpolation)

puis : $(F1 \times C) / D = F2$ et $(F2 \times D) / C = F1$ (décimation par moyennage).

• Exemple : passage de ***F1 = 48 000 Hz*** à ***F2 = 44100 Hz*** :

$$PPCM\,(F1, F2) = 7\,056\,000\ Hz$$

$$48\,000 \times 147 = 7\,056\,000\ Hz\ (C = 147)$$

$$44\,100 \times 160 = 7\,056\,000\ Hz\ (D = 160)$$

Il faudrait donc sur-échantillonner *F1* par *147* (*F2* par *160*) puis décimer par *160* (respectivement par *147*), ce qui n'est pas technologiquement évident ... Les constructeurs rivalisent d'astuce.

Cette opération suppose évidemment le filtrage transversal nécessaire pour être correctement effectuée…

Le passage *32 kHz* ↔ *48 kHz* est évidemment plus simple.

Le *varispeed* (rapport variable des fréquences) peut poser quelques problèmes...

L'emploi d'une autre méthode suppose une perte de dynamique. Celle-ci peut être "camouflée" par emploi de *dithering* et de *noise shaping*.

La conversion de fréquence d'échantillonnage peut donc être destructive.

2.5.0 Remarque sur le lien entre décimation et résolution :

La décimation (réduction de la fréquence d'échantillonnage) peut être effectuée par moyennage des informations. On tombe alors sur le problème du rapport dynamique / bande passante développé § 3.0.0.

L'exemple suivant illustre la possibilité de gain de précision (augmentation de dynamique) possible par ce type de décimation.

Si l'on considère une suite de 16 informations codées sur *1* bit telle que la séquence :

1 0 1 0 0 1 0 1 1 0 0 0 0 1 0 1

et que l'on opère une **décimation par *16***, on obtient une moyenne de :

7/16 = 0.4375
(somme des *1* / nombre d'informations)

Cette moyenne peut se coder sur *4* bits (*16* possibilités) comme :

$(0 \times 2^3) + (1 \times 2^2) + (1 \times 2^1) + (1 \times 2^0)$, c'est à dire : ***0 1 1 1***

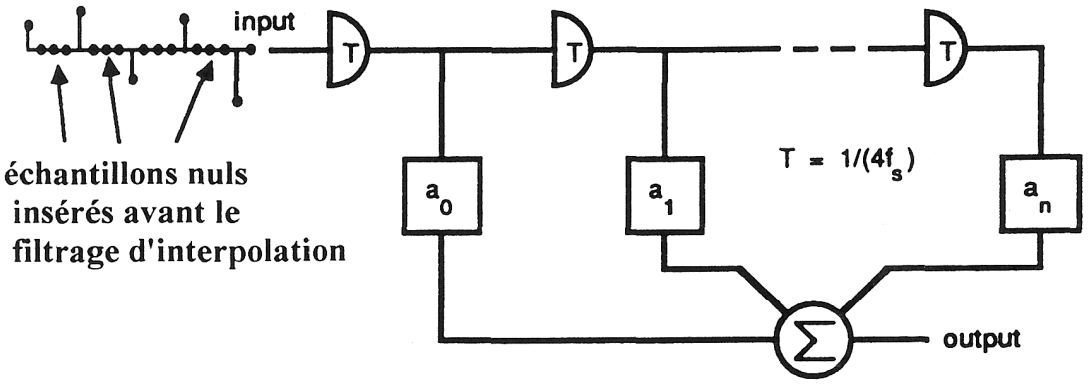
On est donc passé, par décimation, de *16* informations codées sur un bit à une information *4* bits sans perdre de précision. On a ici un échange bande passante (fréquence d'échantillonnage) / résolution (dynamique). Ce procédé permet le passage de 2^n informations codées sur un bit à une information codée sur *n* bits. La fréquence est donc divisée par 2^n pour une liaison parallèle *n* bits, $2^n / n$ pour une liaison série.

2.6.0 Remarque sur le lien entre la décimation et le filtrage transversal :

La décimation peut se décomposer en une étape de filtrage suivie d'une étape de sous-échantillonnage. Si l'accumulation (ou l'intégration selon le cas) se fait par moyennage de *n* échantillons consécutifs, la fréquence de sortie de l'accumulateur (ou de l'intégrateur) peut alors être divisée par *n*. Dans ce cas, les deux phases sont effectuées en une seule opération. Une décimation est donc synonyme de filtrage transversal d'interpolation alors qu'un filtrage d'intégration n'entraîne pas forcément de réduction de la fréquence d'échantillonnage.

Cette remarque justifie la considération séparée de ces deux étapes au cours de ce document.

Structure d'un filtre d'interpolation transversal
(sur-échantillonnage par 4)



On ignore les multiplications par zero pour accelerer le processus

EXEMPLE DE REPONSE IMPULSIONNELLE D'UN FILTRE PASSE BAS NUMERIQUE
(correspond au cas 100 coefficients)

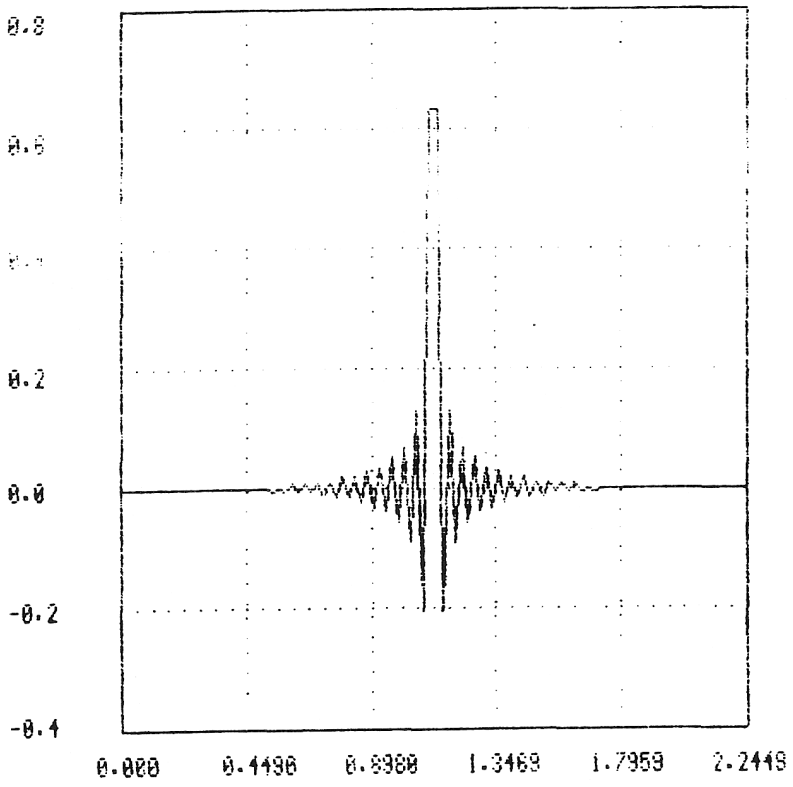


Figure n°18

Quatre exemples de réponses en fréquence de filtres numériques passe-bas
(optimisation Parks - Mc Clellan - Fe = 44,1 kHz)

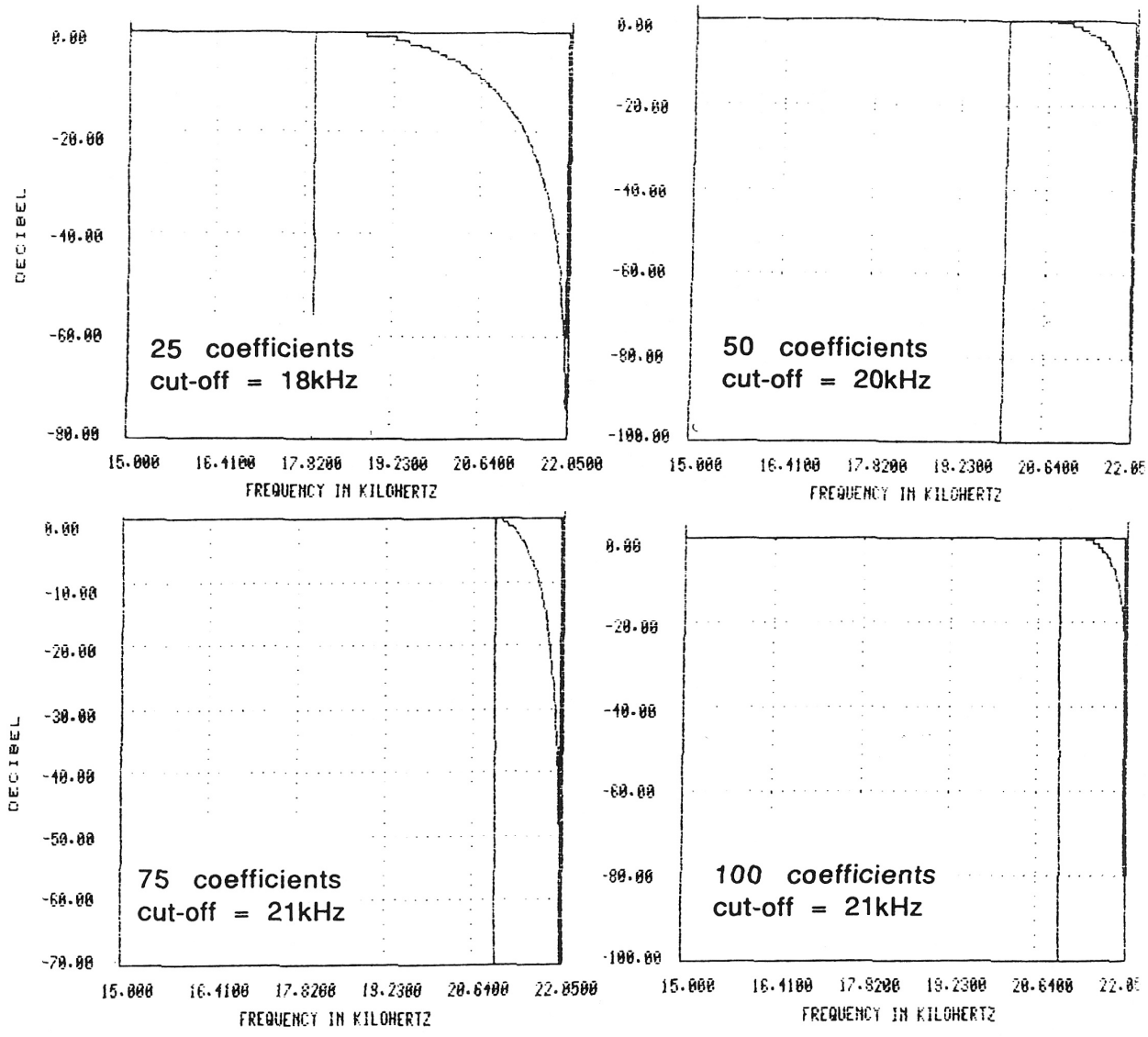


Figure n°19

3.0.0 *Noise shaping :*

La théorie de l'information définit la capacité d'un canal de communication en termes de bande passante et de rapport signal sur bruit. Cela sous-entend qu'à capacité constante (pas de perte d'information), l'un des facteurs puisse compenser l'autre et permet d'envisager d'autres formats.

Intuitivement, cela revient à considérer que plus on a d'échantillons et moins la performance dépend de l'erreur commise sur l'un d'eux (rapport densité d'information / perturbation).

Nous avons vu que, dans un système audionumérique à quantification linéaire, la bande passante utile était déterminée par le choix de la fréquence d'échantillonnage et la dynamique par la résolution (*44.1 kHz* et *16* bits par exemple...)

Nous avons vu, d'autre part, que les relations entre ces paramètres ne sont pas linéaires et qu'au cours d'une conversion numérique-analogique, il était possible d'accroître la fréquence d'échantillonnage par interpolation. Le sur-échantillonnage en acquisition, lui, permet une augmentation de la dynamique disponible par répartition de la puissance de bruit sur une bande spectrale plus large.

Le procédé de relocalisation de la distorsion dans cet espace fréquentiel redondant (hors bande utile) par filtrage est la définition du *noise shaping*.

D'un point de vue technologique, une résolution fine dans les systèmes à quantification parallèle linéaire pose un problème de précision et de temps de conversion (le *16* bits, *48 kHz* approche des limites possibles avec ce genre de procédé).

Une nouvelle stratégie a donc été développée.

Les convertisseurs utilisant le *noise shaping* peuvent être implantés en conversion analogique-numérique comme numérique-analogique. Dans les deux cas, ce seront des convertisseurs fonctionnant à des fréquences très élevées (données fortement sur-échantillonnées) et à quantification réduite. Ces codeurs sont de type récursif et reposent sur les principes de la rétroaction négative (codage différentiel, techniques de prédiction). Ce sont en somme des filtres récursifs…

On passe donc de systèmes où la quantification se fait sur une période d'échantillonnage à la résolution maximale du convertisseur (systèmes de ***Nyquist***), à des systèmes hautement sur-échantillonnés où une quantification plus grossière, une mise en forme du bruit de quantification par filtrage et une étape de décimation par moyennage appropriée, permet d'obtenir de hautes résolutions lors du retour à la fréquence d'échantillonnage standardisée.

Ces derniers utilisent les principes du codage ***ΣDPCM*** ainsi que les possibilités de traitement numérique des signaux.

4.0.0 Types de codage :
(Rappel de différentes formes de codage, cf. *figure n°20* et *21*)

4.1.0 Codages traditionnels :

***PAM* :** *Pulse Amplitude Modulation. E*chantillonné, amplitude continue.

***PCM / MIC* :** *Pulse Code Modulation* / Modulation par Impulsions Codées. Echantillonné et quantifié, systèmes *multibits*.

PCM / MIC linéaire : la quantification est linéaire (pas égaux).
Les amplitudes du signal et de la distorsion sont indépendantes de la fréquence.

***PWM / MLI* :** *Pulse Width Modulation* / Modulation par Largeur d'Impulsion. Signal carré dont le rapport cyclique est modulé (en général par le signal à coder).

4.1.1 Codages différentiels :

***DPCM* / MICD :** *Differential PCM* / MIC différentiel (quantification de la pente du signal). Quantification de la différence entre le signal d'entrée et la moyenne des échantillons précédents (codage récursif - techniques de prédiction).

Ce système de codage requiert des signaux dont la variation maximale est limitée (lien variation ⇔ bande passante ⇔ dynamique - précision de prédiction) et impose une fréquence de fonctionnement d'autant plus élevée que la quantification est grossière (système sur-échantillonné). Au-delà de cette valeur limite, un filtrage à *-6 dB* par octave du signal d'entrée sera nécessaire pour ne pas saturer le quantificateur (distorsion de granulation / distorsion de saturation de pente).

Le *dithering* assure la décorrélation de l'erreur de quantification ; la décimation finale son optimisation.

Ce codage effectue une mise en forme du signal et du bruit (opérateur de dérivation cf. § 5.0.0) et ne permet pas d'obtenir une grande dynamique en conversion. Il est, en revanche, parfois utilisé dans les systèmes de réduction de débit (non sous-échantillonnés) pour lesquels une nouvelle prédiction permet de restituer le signal sous-quantifié (MIC différentiel adaptatif – MICDA -, norme ISO G 722 par exemple)

Le signal *DPCM* peut s'obtenir par différentiation d'un signal *PCM* (un *PCM* par intégration d'un *DPCM*), mais un problème de référence d'échelle risque de se poser si l'on ne connaît pas l'origine des temps (un filtre passe-haut peut être utilisé pour éliminer cette déviation par élimination d'une composante continue éventuelle et des circuits de *reset* sont nécessaires).

***Δ Modulation* :** (*DPCM* sur 1 bit cf. § 5.0.0), même principe, comparaison de l'entrée analogique à la moyenne (analogique ou numérique) des bits précédents. La quantification se faisant sur un bit, on en vient à des systèmes très fortement sur-échantillonnés (la variation maximum permise en une période de sur-échantillonnage ne devant pas avoir une amplitude supérieure au pas de quantification).

On peut retrouver le signal *PCM* par accumulation (intégration et décimation / décimation par moyennage), le signal de départ par filtrage passe-bas d'interpolation analogique du signal *PCM*.

***Sigma DPCM et Sigma-delta* (*ΣΔ*) :** L'intégrateur de décodage est dans ce cas placé devant et /ou dans la boucle du delta-modulateur (codage d'amplitude). Ce système code la différence entre la moyenne de l'entrée et la moyenne de la sortie quantifiée (d'où ΔΣ*, cf. § 5.1.0).

Il n'y a pas ici de mise en forme du signal (tant que l'intégrateur et le différentiateur se compensent), mais seulement du bruit de quantification, d'où l'intérêt de ce type de codeur. Le bruit augmente de ce fait de *6n dB* par octave, *n* étant l'ordre de l'opérateur d'intégration.

Ce sont les *noise shaping converters*.

L'efficacité de la mise en forme du bruit dans la bande utile est déterminée par l'ordre du *noise shaper* (nombre de boucles d'intégrations).

C'est, comme dans le cas du delta-modulateur, un codeur récursif.

Une étape de décimation permet d'obtenir une résolution fine après codage et d'optimiser le *dither*.

La restitution peut se faire, une fois de plus, par filtrage passe-bas d'interpolation analogique (lissage).

* L'appellation d'origine de ce type de codage est le ΔΣ, le procédé ayant été popularisé par ***Inose* & *Yasuda***. Le terme de ΣΔ est apparu dans le "jargon" des laboratoires ***Bell*** (cf. ***J.C. Candy***).
D'autres appellations sont occasionnellement rencontrées.

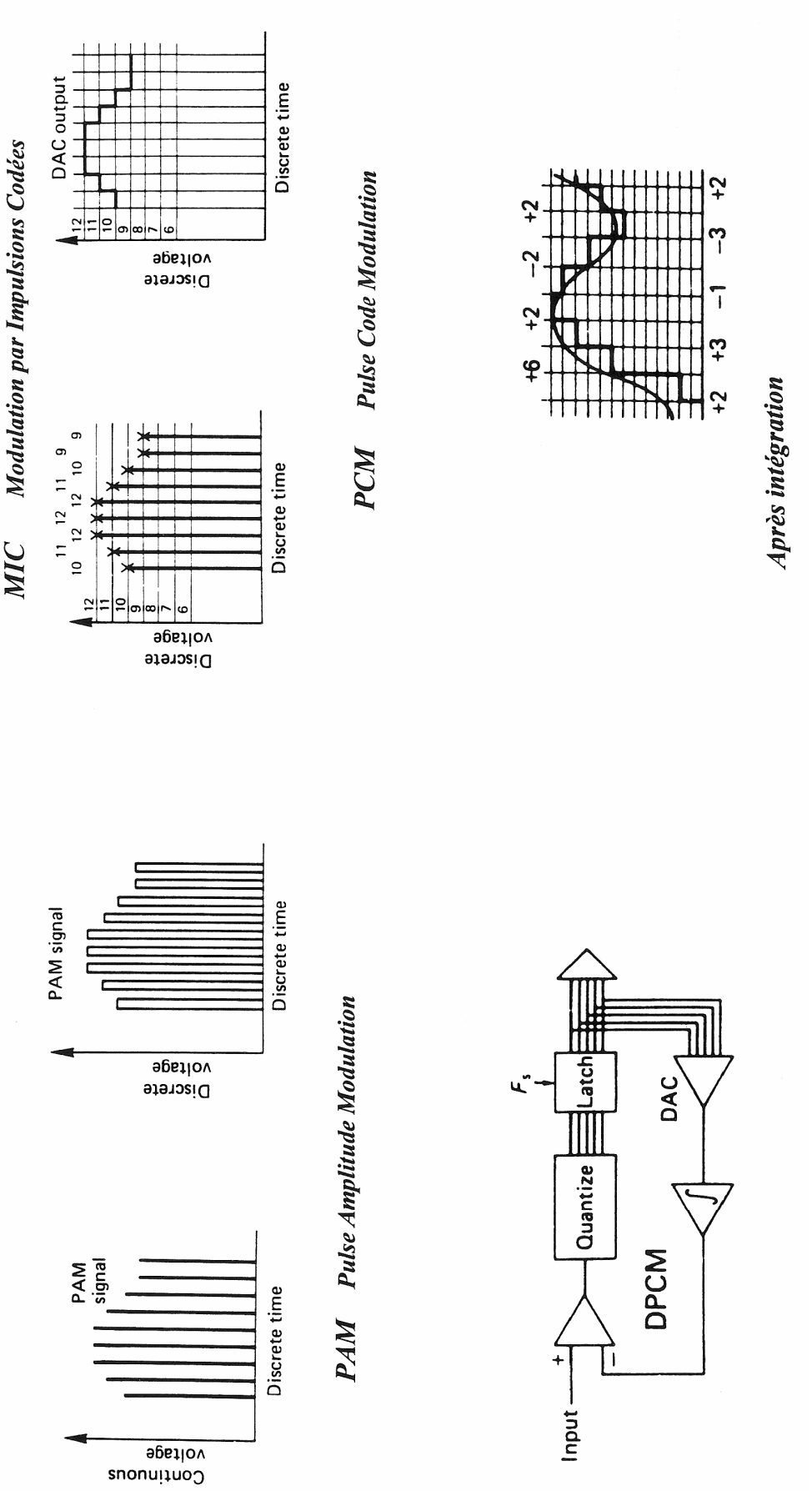
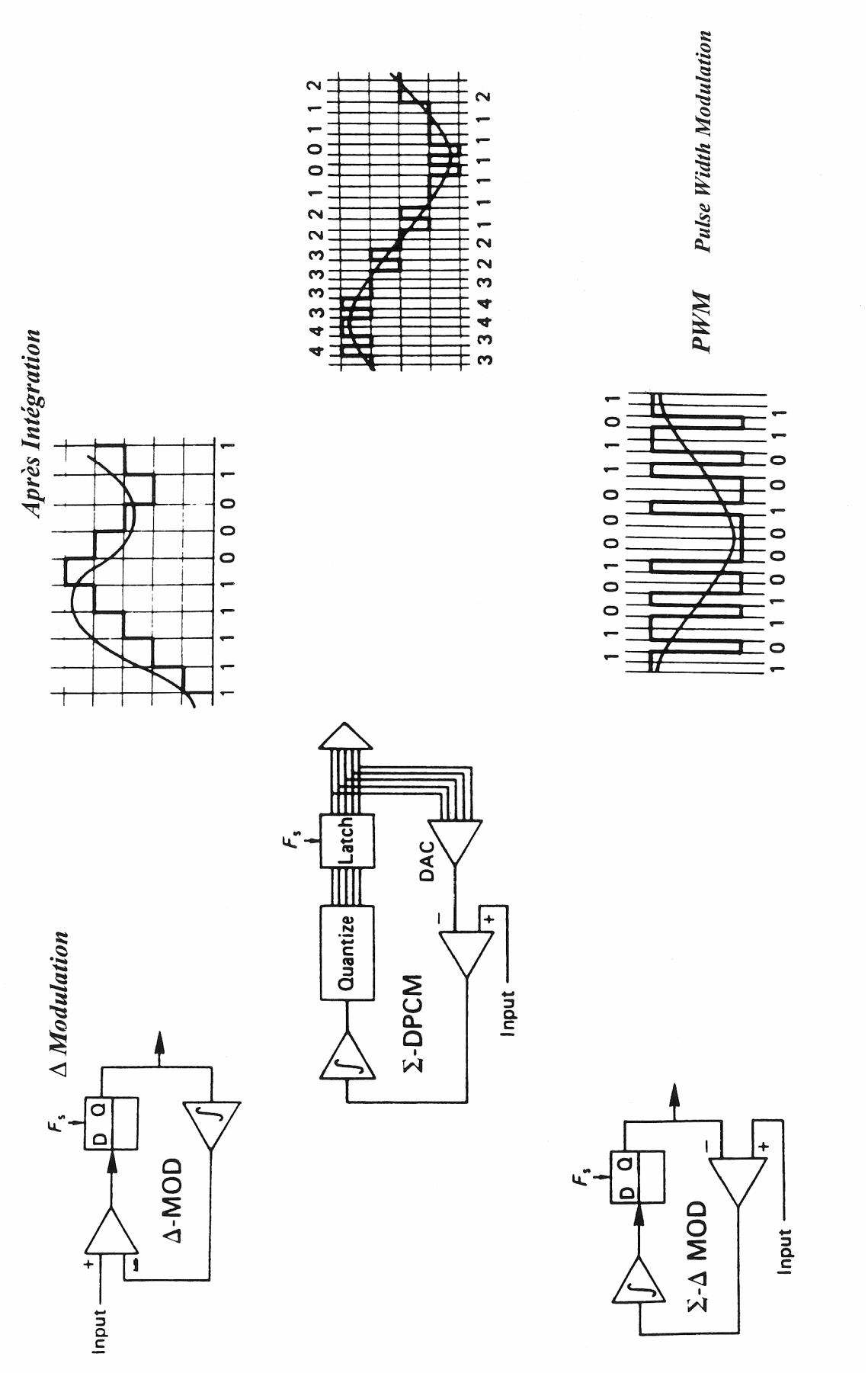

MIC Modulation par Impulsions Codées
DAC output
Discrete voltage
Discrete time
PCM Pulse Code Modulation
Après intégration
PAM signal
PAM Pulse Amplitude Modulation
Continuous voltage
Input
Quantize
Latch
DAC
DPCM

Figure n°20



Codages différentiels *- Figure n°21*

5.0.0 Principe de la delta-modulation :

Le schéma de principe du delta-modulateur et les codages obtenus sont donnés *figure n°22*.

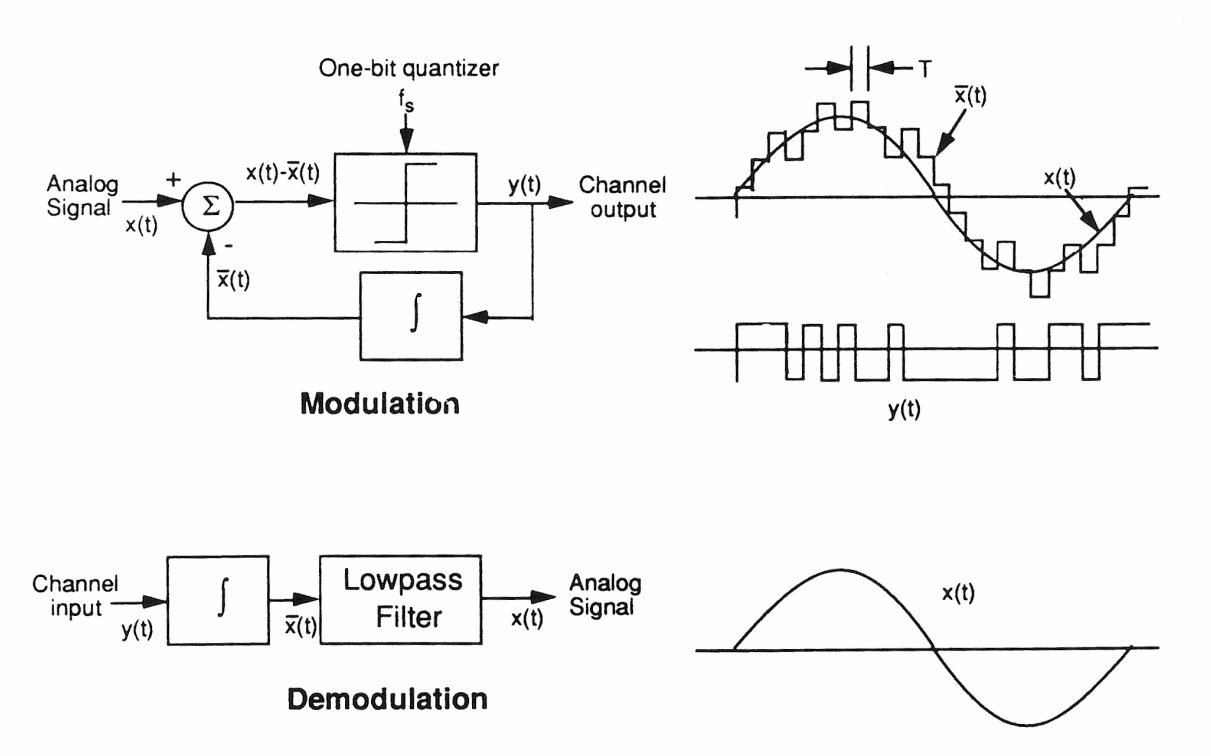


Figure n°22

Ce type de codeur quantifie la pente du signal en faisant la différence entre l'entrée et la moyenne des valeurs quantifiées précédemment (intégrateur dans la boucle de rétroaction / moyenneur dans le cas discret). On peut considérer que l'intégrateur prédit à court terme (une période de sur-échantillonnage) l'évolution du signal d'entrée et que l'erreur de prédiction $x(t)-\bar{x}(t)$ (pente du signal) quantifiée est utilisée pour effectuer la prédiction suivante. Pour être décodée, la sortie *y(t)* sera intégrée (codage d'amplitude), décimée au standard de fonctionnement puis "lissée".

Le signal d'entrée ne devant pas évoluer de plus d'un pas de quantification en une période d'échantillonnage, les performances de ce système ne seront optimisées que pour des signaux dont le spectre d'amplitude décroît à *6 dB* par octave au delà de la cadence imposée par sa fréquence de fonctionnement (elle même imposée par la fréquence maximum du signal à coder et la dynamique souhaitée cf. critère de variation, remarque B).

Dans la plage de fonctionnement (on ne considère dans ce cas que la distorsion de quantification), la prédiction est d'autant plus précise que le système "tourne" rapidement, le signal ne pouvant pas beaucoup évoluer en une période de sur-échantillonnage (bande passante limitée).

L'erreur de prédiction dépend donc de la fréquence de sur-échantillonnage (elle croit proportionnellement à la variation du signal d'entrée) et la dynamique semble pouvoir être déterminée par le choix de celle-ci, les performances du système étant minimales en haut du spectre audible.

La simplicité technologique du codeur autorise une fréquence de fonctionnement très rapide.

Le *dithering*, permettant l'uniformisation statistique de la distorsion de quantification et sa transformation en bruit aléatoire de densité de puissance constante sur toute la bande de fréquence considérée, trouve une fois de plus, un terrain propice. Le bruit de quantification (dit de granulation) peut donc être supposé uniforme (avant sa mise en forme). L'erreur de prédiction, corrélée à la variation du signal, peut ici, varier entre *0* et *Q*.

• Analyse formelle du delta-modulateur :

• On peut, en premier lieu, évaluer les performances du delta-modulateur dans le domaine continu de ***Laplace*** pour le cas de l'acquisition (conversion analogique-numérique, analyse linéaire).

$$Ona : x(p) - A(p)y(p) + Q = y(p)\ i.e.\ y(p) = \frac{x(p)}{A(p)+1} + \frac{Q}{A(p)+1} = \frac{x(p)+Q}{A(p)+1}$$

p étant l'opérateur de ***Laplace*** i. e. $p = j\omega$.

Le signal et le bruit sont ici filtrés par une même fonction de transfert : $\frac{1}{A(p)+1}$

En régime harmonique, on peut choisir *A(p)* comme un opérateur d'intégration ($A(p) = 1/j\omega T$), et l'on retrouve, dans ce cas, la fonction de transfert d'un filtre passe-haut du premier ordre pour lequel la constante de temps *T* (au maximum T_{se}) détermine la fréquence de coupure $Fc = 1/(2\pi T)$, en deçà de laquelle un filtrage à + *6 dB* par octave est effectué.

Si $T = T_{se}$, la fréquence de coupure est $Fc = 1/(2\pi T_{se}) = F_{se}/2\pi$ et l'on a pour $f << Fc$ l'équivalent d'un opérateur de dérivation harmonique.

La dynamique de codage est finalement celle imposée par la résolution de quantification et la fréquence de sur-échantillonnage. Elle dépend donc de la variation du signal d'entrée (l'erreur de quantification étant ensuite "homogénéisée" sur tout le spectre par *dithering*).

La dynamique disponible en sortie de convertisseur (i. e. après décimation finale) est fixée par le nombre de pas de quantification (un dans le cas de la Δ modulation, *n* dans celui du codage *DPCM*), la fréquence de sur-échantillonnage et la fréquence maximum à coder.

Dans la plage audio, le signal et le bruit sont mis en forme par un opérateur différentiel.

Le critère de stabilité du codeur exige que le gain de boucle soit strictement inférieur à *1* et impose donc $|A(p) + Q(p)| < 1$.

L'utilisation de *dither* permet de transformer celui-ci en $|A(p) + Q| < 1$ (en ne considérant que le bruit de granulation, cf. remarque A, et en supposant que la quantification puisse se modéliser de façon linéaire).

• L'analyse discrète du delta-modulateur peut se faire à l'aide des transformées en Z :

L'opérateur de retard unité (une période de sur-échantillonnage) est : Z^{-1}

L'opérateur d'intégration (retard unité et sommation) à une fonction de transfert du type :

$$I(Z) = \frac{1}{1 - Z^{-1}}$$

On a donc :

$$x(Z) - \frac{y(Z) \times Z^{-1}}{1 - Z^{-1}} + Q = y(Z) \; i.e. \; y(Z) = x(Z)\left(1 - Z^{-1}\right) + Q\left(1 - Z^{-1}\right) = \left[x(Z) + Q\right]\left(1 - Z^{-1}\right)$$

où l'on retrouve la même tendance que dans l'analyse continue, dans la mesure où le signal et le bruit de quantification (Q est uniforme - donc indépendante de la fréquence - après *dithering*) sont filtrés par un même opérateur de dérivation $H(Z)$.

$$On\ a\ H(Z) = \left(1 - Z^{-1}\right) qui\ dans\ le\ domaine\ fréquentiel\ peut\ s'écrire, avec\ Z = e^{2j\pi f / f_{se}} :$$

$$\left|H(f)\right| = 2\left|\sin\left(\frac{\pi f}{f_{se}}\right)\right|$$

Ce filtrage n'est pas, dans ce cas, comme le laissait supposer l'approche continue, un filtrage "passif" (le gain pouvant ici être supérieur à *1*). Dans le cas discret, le codeur n'est rien d'autre qu'un filtre numérique récursif (réponse impulsionnelle infinie et fonction de transfert périodique).

On retrouve, tout de même une caractéristique de filtrage passe-haut (arche croissante du sinus) sur l'intervalle *[0 ; fse/2]*.

Remarques :

A : Bruit de granulation et bruit de saturation de pente (cf. *figure n°23*) **:**

Deux phénomènes différents induisent de la distorsion lors du codage par Δ modulateur. Ceci provient du fait que la distorsion de quantification dépend de la variation du signal d'entrée dans ce type de système.

• Si le critère de variation est respecté, la distorsion de quantification est une distorsion, provenant de la discrétisation en "amplitude de variation", de même nature que dans la conversion *PCM*. Dans ce cas, on parlera de **distorsion de granulation** ou de **bruit de granulation**, si l'usage de *dither* le permet.

• Dans le cas de non respect du critère de variation, la trop grande variation du signal d'entrée induit une distorsion d'un autre type que l'on désignera par le terme de **bruit de saturation de pente** (*slope overload noise*). Celle-ci induit une non-linéarité dans le système de codage (supposé linéaire dans notre analyse) qui se traduit par un codage distordu du message et donc une perte de définition (le système ne peut pas suivre…).

Les caractéristiques non-linéaires du quantificateur (dispersion de l'erreur de quantification) font l'objet d'une remarque un peu plus loin et seront prises en compte dans les calculs de dynamique en ΣΔ.

Perceptivement, le bruit de saturation de pente est moins gênant que la distorsion de granulation à niveau de puissance égal (argument corroborant l'usage de *dither* / critères psycho-physiologiques).

B : Critère de variation - Calcul de dynamique :

Sachant que le signal d'entrée est sinusoïdal (du moins on le suppose pour pouvoir effectuer simplement les calculs !) et que le codeur différentiel ne quantifie que sur un bit, on cherche ici la fréquence de sur-échantillonnage permettant une dynamique de *N dB*. Cela revient à expliciter la limitation de la variation maximum du signal d'entrée imposée par la quantification sur *1* bit et à calculer le rapport signal sur bruit **ne tenant compte que du bruit de granulation**.

La mise en forme se faisant sur le signal et le bruit, elle n'intervient pas dans le calcul de dynamique.

La limitation de variation maximum (bruit de granulation seulement) peut s'exprimer en fonction de la dérivée maximum du signal d'entrée comme :

$$A \times 2\pi f_{entrée} = \frac{Q}{T_{se}} \quad i.e. \quad Q = \frac{2\pi f_{entrée} \times A}{f_{se}}$$

Le calcul du bruit de quantification dans la bande utile peut alors être effectué sachant que l'erreur de prédiction est uniformément répartie entre *0* et *Q* (*ditherisation* optimale), que la puissance totale de bruit est répartie sur la bande *[0 ; f_{se}]* et que b^2 est sa valeur quadratique moyenne.

$$\text{On a : } b^2(f_{entrée}) = \int_0^{f_{max}/f_{se}} Q^2 df = Q^2 \frac{f_{max}}{f_{se}} = A^2 \times 4\pi^2 \times \frac{f^2{}_{entrée} \times f_{max}}{f_{se}{}^3}$$

Le calcul de dynamique correspond au rapport des valeurs quadratiques moyennes du signal et du bruit de granulation.

Le signal d'entrée étant sinusoïdal, on a :

$$S/B\,(en\ dB) = 10\log\left(\frac{A^2/2}{A^2 \times 4\pi^2 \times f^2{}_{entrée} \times f_{max}/f_{se}{}^3}\right) = 10\log\left(\frac{1}{8\pi^2}\right) + 10\log\left(\frac{f_{se}{}^3}{f^2{}_{entrée} \times f_{max}}\right)$$

$$Avec\ f_{entrée} = f_{max}\ ,\ on\ a : S/B\,(en\ dB) = -19 + 10\log\left(\frac{f_{se}}{f_{max}}\right)^3$$

Si l'on souhaite avoir une dynamique (ici de codage de pente) équivalente à celle d'un système *PCM* linéaire *16* bits (codage d'amplitude, *N = 98 dB*), il faudra donc choisir une f_{se} de l'ordre de *159 MHz*, pour une f_{max} *=20 000 Hz*. Un facteur ***+20 log n*** est à rajouter dans cette formule de dynamique si la quantification se fait sur *n* pas (cas du MICDA).

C : La quantification se fait ici en même temps que la discrétisation temporelle et induit une dualité continue / discrète du système de codage dans le cas de l'acquisition.

L'étude de la conversion numérique-numérique impose, en revanche, une analyse discrète.

D : La fréquence de fonctionnement imposée par le critère de variation ainsi que la mise en forme du signal et du bruit ne permettent pas d'obtenir de grandes dynamiques pratiquement et condamnent, par conséquent, ce type de codeurs en tant que convertisseurs pour une utilisation audio de haute qualité. Leur capacité de prédiction est utilisée dans les techniques de réduction de débit.

Bruit de granulation et bruit de saturation de pente

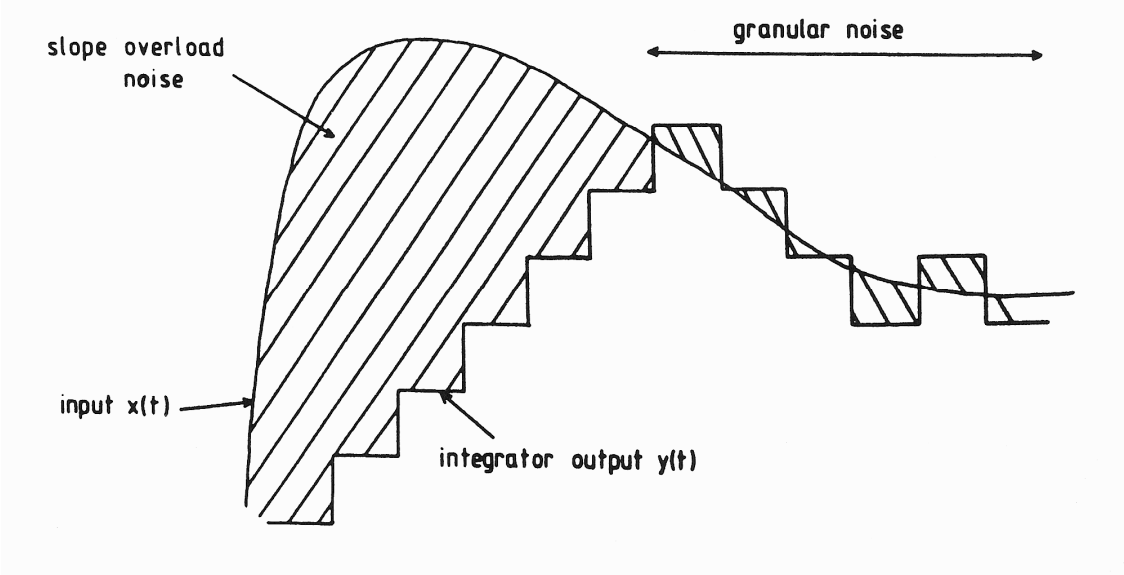


Figure n°23

Néanmoins, on peut imaginer qu'une intégration de même type que celle effectuée dans la boucle de rétroaction et un sous-échantillonnage pourra permettre de gagner en dynamique et de se ramener aux fréquences standards de fonctionnement. Cette étape étant équivalente à un filtrage transversal (multiplication de *Y(Z)* par l'opérateur $(1-Z^{-1})^{-1}$) et un sous-échantillonnage, on retrouvera par ce procédé le signal *X(Z) + Q* sous forme de message *PCM* (codage d'amplitude par compensation du différentiateur constitué par la boucle de codage).

Dans un Δ-modulateur, la dépendance fréquentielle de l'erreur de quantification peut être "homogénéisée" sur toute la bande utile par *dither*, sachant que celui-ci sera moyenné avant la phase de sous-échantillonnage. Reste que celle-ci est un facteur d'augmentation du bruit de granulation. De plus, le critère de variation restreint les possibilités dynamiques (ou de bande passante) du système à des fréquences de fonctionnement envisageables (l'amplitude maximum de l'erreur de prédiction impose celle du bruit résiduel dans la bande utile après *dithering*).

On peut penser s'affranchir de la mise en forme du signal, sans pour autant perdre celle du bruit de quantification (celui-ci étant généré dans la boucle), par intégration du signal d'entrée avant son passage dans le *noise shaper*. Cela permettra d'obtenir un codage d'amplitude, toujours sous forme de signal *PWM,* et non plus de pente du signal d'entrée. De plus l'intégration pré-codage et la différentiation effectuée par la boucle de conversion se compensant, les performances d'un tel système deviendra indépendantes du spectre du signal d'entrée et pourra permettre de réaliser un codage de haute précision du signal à convertir en raison de cette mise en forme sélective (l'erreur de quantification est dans ce cas "homogénéisée" par *dithering* et mise en forme).

• Une analogie de principe peut être faite dans le cas de la conversion inverse par multiplication pré-codage par l'opérateur *(A+1)*.

C'est ce que se proposent les systèmes ***ΣDPCM***.

5.1.0 Principe de fonctionnement des codeurs ΣDPCM :

Les codeurs *ΣDPCM* (le *ΣΔ* étant le cas où la quantification se fait sur un bit) sont actuellement les plus utilisés en conversion analogique-numérique et numérique-analogique. Inventés en 1962, leur utilisation en audionumérique n'a été rendue possible qu'avec le développement des techniques informatiques de traitement du signal et d'intégration à grande échelle des composants (filtrage numérique, *Digital Signal Processing,* techniques de *Very Large Scale Integration*). L'implantation de ces systèmes comprend donc le convertisseur proprement dit (codeur) et le filtrage adapté.

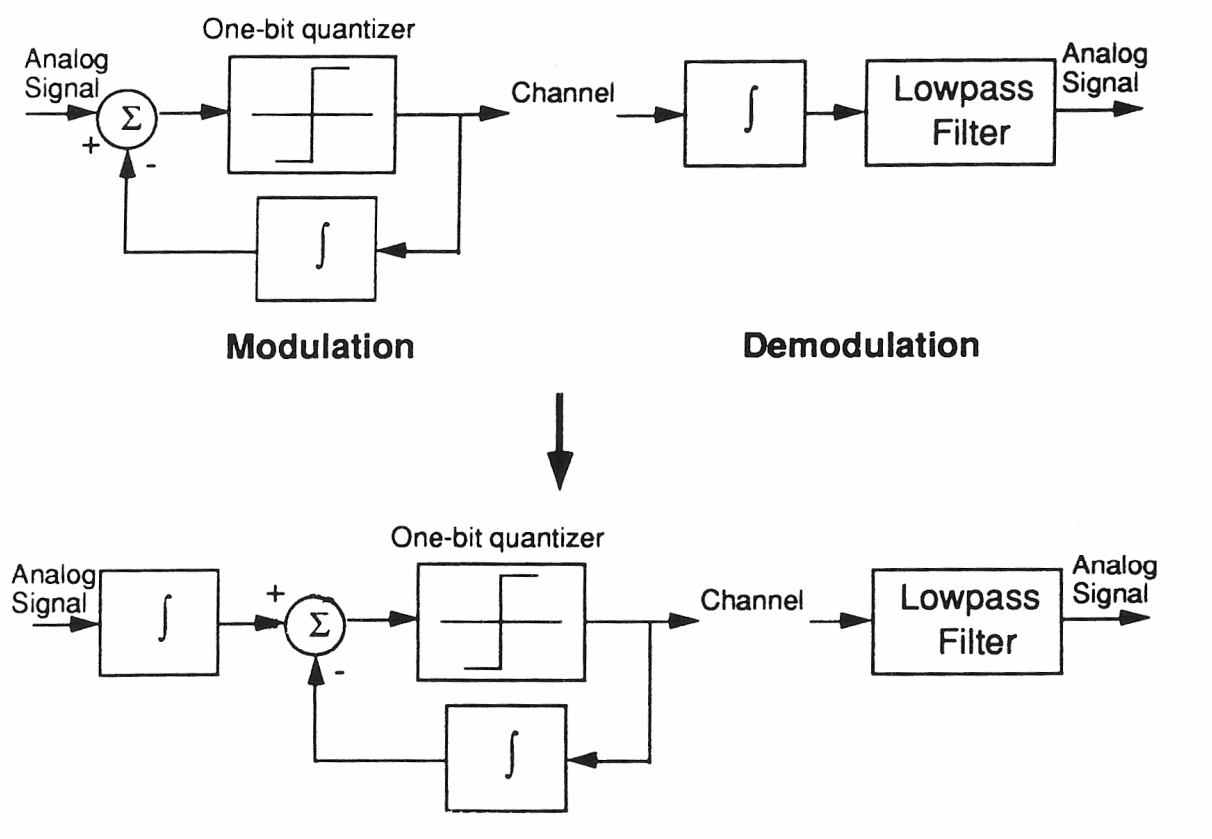
On a vu que la delta-modulation nécessitait deux intégrations, une dans la chaîne de *feed-back* de delta-modulateur, une autre au cours de la démodulation. On peut imaginer placer un intégrateur identique à celui de la boucle en entrée de la chaîne. On a directement, dans ce cas, un système de codage d'amplitude dont les performances ne dépendent pas du spectre du signal d'entrée par compensation de l'intégrateur "pré-boucle-de-codage" et de la différentiation induite par celle-ci.

Le bruit de granulation, lui, reste mis en forme et donc rejeté vers les hautes fréquences par filtrage passe-haut.

Le passage du delta-modulateur au codeur *ΣΔ* est illustré par la *figure n°24*.

On remarque que la propriété de linéarité de l'intégrale peut permettre de regrouper les deux intégrateurs identiques en un seul dans la mesure où l'on considère que seul le signal utile est concerné par la première intégration. On peut illustrer cette propriété (permettant la factorisation) par les calculs suivants en se basant sur le schéma à deux intégrateurs identiques.

Delta Modulation



Passage de la Delta Modulation au Sigma Delta

Figure n°24

- ***Cas continu :***

$$A(p)x(p)-A(p)y(p)+Q=y(p)\ i.e.\ A(p)x(p)+Q=y[A(p)+1]\Rightarrow y(p)=\frac{A(p)}{A(p)+1}x(p)+\frac{1}{A(p)+1}Q$$

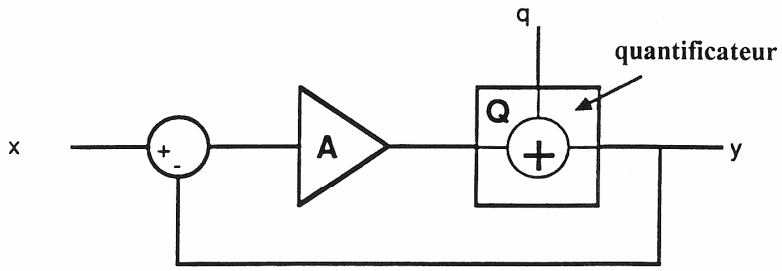
- ***Cas discret :***

$$\frac{x(Z)}{1-Z^{-1}}-\frac{y(Z)\times Z^{-1}}{1-Z^{-1}}+Q=y(Z)\Rightarrow y(Z)=x(Z)+Q\left(1-Z^{-1}\right)$$

La propriété de linéarité n'est bien sûr applicable que dans le cas où le système reste linéaire, ce qui suppose que le bruit de fond en entrée est très inférieur au bruit de quantification, que l'on reste dans les limites du "bruit de quantification seul" et que l'opération de quantification peut être considérée comme linéaire (ce qui est pour le moment le cas dans notre analyse).

On remarque que ces résultats sont équivalents au schéma canonique présenté *figure n°25*.

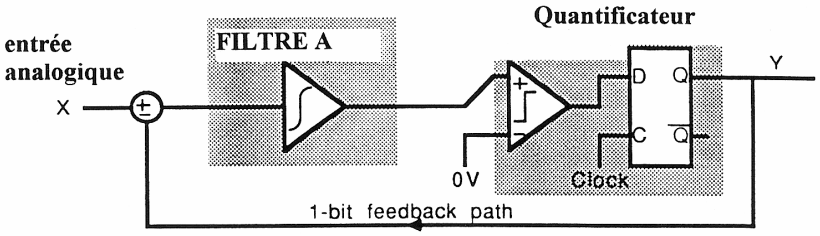
Forme canonique d'un codeur ΣDPCM



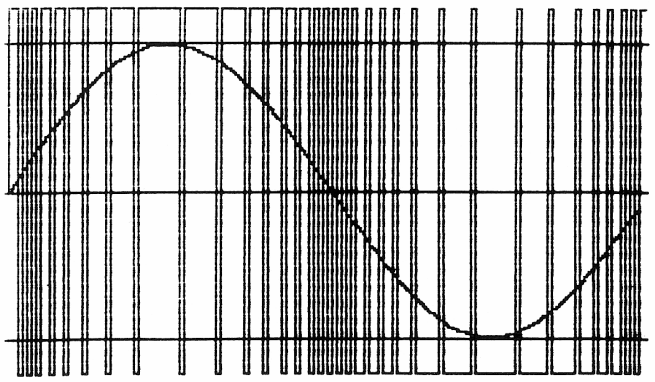
Cas continu $$Y(p) = \frac{A(p)}{A(p)+1} X(p) + \frac{1}{A(p)+1} Q$$

$$Y(Z) = X(Z) + (1 - Z^{-1})Q$$ **Cas discret**

CODEUR ΣΔ DE PREMIER ORDRE



Amplitude



Temps

Exemple de codage obtenu pour une entrée sinusoïdale.
$V_{entrée}$ = 0.8 Sin 2πft

Figure n°25

Un convertisseur numérique-analogique est cependant nécessaire dans la boucle de rétroaction d'un codeur analogique-numérique, si l'on choisit la présentation à intégrateur unique.

Ces deux structures types (un ou deux intégrateurs) sont analytiquement équivalentes dans les limites données (d'où ΣΔ...). On préférera tout de même la présentation à deux intégrateurs, car elle semble plus intuitive physiquement.

Dans la mesure où le système code la différence entre l'intégrale du signal d'entrée et sa prédiction (quantification de l'erreur de prédiction). On peut remarquer que si l'amplitude du signal d'entrée dépasse momentanément les limites du quantificateur (le critère de stabilité restant vérifié), on aura convergence de la prédiction vers le signal d'entrée, malgré une perte de dynamique due à l'apparition de la distorsion de saturation d'amplitude* en sortie de convertisseur et des phénomènes non-linéaires qui en découlent.

* **le critère de variation** est ici remplacé par un **critère d'amplitude : *Amax* = *Q***.

Le critère de stabilité reste le même que précédemment, à savoir que le gain de boucle doit être strictement inférieur à l'unité : $|A(p) + Q| < 1$, (*Q* étant uniformisée par *dithering* et supposée additive par l'analyse linéaire).

Le spectre du bruit de quantification (bruit suppose *dither*), comme dans le cas de la delta-modulation, est ici mis en forme et rejeté hors de la bande utile par filtrage passe-haut (cf. § 5.1.0). Cette propriété est à l'origine de l'appellation de *noise shaping coders* (regroupant les systèmes *ΣDPCM*) et de l'utilisation de ceux-ci en traitement du signal audio.

L'ordre de l'opérateur d'intégration détermine l'ordre de la mise en forme de bruit.

L'appellation ΣΔ reflète les deux opérations structurant le système (intégration et différentiation).

• On remarque que le cas continu impose un filtrage passe-bas du signal. La fréquence de coupure étant $Fc=1/2\pi T$ (*T* étant la constante temporelle d'intégration), on peut considérer que la bande utile reste dans la bande passante du filtre, si le système "tourne" assez rapidement (si $T = Tse$, on a $Fc = fse/2\pi$).

Ce filtrage est dû à la multiplication du signal d'entrée par l'opérateur *A* et non par *(A+1)*. Il n'a donc pas d'incidence dans la bande utile si $f_{entrée} << Fc$.

Cet opérateur de filtrage disparaît dans le modèle discret (dans ce cas $T=T_{se}$) car la multiplication par $1/(1-Z^{-1})$, intégration pré-boucle de codage correspond, dans ce cas, à l'inverse de la mise en forme effectuée par le *noise shaper*. La fonction de mise en forme est, dans ce cas, périodique, de période f_{se}.

5.2.0 Conversion analogique-numérique :

La structure globale du codeur reste la même pour des entrées analogiques continues ou discrètes, bien que l'étude qui s'y rapporte et les résultats obtenus diffèrent.

Le codeur analogique-numérique est constitué d'un étage différentiateur, d'un intégrateur, d'un quantificateur synchrone, et d'un *CNA* dans la boucle de rétroaction. Dans le modèle proposé, la distorsion de quantification (supposée uniforme par *dithering*) est découplée du quantificateur (supposé dans ce cas comme parfait) et s'ajoute à la valeur quantifiée (analyse linéaire).

La séquence de sortie peut s'exprimer comme :

$$y(p) = \frac{A(p)}{A(p)+1}x(p) + \frac{Q}{A(p)+1}$$

Et l'on définit la fonction de mise en forme du bruit : $Df = (A(p)+1)^{-1}$

Si A est choisi beaucoup plus grand que 1 dans la bande audio (sachant qu'il doit être inférieur à $(1-Q)$ dans la bande totale d'après le critère de stabilité), l'expression peut se réduire à :

$$y(p) = x(p) + Q \times Df$$

Le signal de sortie est donc le même que celui d'entrée auquel est ajoutée la distorsion de quantification filtrée par Df. Le but recherché étant de relocaliser Q hors de la bande utile, on a vu que Df devait avoir les caractéristiques d'un filtre passe-haut (dont la fréquence de coupure est déterminée par la constante temporelle d'intégration T).

De plus, l'opérateur A étant dans la boucle de codage (dont le gain de rétroaction est unité), il doit satisfaire aux exigences de stabilité de celui-ci : $|A(p) + Q| < 1$ (Q ne représentant que le bruit de granulation après *dithering*).

L'intégrateur répondant à ces exigences, $A(p)$ sera choisi de la forme $A(p) = 1/j\omega T$, T étant la constante temporelle d'intégration (une période de sur-échantillonnage au maximum) et ω la pulsation.

On a donc, pour un *noise shaper* du premier ordre :

$$|Df| = \frac{\omega T}{\left[1+(\omega T)^2\right]^{1/2}} \approx \omega T \quad pour\ \omega T << 1\ i.e.\ f_{utile} << F_C = 1/2\pi T$$

La mise en forme du bruit peut donc se représenter, en première approximation, comme un filtrage passe-haut du premier ordre (circuit différentiateur, pente à $+6$ *dB* / octave) de fréquence de coupure $F_C = 1/2\pi T$ comme dans le cas de la Δ-modulation (cf. *figure n°26*).

Sigma-delta et noise shaping du premier ordre

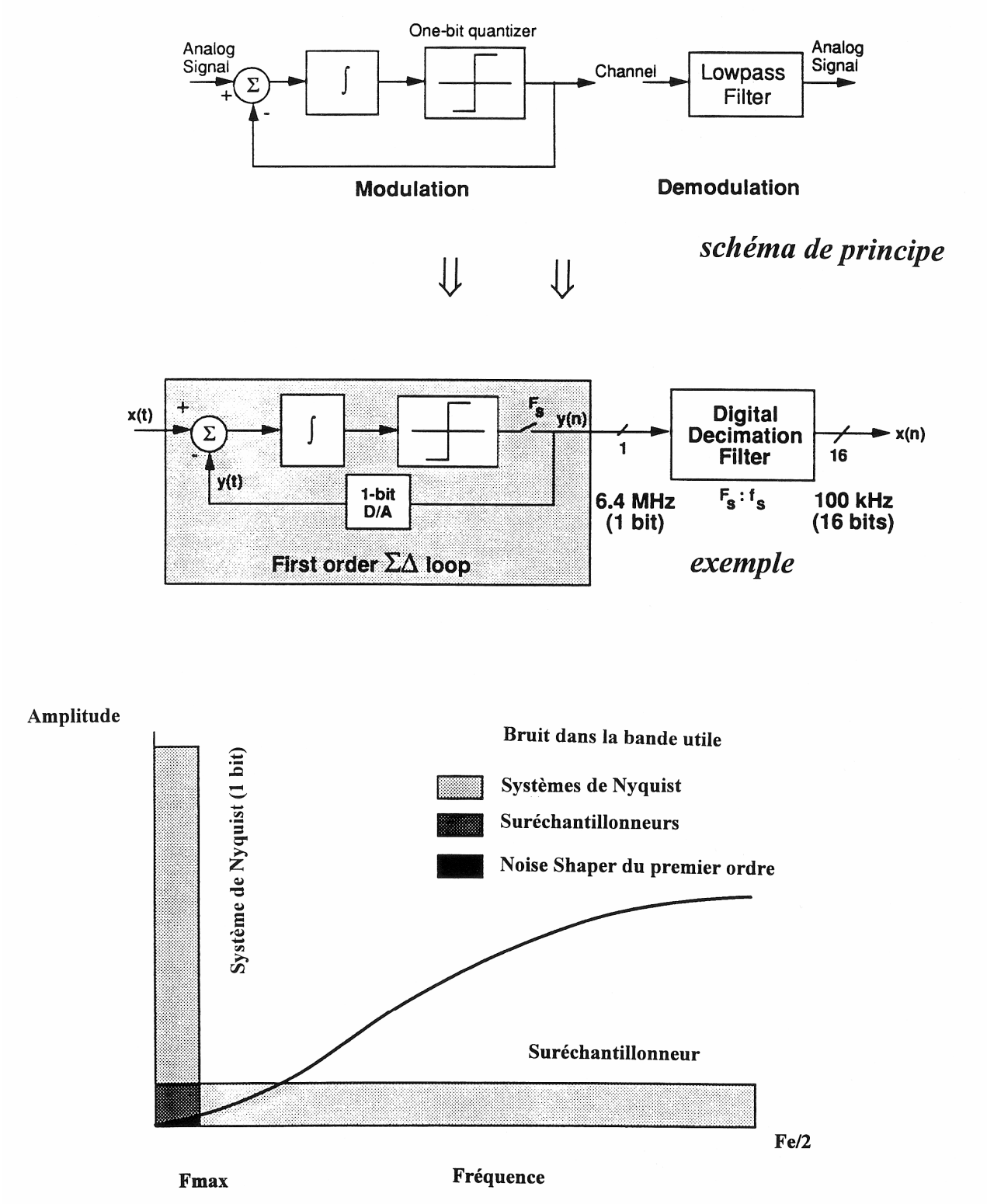


Figure n°26

La fonction de transfert du signal est un filtrage passe-bas dont la fréquence de coupure est aussi $F_c=1/2\pi T$, la plage de fréquence utile reste dans la bande passante de ce filtre d'après le choix de T qui, comme on l'a vu, est vérifié dans le cas où $T=T_{se}$, si le système fonctionne assez rapidement.

Si le critère d'amplitude n'est pas respecté, la distorsion de saturation d'amplitude s'ajoute au bruit de quantification et entraîne une non-linéarité du système.

En pratique, la pente de *+ 6 dB* / octave de mise en forme de bruit est insuffisante pour obtenir un rapport signal sur bruit satisfaisant en haut du spectre audio. Des performances de haute qualité supposent un facteur de sur-échantillonnage prohibitif.

Le problème peut être résolu en accroissant l'ordre de l'opérateur d'intégration dans le modulateur sigma-delta (par mise en cascade de cellules d'intégration), la pente du *noise shaper* étant pour un filtre du énième ordre de *6n dB* par octave.

Dans la représentation à deux intégrateurs, le *noise shaping* d'ordre supérieur suppose l'élévation de l'ordre des deux opérateurs.

Le critère de stabilité étant mis en défaut par la mise en cascade d'étages intégrateurs, on décompose l'asservissement en *n* boucles de premier ordre que l'on somme en fin de boucle. La fonction de transfert de l'ensemble est donc tempérée par l'introduction de *(n+1)* zéro et ramène la stabilité de l'ensemble à celle d'un système du premier ordre dans certaines plages de fonctionnement.

Le critère de stabilité réel du système devient dans ce cas :

$|A(p) + Q(p)| < 1$, sachant que $Q(p)$ n'est ni uniforme, ni linéaire dans les cas réels.

Ce procédé est illustré par la *figure n°27*, des détails de calcul sont donnés au § 5.3.0.

Exemple de codeur à mise en forme de bruit d'ordre n

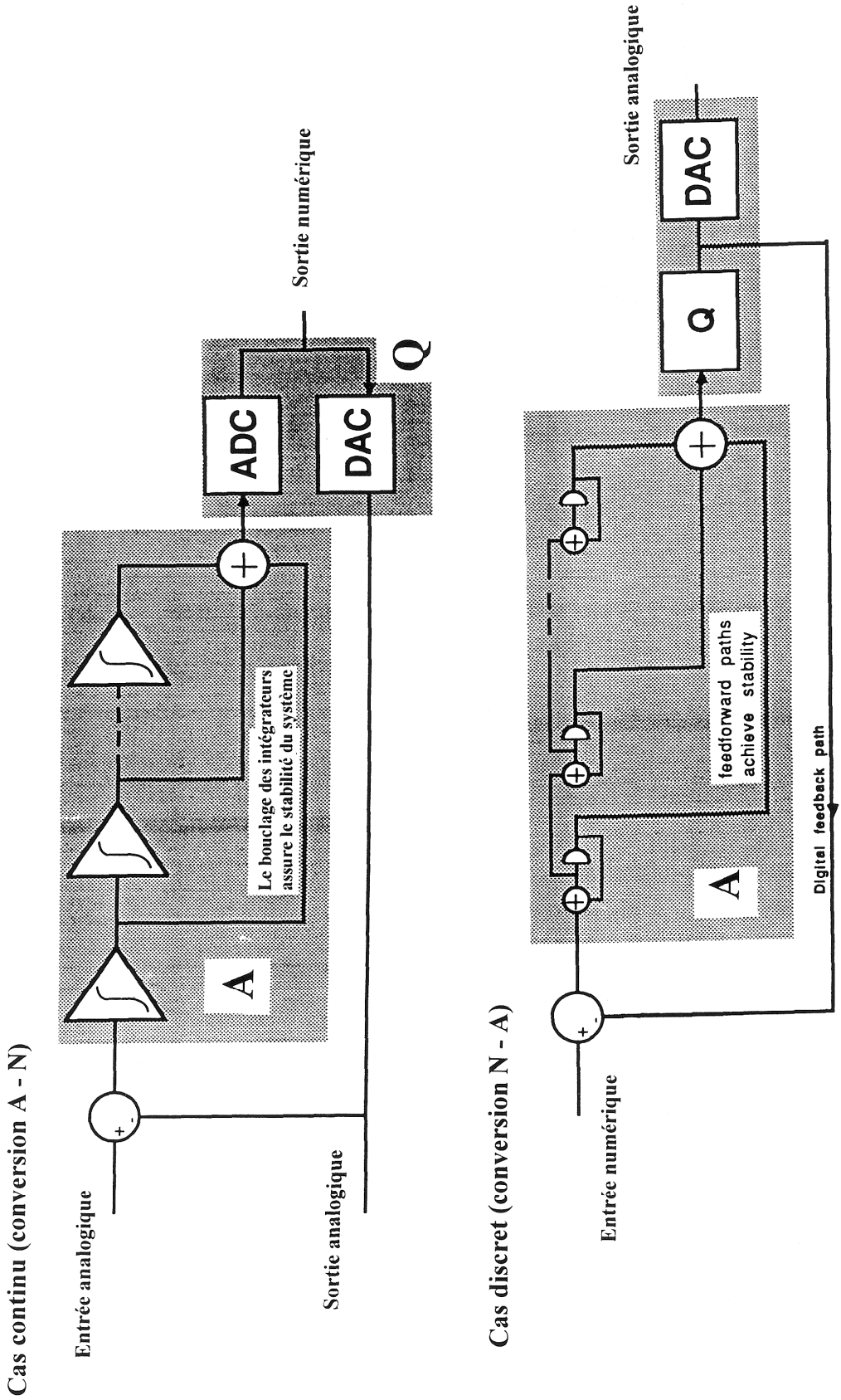


Figure n°27

5.2.1 Conversion numérique-numérique :

Pour une conversion numérique-numérique, les opérations d'intégration et de quantification doivent être transférées dans le domaine numérique et donc transformées en processus arithmétiques. L'intégration peut se faire par accumulation (gain unité, *feed-back* positif décalé d'une période de sur-échantillonnage i. e. sommation séquentielle) et la quantification (où plutôt "requantification") par approximation du résultat à la précision souhaitée.

On a dans ce cas :

$$y(Z) = x(Z) + Q(1 - Z^{-1})$$

$$i.e.\ DZ = (1 - Z^{-1})\ qui\ dans\ le\ domaine\ fréquenciel\ peut\ s'écrire,\ avec Z = e^{2j\pi f / f_{se}} :$$

$$DZ = 1 - \cos \omega T_{se} + j \sin \omega T_{se}$$

$$|D f| = \sqrt{2(1 - \cos \omega T_{se})} = \left| 2 \sin\left(\frac{\omega T_{se}}{2}\right)\right| = \left| 2 \sin\left(\frac{\pi f}{f_{se}}\right)\right| = \left| 2 \sin\left(\frac{\pi f}{R f_c}\right)\right|$$

f_c étant la fréquence de ***Nyquist*** et R le facteur de sur-échantillonnage.

• Pour un *noise shaper* d'ordre n on aura (cf. § 5.3.0):

$$|Df| = \left| 2 \sin\left(\frac{\pi f}{R f_c}\right)\right|^n$$

La représentation graphique de cette fonction est donnée *figure n°29* pour n variant de *1* à *6*.

Une représentation "plus large" des effets du *noise shaping* est donnée par la *figure n°30*.

On remarque que :

Si $f << R\ f_c$, $|Df| = [2\pi f / R\ f_c]^n$, la forme de Df est similaire au cas analogique pour $T = Tse$.
Si $f = R\ f_c/6$, $|Df| = 1$ quel que soit n.
Si $f > R\ f_c/6$, $|Df| > 1$ i.e. le spectre de bruit est amplifié.

et Df est maximum à $R\ f_c/2$ où $|Df| = 2^n$.

L'utilisation de tels convertisseurs est donc optimisée si la bande utile reste en dessous de $f_{se}/6$, le gain de bruit étant inférieur à *1*.

Une étape de décimation par moyennage peut ensuite se faire pour gagner en résolution (dynamique) et se ramener aux fréquences standards de fonctionnement. Le bruit rejeté en haute fréquence au delà de $f_c/2$ est filtré, la dynamique augmentée et le *dither* optimisé.

Le bruit de quantification étant mis en forme, l'erreur admissible dans ces systèmes peut être plus grande qu'en *PCM* classique. Il est de plus moyenné au cours de la décimation finale. On voit apparaître ici une notion de psycho-acoustique dans l'appréciation des performances de ces codeurs.

Remarque : Dans cette analyse des codeurs *ΣDPCM*, la distorsion de quantification a été considérée comme un procédé additif (analyse linéaire). La non-linéarité du quantificateur devra être prise en compte pour valider cette approche théorique.

Une simulation sur ordinateur confirme les résultats attendus pour des *noise shapers* d'ordres pas trop élevés.

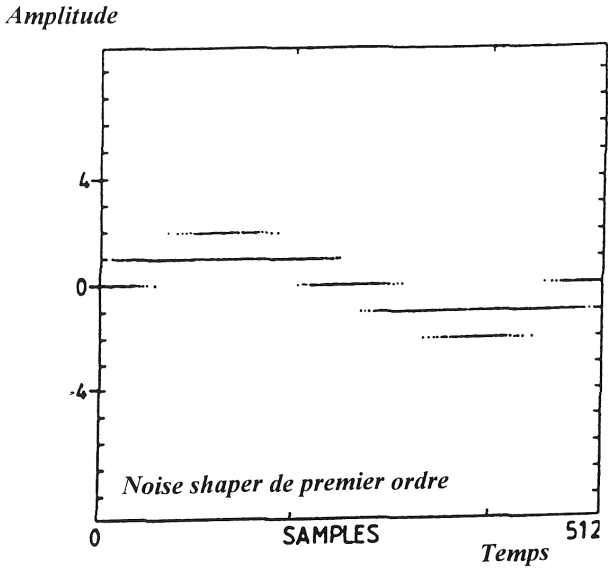
Néanmoins des divergences semblent apparaître si l'on observe la structure temporelle de la séquence de sortie d'un codeur numérique-analogique d'ordre élevé ($n > 2$).

On observe, par exemple, pour un convertisseur ΣΔ du quatrième ordre ($f_{se} = 8.82\ MHz$), et un signal d'entrée sinusoïdal de *20 kHz* dont l'amplitude est légèrement supérieure au pas de quantification ($Q\sqrt{2}$ i.e. non respect du critère d'amplitude), que la séquence de sortie obtenue contient du bruit dont l'amplitude se disperse entre $-8Q$ et $+8Q$. A première vue, la différence des amplitudes semble remettre en question la stabilité de la boucle de codage. Une étude plus approfondie de ce phénomène montre que la divergence du système n'est pas en cause et que le signal codé obtenu en limitant la séquence de sortie à *30 kHz* peut avoir un rapport signal sur bruit supérieur à *100 dB*. Le bruit observé est un signal chaotique dû à la combinaison de non-linéarité dans la boucle de codage et de sous-quantification (cf. *figure n°28*).

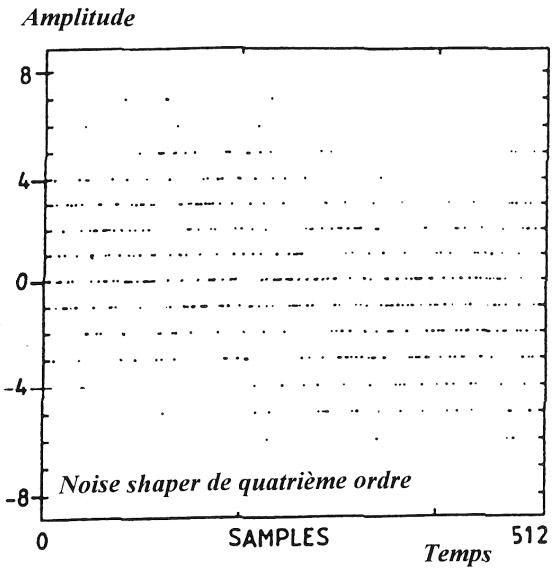
Le fait que le bruit mesuré ait une amplitude supérieure au pas de quantification peut se justifier par l'observation de Df. Le gain que l'on observe si $f > R\,f_c/6$ peut être rapproché de la dispersion du bruit sur *4* bits et comparée à Df à $f = R\,f_c/2$ à laquelle on ajouterait la distorsion de requantification due à l'erreur de Q. Cette propriété peut être exploitée et assurer un rôle de *dither*.

Ce phénomène montre que la distorsion de quantification est non seulement rejetée vers les hautes fréquences mais aussi décorrélée du signal par la structure de la boucle de codage elle même. L'effet de "chaotisation" ne dépendant que de l'ordre du *noise shaper*. La conversion produit, dans ce cas, un bruit de quantification sans nécessiter de *dither*.

Phénomènes de non-linéarité dans les codeurs ΣΔ



La dispersion augmente avec l'ordre du Noise Shaper.



Le signal d'entrée à une fréquence de 20 KHz et une amplitude de $Q\sqrt{2}$

Figure n°28

5.3.0 *Noise shaping* d'ordre élevé :

On analysera rapidement ici une méthode permettant d'obtenir un *noise shaper* d'ordre supérieur à un.

Un schéma de principe d'un codeur de deuxième ordre est proposé par la *figure n°31*.
On a ainsi :

$$Y(Z) = W(Z) + Q(Z)W(Z) = \frac{1}{1-Z^{-1}} V(Z)V(Z) = S(Z) - Z^{-1}Y(Z)$$

$$S(Z) = \frac{1}{1-Z^{-1}} R(Z)R(Z) = X(Z) + Z^{-1}Y(Z)$$

$$\Rightarrow Y(Z) = X(Z) + Q(Z)\left(1-Z^{-1}\right)^2 \; i.e.\; D_2(Z) = \left(1-Z^{-1}\right)^2 = \left\{D_1(Z)\right\}^2$$

$$On\ a\ donc : \left|D_2(f)\right| = 4\sin^2\left(\frac{\pi f}{f_{se}}\right)$$

Les courbes représentatives des fonctions D_i sont données aux *figures n°29* et *30*.

La qualité de ce type de technologie admet donc des limites théoriques reposant sur de nouveaux facteurs et permet d'envisager des définitions supérieures à celles du format *16* bits - *48 kHz*.

La tendance est, actuellement, à l'utilisation de ce type de codeurs (capables de coder les signaux de faibles amplitudes) en association avec des filtres d'interpolation, d'intégration et de décimation évolués. Ceci supposant une *ditherisation* optimale lors de la quantification dans les cas de mise en forme d'ordre faible.

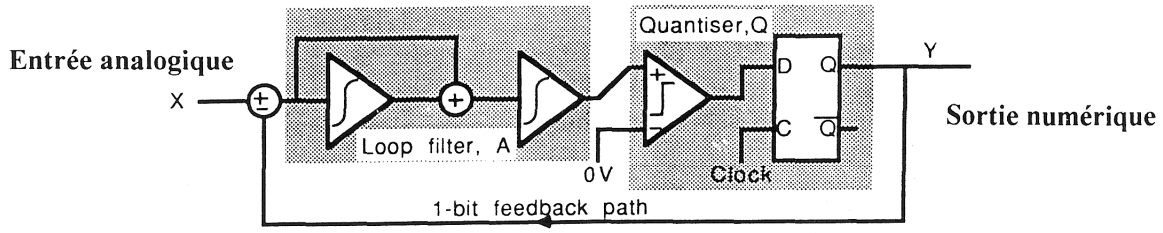
Les avantages du *noise shaping* et du sur-échantillonnage classique peuvent être combinés.

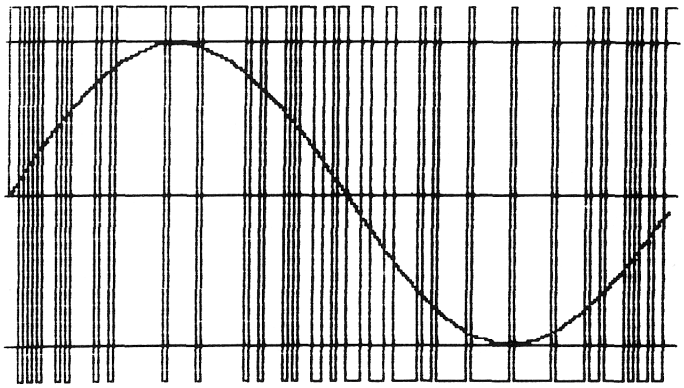
Philips à usé de la possibilité d'utiliser une résolution inférieure de deux bits (quantification en *14* bits), permise par un facteur de sur-échantillonnage *R=4* (gain de 1 bit cf. § 2.0.0) et un circuit de noise shaping du premier ordre (gain du second bit). La qualité finale étant perceptivement comparable à celle d'un *16* bits classique.

On a vu que le *noise shaping* n'était optimisé que pour des fréquences de sur-échantillonnage très rapides et que son avantage principal était le filtrage à *+6n dB* par octave du bruit de quantification (mise en forme). Les courbes psycho-acoustiques de ***Fletcher*** et ***Munson*** entrent alors en ligne de compte dans l'appréciation des performances de tels systèmes. Un bruit quantitativement plus important en haute fréquence peut se révéler moins gênant qu'un bruit de puissance inférieure centré sur une bande de fréquence plus sensible. Les effets de masque peuvent aussi être pris en compte dans les algorithmes de *noise shaping* (*super bit mapping , psychoacoustically optimised noise shaping...*).

Néanmoins, ces systèmes ne se justifient vraiment que lorsque l'on en arrive à des résolutions élevées (*16* bits ou plus).

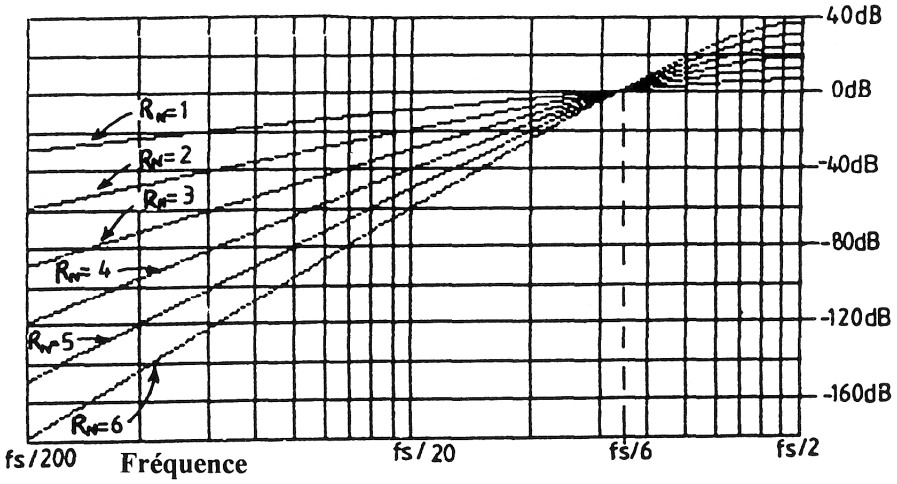
Codeur ΣΔ de deuxième ordre





Exemple de codage obtenu pour une entrée sinusoïdale.
$V_{entrée}$ = 0.8 sin 2πft
On remarque que les bas niveaux sont codés plus précisément qu'au premier ordre.

$1 \leq n \leq 6$



Représentations de Df en fonction de la fréquence pour des noise shapers d'ordre n

Figure n°29

Courbes de mise en forme de bruit pour du noise shaping d'ordre élevé

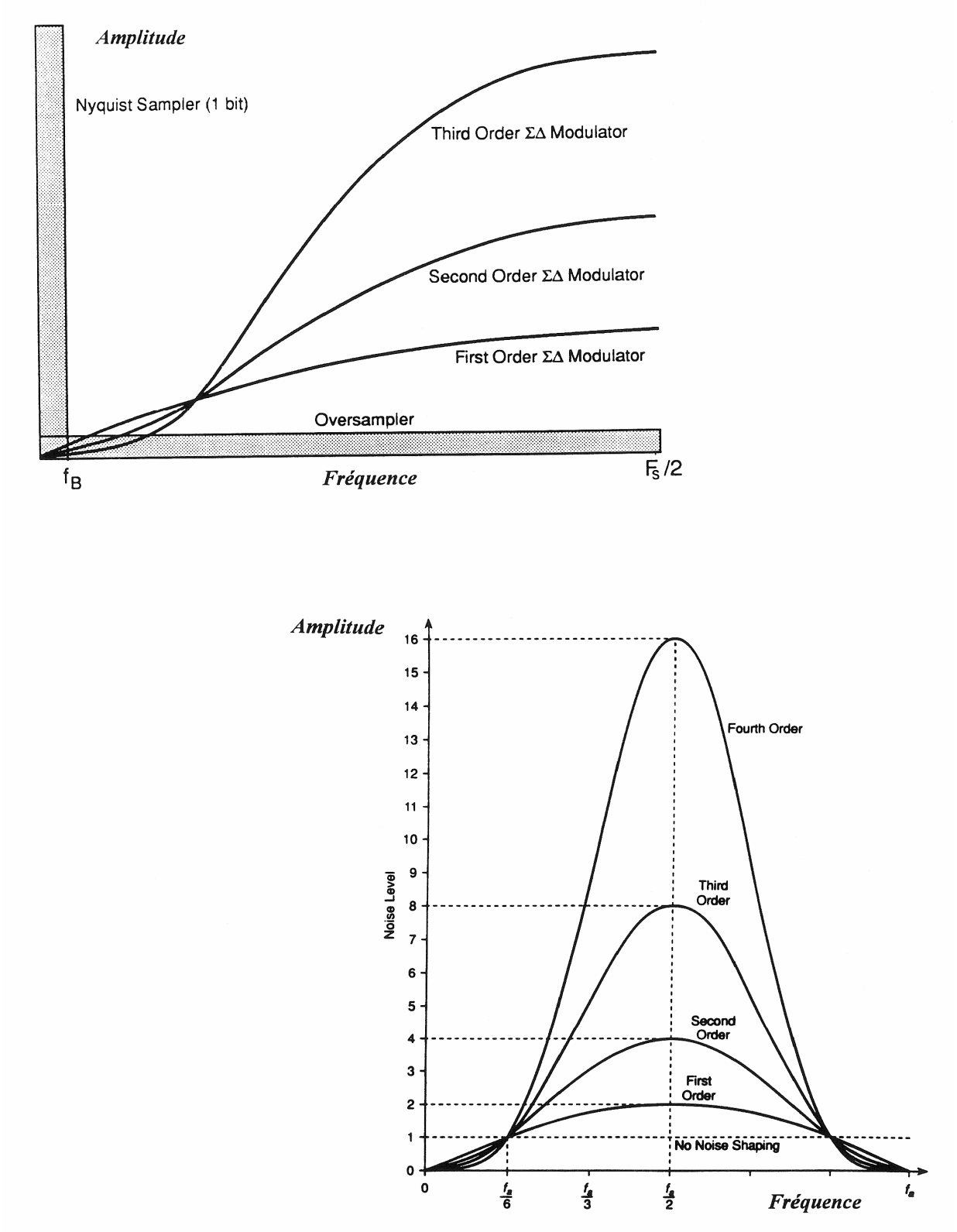
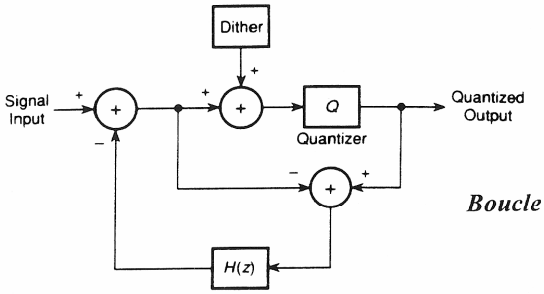


Figure n°30

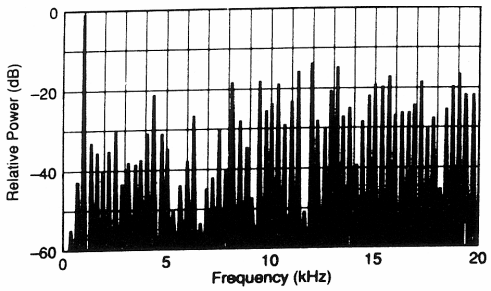
Noise shaping et dither

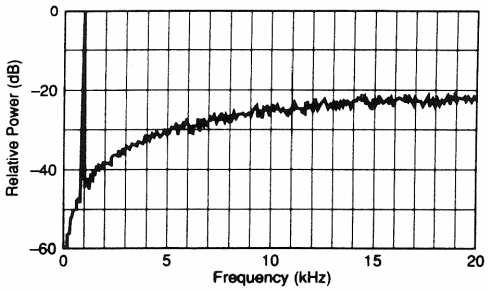


Boucle de codage avec dithering

Effet du moyennage du dither sur le spectre de bruit

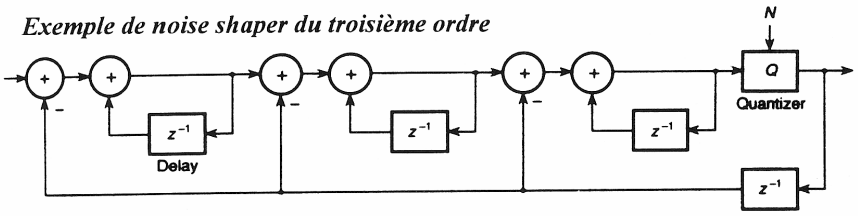
Non Moyenné

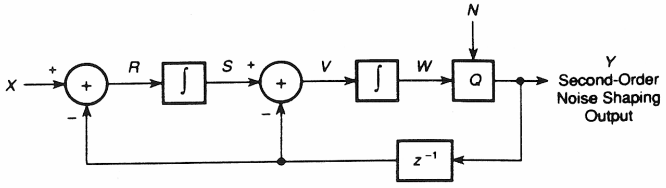




Après Moyennag

Exemple de noise shaper du troisième ordre





Noise shaper du second ordre détaillé au chapitre 5.3.0

Figure n°31

De plus, ces codeurs permettent de s'affranchir des impératifs du filtrage anti-repliement et des circuits suiveurs-bloqueurs.

Le *jitter* reste cependant problématique et peut dégrader les performances de ces systèmes par l'augmentation de la puissance de bruit qu'il introduit.

L'évolution des codeurs allie donc les avantages du sur-échantillonnage et du *noise shaping* (répartition de la puissance de bruit sur une bande spectrale plus large et mise en forme de celui-ci).

Les convertisseurs à *noise shaping* sont utilisés aussi bien en codage (systèmes ***one bit***) qu'en *noise shapers* simples (ils n'interviennent donc plus, dans ce cas, comme convertisseurs proprement dits) en appoint de systèmes de traitement du signal en cas de sur-échantillonnage (troncatures et opérations diverses). Ils sont alors exploités pour leurs propriétés de filtrage (mise en forme) et de décorrélation du bruit.

On peut obtenir par cette méthode des codeurs extrêmement précis qui dépassent largement les performances des systèmes *PCM* linéaires dans les limites données.

Ce type de technologie s'impose donc pour les systèmes de conversion de résolution supérieure à *16* bits.

Le bruit dépend donc de la fréquence par l'intermédiaire de la mise en forme effectuée et non du niveau ou de la fréquence du signal d'entrée.

Remarque : La relation historique entre la delta-modulation et les systèmes sigma-delta ainsi que la ressemblance de leur structure (pouvant justifier une démarche didactique) tendent à obscurcir le fait que ces deux méthodes ont un intérêt de nature totalement différente. La delta-modulation est utilisée en codage à débit réduit (et donc pour les opérations de stockage ou de transmission) alors que les systèmes ΣΔ sont exploités pour leurs capacités de *noise shaping* en conversion ou en traitement du signal.

5.4.0 : Codeurs ΣΔ

La technologie ΣΔ est un codage *ΣDPCM* à quantification sur un bit (fortement sur-échantillonné). Les appellations "***one bit***" et "***bitstream***" ont été proposées par les constructeurs.

Le signal de sortie d'une boucle ΣΔ est un train binaire sériel modulé en largeur d'impulsion (*PWM*), la modulation étant proportionnelle au signal d'entrée.

La quantification sur un bit entraîne des problèmes de stabilité de la boucle de codage et limite l'ordre de l'opérateur d'intégration vers le cinquième ou sixième ordre actuellement (non-linéarité des intégrateurs et du quantificateur, problèmes de stabilité du système).

Le fonctionnement des codeurs ΣΔ peut s'analyser par analogie avec un système à modulation de fréquence un peu particulier. On peut imaginer que le signal d'entrée module un *VCO* (*Voltage Control Oscillator*) qui pour chaque passage à zéro positif délivre une impulsion de durée constante (la densité d'impulsion est alors proportionnelle au signal d'entrée).

Le moyennage de ce train d'impulsions permet d'obtenir un signal *PWM* dont l'aire des impulsions est proportionnelle à l'amplitude du signal d'entrée.

La quantification peut s'imaginer, dans cet exemple, comme une relocalisation des impulsions *PWM* obtenues aux instants d'échantillonnage les plus proches. Le codeur ΣΔ peut alors être considéré comme un système de modulation de fréquence dont le temps est quantifié.

Le codage obtenu et les schémas de principes des codeurs ΣΔ du premier et second ordre sont donnés *figures n°25* et *29*. On remarque que le codeur du second ordre est plus efficace pour le codage des bas niveaux, en observant la largeur des impulsions de codage.

Leur principe est évidemment le même que celui des systèmes *ΣDPCM* étudiés précédemment.

Un exemple de structure de *CNA* ΣΔ Philips (sur-échantillonnage par 256 et noise shaping du second ordre) est présenté *figure n°32*.

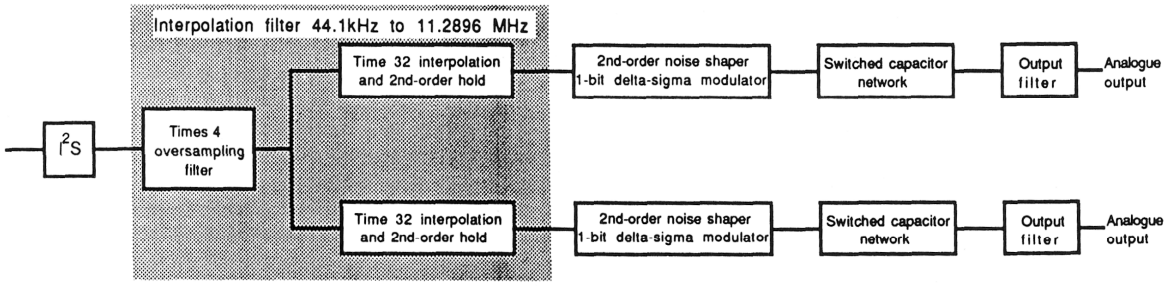


Figure n°32

5.4.1 Remarque sur le codage des bas niveaux :

Dans les cas où la différence entre la prédiction et la somme du signal d'entrée est très faible, le comparateur de quantification oscille entre les deux états de sortie. Cette oscillation est due au fait qu'il code, dans ces conditions, le bruit inhérent au système, voire le *dither*. Ce phénomène est synonyme de perte de dynamique.

Il se pose ici un problème d'ordre de grandeur si l'on veut évaluer l'effet produit car il dépend du paramétrage du système (fréquence de fonctionnement et *noise shaping* effectué).

Un *noise shaping* d'ordre élevé peut, en tout cas, permettre de minimiser cet effet par réduction du bruit dans la bande utile, ainsi que par le gain de précision de codage des bas niveaux que permet l'intégration d'ordre élevé. Ceci, dans les plages de fonctionnement optimales, c'est-à-dire, quand le niveau maximum du signal d'entrée est adapté au quantificateur et que le système reste stable.

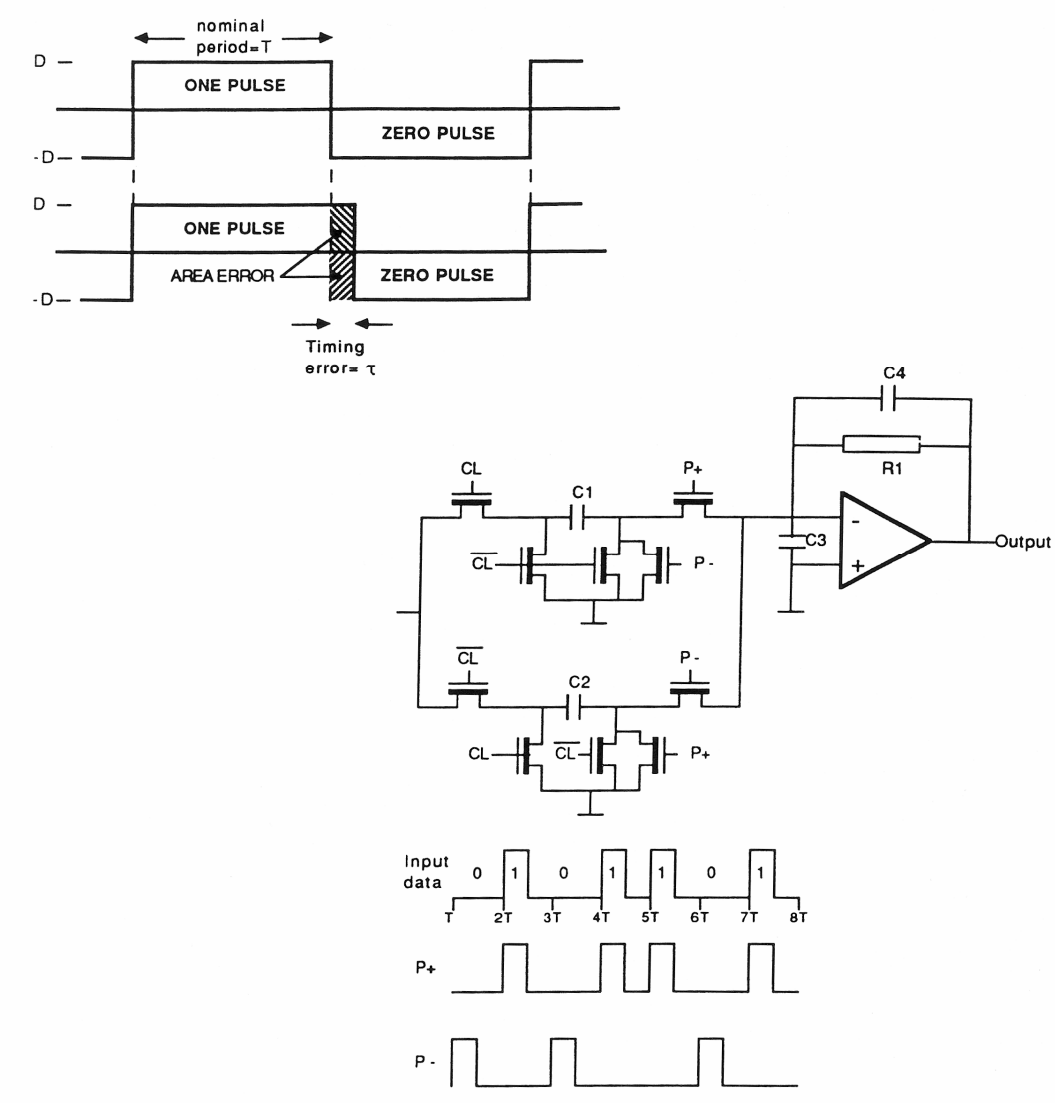
5.4.2 Effet du *jitter* sur les *CNA* ΣΔ :

Le *jitter* introduit une erreur de reconstitution du signal par modification de la largeur d'impulsion des informations codées. L'effet est présenté *figure n°33*.

Un intégrateur analogique de décodage dont la constante de temps serait proportionnelle à la période d'échantillonnage instantanée peut apporter une solution à ce problème.

Pour effectuer cette opération un réseau du type de celui de la *figure n°33* en bas (réseau à capacités commutées) peut être utilisé. Dans ce cas, les impulsions faussées par le *jitter* sont moyennées sur un temps proportionnellement plus long et garde donc leur valeur initiale.

Jitter et conversion numérique analogique



Exemple de réseau à capacités commutées permettant d'effectuer l'interpolation analogique sur la période d'échantillonnage instantanée (en sortie du CNA) et donc de s'affranchir en partie des problèmes dus au *jitter* en restitution

Figure n°33

5.4.3 *Jitter* en acquisition :

On a vu que le *jitter* en acquisition entraînait une distorsion proportionnelle au spectre d'entrée dans les systèmes *PCM* linéaires (erreur plus importante en haut du spectre audio à amplitude constante).

Dans les systèmes différentiels, son action est analogue et par conséquent proportionnelle à la fréquence d'entrée. Le bruit étant mis en forme dans les systèmes *ΣDPCM*, il reste à déterminer dans quelle mesure la distorsion due au *jitter* prend le pas sur la mise en forme de bruit, sachant que cette erreur est moyennée par la décimation finale.

En cas de fonctionnement non-linéaire stable, on peut supposer que l'erreur faite est répartie sur toute la bande utile.

5.5.0 Analyse du bruit dans les codeurs ΣΔ - Calcul de dynamique théorique :

On a vu que l'attrait de ce type de codeur résidait dans leur simplicité d'implantation et la précision de leurs performances. S'agissant de systèmes fortement sur-échantillonnés, ils évitent les problèmes dus au filtrage anti-repliement ainsi que ceux du circuit suiveur-bloqueur.

Le gain de dynamique peut s'expliquer en raison de deux facteurs principaux, à savoir la répartition du bruit sur une bande de fréquence très largement supérieure à la bande utile (mêmes conséquences que dans les systèmes à sur-échantillonnage classiques) et par la mise en forme de celui-ci par filtrage passe-haut du *n*ième ordre le rejetant hors de la bande spectrale utile.

Ces techniques codent donc le signal entrant en signal *PWM*, qui sera ensuite décimé par moyennage au standard de fonctionnement.

On peut effectuer l'analyse du bruit dans un système sigma-delta en ne considérant que le bruit de granulation du système (bruit de quantification B_Q / analyse linéaire) et sa mise en forme. L'usage d'un *dither* approprié permet de faire l'hypothèse de la constance de sa densité spectrale de puissance sur la bande audio.

On a vu que la mise en forme de bruit avait une caractéristique de filtrage passe-haut par multiplication du bruit de quantification par un opérateur de type $Df = (A+1)^{-1}$ pour un *noise shaper* du premier ordre dans le cas de la conversion analogique-numérique.

La puissance totale du bruit de granulation du système peut donc s'écrire comme :

$$B^2{}_{(Q+noiseshaping)} = B_Q{}^2 \times |Df|^2 = B_Q{}^2 \left[\frac{1}{|1+A(f)|^2} \right]$$

qui montre que le bruit de granulation est pondéré par la fonction de mise en forme sur la bande de fréquence totale (utile + redondante).

On peut évaluer ce bruit dans la bande audio comme :

$$B^2{}_{(Q+noiseshaping)\,Audio} = \frac{B_Q{}^2}{f_{se}} \int_0^{f_{max}} |Df|^2 df = \frac{B_Q{}^2}{f_{se}} \int_0^{f_{max}} \frac{1}{|1+A(f)|^2} df$$

où $\frac{B_Q{}^2}{f_{se}}$ est la puissance du bruit de quantification par unité de fréquence et $f_{\max} = 20\,000\ Hz$.

Supposant que le signal d'entrée est sinusoïdal, de type $S = B_Q \sin(2\pi ft)$ et donc de valeur quadratique moyenne $\frac{B_Q{}^2}{2}$ (son amplitude ne pouvant dépasser B_Q si l'on souhaite rester dans les limites d'un bruit de granulation i.e. $B_Q = Q$), et que l'on reste dans la bande de fréquences utiles ($f << f_{se}$, approximation permettant de simplifier le calcul) où l'on a vu que $|Df| = \frac{2\pi f}{f_{se}}$.

Le rapport signal sur bruit peut alors s'exprimer comme :

$$S/B(dB) = 10\log\frac{3}{4\pi^2} + 10\log\left(\frac{f_{se}}{f_{max}}\right)^3 \ i.e.\ S/B(dB) = -14 + 10\log(2R)^3$$

R étant le facteur de sur-échantillonnage.

Un raisonnement analogue peut être fait à l'ordre n avec :

$$|Df|^{2n} = \left(\frac{2\pi f}{f_{se}}\right)^{2n}$$

et l'on a :

$$S/B_{ordre\,n}(dB) = 10\log\left[\frac{2n+1}{2(2\pi)^{2n}}\right] + 10\log(2R)^{2n+1}$$

Ce calcul peut illustrer les cas continu et discret dans la mesure où l'approximation faite sur $|Df|$ se retrouve dans les deux cas (en analogique pour $f << f_{coupure} = f_{se}/2\pi$ si $T = T_{se}$, en numérique pour $f << f_{se}$ / premier terme du développement limité du sinus). Il suppose néanmoins un fonctionnement linéaire du système et un signal d'entrée sinusoïdal.

En fonctionnement linéaire, à l'ordre n et sans restriction de domaine d'étude, les mêmes opérations peuvent être faites dans le domaine continu ou discret.

Le cas continu impose de prendre une puissance de bruit de la forme :

$$B^2_{(Q+\text{noise shaping d'ordre n})\,Audio} = \frac{B_Q{}^2}{f_{se}} \int_0^{f_{max}} \left| \frac{1}{A(f)+1} \right|^{2n} df$$

et le cas discret :

$$B^2_{(Q+\text{noise shaping d'ordre n})\,Audio} = \frac{B_Q{}^2}{f_{se}} \int_0^{f_{max}} \left| 2\sin\left(\frac{\pi f}{f_{se}}\right) \right|^{2n} df$$

• Un tableau de résultats peut illustrer les calculs détaillés :

R	***Ordre du noise shaper***	***Dynamique (dB) d'après les calculs simplifiés détaillés***	***Dynamique (dB), modélisation du codeur (Hawksford)***
50	*1*	*46*	*48*
100	*1*	*50*	*53*
200	*1*	*64*	*57*
50	*2*	*72*	*66*
100	*2*	*87*	*79*
200	*2*	***102***	*92*
50	*3*	***97,5***	*85*
100	*3*	***118,5***	***105***
200	*3*	***139,5***	***122***
50	*4*	***123***	***100***
100	*4*	***150***	***125***
200	*4*	***177***	***151***

La quatrième colonne présente les résultats obtenus par simulation informatique du codeur en régime sinusoïdal tenant compte des problèmes de dispersion de l'erreur de quantification posée par le *noise shaping* d'ordre élevé. Cette simulation ne suppose pas l'utilisation de *dither*.

La différence des résultats obtenus est donc due à l'approximation faite sur Df, et aux hypothèses quant à la linéarité du système.

• Si la non-linéarité du quantificateur est prise en compte, en ce qui concerne le phénomène de dispersion de l'erreur de quantification, la quantité B_Q devra s'exprimer en fonction de celle-ci. La dispersion se faisant sur w pas, on aura $B_Q=wQ$.

Le signal ayant toujours une amplitude identique (égale à Q au maximum), l'expression du rapport signal sur bruit sera modifiée (à la baisse) et dépendra donc de w (sachant que si $w=1$, on se ramène à l'étude précédente).

Cette analyse est proposée par ***Darling*** et ***Hawksford*** dans leur article : "*Oversampled Analog-to-Digital Conversion for Digital Audio Systems*", JAES, Vol. 38, n°12, Décembre 1990.

Faisant les mêmes approximations sur *Df* que précédemment, le bruit de quantification dans la bande audio peut alors s'exprimer comme :

$$B^2{}_{(wQ+\text{noise shaping d'ordre } n)\,Audio} = w^2Q^2 \times \frac{1}{2n+1} \times (2\pi)^{2n} \times \left(\frac{1}{2R}\right)^{2n+1}$$

et l'expression du rapport signal sur bruit devient :

$$S/B_{ordre\ n}(dB) = 10log\left[\frac{2n+1}{2w^2(2\pi)^{2n}}\right] + 10log(2R)^{2n+1}$$

On trouve :

R	***Ordre du noise shaper n***	***w***	***Dynamique (dB)***
50	*1*	*5*	*32*
100	*1*	*5*	*41*
200	*1*	*5*	*50*
50	*2*	*7*	*55*
100	*2*	*7*	*70*
200	*2*	***7***	*85*
50	*3*	***11***	*77*
100	*3*	***11***	***98***
200	*3*	***11***	***119***
50	*4*	***15***	***99***
100	*4*	***15***	***126***
200	*4*	***15***	***153***

Cette analyse semble confirmer les résultats de la simulation pour les ordres élevés de mise en forme de bruit. Pour les faibles valeurs de *n*, l'écart constaté est dû au fait que la modélisation informatique n'utilise pas de *dither*, se servant du "*dithering* naturel" produit par les phénomènes de non-linéarité aux ordres élevés de *noise shaping*. L'hypothèse d'un B_Q constant sur la bande utile n'est donc plus valable et peut expliquer (avec les approximations faites) les divergences constatées.

La dynamique de ce type de système est donc fonction de l'ordre du *noise shaping* effectué et de la fréquence de sur-échantillonnage.

6.0.0 Conclusions :

Les techniques de conversion récursives sur-échantillonnées de type ***ΣDPCM*** sont les plus utilisées actuellement, en raison des performances qu'elles proposent et de leur simplicité technologique. Elles ont aussi leur place dans tous les types de traitements numérique des signaux qui font appel à une requantification (filtrage, conversion de fréquence, changement de résolution, erreur de calcul...).

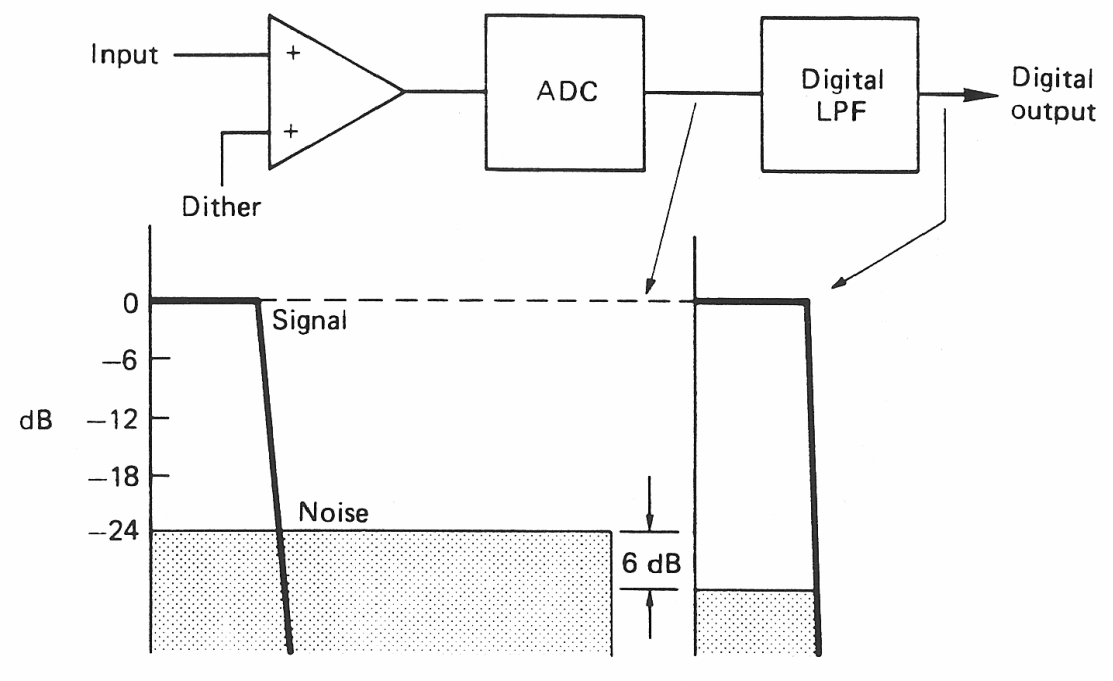
L'intégration des composants à grande échelle (*VLSI*) permet une mise au point précise des structures complexes de traitement numérique du signal, et une simplification en conséquence de la partie de traitement analogique. L'emploi des réseaux à capacités commutées peut permet ainsi de s'affranchir, en partie, des problèmes de *jitter* en restitution.

Une conversion de haute qualité devient alors possible, et repousse les limites technologiques un peu plus loin (résolutions supérieures à 20 bits), sachant que l'on peut manipuler des dynamiques de l'ordre de 110 dB.

La qualité du résultat obtenu repose finalement sur la précision d'implémentation des circuits numériques, la résolution des calculateurs, le soin apporté à la conception des étages analogiques, et la précision d'horloge.

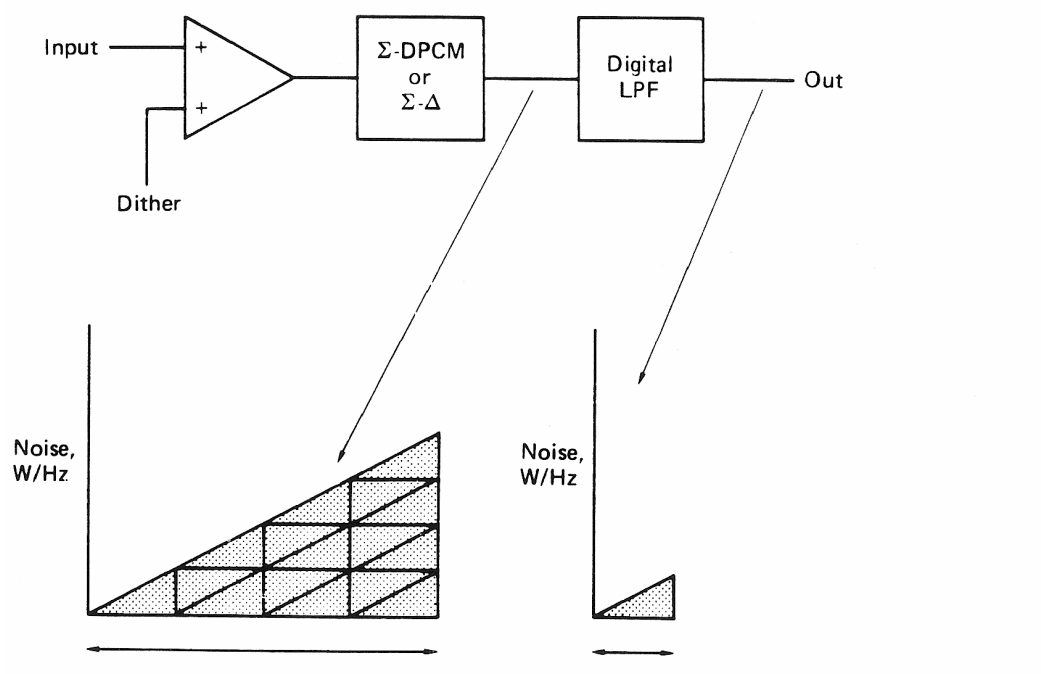
Le temps de traitement des codeurs ***ΣDPCM*** est de l'ordre de 1,2 ms en 20 bits actuellement. Face aux impératifs de synchronisation audionumérique, et à leur utilisation en exploitation audio, ce temps de processing peut poser problème, et impose une (re)-synchronisation du message numérique aux différentes étapes de la chaîne de production, ou de post-production.

Conversion classique : effet sur la puissance de bruit du sur-échantilloonnage par 4



La puissance de bruit est divisée par 4 (-6 dB)

Conversion ΣDPCM ou ΣΔ :



La puissance de bruit est divisée par 16 (-12 dB). Le bruit est mis en forme.

Principes technologiques de la conversion

PRINCIPES TECHNOLOGIQUES DE LA CONVERSION

7.1.0 Conversion numérique-analogique :

On commencera l'étude des principes de conversion par les *CNA* car ils sont souvent utilisés dans les boucles de rétroaction des convertisseurs analogiques-numériques.

De plus, la série de tests perceptifs prévue après l'étape de mesures physiques des systèmes de conversion suppose la détermination d'un convertisseur numérique-analogique de référence.

La conversion de messages binaires discrets en tension analogique continue peut se faire par les différents procédés suivants. On ne s'attachera, ici, qu'aux principes généraux de fonctionnement afin d'en mettre en évidence les limites. Cet exposé présente les systèmes les plus souvent utilisés et n'a évidemment pas la prétention d'être exhaustif.

7.1.1 Conversion par sommation de courants pondérés :

Le message *PCM* est décodé par sommation de sources de courants dont l'intensité est proportionnelle au nombre de pas de quantification (cf. *figure n°34*).

Les limites technologiques sont assez évidentes si l'on considère que, dans le cas d'une résolution assez fine, la précision des courants pondérés doit être supérieure à $1/2^n$ (en *16* bits cela équivaut à une précision de l'ordre de *0.0015 %*...). Les phénomènes de dérive en temps et température des composants rendent vite cette solution inenvisageable technologiquement et son utilisation inimaginable pour des applications audio sans quelques astuces.

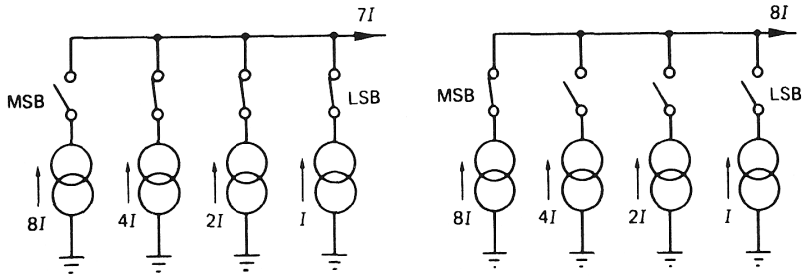
• L'obtention des sources de courant stables et précises peut ce faire par *dynamic element matching*. Dans ce cas, la source de courant de référence sera *n* fois divisée par deux en la faisant débiter à travers deux résistances de même valeur nominale. Sachant que la précision des composants (tolérances et dérives) entraîne une division approximative, on aiguille ensuite séquentiellement chaque branche sur chaque sortie par l'intermédiaire de commutateurs synchronisés sur une horloge stable dont le rapport cyclique est précisément de *50 %*. Chaque sortie est ensuite moyennée par un réseau capacitif classique.

Ce principe est présenté *figures n°35* et *36*.

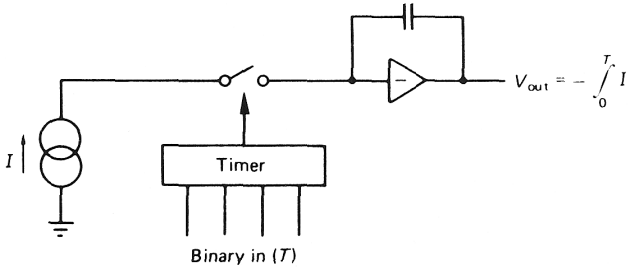
Ces éléments peuvent ensuite être mis en cascade pour une division par 2^n.

Cette méthode à l'avantage de ne pas nécessiter de réglage.

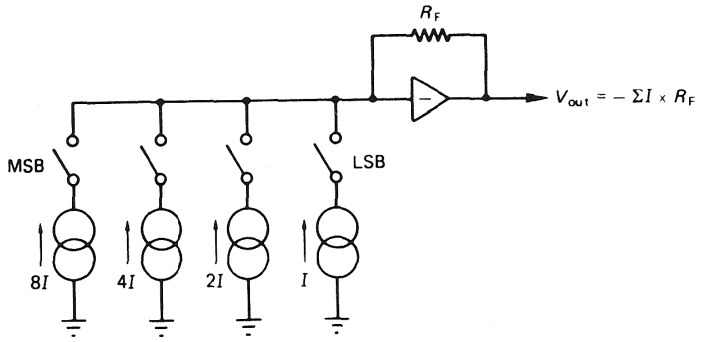
PRINCIPES TECHNOLOGIQUES ELEMENTAIRES DE LA CONVERSION NUMERIQUE-ANALOGIQUE



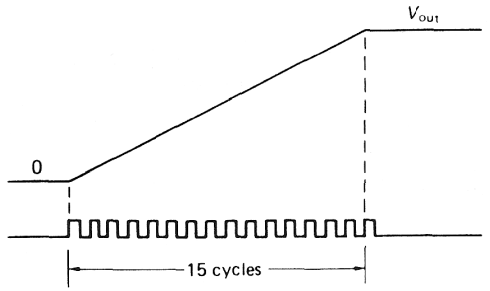
Courants Pondérés



Rampe d'Intégration Simple

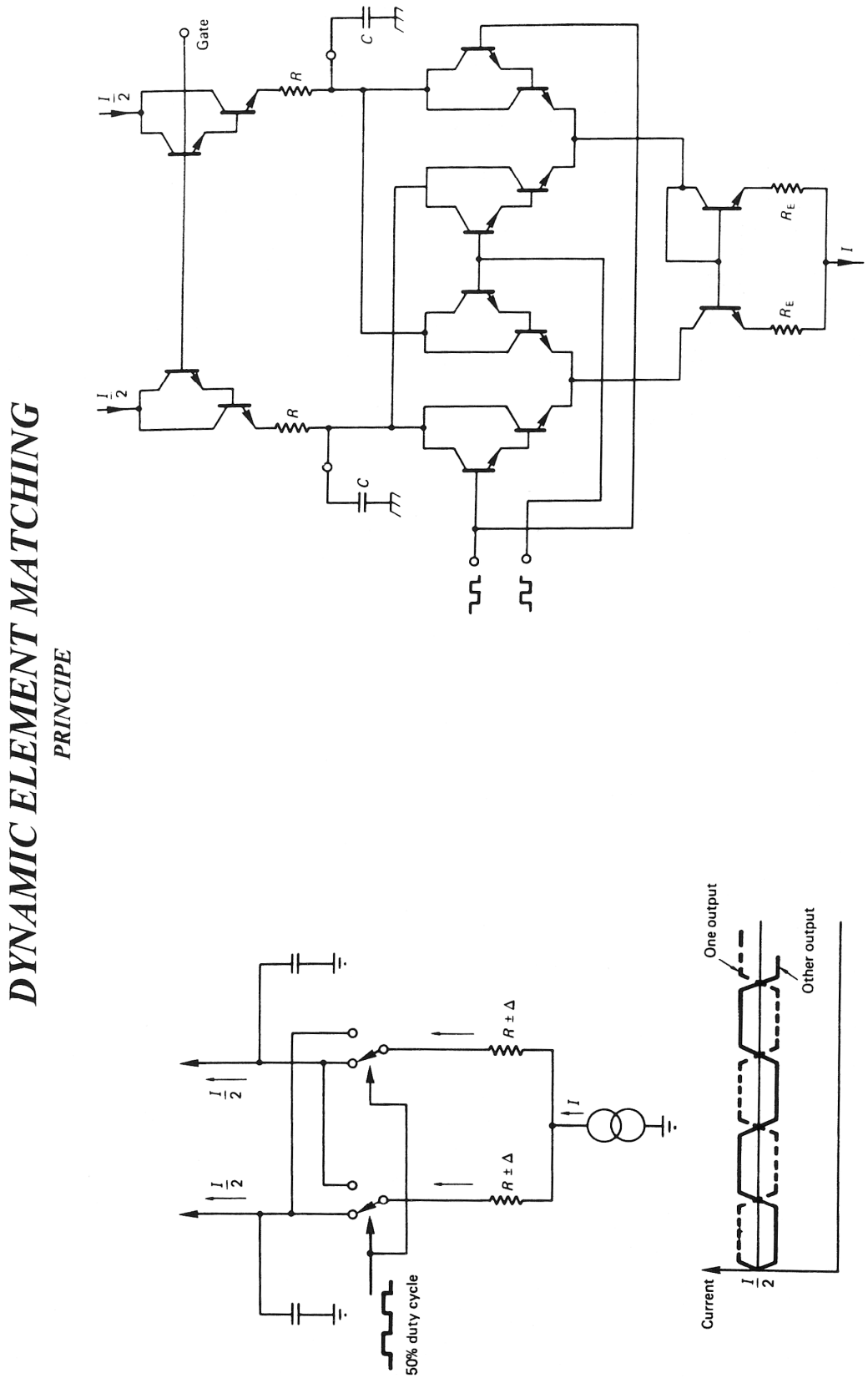


Exemple pour un message d'entrée 0 1 1 1



Rampe obtenue pour un message d'entrée 1 1 1 1 (ie15 cycles d'horloge)

Figure n°34



La moyenne du courant dans les deux branches sera identique si le rapport cyclique est précisément de 50 % . A droite, un exemple d'implantation monolithique. Remarque : la fréquence d'horloge est arbitraire.

Figure n°35

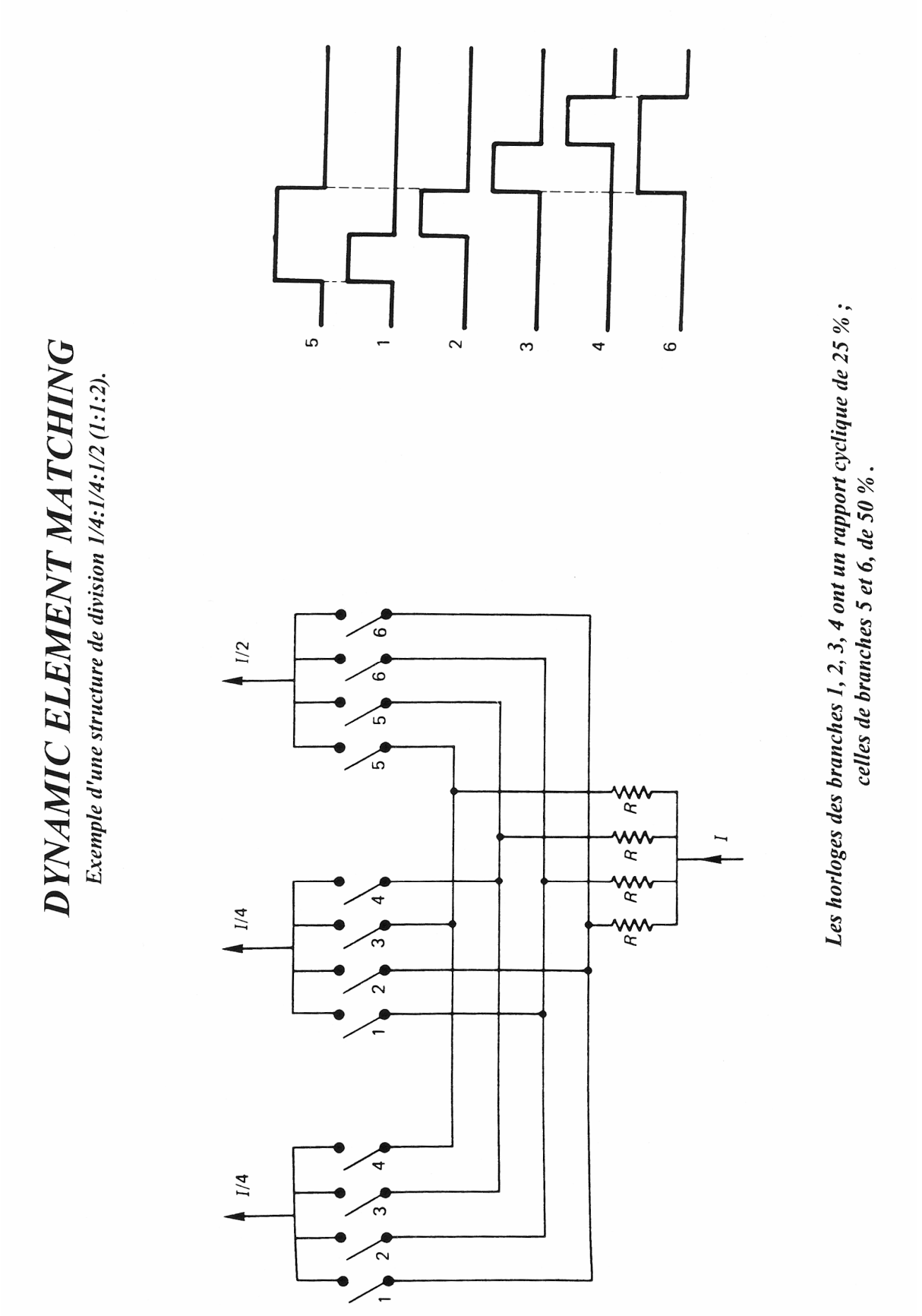

DYNAMIC ELEMENT MATCHING
Exemple d'une structure de division 1/4:1/4:1/2 (1:1:2).
Les horloges des branches 1, 2, 3, 4 ont un rapport cyclique de 25 % ;
celles de branches 5 et 6, de 50 % .
5
1
2
3
4
6
I/2
I/4
I/4
R
I

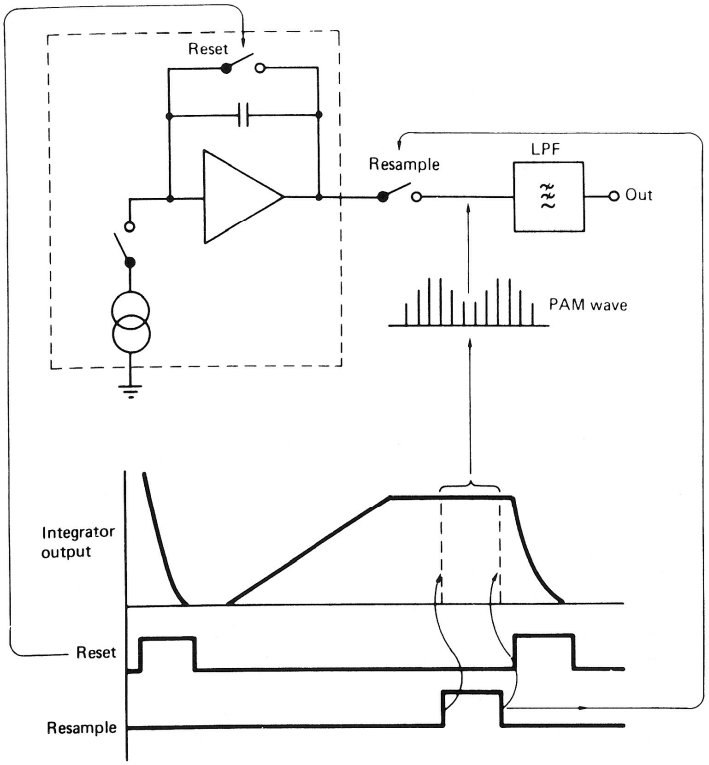
Figure n°36

7.1.2 Conversion par intégration d'un courant fixe :

Cette fois-ci c'est le temps d'intégration d'un courant fixe qui est pondéré par le message binaire (cf. *figure n°37*). Ce procédé utilise une horloge de comptage 2^n fois plus rapide que la fréquence d'échantillonnage (*3 GHz* en *44.1 kHz*), un intégrateur et un circuit de maintien de hautes performances capable de se charger/décharger très rapidement sans dérives. Ces limites supposent quelques raffinements pour l'utilisation de ce type de convertisseur pour l'audio.

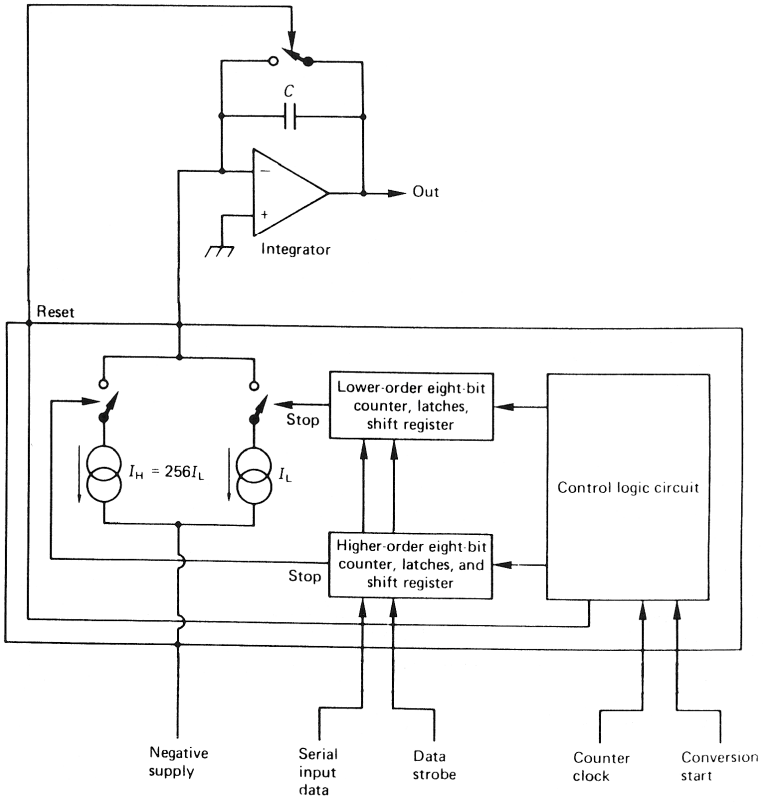
Un exemple de convertisseur 16 bits à double source de courant est présenté *figure n°38*. Les *MSB* et *LSB* sont, dans ce cas, traités par des circuits différents et la fréquence de comptage peut être réduite en conséquence. Ce type de conversion nécessite, une fois de plus, une grande précision des sources de courrant qui peut être obtenue par intégration des composants. Un circuit échantillonneur-bloqueur (ré-échantillonnage) complète le système en permettant de garder la valeur trouvée en sortie d'intégrateur, une fois les sources de courant déconnectées. Le condensateur est ensuite déchargé pour permettre la conversion suivante.

Les phases de conversion, ré-échantillonnage, et de décharge du condensateur doivent se faire au cours de la période d'échantillonnage. Une horloge de *20 MHz* permet une période de *12,8 µs* pour l'intégration (rampe), et de *8 µs* pour le ré-échantillonnage et la décharge du condensateur lors d'une conversion *16 bits*, *48 KHz*.



Intégrateur à rampe simple. La sortie n'étant stable qu'après la fin de la rampe, un interrupteur sera donc nécessaire pour isoler la rampe du reste du circuit. Celui-ci peut être utilisé pour obtenir un signal PAM si on le souhaite.

Figure n°37



Intégrateur à double rampe. L'interpolateur de lissage est ici extérieur (Sony CX-20017)

Figure n°38

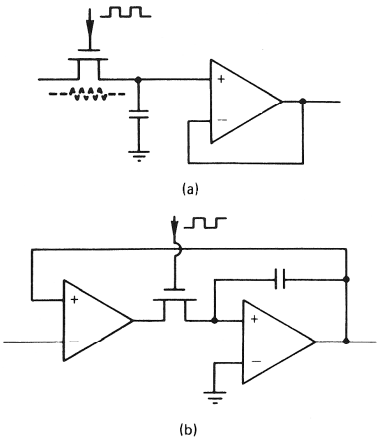
7.2.0 Conversion analogique-numérique :

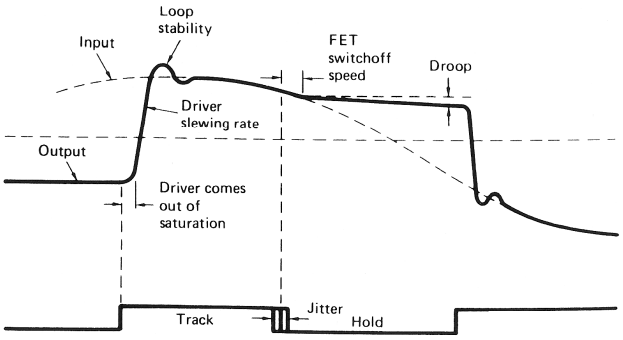
7.2.1 Circuit suiveur-bloqueur :

L'étape de conversion analogique-numérique nécessite, comme on l'a vu dans les principes généraux, une étape de maintien. Cette étape est prise en charge par le circuit suiveur-bloqueur. Un schéma de principe est donné à la *figure n°39*. Quand l'interrupteur est fermé, la sortie suit l'entrée et se bloque sur une valeur à l'ouverture de celui-ci. Le temps de charge du condensateur dépend de la résistance du *FET* et peut être minimisé en incluant celui-ci dans une boucle de rétroaction divisant sa résistance par le gain en boucle ouverte du circuit suiveur.

La réalisation d'un circuit suiveur-bloqueur de bonne qualité est extrêmement difficile en regard de la précision demandée par les applications audio. La qualité du circuit de maintien ne pourra être optimisée que pour des systèmes de résolutions moyennes, le temps de conversion demandé par un système *PCM* linéaire *16 bits*, *48 KHz* constituant une limite technologique.

La question de stabilité d'horloge à été évoquée précédemment. Nous avons vu que pour éviter les erreurs d'acquisition dues au *jitter*, une précision de l'ordre de quelques picosecondes peut être nécessaire.





(a) Schéma de principe
(b) Optimisation du circuit

Paramètres sensibles du circuit suiveur-bloqueur

Figure n°39

7.2.2 Quantificateur :

L'étape de quantification peut se faire par différentes méthodes dont on passera en revue les principales.

7.2.3 Convertisseurs « *flash* » :

La conversion « *flash* » est la méthode la plus rapide pour les conversions *PCM* et *DPCM*.

Le principe est présenté à la *figure n°40* et une réalisation en *8* bits à la *figure n°41*. Le voltage de référence est divisé par une chaîne de résistances et permet d'obtenir les tensions de référence secondaires correspondant à chaque niveau de quantification.

La référence de départ (correspondant au niveau maximum) peut varier et détermine la sensibilité d'entrée du système. La conversion se fait ensuite par comparaison entre les références secondaires et le signal d'entrée.

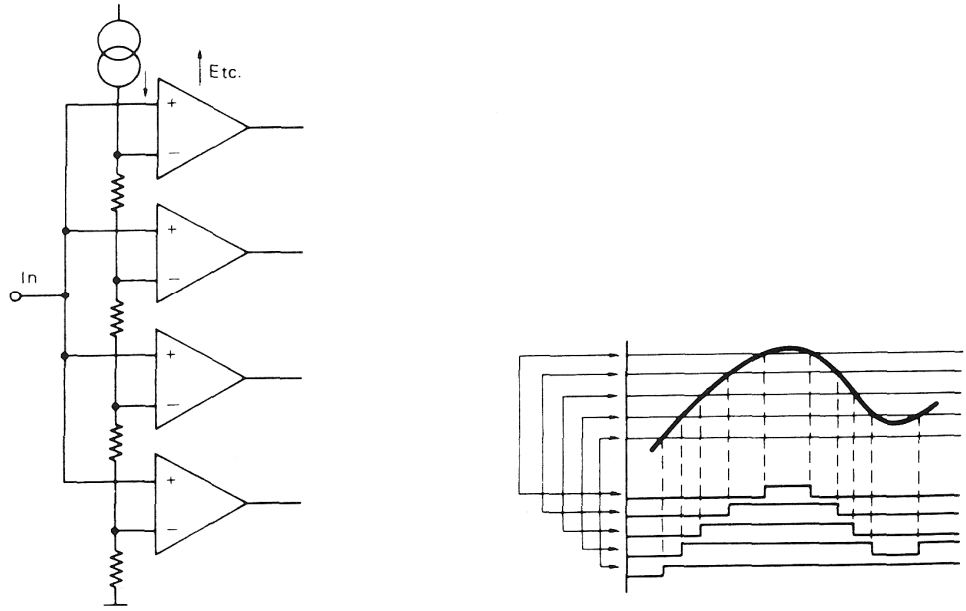
Ce procédé requiert un comparateur par niveau de quantification et un codage de la sortie de l'ensemble du système. On se rend vite compte qui si la résolution augmente, la circuiterie devient très importante et impose une intégration de l'ensemble.

Un convertisseur *n* bits nécessitant 2^n comparateurs, l'utilisation de ce type de technologie n'est envisageable que pour des résolutions basses ou moyennes (cas des technologies *DPCM*, *ΣDPCM* et des convertisseurs vidéo). Actuellement on arrive à des résolutions de l'ordre de 10 bits).

Le signal d'entrée ayant à alimenter un grand nombre de comparateurs, il sera nécessaire de prendre des précautions d'adaptation d'impédance qui s'imposent (*driver* basse impédance).

L'intérêt de ce type de conversion réside dans le traitement parallèle des informations et, par conséquent, dans la rapidité du processus. L'étape de maintien n'est plus forcément nécessaire (on évite les problèmes mentionnés). Ces propriétés expliquent leur utilisation dans les systèmes sur-échantillonnés.

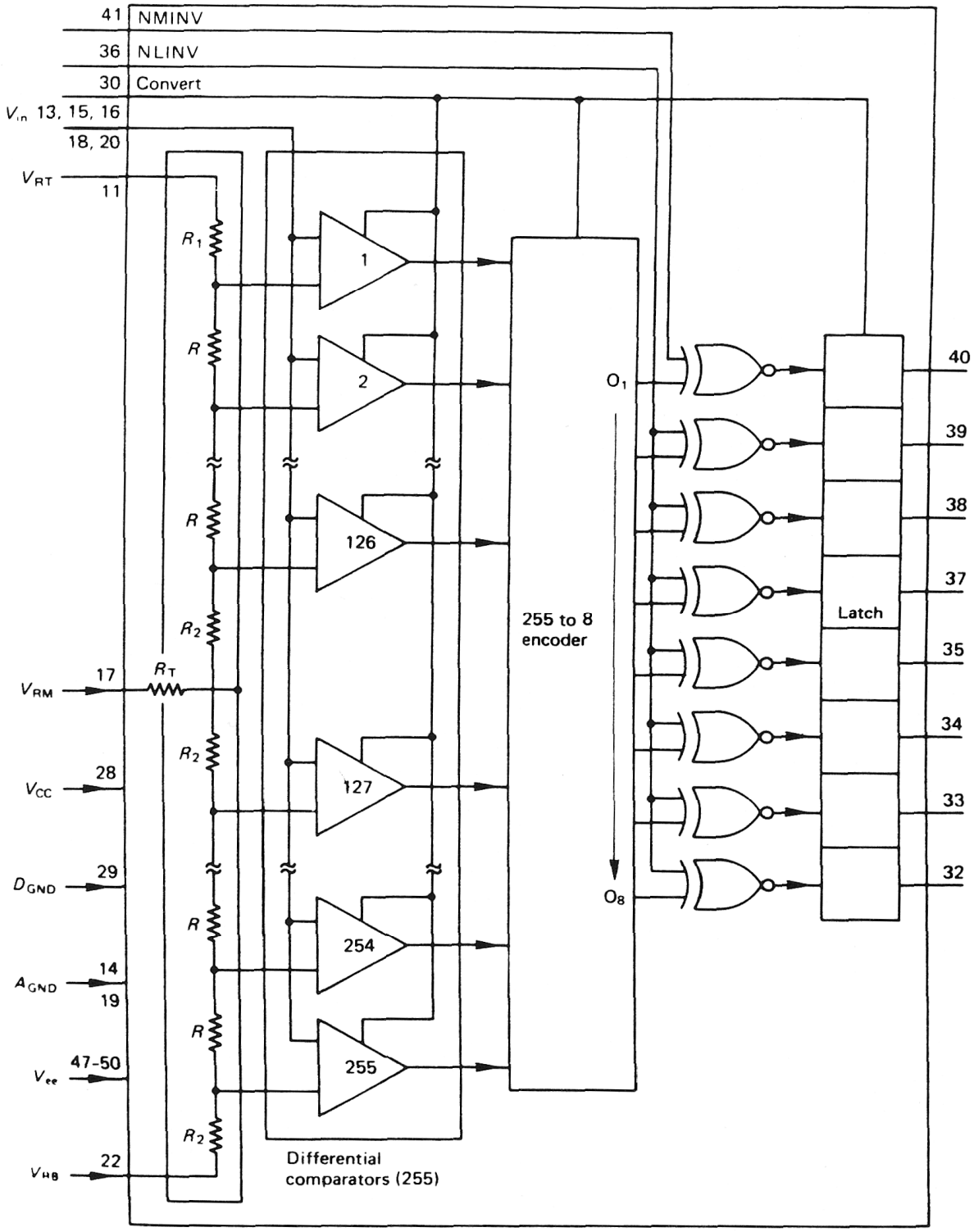
Convertisseur « flash », Principe et codage obtenu



Chaque pas de quantification a son comparateur propre. Un codage du résultat obtenu est nécessaire pour se ramener à un codeur binaire.

Figure n°40

Convertisseur « flash » 8 bits



Convertisseur analogique-numérique flash 8 bits typique (application vidéo et *ΣDPCM*)

Figure n°41

7.2.4 Convertisseurs à rampe :

◊ Rampe simple :

La méthode la plus « primaire » consiste à comparer la tension d'entrée à celle obtenue par conversion numérique-analogique de la sortie d'un compteur. La sortie du comparateur est alors utilisée pour arrêter le comptage. Ce principe est présenté à la *figure n°42*.

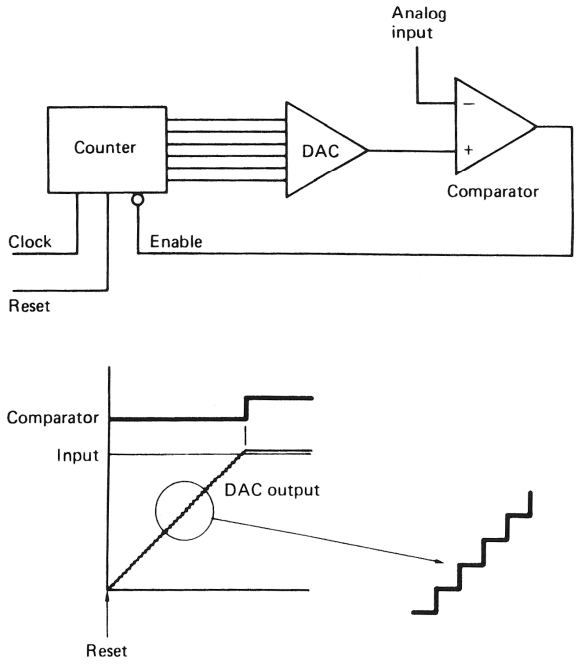
La lenteur du procédé et les défauts introduits par l'utilisation d'un *CNA* sont rédhibitoires et en condamne l'utilisation pour des applications de haute résolution.

Ces convertisseurs sont parfois appelés "convertisseurs à rampe numérique".

Une autre méthode reposant sur le même principe est la conversion dite "à rampe analogique". Dans ce cas, ce n'est pas un compteur à travers un *CNA* qui génère la rampe mais l'intégration d'une tension de référence. La comparaison se fait, dans ce cas, entre l'entrée à convertir et la sortie de l'intégrateur. Un compteur permet d'obtenir la valeur binaire représentant le temps d'intégration.

La précision de la conversion dépend de la linéarité et de la stabilité de la rampe générée ; le temps de conversion de la fréquence de comptage.

CONVERTISSEUR A RAMPE NUMERIQUE



Le convertisseur analogique-numérique à rampe numérique simple compare la sortie du DAC à l'entrée analogique. Le comptage est stoppé quand la rampe atteint la valeur d'entrée.

Figure n°42

◊ **Rampe double :**

Le principe reste le même, mais c'est le temps de décharge du condensateur qui est compté. L'intérêt réside dans le fait que la valeur trouvée ne dépend pas des valeurs *RC* de l'intégrateur (à condition qu'elles restent constantes pendant toute la durée de conversion). Elles peuvent néanmoins varier à long terme (tant que la période de variation est grande devant le temps de conversion).

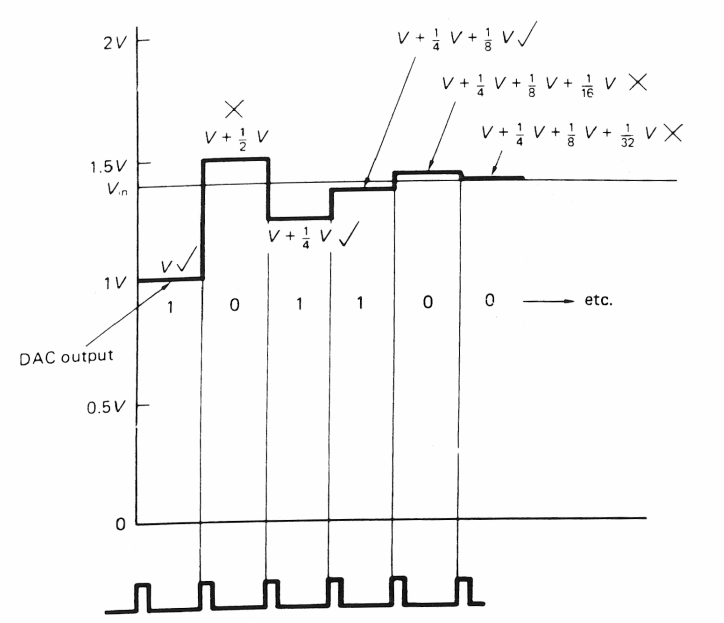
Si la tension d'entrée varie au cours de la conversion, on obtient alors une valeur binaire qui correspond à la moyenne de la tension d'entrée sur le temps d'intégration de la première rampe.

Dans tous ces systèmes dits "à rampe" un circuit suiveur-bloqueur doit être utilisé.

7.2.5 Convertisseurs à approximations successives :

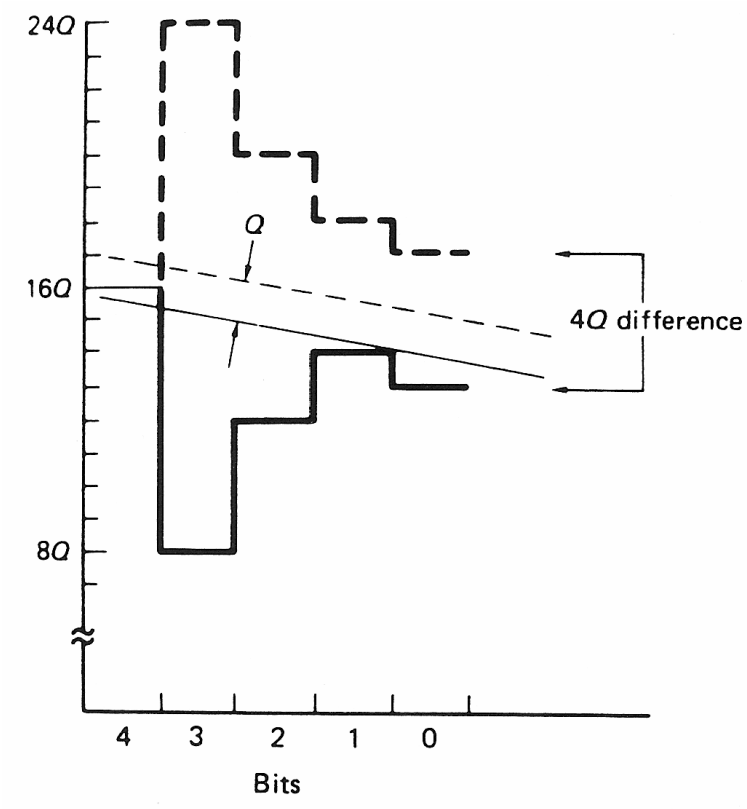
La conversion se fait séquentiellement par comparaisons successives du signal d'entrée et des tensions associées aux bits de codage (du *MSB* vers le *LSB*). Si l'entrée est plus grande que la moitié de l'échelle totale (*Vmax/2*) à la première comparaison, alors la valeur *1* est conservée et l'intervalle *[Vmax/2; Vmax]* et sert de base à la comparaison du bit suivant. Le processus se termine après *n* comparaisons séquentielles (cf. *figure n°43*).

La conversion par approximations successives a donc l'avantage d'être une méthode rapide, mais présente l'inconvénient de traiter le *LSB* (le plus sensible aux imperfections de maintien) à la fin du processus. Cet inconvénient peut dans certains cas se traduire par une sorte de divergence du système. Ce phénomène est schématisé, *figure n°44*, pour deux valeurs d'entrée différant d'un pas de quantification (autour d'une valeur de référence). La dérive du système de maintien se traduit ici par la non monotonie du système de conversion.



L'approximation successive compare l'entrée aux références en commençant par la plus grande. Si l'entrée est en dessous de *V,* la valeur conservée est 1 ; si elle est au dessus (*X)*, on conserve 0.

Figure n°43



Effet de dérive du système pour deux signaux bloqués différant d'un pas *q* de quantification autour d'une référence. Le résultat de la conversion, dans cet exemple, diverge de *4q*. L'instabilité du maintien peut donc détruire la monotonie du convertisseur.

Figure n°44

7.2.6 Convertisseurs à double source de courant (expansion résiduelle) :

La décharge du condensateur de la cellule de maintien de l'échantillon est ici utilisée comme rampe d'intégration. Le principe (pour une conversion en *16* bits) consiste à décharger cette capacité séquentiellement par deux courants de rapport *127/1*, et de quantifier par comptage ce temps de décharge. Le courant le plus faible permet de coder les *7 LSB,* et la somme des deux les *9 MSB*. Le principe est représenté à la *figure n°45*. Le codage des *9 MSB* se fait en utilisant les deux sources en série (*128I*), et en comptant le temps de décharge jusqu'à la référence de *128Q* du premier comparateur (*Q* étant le pas de quantification). Puis, la source la plus élevée est coupée et le comptage des *7 LSB* peut se faire jusqu'à la décharge totale du condensateur de maintien.

C'est, en fait, une conversion à rampe d'intégration double un peu particulière.

Temps de conversion : Comptage des *MSB* sur *9* bits, des *LSB* sur *7 (512 + 128 = 640)*, si l'on considère que le circuit de blocage utilise *25 %* de la période d'échantillonnage, une conversion à *48 kHz* nécessitera une horloge de comptage à 40 *MHz*...

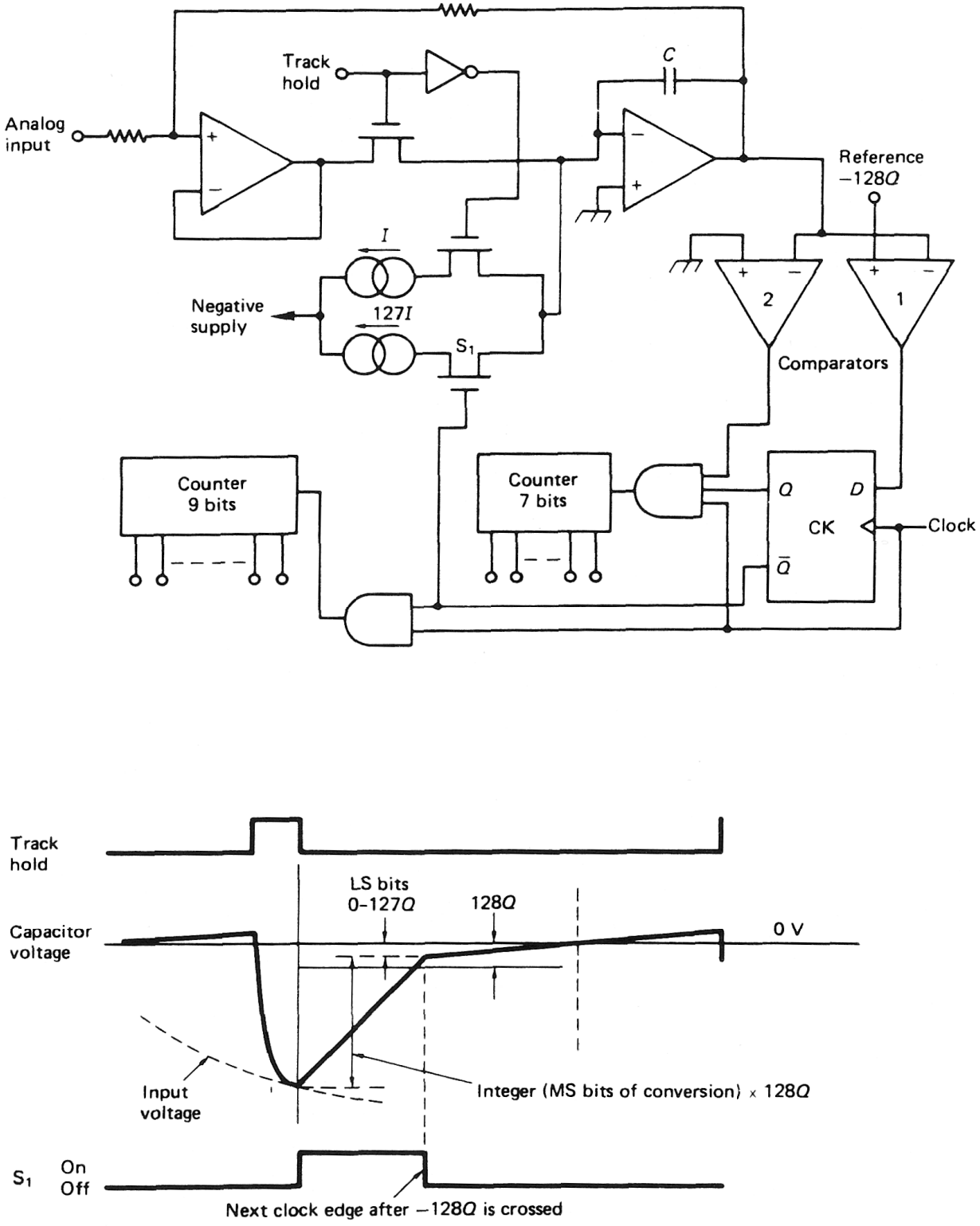

Track
hold
C
Analog
input
Reference
−128Q
I
Negative
supply
127I
S1
2
1
Comparators
Counter
9 bits
Counter
7 bits
Q
D
CK
Clock
Track
hold
LS bits
0-127Q
128Q
Capacitor
voltage
0 V
Input
voltage
Integer (MS bits of conversion) x 128Q
S1
On
Off
Next clock edge after −128Q is crossed

Figure n°45

• Un autre type de conversion par expansion résiduelle peut être fait en utilisant deux convertisseurs *flash* de *n* bits. Le signal maintenu est alors converti une première fois sur *n* bits et stocké dans un registre. Il est ensuite reconverti dans le sens inverse (par un convertisseur *N/A* de référence) et soustrait au signal d'entrée. La différence est ensuite numérisée à son tour sur *n* bits et stockée dans un registre différent. On a donc deux conversions successives, équivalentes à une conversion simple quantifiée sur ***2n-1*** bits, un bit étant perdu pour cause de non-linéarité du convertisseur *flash*.

Moins rapide que la conversion flash simple, l'avantage de cette méthode réside dans le fait que chacun des convertisseurs pourra avoir une précision deux fois moindre que la précision finale envisagée.

8.0.0 Perspectives :

Codage spectral : Le codage spectral se propose d'enregistrer la transformée de Fourier rapide (*Fast Fourier Transform*) du signal sur quelques bandes spectrales bien définies (pondérations psycho-acoustiques, bandes critiques...). La restitution peut ensuite se faire par (re)synthèse additive. Cette technique en est, pour le moment, au stade expérimental, les temps de calcul étant prohibitifs...

Critères de choix et spécifications

CRITERES DE CHOIX ET SPECIFICATIONS

En raison des nombreuses applications nécessitant l'emploi de convertisseurs, une palette très large de composants est proposée par les constructeurs. Leurs spécifications peuvent, par conséquent, être très différentes (en terme de bande passante, dynamique, temps de conversion...), en fonction du secteur d'application auxquels ils sont destinés.

On ne parlera bien évidemment ici que de leur utilisation en acquisition et restitution de signal audio.

On peut diviser les caractéristiques de fonctionnement de ces composants en deux catégories : statique et dynamique.

9.1.0 Caractéristiques statiques (cf. *figures n°46* et *47*) :

• **Résolution :** Comme nous l'avons vu précédemment, la résolution détermine l'écart entre deux pas de quantifications consécutifs dans les systèmes *PCM* linéaires. Elle est fonction du nombre de bits sur lequel l'information d'amplitude est codée (précision du codage), sachant qu'un système de quantification sur n bits permet un codage sur 2^n niveaux (le plus petit incrément est par conséquent de $1/2^n$ soit *1 Least Significant Bit*).

Elle peut être exprimée soit en pourcentage de la pleine échelle soit en nombre de bits.

Un convertisseur *16* bits aura donc $2^{16} = 65\,536$ niveaux ($1{,}523\ 10^{-3}$ % de la pleine échelle) et pourra donc détecter une variation de *152.3 µV*, si la référence est de *10 V*.

Les précisions demandées deviennent très grandes dès que l'on améliore la résolution et requièrent une technologie adaptée.

La résolution seule ne permet pas de juger de la précision d'un circuit dans la mesure où sa "résolution réelle" dépend de la linéarité de celui-ci.

• **Précision :** La précision est ici entendue en tant que précision absolue. Elle représente l'écart entre la valeur mesurée en sortie de convertisseur et la valeur théorique que l'on aurait obtenu avec un modèle parfait (pas d'erreur de conversion). Elle cumule donc les erreurs de quantification, d'*offset*, de gain, de *jitter*...

La précision donne, de ce fait, une idée globale des performances du circuit, mais peut s'avérer trop limitative dans certains cas. On y préférera un détail des performances séparées pour chaque type d'erreur afin de pouvoir en effectuer une interprétation adaptée à chaque type d'utilisation.

• **Erreur de quantification :**

◊ L'erreur de quantification représente la déviation maximale d'un système par rapport à la caractéristique de quantification idéale (transfert droit ≡ transparence).

On a vu précédemment qu'elle avait une valeur absolue maximale de $Q/2$ dans un système de type Nyquist (Q étant le pas de quantification) sans offset (*1/2 LSB* par défaut).

Elle représente, dans ce cas, une fraction de ($1/2^{n+1}$) de la pleine échelle.

◊ Dans les systèmes de type *ΣDPCM* la valeur de l'erreur de quantification peut varier entre zéro et Q. Elle sera ensuite moyennée par la décimation.

• **Erreur de gain ou d'échelle :** L'erreur de gain est l'erreur de pente par rapport au cas idéal (première bissectrice du graphique tension d'entrée analogique, amplitude quantifiée, caractéristique de quantification). Cette erreur est la différence entre l'amplitude maximale théorique du convertisseur et celle que l'on mesure. Elle se traduit par une rotation de la caractéristique de quantification autour de l'origine et est exprimée en pourcentage de la pleine échelle ou en fraction de *LSB*.

Cette erreur est en général due à une erreur d'alignement de l'entrée analogique ou à une incertitude sur la (les) référence(s) de tension du système.

La plupart des circuits permettent de régler ce paramètre par un potentiomètre extérieur.

La dérive thermique des composants électroniques peut aussi en être une cause.

• **Erreur d'*offset* :** L'erreur d'*offset* est représentée par la tension d'entrée pour laquelle la sortie du convertisseur est nulle (ajout d'une composante continue). Elle est, en général, causée par l'*offset* d'un étage préamplificateur, ou d'un comparateur interne ou externe au circuit. Elle se traduit par une translation verticale de la caractéristique de quantification et est, une fois de plus, exprimée en pourcentage de la valeur pleine échelle ou en fraction de *LSB*. Il sera toujours possible de régler ce paramètre soit par un potentiomètre externe associé au convertisseur, soit par l'intermédiaire d'un amplificateur extérieur équipé d'un réglage d'*offset*.

La dérive thermique des composants électroniques peut aussi en être une cause.

Dans le cas d'une conversion analogique-numérique associée à un traitement par microprocesseur, une acquisition de l'*offset* éventuel est possible avant traitement et permet d'en effectuer la correction en numérique (le réglage analogique étant susceptible de dériver).

• **Erreur d'hystérésis :** Les transitions de la caractéristique de quantifications peuvent être différentes suivant le sens de parcours de celle-ci. Ce phénomène est, en général, dû à l'hystérésis des comparateurs utilisés en tant que quantificateurs (approximations successives...).

• **Monotonie :** L'évolution en sortie du convertisseur doit être similaire à celle de l'entrée. Cette similitude est désignée sous le terme de monotonie. La non-monotonie est un cas particulier de non-linéarité différentielle (inégalité des pas de quantification, précision des références de tension).

• **Linéarité (ou linéarité intégrale):** L'erreur de linéarité n'inclut pas les erreurs de quantification, d'*offset* ou de gain, mais s'additionne à celles-ci. Elle se traduit par une déformation de la caractéristique de quantification et entraîne une inégalité des pas de quantification.

Une non-linéarité de *± 1/2 LSB* peut provoquer l'annulation d'un pas de quantification et conduit donc à un code manquant. En revanche, une non-linéarité inférieure à *± 1/2 LSB* garantit la monotonie de la fonction de transfert.

Elle peut être exprimée en pourcentage de la valeur pleine échelle ou en fraction de *LSB*.

On passe d'une valeur à l'autre par le calcul suivant (*n* étant le nombre de bits de résolution) :

$$Valeur\,en\ LSB = \frac{2^n \times non - linéarité\ en\ \%}{100}$$

Les "pourcentages" sont en général donnés en **ppm** (*parts per million* i.e. 10^{-4} %).

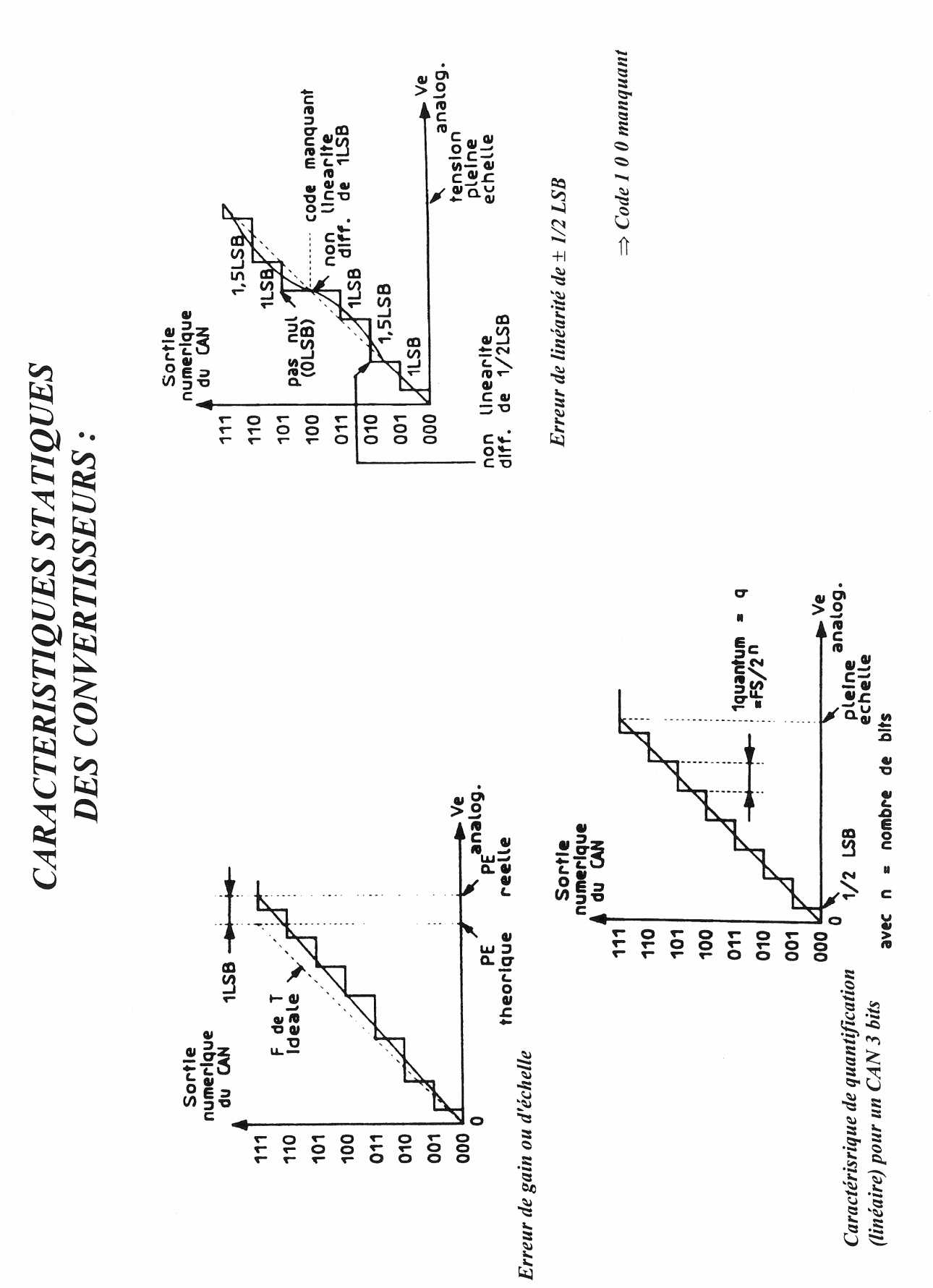

CARACTERISTIQUES STATIQUES
DES CONVERTISSEURS :
Sortie numerique du CAN
111
110
101
100
011
010
001
000
0
1LSB
F de T Ideale
PE theorique
PE reelle
Ve analog.
Erreur de gain ou d'échelle
Sortie numerique du CAN
1,5LSB
1LSB
pas nul (0LSB)
code manquant
non linearite diff. de 1LSB
1LSB
1,5LSB
1LSB
Ve analog.
tension pleine echelle
non linearite diff. de 1/2LSB
Erreur de linéarité de ± 1/2 LSB
⇒ Code 1 0 0 manquant
Sortie numerique du CAN
1quantum = q =FS/2^n
1/2 LSB
Ve analog.
pleine echelle
avec n = nombre de bits
Caractérisrique de quantification (linéaire) pour un CAN 3 bits

Figure n°46

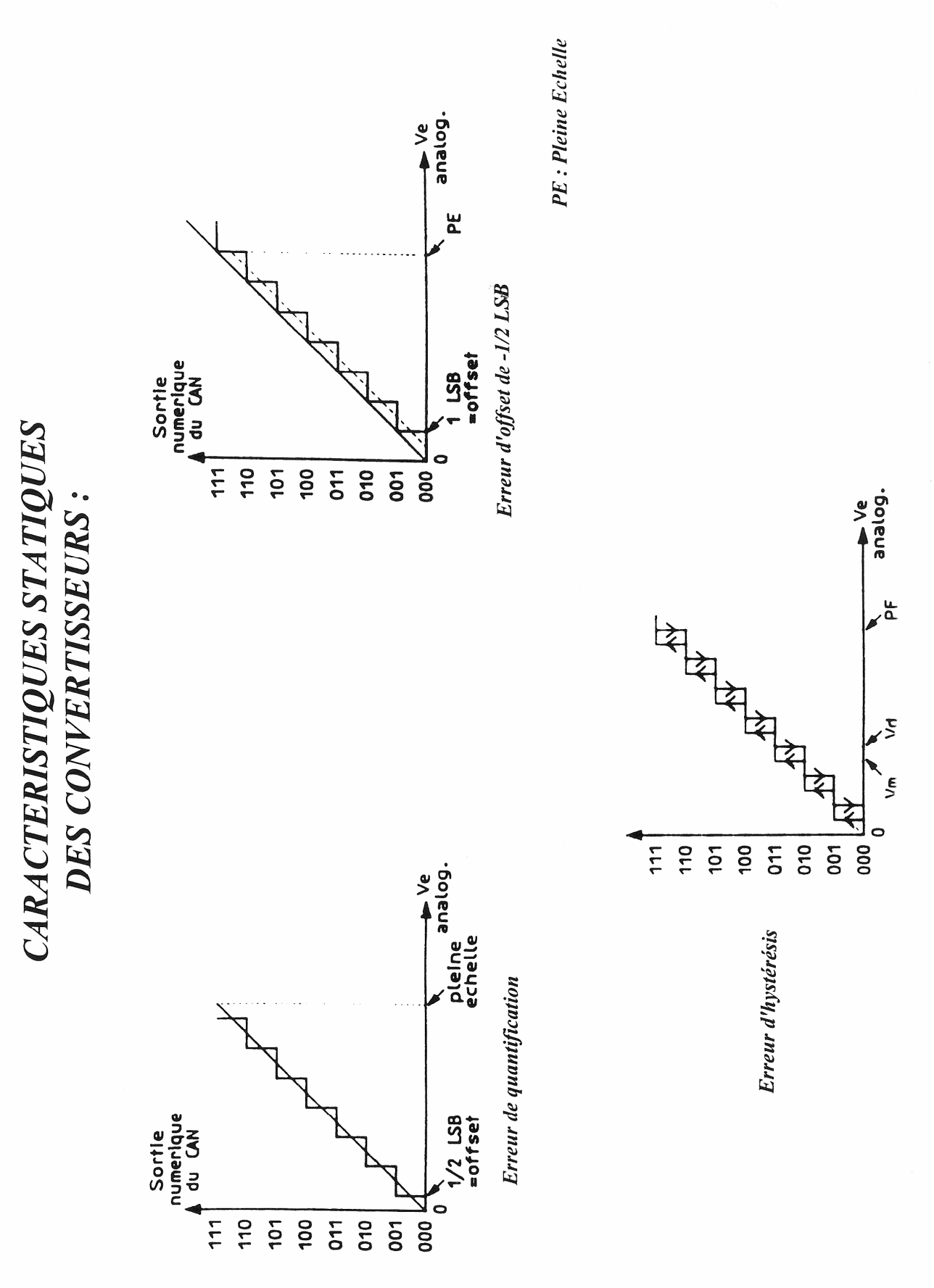

CARACTERISTIQUES STATIQUES
DES CONVERTISSEURS :
Sortie numerique du CAN
111
110
101
100
011
010
001
000
0
1/2 LSB =offset
pleine echelle
Ve analog.
Erreur de quantification
Sortie numerique du CAN
1 LSB =offset
PE
Ve analog.
Erreur d'offset de -1/2 LSB
PE : Pleine Echelle
Vm
Vm
PF
Ve analog.
Erreur d'hystérésis

Figure n°47

9.2.0 Mesure des caractéristiques dynamiques :

Aucune norme n'a encore été proposée afin d'assurer l'homogénéité des caractéristiques dynamiques des convertisseurs. Cela entraîne une différence de spécifications en fonction des constructeurs et peut poser quelques problèmes lorsqu'il s'agit de faire un choix...

Reste à tester soi-même les différents circuits par une méthode fiable et reproductible.

Deux types de chaînes de mesure sont alors envisageables. La méthode analogique, consistant à placer derrière le *CAN* à tester un *CNA* de référence et à effectuer l'analyse par les outils analogiques traditionnels et la méthode numérique. On préférera cette dernière, car elle a le grand avantage de s'affranchir des problèmes analogiques (spécifications du *CNA* de référence, circuiterie d'antiparasitage...) L'environnement du composant sous test n'en sera pas moins soigné afin de ne pas fausser les mesures. Il est nécessaire pour cela de disposer d'un générateur de fonction de très grande précision (performances en dynamique et distorsion supérieures à celles du circuit sous test) et d'un analyseur de spectre de haute qualité permettant de s'affranchir des artefacts de mesure en ce qui concerne les paramètres d'analyse numérique du signal (transformée de ***Fourier*** rapide de bonne qualité, possibilités sophistiquées de fenêtrage ou d'échantillonnage cohérent $\Rightarrow$ table d'onde...)

On rappelle qu'un échantillonnage cohérent utilise une fenêtre temporelle d'observation rectangulaire d'un nombre entier de périodes du signal à analyser et qu'il permet de ce fait de s'affranchir des pondérations dues à l'observation (seule la composante continue étant conservée). Dans le cas d'échantillonnage non-cohérent (signal non périodique par exemple), les lobes secondaires du spectre de la fenêtre interfèrent sur celui du signal étudié et fausse en conséquence les résultats de mesure (effets de bords...) Un fenêtrage temporel approprié peut permettre, dans cette situation, de minimiser ces phénomènes.

Les fenêtres de mesure, conventionnellement employées en traitement numérique du signal, pondèrent la fenêtre rectangulaire d'observation par des développements trigonométriques de type :

$$f(kT_e) = \sum_i a_i \cos\left(\frac{2\pi ki}{N}\right), \text{ } k \text{ étant un entier naturel.}$$

NT_e étant la largeur totale de la fenêtre d'observation avant pondération, i un indice qui, dans cet exemple, est tel que $0 \leq i \leq 3$ et les a_i les coefficients, tels que : $\sum_i |a_i| = 1$.

On donne ici les coefficients a_i pour quelques-unes des fenêtres classiques :

Coefficient	*Hann*	*Hamming*	*Blackman Harris*
a_0	*1/2*	*0,54*	*0,35875*
a_1	*-1/2*	*-0,46*	*-0,48829*
a_2	*0*	*0*	*0,14128*
a_3	*0*	*0*	*-0,01168*

Il faut, de plus, remarquer que plus les lobes secondaires de la fenêtre d'observation sont de faible amplitude et plus la largeur du lobe principal de celle-ci est importante. Cela doit être pris en compte lors du choix du nombre d'échantillons sur lequel s'effectue le calcul de la transformée de Fourier discrète. La largeur du lobe principal est, ici, inversement proportionnelle à *N*.

On résumera donc les impératifs de mesure par *FFT* comme suit :

◊ Quand le signal d'entrée pourra être maîtrisé (échantillonnage cohérent), on le choisira tel que :

$$\frac{f_{signal}}{f_{échantillonnage}} = \frac{M}{N} = Constante$$

où *N* est le nombre de périodes d'échantillonnage d'observation et *M* un nombre premier de périodes (algorithme de calcul).

Et il n'est pas, dans ce cas, nécessaire de pondérer la fenêtre d'observation.

◊ Autres cas : utilisation d'une fenêtre de pondération adaptée au circuit testé, afin que les erreurs d'analyse (amplitude des rebonds spectraux de la fenêtre utilisée) soient toujours négligeables par rapport à la dynamique théorique du convertisseur à tester.

• **Remarque sur le fenêtrage en *TFD* (Transformée de *Fourier* Discrète) :** La dynamique de mesure dépend, dans le cas d'un échantillonnage non-cohérent, de l'amplitude des rebonds spectraux de la fenêtre d'acquisition. Ceux-ci se repliant par périodisation spectrale due à la discrétisation temporelle, ils déterminent, de fait, la dynamique de mesure.

Dans le cas d'un système de haute résolution, l'utilisation de fenêtres sophistiquées s'impose. On remarque que dans le cas de l'utilisation d'une fenêtre de ***Blackmann Harris***, l'amplitude des rebonds spectraux est inférieure à - *92 dB* et autorise donc une étude des systèmes de conversion jusqu'à *15* bits de résolution. Au-delà, un fenêtrage différent devra être utilisé (fenêtre de plus grande dynamique, ou augmentation de l'ordre du développement limité par exemple…)

La largeur de la fenêtre d'acquisition étant proportionnelle à l'amplitude des rebonds spectraux, le calcul de *TFD* devra se faire sur un nombre de points précis, ce qui n'est pas toujours prévu dans les analyseurs de spectre, l'implantation de la transformation de *Fourier* rapide (*FFT)* reposant sur l'utilisation des puissances de deux.

Le même problème de nombre de points se pose en échantillonnage cohérent car il varie avec la fréquence.

• **Remarque sur la *TFD* :** La transformée de ***Fourier*** d'un signal discret étant discrète, la résolution en fréquence est donc fonction d'un pas Δf. Le choix de $\Delta f = 1/NT_e$ simplifie énormément les calculs et est, en général, adopté dans les algorithmes de calcul de *FFT*.

Il en résulte que $F_e = N\Delta f = 1/T_e$ et qu'il existe donc N valeurs spectrales différentes (spectre périodique de période N).

On a donc :

$$X^*(k\Delta f) = \frac{1}{N}\sum_{n=0}^{N-1} x^*(nT_e)e^{-2j\pi nk/N}$$

Avec $\Delta f = \frac{1}{NT_e}$ le pas en fréquence, T_e le pas temporel, $T_0 = NT_e$ la durée d'observation temporelle, X^* le spectre échantillonné et x^* le signal temporel discrétisé.

On peut, à présent, passer à la description des caractéristiques dynamiques.

9.2.1 Caractéristiques dynamiques :

• **Rapport signal sur bruit :** Comme on l'a vu, la dynamique des systèmes de conversion dépend de la résolution de quantification et peut être exprimée, avec les réserves étudiées précédemment, par :

$$\frac{S}{B}(en\ dB) = 10\log\frac{(Valeur\ éfficace\ \ du\ \ fondamental)^2}{\sum_{i=1}^{N/2}(Valeurs\ éfficaces\ \ des\ composantes\ résiduelles)^2}$$

La composante continue d'indice zéro n'est pas prise en compte dans le calcul du bruit; N est le nombre d'échantillons sur lequel est calculée la *FFT*.

En fonction du système testé, il peut être intéressant d'effectuer cette mesure à différentes fréquences et à différents niveaux.

• **Distorsion** (cf. *figure n°48*) **:** La distorsion est due aux non-linéarités du système de conversion. On s'attachera ici à caractériser la distorsion harmonique totale (*THD*) et la distorsion par inter-modulation (*IMD*).

La distorsion devra être mesurée à différentes fréquences dans le spectre utile, car on a vu qu'elle pouvait dépendre du spectre d'entrée.

◊ **Distorsion harmonique totale :** On effectuera cette mesure de façon classique en mesurant le rapport entre la puissance des harmoniques dues à la non-linéarité du système et celle du signal de test.

Ce rapport peut s'exprimer comme :

$$THD\left(endB\right) = 20log\left(\frac{\sqrt{V_2^{\,2} + V_3^{\,2} + ... + V_n^{\,2}}}{V_1}\right)$$

Les indices utilisés sont les rangs des harmoniques sachant que le fondamental a ici l'indice n°*1*. Les valeurs *V* sont les amplitudes efficaces des harmoniques de rang *n*.

Cette mesure peut s'effectuer au rang *n* désiré sachant que selon la fréquence de test, le nombre d'harmoniques audibles peut varier...

On pourrait caractériser de la même manière la **distorsion totale** en calculant ce rapport sur tous les pas fréquentiels du spectre utile.

◊ **Distorsion par inter-modulation :** On désigne sous le terme de distorsion par inter-modulation, le phénomène de combinaison linéaire possibles (de rang *m* et *n*) en sortie d'un système non-linéaire quand plusieurs signaux de fréquences différentes sont appliqués à son entrée.

Pour deux signaux d'entrée de fréquences f_1 et f_2, on peut avoir l'apparition des composantes :

$$f_{mn} = mf_1 \pm nf_2$$

L'*IMD* est, en général, spécifiée dans les trois cas suivants :

✧ *IMD* de deuxième ordre : $IMD_2\left(endB\right) = 20log\frac{\sqrt{V^2{}_{(f_1+f_2)} + V^2{}_{(f_1-f_2)}}}{\sqrt{V^2{}_{f_1} + V^2{}_{f_2}}}$

✧ *IMD* de troisième ordre : $IMD_3\left(endB\right) = 20log\frac{\sqrt{V^2{}_{(2f_1+f_2)} + V^2{}_{(2f_1-f_2)} + V^2{}_{(f_1+2f_2)} + V^2{}_{(f_1-2f_2)}}}{\sqrt{V^2{}_{f_1} + V^2{}_{f_2}}}$

✧ *IMD* totale : dans ce cas, toutes les composantes dans le spectre utile sont considérées.

Distorsion harmonique et distorsion par inter-modulation

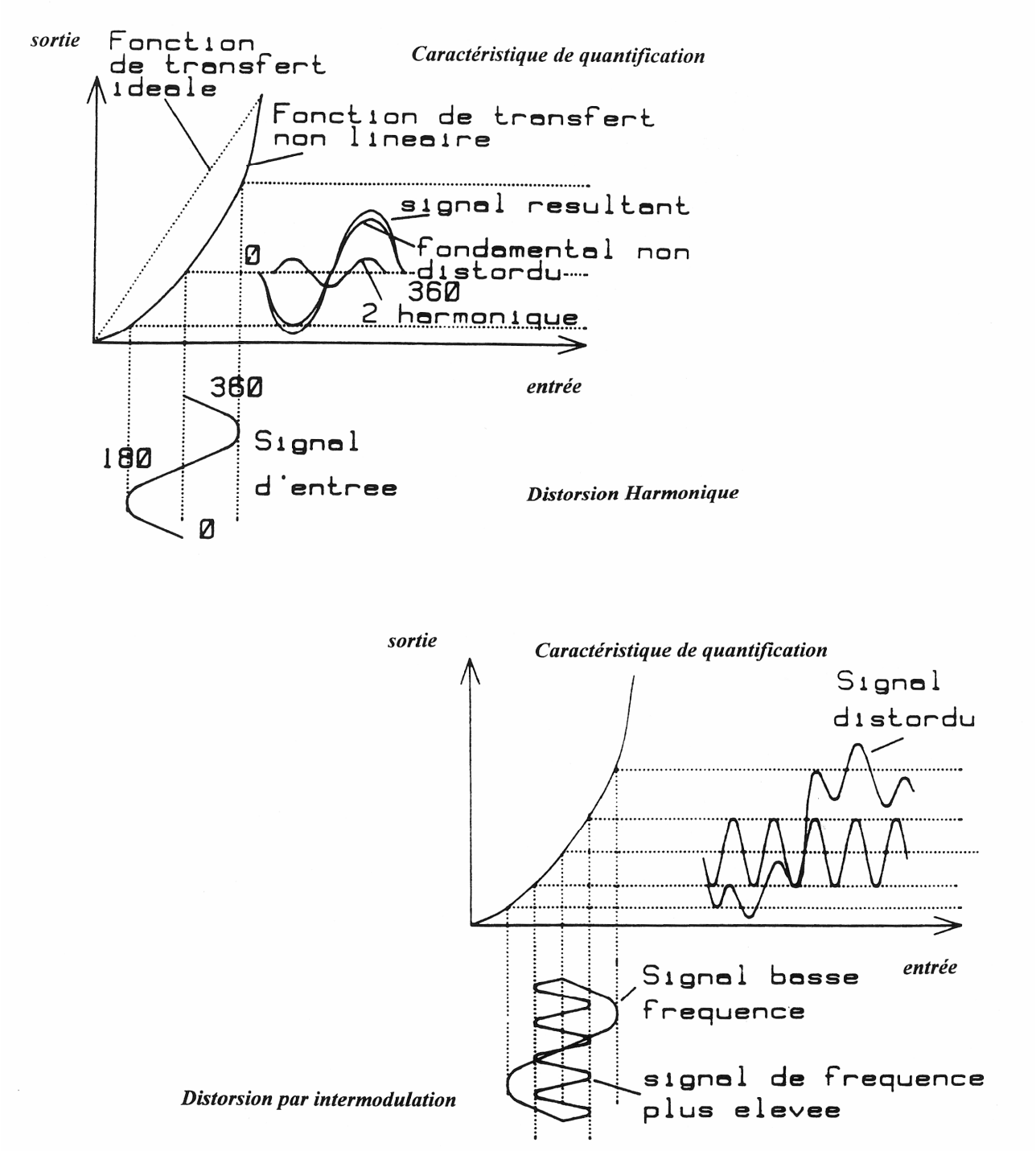


Figure n°48

• **Nombre réel de bits (*Effective Number Of Bits)* :**

On exprime la formule de la dynamique théorique, telle que :

$$n = \frac{S/B\,(en\ dB) - 1.76}{6.02},$$

L'*ENOB* sera calculée à partir de la mesure de la dynamique, pour une sinusoïde dont le spectre balaie la bande utile en entrée :

$$ENOB(f) = \frac{S/B(f)_{(en\ dB)\ mesuré} - 1.76}{6.02}$$

Pour les systèmes de type *ΣDPCM*, il faudra partir de la formulation théorique adaptée et procéder de façon analogue.

• **Linéarité intégrale en régime dynamique :** Lorsqu'une sinusoïde de fréquence donnée est appliquée en entrée de convertisseur et que plusieurs millions d'échantillons sont enregistrés, un histogramme visualisant la fréquence d'occurrence de chacun des 2^n codes générés par le convertisseur peut être tracé. Tout pic apparaissant dans cet histogramme est caractéristique d'un problème de non-linéarité.

On peut exprimer l'indice de non-linéarité de la manière suivante :

$$INL(i) = \left[\frac{V(i) - V(o)}{V(pe) - V(o)} \times 2^n \right] - i$$

Avec *INL*(*i*) l'indice de non-linéarité intégrale au code *i*, *V(pe)* et *V(o)* les estimations des valeurs pleine échelle et d'*offset* et avec *V(i)* l'estimation de la valeur pour le *i*ème code telle que :

$$V(i) = -A\cos\left(\frac{\pi \sum_{n=0}^{i} V(n)}{N} \right)$$

où *A* est l'amplitude maximum de la sinusoïde d'entrée, *N* le nombre d'échantillons considérés pour la réalisation de l'histogramme et *V(n)* l'estimation de la valeur pour le *n*ième code.

La mesure de linéarité peut aussi se faire de façon classique (rampe analogique en entrée, sortie analogique) mais superpose les problèmes de linéarité des deux convertisseurs A /N et N / A.

Reste dans ce cas à s'assurer que le convertisseur N /A de mesure est plus linéaire que le *CAN* sous test...

Remarque : la linéarité intégrale (déformations de la caractéristique de quantification idéale) peut se subdiviser en linéarité différentielle (égalité des pas de quantification) et en monotonie. La monotonie est un cas particulier de non-linéarité différentielle.

Une non-linéarité provoque, quoi qu'il en soit, de la distorsion...

Annexe :

Draft AES-11-19XX

STANDARDS AND RECOMMENDED PRACTICES

Call for Comment on DRAFT AES Recommended Practice for Digital Audio Engineering— Synchronization of Digital Audio Equipment in Studio Operations

This document was developed by a writing group of the Audio Engineering Society Standards Committee (AESSC). The draft is being concurrently submitted for approval by the accredited Standards Committee S4 on Audio Engineering affiliated with the American National Standards Institute, by the public via American National Standards Institute *Standards Action*, and by the membership of the Audio Engineering Society via this publication. Committee members from various countries will be submitting the draft via their respective national committees to the International Electrotechnical Commission.

There are no existing international standards covering the subject of this document. IEC and ISO documents were used and referenced throughout its development.

This Recommended Practice will be approved by the AES after any adverse comment received within 3 months of this publication has been resolved and any appeal has been settled.

Please send comments to Standards Secretariat, Audio Engineering Society, Inc., 60 East 42nd Street, New York, NY 10165, USA.

This document was balloted to ANSI in the form published herein. Subsequently, a decision was made by the AESSC to use ANSI-style mandatory wording (i.e., "shall") for all mandating sections in place of verbs "must" or "is." The final document shall reflect this decision.

AES11-19XX
(ANSI S4.44-19XX)

AES Recommended Practice for Digital Audio Engineering— Synchronization of Digital Audio Equipment in Studio Operations

Published by
Audio Engineering Society, Inc.

Abstract

This recommendation provides a systematic approach to the synchronization of digital audio signals. Recommendations are made concerning the accuracy of sample clocks as embodied in the interface signal and the use of this format as a convenient synchronization reference where signals must be rendered cotimed for digital processing. Two techniques, the use of Genlock or a Masterclock, are proposed. Synchronism is defined and limits are given which take account of relevant timing uncertainties encountered in an audio studio.

(ANSI 4.44-19XX)

Foreword

This foreword is not a part of the AES Recommended Practice for Digital Audio Engineering -- Synchronization of Digital Audio Equipment in Studio Operations, AES11-19XX (ANSI S4.44-19XX).]

This document provides operating standards and guidance for users needing to synchronize digital audio signals. This is an essential requirement in studios for the handling of remote program sources. The development of a working practice for this aspect of system engineering follows from the standardization of sampling frequency and the international agreement on the serial transmission format for the professional environment.

A working group was established in 1984 by the Subcommittee on Digital Audio of the AES Standards Committee to consider the topic with the possibility of formulating a policy on behalf of the industry. An approach was made to some 60 manufacturers of equipment to seek their advice and comment. Meetings were attended by engineers able to represent views from the SMPTE, EBU, and IEC. The final conclusions endorse the AES3-1985 and AES5-1984 standards and seek to address, primarily, the principles to be applied in synchronizing operations, thus allowing for future developments affecting digital audio systems.

During the course of this work the biannual meetings have been regularly supported by representatives of broadcasting, recording studios, and equipment manufacturers worldwide. A European-based subgroup met at various times for more detailed study. Input documents were provided by N. Gilchrist, A. Komly, R. Lagadec, P. Lidbetter, A. Weisser, J. Wilkinson, and the chairman. The following individuals have contributed to the work leading up to these proposals: K. Altmann, G. Barton, B. Blüthgen, S. Busby, R. Cabot, R. Caine, L. Fielder, R. Finger, N. Gilchrist, A. Griffiths, R. Hankinson, D. Hayes, K. Hoffner, T. Holman, Y. Ishida, A. Komly, R. Lagadec, P. Lidbetter, B. Locanthi, S. Lyman, G. McNally, J. Nunn, K. Tanaka, N. Taylor, D. Walstra, A. Weisser, J. Wilkinson.

W. T. SHELTON, *Chairman*,
AES Standards Committee Working Group on Synchronization
1989 May

Contents

AES Recommended Practice for Digital Audio Engineering— Synchronization of Digital Audio Equipment in Studio Operations

1. Scope

1.1 This document provides a recommended practice for manufacturers and users of digital audio equipment aimed at promoting economical and efficient methods for synchronizing interconnected digital audio equipment.

1.2 Synchronization of digital audio signals is a necessary function for the exchange of signals between equipments. The objective of synchronization is primarily to time-align sample clocks within digital audio signal sources. The recommendations address only essential aspects necessary for successful studio operation.

1.3 The recommendations make use of the two-channel digital audio interface standard for professional use (AES3-1985 [1], CCIR Rec. 647 [2]). It is expected that the recommendations shall be adopted for synchronizing all other digital audio interfaces.

1.4 The document addresses two groups of parameters. The first concerns the performance requirements for the successful interchange of digital audio data between equipments (Section 5). The second concerns the minimum performance requirements for the regeneration of clocks used for analog-to-digital and digital-to-analog conversion (Section 6).

2. Definitions

2.1 Synchronism. Identical frame frequencies for two digital audio signals (i.e., both signals have the same number of frames over a defined period of time). The successful interconnection of digital audio equipment further requires that the timing difference between two signals be maintained within a recommended timing tolerance on a sample-by-sample basis (defined in Section 5). Where the timing difference exceeds the recommended tolerance, one of the signals will need to be retimed, even though the frequencies are identical. Note that this definition specifically excludes alignment of the block-formatted data.

2.2 Frame. A sequence of two subframes, each carrying audio sample data for each of two channels, and transmitted in one sample period.

2.3 Timing Reference Point. For the purpose of this document, the initial transition of the form 2 preamble of the frame of a digital audio signal as specified in

3. Area of Application

Items of stand-alone digital audio equipment interconnected via analog inputs and outputs require no consideration in this document.

The primary area of application is the digital interconnection of digital audio equipment wholly contained within the studio environment. There is a further application in which signal sources and destinations external to the studio environment interface with the equipment within the studio environment.

3.1 Digital Interconnections within the Studio Environment. Digital audio equipment within a self-contained area, such as a studio or studio center, exchange digital signals with the timing from all items of equipment controlled.

3.2 Digital Interconnections Involving Sources and Destinations outside the Studio Environment. Digital interconnections involving equipment, either local or remote, with timing not under the control of the studio or studio center.

4. Modes of Operation

4.1 It is recommended that equipment provide the ability to lock an internal sample clock generator to a digital audio reference signal. It is advisable to provide a separate input socket reserved for the use of the digital audio reference signal.

4.2 It is recommended that equipment be synchronized by one of the two methods 4.2.1 or 4.2.2.

4.2.1 The use of the Digital Audio Reference Signal, which ensures that all input–output equipment sample clocks are locked to that reference. This method is preferred for normal studio practice.

4.2.2 The use of the embedded sample rate clock within a digital audio input signal, which may be program, that enables the input–output sample rate clock to be locked. This method may increase the timing error between items of equipment in a cascaded implementation.

4.3 The Digital Audio Reference Signal shall be distributed in a star configuration with a maximum of four receivers per spur circuit, in accordance with AES3-1985.

4.4 When connecting external signals to an otherwise

following two conditions will apply.

4.4.1 Where the incoming signal is identical in sample frequency, but is out of phase with the Digital Audio Reference Signal, digital audio frame alignment will be necessary.

4.4.2 Where the incoming signal is not identical in sample frequency, sample rate conversion will be necessary.

4.5 In the case of a combined video and audio environment, the source of the Digital Audio Reference Signal shall be locked to the video source so that the mathematical relationships given in Table 1 are obtained precisely.

5. Recommendations for Equipment Synchronization Practice

5.1 Digital Audio Reference Signal

5.1.1 The Digital Audio Reference Signal shall have the format and electrical configuration of the two-channel AES/EBU interface and use the same connector as given in AES3-1985. However, the basic structure of the AES/EBU format, where only the preamble is active, is acceptable as a digital audio synchronizing signal.

In vari-speed applications, or otherwise when the sample frequency does not conform with AES5-1984 [3] (CCIR 646 [4]), 5.2 does not apply.

5.1.2 The Digital Audio Reference Signal may be categorized as either grade 1 or grade 2, as follows and as detailed in 5.2.

5.1.2.1 A grade 1 reference signal is a high-accuracy signal intended for synchronizing systematically a multiple-studio complex and may also be used for a stand-alone studio.

5.1.2.2 A grade 2 reference signal is the recognized accuracy signal intended for synchronizing within a single studio, where there are no technical or economic benefits in working to grade 1 standards.

5.1.3 A digital audio reference signal, which has the prime purpose of studio synchronization, shall be identified as to its intended use by byte 4 bits 0, 1 of channel status as follows:

00 = default
01 = grade 1
10 = grade 2
11 = reserved for future use.

5.1.4 A digital audio reference signal shall be identified in channel status as nonaudio when it contains other data rendering it unusable as a normal audio signal.

5.2 Sample Frequency Tolerances in Equipment

5.2.1 Sample frequency tolerances in equipment are specified by the long-term frequency drift of internal oscillators when in free-running mode. This recommendation provides for two levels of equipment frequency tolerance, previously defined in 5.1 as grade 1 and grade 2.

5.2.1.1 Grade 1. A grade 1 reference signal shall maintain a long-term frequency accuracy within ± 1 ppm. Equipment designed to provide grade 1 reference signals shall only be required to lock to other grade 1 reference signals.[1]

5.2.1.2 Grade 2. The normal equipment free-running frequency tolerance shall be less than ± 10 ppm as specified in AES5-1984.

Frequency tolerances are necessarily a function of the environment in which the equipment is operated. Reference should therefore always be made to the manufacturer's recommended operating conditions.

5.2.2 The capture range of equipment oscillators designed to lock to external inputs should be a minimum of:

5.2.2.1 ± 2 ppm for grade 1 equipment.

5.2.2.2 ± 50 ppm for grade 2 equipment and other apparatus of lower performance.

5.3 Equipment Timing Relationships

5.3.1 The timing reference point (see 2.3) is used to define the timing relationship between the Digital Audio Reference Signal and digital audio input and output signals. An item of equipment is deemed to be synchronized when the following two conditions are met in both the static and the dynamic modes.

5.3.1.1 The difference between the timing reference points of the digital audio synchronizing signal and all output signals, at the equipment connector points, shall be less than ± 5% of the digital audio frame period.

5.3.1.2 The difference between the timing reference points of the digital audio synchronizing signal and all input signals, at the equipment connector points, shall be less than ± 25% of the digital audio frame period.

[1] Even where the high accuracy specified in 5.2.1.1 has been implemented for individual equipment sample clocks, if these are free running, synchronism between independent equipment (or other similarly independent processes such as film or video) is not maintained.

Table 1. Audio/video synchronization.

Sample rate (kHz)	Samples per TV/film frame			
	24 Hz motion picture	25 Hz PAL/SECAM	30 Hz 525 mono	30/1.001 NTSC
32	4000/3	1280	3200/3	16016/15
48	2000	1920	1600	8008/5
44.1	3675/2	1764	1470	147147/100

5.3.2 It is desirable that where a definable delay exceeding 1 digital audio frame period is introduced, between the input and the output of the equipment, the delay or range of delay should be specified in units of 1 frame. Where the delay time is known, the equipment connector panel shall be clearly labeled with the value of the delay.

5.3.3 Table 2 specifies tolerances in 5.3.1.1 and 5.3.1.2 as absolute values for the sample frequencies defined in AES5-1984.

5.4 System Practice. Good engineering practice requires that timing differences between signal paths be minimized to avoid timing errors accumulating with a risk of loss of synchronism.

The timing tolerance permitted for synchronous signals under the definition given in 2.1 is less than ±25% of the digital audio frame period. However, it is desirable that system synchronization be instrumented to the closest limits possible so that subsequent timing offsets have minimum effect.

5.4.1 Timing Differences

5.4.1.1 Instrumental Delays in Apparatus, Residual Errors in Phase-Lock Loops, and Cable Delay in the Transmission Path. Variations in delay from these causes may result from changes in the configuration of the audio system.

5.4.1.2 Clock Jitter. Jitter noise may be either random or in the form of modulation, which at frequencies less than sample rate will cause a timing error to accumulate according to the amplitude and frequency of the modulation waveform. It should be noted that the interface specification permits a jitter level, sample to sample, of ±20 ns, irrespective of its character.

6. Clock Specifications for Audio Sampling Clocks

6.1 Timing Precision. In order to obtain the best performance from analog-to-digital and digital-to-analog converters the sample conversion clock, when externally locked to the Digital Audio Reference Signal, is required to have increased timing accuracy above that specified for grade 1 and 2 signals in the areas of random jitter and jitter modulation.

6.2 Sample Clock Parameters

6.2.1 For the sample conversion clock, the following conditions shall be met, which are further explained in Fig. 1.

6.2.1.1 To limit clock modulation to insignificant levels of wow and flutter, the peak sample clock modulation, sample to sample, shall be less than ± 1 ns at all modulation frequencies above 40 Hz. This figure applies to conversion systems with a resolution of 16

Table 2. Synchronization of digital audio: limits.

Professional sampling frequency (kHz)	$\frac{1}{f_s}$	Synchronization Window (μs) ±25% (permitted variation)	±5% (in lock)
32	31.25	±7.8	±1.6
44.1	22.68	±5.7	±1.1
48	20.83	±5.2	±1.0

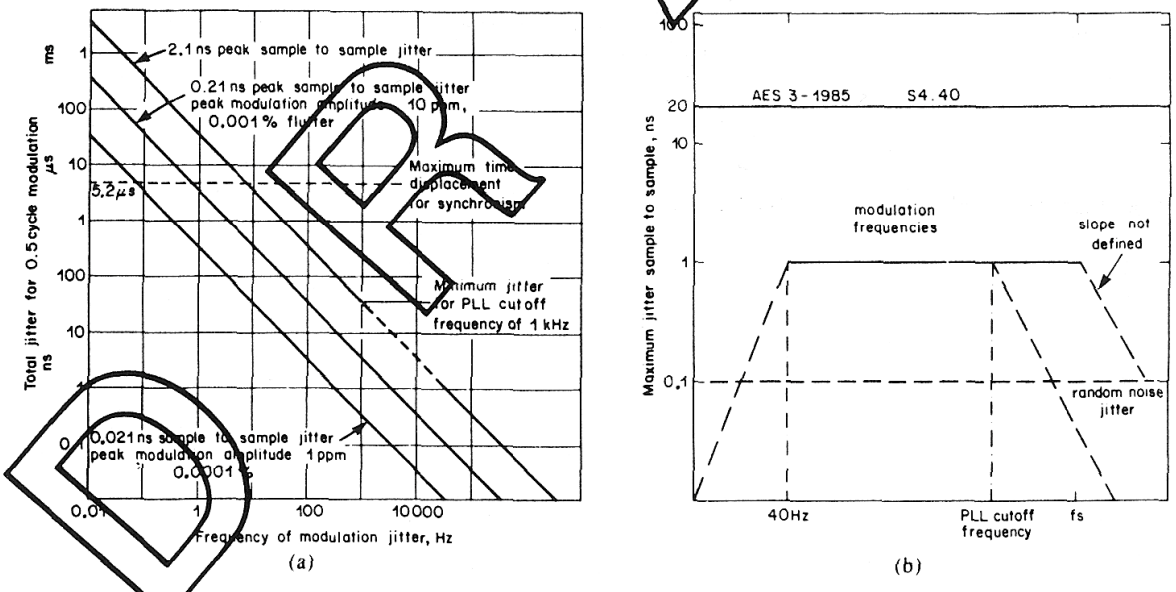


Fig. 1. (a) Effects of jitter modulation in digital audio sample clocks. (b) Jitter limits for sample clocks in analog-to-digital

bits or more.

6.2.1.2 Clock jitter, sample to sample, should ideally as a target be less than ±0.1 ns per sample clock period so as to provide a satisfactory limit to the noise modulation introduced into the program material. This figure refers to 16-bit-per-sample conversion systems and should be scaled appropriately for higher resolution systems.

7. References

[1] AES3-1985, "AES Recommended Practice for Digital Audio Engineering—Serial Transmission Format for Linearly Represented Digital Audio Data," *J. Audio Eng. Soc.* (*Standards and Information Documents*), vol. 33, pp. 975–984 (1985 Dec.).

[2] CCIR Recommendation 647, "A Digital Audio Interface for Broadcasting Studios," *Green Book*, vol. 10, pt. 1, CCIR, Dubrovnik (1986).

[3] AES5-1984, "AES Recommended Practice for Professional Digital Audio Applications Employing Pulse-Code Modulation—Preferred Sampling Frequencies," *J Audio Eng. Soc.* (*Standards and Recommended Practices*), vol. 32, pp. 781–785 (1984 Oct.).

[4] CCIR Recommendation 646, "Source Encoding for Digital Sound Signals in Broadcasting Studios," CCIR (1986).

Bibliographie

BIBLIOGRAPHIE :

• ***Robert W. Adams :*** *"An IC Chip for 20-bit A/D Conversion", J. Audio Eng. Soc., Vol 38, No 6, 1990 June.*

• ***Bob Adams :*** *"Beat the Clock", Studio Sound, August 1993.*

• ***J.Y. Bedu :***

"Spécification et critères de choix des CAN'S", Electronique Radio Plans N° 521/522.
"Les CAN'S Sigma-delta à coeur ouvert", Electronique Radio Plans N°525.
"Les CAN'S Delta Sigma", Electronique Radio Plans N° 526.

• ***P. Buser / M. Imbert :*** *Neurophysiologie Fonctionnelle III, Audition, Collection méthodes, Hermann Paris 1987.*

• ***Timothy F. Darling / Malcom O. J. Hawksford :*** *"Oversampled Anolog-to-Digital Conversion for Digital Audio Systems", J. Audio Eng. Soc., Vol. 38, No. 12, 1990 December.*

• ***Michael Gerzon :*** *"Measure for Measure", Studio Sound, 1992.*

• ***Steven Harris :*** *" How to Achieve Optimum Performance from Delta-Sigma A/D and D/A Converters", J. Audio Eng. Soc., Vol 41, No. 10, 1993 October.*

• ***Steven Harris :*** *"Measurements Techniques for Debugging ADC and DAC Systems", AES 11th International Conference.*

• ***Max W. Hauser :*** *"Principles of Oversampling A/D Conversion", J. Audio Eng. Soc., Vol. 39, No. 1/2, 1991 Juanuary/February.*

• ***Malcom Omar Hawksford :*** *Tutorial "An Introduction to digital audio", Image of audio, Proceedings of the 10th International Conference, 7th - 9th September 1991.*

• ***Stanley P. Lipshitz / John Vanderkooy, Robert A. Wannamaker :*** *"Minimally Audible Noise Shaping", J. Audio Eng. Soc.,Vol 39, No. 11, 1991 November. "Quantisation and Dither : A Theoretical Survey", J. Audio Eng. Soc., Vol 40, No. 5, 1992 May.*

• ***Robert C. Maher :*** *"On the Nature of Granulation Noise in Uniform Quantization Systems", J. Audio Eng. Soc., Vol 40, No. 1/2, 1992 January/February.*

• ***C. Morillon :*** *Polycopiés C.N.A.M. "Instrumentation scientifique B1 et B2".*

• ***Ken C. Pohlmann :*** *Advanced Digital Audio, SAMS 1991.*

• ***Francis Rumsey :***

Digital Audio Operations, Focal Press, 1991.
"Maintaining Digital Audio Quality", Studio Sound, February 1991.
"Digital Audio Synchronization", Studio Sound, March 1991.
"Digital Audio Problem Solvers", Studio Sound, July 1991.
"The Bit Bottleneck", Studio Sound, October 1992.
"The Bit Budget", Studio Sound, February 1993.
"The Pleasure of Measurement", Studio Sound, August 1993.

• ***David Smith :*** *"Sony Classical's Don Carlo", Studio Sound, December 1992.*

• ***Robert A. Wannamaker :*** *" Psychoacouysticaly Optimal Noise shaping", J. Audio Eng. Soc., Vol 40, No. 7/8, 1992 July/August.*

• ***John Watkinson :*** *The Art of Digital Audio, Focal Press, 1991.*

• ***AES Standard and Recommended Practices :*** *"Call for Comment on DRAFT AES Recommended : Practice for Digital Audio Engineering - Synchronization of Digital Audio Equipments in Studio Operations", AES 11-19XX (ANSI S4.44 -19XX).*

• ***AES Standard method for digital audio engineering -*** *Measurements of digital audio equipment. AES17-1991 (ANSI S4.51-1991).*

• ***Audio Precision :***

System one user's manual, September 1991.
DSP user's manual, October 1992.

• ***Audio Recording****, Studio Sound, August 1992 + Différents articles et critiques de convertisseurs.*

• ***Motorola Inc., 1990 :*** *"Principles of Sigma-Delta-modulation for Analog to Digital Converters".*