NEUVIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 16 au 20 MAI 1983

AMELIORATION DE LA PRECISION DE CALCUL DE LA FFT PAR QUANTIFICATION A REFERENCE STOCHASTIQUE

F. CASTANIE - D. WAN (*)

G.A.P.S.E. (Groupe d'Analyse des Processus Stochastiques en Electronique)
INP-ENSEEIHT, 2, rue Ch. Camichel, 31071 TOULOUSE Cedex

RESUME

La communication est consacrée aux performances du calcul de Transformée de Fourier Rapide lorsque la Quantification à Référence Stochastique est utilisée.

Après un bref rappel du principe et des propriétés cette méthode de quantification, on montre qu'elle conduit à superposer un bruit au calcul idéal. On compare ses performances en moyenne quadratique avec la quantification déterministe. On montre enfin que l'erreur au sens d'un intervalle de confiance est divisé sensiblement par 2. Les applications possibles sont évoquées.

Mots clés : Quantification, longueur de mot finie, Transformée de Fourier Rapide.

SUMMARY

The paper deals with the accuracy of FFT computation when Random Reference Quantization is used. After a brief survey of principle and properties of this quantization method, it is shown that it results in adding a noise to the ideal computation. The comparison in a RMS sense with the conventional method is achieved.

It is then shown that in a confidence interval sense, the error is divided nearly by two. Possible applications are examined.

Key words : Quantization, Finite wordlength, FFT.

(*) Ingénieur de l'Institut Microélectronique de SHENSI (Chine)

AMELIORATION DE LA PRECISION DE CALCUL DE LA FFT PAR QUANTIFICATION A REFERENCE STOCHASTIQUE

1- INTRODUCTION

Le principe de la Quantification à Référence Stochastique (QRS) (cas particulier simple de la Quantification Aleatoire (5) connu depuis fort longtemps) a, en général été appliqué à la conception d'estinateurs ou de calculateurs spécialisés [4] [6] , avec pour motivation essentielle de réduire la longueur de mot à l'entrée du système et dans le traitement, jusqu'à des valeurs très basses (typiquement de 4 à 1 bit) .

Appliqué aux quantifications internes des calculateurs numériques conventionnels, le principe de QRS devient une méthode d'arrondi des calculs intermédiaires, permettant le contrôle des longueurs de mot sans introduire -en un certain sens- d'effet non-linéaire (arrondis non biaisés). Ce qui est quelquefois appelé "Arrondi Aléatoire" dans la littérature est une version simplifiéeet partiellement biaisée de la QRS.

L' application du principe aux algorithmes de filtrage digital [7] a montré que certains effets non-linéaires disparaissent (cycles limites par exemple) et qu'il conduit en général à une réduction de la longueur de mot et /ou à une augmentation de la précision du calcul.

La présente communication résume les principaux résultats concernant l'application de la QRS au calcul de Transformée Discrète de Fourier par algorithme de FFT en base 2 [8] .

2- RAPPEL DES PRINCIPES DE LA QRS

Le concept de base de la quantification aléatoire réside dans l'élargissement de la notion de Quantificateur au cas où les paramètres d'entrée (seuils) de cet opérateur non linéaire sont des variables aléatoires [9] . La Quantification à Référence Stochastique est le cas particulier où tous les seuils se déplacent aléatoirement en conservant leurs écarts constants : seule la position de l'origine de la fonction de transfert est aléatoire, il est aisé de voir que le QRS est identique au Quantificateur Déterministe (QD) lorsque celui-ci reçoit à l'entrée la somme du Signal d'entrée et d'un bruit de référence, statistiquement parfaitement déterminé, noté source auxiliaire (SA).

La figure 1 donne le schéma de principe (qui est aussi le schéma pratique pour les interfaces d'entrée) et son application aux quantifications internes (arrondis)

Les grandeurs essentielles définissant un QRS pour la présente étude sont ses fonctions de transfert moyen d'ordre 1 et 2 $\lambda_1(x)$ et $\lambda_2(x)$ définis par

$$E[XQ] = E[\lambda_1(X)]$$
$$E[XQ^2] = E[\lambda_2(X)] \qquad (1)$$

On choisira dans les applications linéaires

$$\lambda_1(x) = x \qquad (2)$$

tout en minimisant l'accroissement de variance

$$\Delta\sigma^2 = Var(XQ) - Var(X) \qquad (3)$$

On montre que ceci conduit à choisir une S.A. équirépartie sur un pas de quantification Δ ; $\lambda_2(x)$ est alors une interpolée linéaire de la parapole x^2.

D'autre part, on a l'inégalité :

$$0 \leq \Delta\sigma^2 \leq \Delta^2/4 \qquad (4)$$

et l'approximation $\Delta\sigma^2 \sim \Delta^2/6$ (5)

Le"bruit" de quantification $\xi Q = XQ - X$ est strictement non corrélé avec le Signal X quel que soit Δ, lorsque Eq (2) est vérifiée. (ce n'est pas ici une approximation)

3 - APPLICATION AU CALCUL DE FFT EN VIRGULE FIXE

Le calcul d'une FFT en base 2 sur N points peut se résumer par la formule du papillon

$$R(X_i^m) = R(X_i^{m-1}) + W_R.R(X_j^{m-1}) - W_I(X_j^{m-1}) \quad a)$$
$$I(X_i^m) = I(X_i^{m-1}) + W_I.R(X_j^{m-1}) + W_R(X_j^{m-1}) \quad b) \qquad (6)$$

où - X_i^m est la valeur au noeud i (i = o, ...,N-1) de la colonne m (m = 1, ..., $\log_2 N$) du graphe de calcul de FFT

- R(.) et I(.) sont les parties Réelle et Imaginaire
- W_R et W_I sont les parties Réelle et Imaginaire du noyau choisi (implicitement fonction de i, j, m,)

Les indices i et j sont significatifs de l'algorithme choisi.

Si nous implantons l'Eq(6) en virgule fixe, le cheminement des calculs nécessite l'introduction d'Arrondis et d'un recalage avec arrondi . (figurés figure 2 pour l'Eq 6a) .

Avec le recalage habituel (facteur 1/2) on obtient une formulation qui permet d'évaluer la propagation des erreurs

$$R(x_i^m) = \frac{1}{2}\left[R(x_i^{m-1}) + W_R\,R(X_j^{m-1}) + \varepsilon_1 - W_I\,I(X_j^{m-1}) + \varepsilon_2\right] + \varepsilon_p \qquad (7)$$

où ε_1, ε_2, ε_p, sont les erreurs d'arrondi et recalage, avec une formulation identique pour $I(X_i^m)$.

Les propriétés de la QRS (accroissement de variance finie, non corrélation Signal-Bruit Cf §2) permettent de calculer la propagation de la variance de calcul en tout point.

$$Vi_R^m = 1/4\,(Vi_R^{m-1} + W_R^2\,V_{jR}^{m-1} + W_I^2\,V_{jI}^{m-1} + 2\sigma Q^2) + \sigma p^2 \qquad (8)$$

où $Vi_R^m = Var\,R(x_i^m)$, $V_{iI}^m = Var\,I(x_i^m)$, σQ^2 et σp^2

l'apport de variance dû aux opérateurs Q et P de la fig. 2, et une formulation semblable pour V_{iI}^m
Pour une FFT de type habituel $W_R^2 + W_I^2 = 1$
Si les Variances de rang 0 sont identiques $V_{iR}^o = V_{iI}^o$ (variances dues à l'interface d'entrée) il vient :

$$Vi_R^m = Vi_I^m = V^m \quad \forall i \qquad (9)$$

(i.e les variances le long d'une colonne sont égales)

AMELIORATION DE LA PRECISION DE CALCUL DE LA FFT PAR QUANTIFICATION
A REFERENCE STOCHASTIQUE

On obtient une équation de récurrence

$$V^m = \frac{1}{2} V^{m-1} + \frac{1}{2}.\sigma Q^2 + \sigma p^2 \qquad (10)$$

pour $m = M = \mathrm{Log}_2 N$, on obtient la variance en sortie

$$V^M = \frac{1}{2^M} V^0 + (\frac{1}{2}\sigma Q^2 + \sigma p^2).\, 2(1 - 2^{-M}) \qquad (11)$$

Avec les ordres de grandeurs habituels pour N,M, il vient

$$V^M \sim \frac{V^0}{N} + \sigma Q^2 + 2Q\sigma^2 \sim \sigma Q^2 + 2\sigma p^2 \qquad (12)$$

L'Eq (11) est valide dès lors que les sources de bruit sont non corrélées avec les signaux : C'est vrai pour la QRS, et en général considéré comme "approximativement" vrai pour la QD

On obtient ainsi [8]

QD: $\sigma Q^2 + 2\sigma p^2 \sim \Delta^2/3$ si l'on admet les approx. de SHEPPARD (13)

QRS: $\sigma_\varphi^2 + 2\sigma_P^2 \sim 5\Delta^2/12$ Si l'on admet l'Eq(5)

Les travaux classiques [1] [2] caractérisent l'erreur globale sous forme d'un rapport Signal/Bruit γ (valeurs RMS)

Si l'on compare ces deux rapports pour QRS et QD, il vient $\frac{\gamma QD}{\gamma QRS} = 1{,}12$ (14)

ceci montre que la FFT-QRS est légèrement plus "bruitée" que la FFT-QD

Cependant, cette analyse est très simpliste. En effet, considérons deux calculs de FFT-QRS et -QD sur un même signal (sur des mots très courts, pour visualiser les phénomènes. (Cf figures 3a) et 3b)

On voit tout d'abord que la QD fait apparaitre le caractère déterministe de son erreur (raies parasites symétriques) tandis que la QRS produit un "vrai" bruit de calcul.

La première amélioration que suggère cette constatation, est d'effectuer la moyenne des parties symétriques du spectre : le résultat est inchangé pour la QD, et la variance de la QRS est divisé par 2

Au prix de cet accroissement peu important de calcul, on obtient $\frac{\gamma QD}{\gamma QRS} \sim 0.8$ (15)

Mais il est plus important de remarquer que la distribution des erreurs est tout à fait différente : elle est sensiblement gaussienne pour la QD, et proche d'une distribution uniforme pour la QRS.

La mesure quadratique d'une dispersion n'est donc pas significative

La fig. 4 donne la fonction répartition des erreurs QD et QRS sur le signal réel [8]

On remarque qu'au sens d'un intervalle de confiance de niveau de confiance α (noté $\Delta(\alpha)$) proche de l'unité

$$\Delta(\alpha)_{QRS} \ll \Delta(\alpha)_{QD} \qquad (16)$$

Par exemple pour $\alpha = 0.95$

$$\Delta(0.95)_{QRS} \sim \frac{1}{2}\Delta(0.95)_{QD}$$

Les applications à des calculs sur signaux réel ont montré 8 que la "dynamique" pratique de la FFT QRS est bien de 3dB supérieur à la FFT-QD

La figure 5 résume cette propriété en donnant le rapport $\frac{\delta\Delta}{\Delta QD} = \frac{\Delta QD - \Delta QRS}{\Delta QD}$ expérimental pour $\alpha = 0.9$ et ~ 1, et pour différentes longueurs de mot ; on voit que l'écart entre les dispersions passe d'environ 0.5 pour des mots courts à des valeurs qui tendent vers 0 pour des mots longs.

Influence de la Quantification d'entrée

L'Eq (11) montre que V^0 a peu d'effet sur la variance de sortie.

Il est possible de montrer que la quantification d'entrée peut être effectuée sur des mots plus courts que ceux utilisés dans le calcul ; ce "raccourcissement" peut être de l'ordre de $\frac{1}{2}\mathrm{Log}_2 N$ (N nombre d'échantillons) sans augmenter notablement le bruit en sortie de calcul ; pour N=1024, l'interface d'entrée peut donc présenter des mots de 5 à 6 bits plus courts que ceux de l'Unité de calcul

4 - CONCLUSIONS

L'application de la QRS au calcul de FFT apporte en principe une augmentation de précision dont l'ordre de grandeur est 2. Cette amélioration est obtenue au prix du remplacement de la quantification d'entrée et de l'arrondi classiques par leurs versions Stochastiques. Celles-ci nécessitant une Source Auxiliaire et une opération binaire élémentaire supplémentaire (cf fig. 1), il est pertinent de s'interroger sur l'intérêt pratique du principe.

Si l'on excepte les cas particuliers de performances extrêmes en QD que l'on essaie d'améliorer encore, l'argument de l'augmentation de précision ne justifiera sans doute pas l'implantation pratique d'une telle méthode.

Par contre, la nature de l'erreur de calcul est réellement aléatoire à moyenne nulle pour la QRS, et peut permettre l'amélioration sensible du *traitement* du spectre; dans le courant de cette communication nous avons utilisé cette propriété pour "moyenner" les deux parties symétriques du spectre, pour reduire le bruit de calcul.

Mais il est évident que dans de nombreux problèmes la nature aléatoire de l'erreur est un avantage. Citons par exemple :-possibilité de moyenner des calculs successifs sur le même signal et d'augmenter autant qu'on le veut la precision

- levé de doute entre raies faibles et artefacts liés au calcul (qui apparaissent en QD)
- traitement adapté de formes spectrales, considérées en QRS comme des formes altérées par du "bruit blanc", et non déformées d'une façon inconnue déterministe (comme en QD).

Enfin, pour le concepteur d'une machine spécialisée la QRS offre un degré de liberté supplémentaire: La précision globale sera la même pour 1 calcul sur q bits que pour une moyenne de 2^{q-n} calculs sur n bits (n < q)

BIBLIOGRAPHIE

1 WELCH P.D. : "A fixed Point Fast Fourier Transform Error Analysis" IEEE Trans. on A.E., AU-17 n° 2 Sept. 1969

2 OPPENHEIM A.V. WEINSTEIN C.J. : "Effects of finite Register Length in Digital Filtering and Fast Fourier Transform". Proc. IEEE, 60, n°8, Aug. 1972

3 F. CASTANIE, J.C. HOFFMANN : "Transformateur stochastique de Fourier" Proc. of Adv. Sign. Proc. Tech. Journées d'Etudes de l'EPFL - Lausanne Oct. 1975

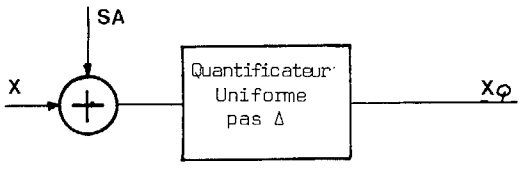
4 B. CHABERT, J. MAX, VIDAL-MADJAR : "Application à l'étude de l'ionosphère de l'optimisation du codage dans le calcul des fonctions de corrélation" Proc. 5e Colloque GRETSI Nice 1975.

5 F. CASTANIE : "Signal Processing by Random Reference Quantizing" Signal Processing 1, 1, pp. 27-43.

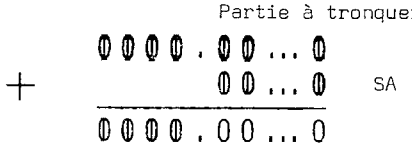
6 F. CASTANIE : "Stochasting Computing : A bridge between Signal Processing and Digital Computing" pp. 467-482, Proc. of EUSIPCO 80. North Holland pub.

7 D. DUBE : "La Quantification à Référence stochastique Appliquée au filtrage Numérique" Thèse Doct. Ing. Toulouse, FEV. 81

8 D. WAN : "La Quantification à Référence Stochastique : Application au Calcul Rapide de Transformée de Fourier" Thèse Docteur Ing. Toulouse Nov.82



1a) Schéma de principe et implantation analogique



0 = variable binaire

1b) Application à l'arrondi Stochastique

Figure 1

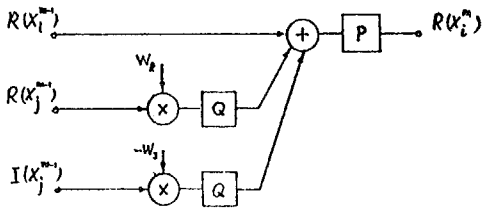


Figure 2 Opérateur arithmétique de papillon élémentaire.

Q : arrondi
P : recalage + arrondi

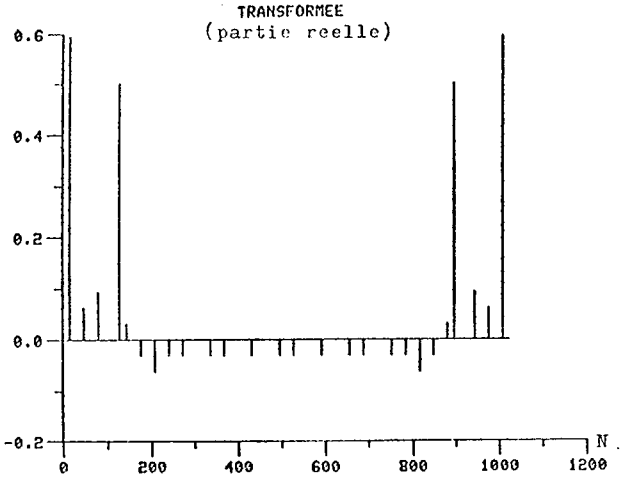


Figure 3a) FFT avec Arrondi Déterministe
b = 4 bits
N = 1024

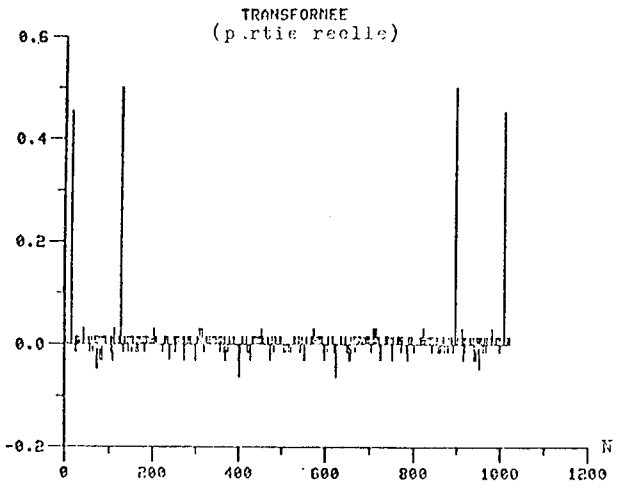


Figure 3b) FFT avec Arrondi Stochastique
b = 4 bits
N = 1024

AMELIORATION DE LA PRECISION DE CALCUL DE LA FFT PAR QUANTIFICATION
A REFERENCE STOCHASTIQUE

Fonction de répartition des erreurs de calcul FFT (Signal entrée = bruit blanc)

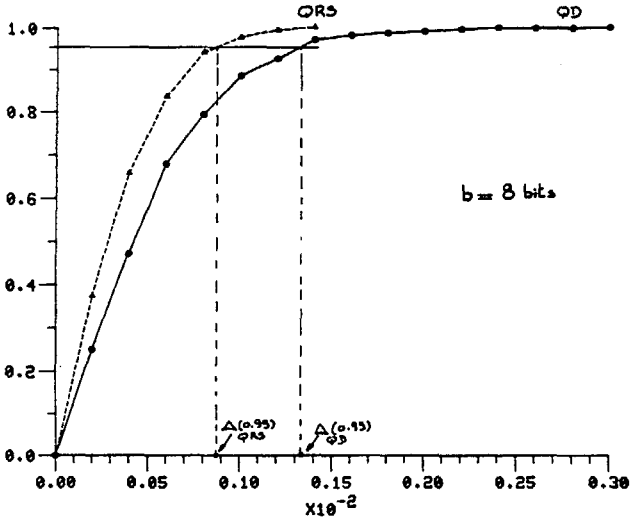


Figure 4a)

a = longueur de mots 8 bits

"Gain" relatif en intervalle de dispersion en fonction de la longueur de mots

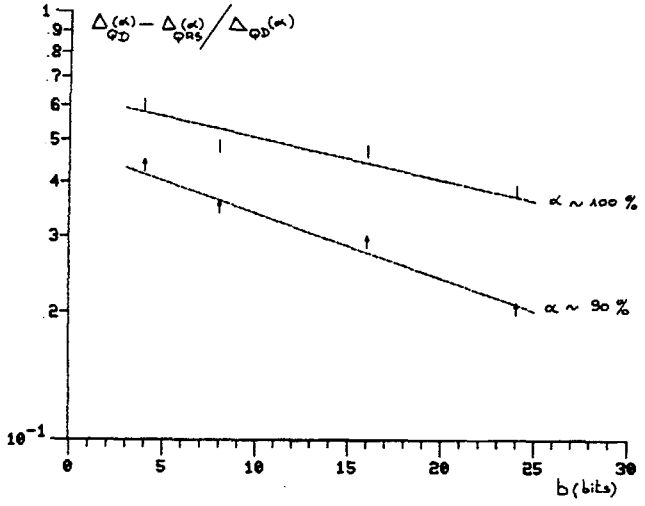


Figure 5

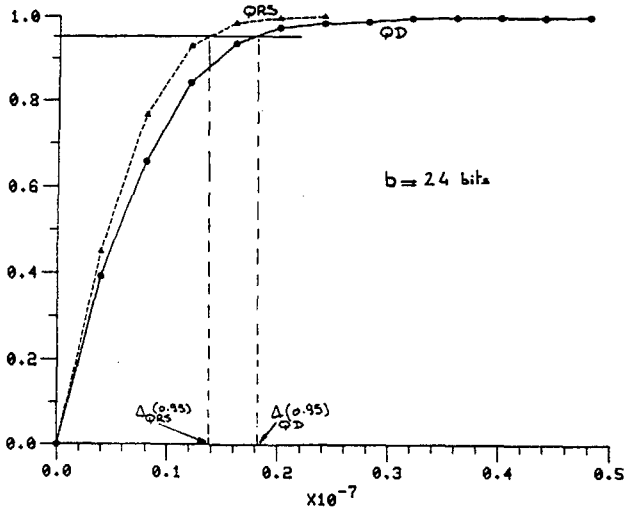


Figure 4b)

b = longueur de mots 24 bits