HUITIEME COLLOQUE SUR LE TRAITEMENT DU
SIGNAL ET SES APPLICATIONS

NICE du 1er au 5 JUIN 1981

Implantation de filtres digitaux récursifs à deux dimensions

FINITE REGISTER LENGTH EFFECTS IN THE IMPLEMENTATION OF TWO-DIMENSIONAL RECURSIVE DIGITAL FILTERS

A.N. Venetsanopoulos and S.H. Mneney

Department of Electrical Engineering
University of Toronto, Toronto, Canada

RESUME

Les éffets de la précision limitée sont considérés pour les filtres digitaux récursifs à deux dimensions. Des erreurs sont introduites dans la quantification de l'entrée, dans les coefficients et dans l'évaluation des multiplications des implantations de filtres. Des modèles de filtres à longueur de registres fini sont développés et différentes formes de réalisation sont montrées. Une approche statistique est adoptée, ce qui permet de calculer la distribution d'erreurs à la sortie de filtres. L'arrondissement et la coupure sont comparés comme méthodes de quantification. Différentes formes d'arithmétique à point fixe sont considérées et quelques examples sont donnés.

SUMMARY

The effects of finite precision are considered in two-dimensional recursive digital filters. Errors are introduced in quantizing the input, the coefficients and in the evaluation of the multiplications of the filter implementations. Finite register length models of the filters are developed and block diagrams are given for various forms of realization. A statistical approach is utilized, which allows us to compute the statistics of the error at the output of the filters. Rounding and truncation are compared, as methods of quantization. Various forms of fixed point arithmetic are considered and some examples are given.

* Les recherches rapportées dans cet article ont été supportées par une subvention du Conseil de Recherches en Sciences Naturelles et en Génie Canada (Subvention A7393).

* The research reported in this paper was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada under Grant A 7397.

Implantation de filtres digitaux récursifs à deux dimensions

FINITE REGISTER LENGTH EFFECTS IN THE IMPLEMENTATION OF TWO-DIMENSIONAL RECURSIVE DIGITAL FILTERS

I. INTRODUCTION

The memory requirement and speed of implementation of two-dimensional digital filters depends, to a large extent, on the register length used. Long register length filters produce more accurate results but have a longer implementation time and require more memory. Short register length filters, on the other hand, have lower running cost, but may result in an intolerable amount of error [1]. Errors are caused through input quantization and by rounding or truncation of signal values and coefficients to fit the register length exactly. In this paper an analysis is presented, which allows us to determine the statistics of such errors at the filter output, for different methods of realization. From such statistics the mean square error, the signal-to-noise ratio or error bounds can be deduced. The ensuing analysis is based on recursive implementation, which is known to be more efficient in terms of memory requirement and speed of implementation than its counterparts (direct convolution and FFT implementation [2]).

Quantization error bounds have been developed in [3], for direct form two-dimensional digital filters. The accumulation roundoff error was analysed for a special direct form filter in [4], in which an error bound was obtained. However previous published work has not taken into account the effects of coefficient and input quantization to the output signal and was not concerned with error models that are easily extended to compute the error statistics at the output of higher order filters. The designer is usually interested in an efficient method to predict the filter quality in terms of known filter parameters or a method that will allow choice of the parameters for a desired performance. To our knowledge such considerations are not found in the literature and very little exists in the analysis of quantization errors in two-dimensional digital filters. In this paper fixed point arithmetic is used and the different forms, in which the negative numbers are represented, are considered. Distinction between rounding and truncation is made. Different forms of realization of the two-dimensional digital filters are studied.

The statistical properties of the quantization errors are discussed in Section II. The three different types of errors are separately treated in Sections III, IV and V. Simulation results are given in Section VI.

II. PROPERTIES OF QUANTIZATION ERRORS

To simplify the analysis some assumptions are made. These assumptions and their limitations are discussed below:

1) The sequence of error samples is a sample sequence of a stationary random process.
2) The quantization process is a white noise process. The random variables representing the error process are uncorrelated, independent of the sampling rate.
3) The error sequence is uncorrelated with the sequence of exact samples. This is intuitively true if the amplitude of the exact samples varies in a random fashion, but is generally not a good assumption for one's complement and two's complement representations.
4) The quantization error has a uniform density function. This implies that the signal is equally likely to be anywhere within a quantization interval.

With these assumptions, it follows that signals being discussed exclude impulse, step or sinusoidal signals. Images and geophysical signals are examples of signals with the previous properties.

Quantization errors are caused by either truncation or rounding each mode resulting in a different error effect. Rounding is defined as follows. If the wordlength is b_1+1 bits and the available register length is b_2+1 bits add 1 to the b_2+1 bit if the b_2+2 bit is 1 and take no action otherwise. In the case of truncation of the b_2+1 most significant bits are taken from the b_1+1 bits and the rest are omitted.

As in [1] fixed point arithmetic involves binary numbers with a fixed binary point. Bits to the left of the binary point represent integers and those to the right represent fractions. Negative numbers are represented in any of three forms of fixed point arithmetic representation. These forms, which are defined below, result in different error effects at the filter output.

1) Sign-and-magnitude representation; the leading binary digit represents the sign. 0 represents + and 1 represents - . The remaining b of the b+1 bits represent the magnitude which is a fraction.
2) One's complement representation; positive numbers are represented as in sign-and-magnitude representation. Negative numbers are obtained by subtracting the magnitude from $2-2^{-b}$, where b+1 is the register length.
3) Two's complement representation; positive numbers are represented as in sign-and magnitude representation and negative numbers are obtained by subtracting the magnitude from 2.

As shown in [5], each of these arithmetic representations has its own error characteristics. On the basis of the three assumptions made and for the types of arithmetic representation the quantization error properties are easily deduced and summarized in Table I.

III. ROUNDOFF ACCUMULATION ERROR

A. Direct Form Realization (Fig. 1a)

Consider an infinite register length filter, with the quantized input represented by $\bar{x}_{nm}$. Its output y_{nm} is given by (1).

$$y_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} a_{ij}\bar{x}_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} b_{ij}y_{n-i,m-j} \qquad (1)$$

If q denotes a quantity quantized after multiplication, the output $\bar{y}_{nm}$ from the finite length register filter is given by

$$\bar{y}_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} (a_{ij}\bar{x}_{n-i,m-j})_q - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} (b_{ij}\bar{y}_{n-i,m-i})_q \qquad (2)$$

Let f_{nm} be the error at the output

$$f_{nm} \triangleq y_{nm} - \bar{y}_{nm} \qquad (3)$$

From (1)-(3) and Table I the statistics of the error at the output of a direct realization are deduced and summarized in Table II. α and β correspond to the number of coefficients in the numerator and the denominator of the transfer function $H[z_1,z_2]$, of the two-dimensional digital filter, which are neither 0 nor 1. h_{ij} corresponds to the unit pulse response of the filter with transfer function $\frac{1}{D(z_1, z_2)}$

B. Parallel Form Realization (Fig. 1b)

While general two-dimensional filters are not factorable some recent design techniques have considered special classes of filters with a denominator of the transfer function factorable into lower order polynomials [6,7]. It is sometimes possible to express the transfer function as a sum of partial fractions and realize the filters by parallel realizations. In this analysis a second order filter was used as a basic building block.

The statistics of the filter output are shown in Table III.

C. Cascade Realization (Fig. 1c)

If the basic building blocks are cascaded the statistics of the error at the output of a cacade realization are summarized on Table IV, where $g^{i}_{k\ell}$ is the unit pulse response of the filter after the ith noise.

IV. ERRORS DUE TO COEFFICIENT QUANTIZATION

A. Direct Form Realization (Fig. 2a)

With infinite precision arithmetic and infinite length registers the filtering process is represented by the recursion relation

$$\omega_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} a_{ij} x_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} b_{ij}\omega_{n-i,m-j} \qquad (4)$$

With finite length registers the implementation of the filtering process introduces errors due to the quantization of the coefficients, which manifest as a deviation of the transfer function from the ideal one. Let the result of quantizing the coefficients a_{ij} and b_{ij} be

$$\bar{a}_{ij} = a_{ij} - \mu_{ij}$$
$$\bar{b}_{ij} = b_{ij} - \eta_{ij} \qquad (5)$$

The finite precision recursion equation then becomes

$$y_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} \bar{a}_{ij} x_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} \bar{b}_{ij} y_{n-i,m-j} \qquad (6)$$

A multiplication roundoff term has not been added, as this has already been taken care of in Section III. At the filter output the error due to coefficient quantization is given by

$$f_{nm} = \omega_{nm} - y_{nm} \qquad (7)$$

The assumptions made in Section II also hold for the coefficient quantization error. The validity of these assumptions is not as good. This is because for low order filters the coefficients are few and the errors occupy only a few points on the quantization interval. In this section the input x_{nm} is assumed to be zero-mean and wide sense stationary (w.s.s). It is easy to show that in such a case, the output y_{nm} of a linear time invariant system is also zero-mean and w.s.s. [8]. The statistics of the error at the output as summarized in Table V. In this table

$\sigma_x^2 \triangleq$ variance of input signal

$\sigma_w^2 \triangleq$ variance of output signal $= \sigma_x^2 \sum_{i=0}^{N_A}\sum_{\substack{j=0\\ i+j\neq 0}}^{M_A} \frac{a_{ij}^2}{\Delta}$

$\Delta \triangleq \sum_{i=0}^{N_B}\sum_{j=0}^{M_B} b_{ij}^2 \quad (i+j\neq 0)$

B. Parallel Form Realization (Fig. 2b)

The basic building block is now the two-dimensional direct form realized filter with $M_A=N_A=M_B=N_B=1$. The error statistics are summarized in Table VI. α_k and β_k correspond to the number of coefficients in the numerator and the denominator of the k^{th} component, of the transfer function, in partial fraction form, that are not equal to zero or one.

c. Cascade Realization (Fig. 2c)

The basic building block is similar to that used in parallel form realization. The statistics of these errors are omitted here.

V. INPUT QUANTIZATION ERROR

Quantization of the input occurs at the analog to digital converter. Further quantization may occur if the filter register length is shorter than that of the quantizer. The quantization error e_{nm} is given by

$$e_{nm} = x_{nm} - \bar{x}_{nm}$$

where x_{nm} is the input before quantization and $\bar{x}_{nm}$ is the quantized input. The error e_{nm} has the properties discussed in Section II and the statistics of the error are shown in Table I.

A. Direct Form Realization (Fig. 3a)

When the input is unquantized and the effects of coefficient and accumulation roundoff errors are neglected the output is given by

$$\omega_{nm} = \sum_{i=0}^{M_N}\sum_{j=0}^{M_A} a_{ij} x_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} b_{ij}\omega_{n-i,m-j} \qquad (8)$$

when the input is quantized, the output becomes

$$y_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} a_{ij} \bar{x}_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} b_{ij} y_{n-i,m-j} \qquad (9)$$

The error at the output of the filter is given by

$$f_{nm} \triangleq \omega_{nm} - y_{nm} = \sum_{i=0}^{N_A}\sum_{j=0}^{M_A} a_{ij} e_{n-i,m-j} - \sum_{\substack{i=0\\ i+j\neq 0}}^{N_B}\sum_{j=0}^{M_B} b_{ij} f_{n-i,m-j} \qquad (10)$$

The noise statistics at the output are summarized in Table VII.

B. Paralell Realization (Fig. 3b)

The basic building block is similar to the direct realized filter with $M_A=N_A=N_B=M_B=1$. The error statistics at the output are given in Table VIII.

C. Cascade Realization (Fig. 3c)

When the basic building blocks are cascaded the result can be obtained in a similar fashion. The statistics of these errors are omitted here.

IV. SIMULATED RESULTS AND APPLICATIONS

Error expressions for the three modes of quantization were obtained for the three basic realization structures. The variance of the truncation error was computed theoretically. Similar results were also obtained through simulation for particular filters. These are compared in Figs. 4 and 5, where the error variance at the filter output is plotted against the

register length. The close agreement between the theoretical and the simulated results indicates the validity of the error models developed. An example is shown in Fig. 6.

VI. CONCLUSIONS

The effects of finite precision in the implementation of two-dimensional recursive digital filters were analyzed. The statistical properties of the quantization errors, at the output of the filters were deduced. These properties include the mean and the variance, from which the mean square error, the signal-to-noise ratio as well as error bounds can be computed. Tables II to VIII give such properties for different forms of realizations, when fixed point arithmetic is used with its different representations. The error variance and the statistical mean, introduced by rounding, are the same for the various representations of fixed point arithmetic, for each realization. Truncation results in a larger error variance in sign-and-magnitude and one's complement representation. The error variance for two's complement representation is the same in rounding and truncation.

Simulated and theoretical values of the variance were plotted against the register length used. The close agreement indicates the validity of the error models developed. For each realization the variance depends on the register length. It is therefore possible to design a filter with a pre-determined error-variance, for a required register length on the basis of such diagrams.

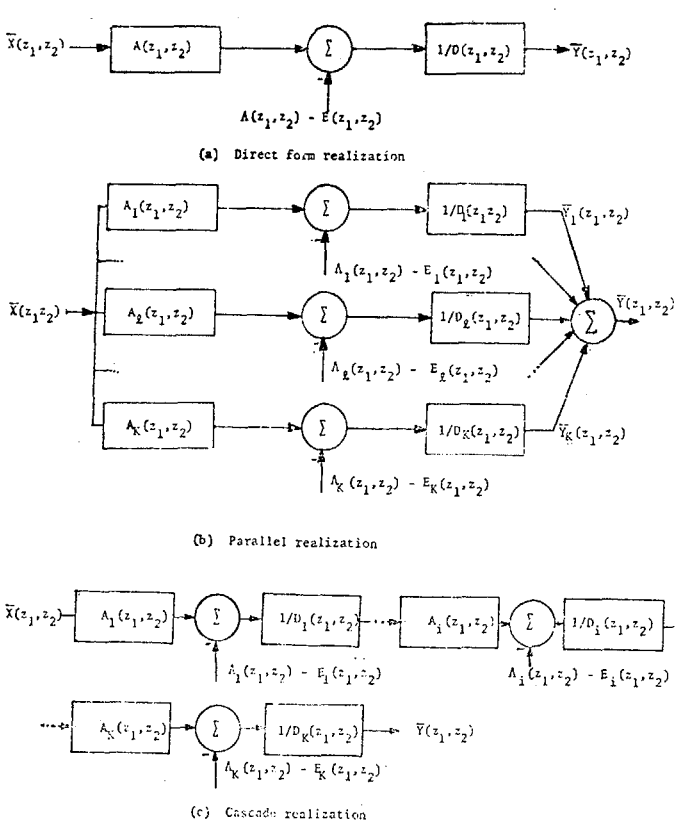

Figure 1. Block diagrams for recursive filtering including roundoff accumulated error.

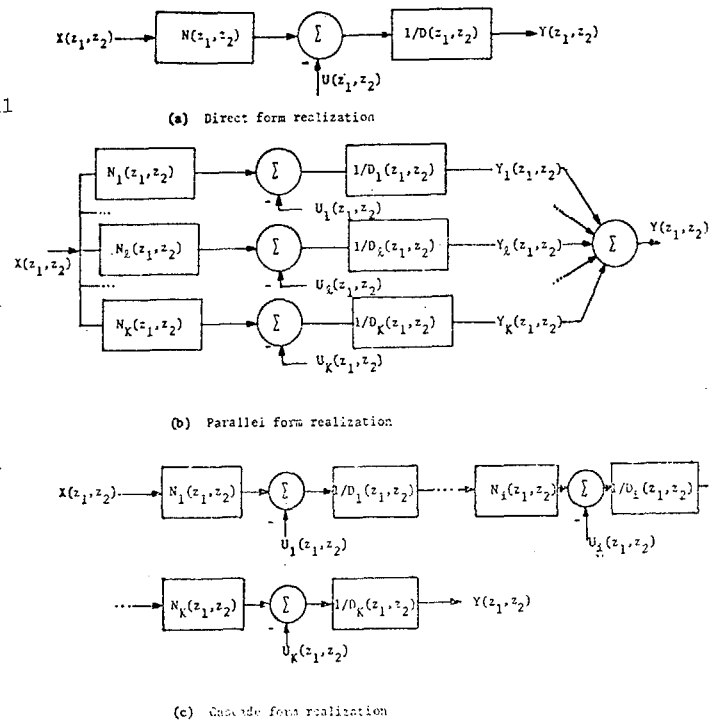

Figure 2. Block diagram representation for recursive filtering including coefficient quantization error.

REFERENCES

[1] B. Liu, "Effects of finite-word length on the accuracy of digital filters - a review", IEEE Trans. Circuit Theory, vol. CT-18, pp.670-677, Nov. 1971.

[2] E.L. Hall, "A comparison of computations for spatial frequency filtering", Proc.IEEE, vol.60, pp. 887-891, July 1972.

[3] Ming-Duenn Ni and J.K. Aggarwal, "Two-Dimensional digital filtering and its error analysis", IEEE Transactions on Computers, vol. C-23, no.9, September 1974.

[4] G.A. Maria and M.M. Fahmy, "Bounds for the amplitude of quantization error in forced 1st order two-dimensional digital filters", Circuit Theory and Applications, vol. 6, 221-233 (1977).

[5] A.N. Oppenheim and R.W. Schafer, "Digital Signal Processing", Prentice Hall, 1975.

[6] J.M. Costa and A.N. Venetsanopoulos, "Design of circularly symmetric two-dimensional recursive filters", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-26, no.4, pp. 290-304, August 1978.

[7] D.J. Goodman, "A design technique for circularly symmetric low-pass filters", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-26, no.4, pp. 290-304, August 1978.

[8] A. Papoulis, "Probability, Random Variables and Stochastic Processes", McGraw-Hill Book Company, 1965.

[9] A.N. Oppenheim and C.J. Weinstein, "Effects of finite register in digital filtering and the Fast-Fourier transform", Proc. IEEE, vol. 60, pp. 957-976, Aug. 1972.

[10] W.B. Davernport, "Probability and Random Processes", McGraw-Hill Book Company, pp. 208-267.

[11] L.R. Rabiner and B. Gold, "Theory and Applications of Digital Processing", Prentice Hall,1975.

[12] J.M. Costa and A.N. Venetsanopoulos, "Recursive implementation of factorable two-dimensional digital filters", Canadian Electrical Engineering Journal, vol.4, no.3,pp.33-40,July 1979.

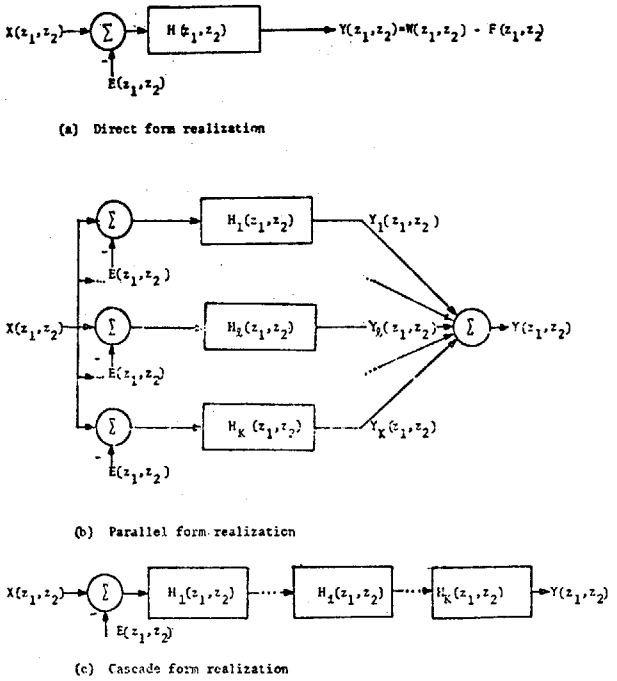

Figure 3. Block diagram for recursive filtering including input quantization error.

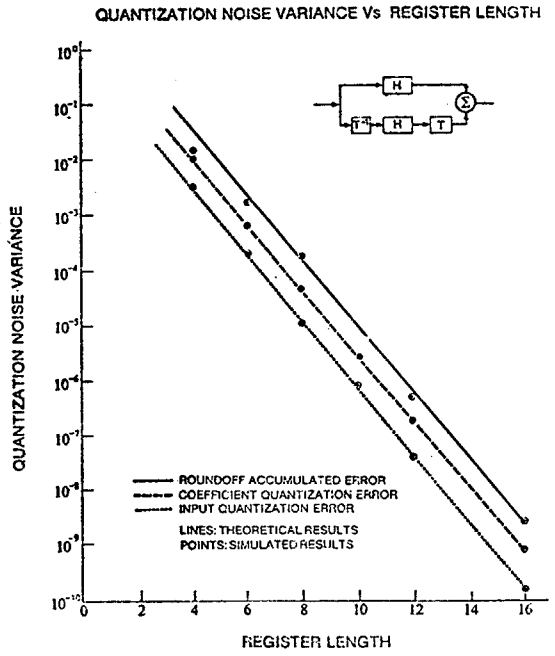

Figure 5

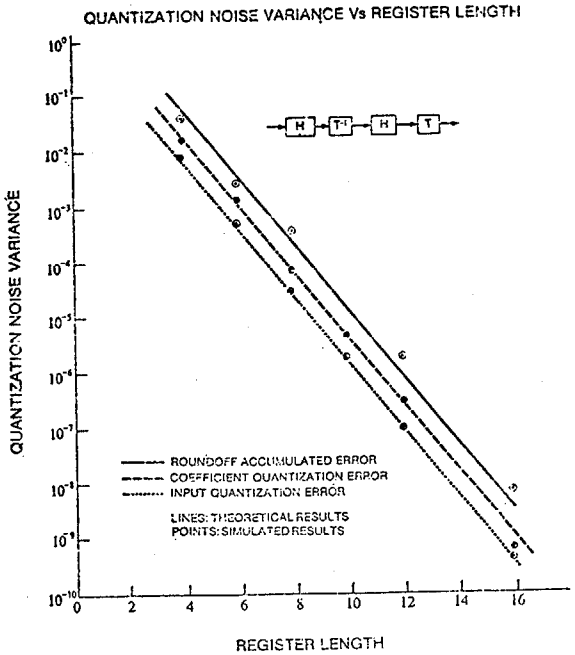

Figure 4

TABLE I

STATICS OF THE QUANTIZATION ERROR

Representation		Rounding	Truncation
Sign-And-Magnitude	$\bar{e}$	0	0
	σ^2_e	$q^2/12$	$q^2/3$
One's Complement	$\bar{e}$	0	0
	σ^2_e	$q^2/12$	$q^2/3$
Two's Complement	$\bar{e}$	0	$q/2$
	σ^2_e	$q^2/12$	$q^2/12$

A. ROUNDOFF ACCUMULATION ERROR

TABLE II

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A DIRECT FORM FILTER

Representation		Rounding	Truncation
Sign-And Magnitude	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$	$\frac{q^3}{3}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$
One's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$	$\frac{q^2}{3}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$
Two's Complement	$\bar{f}$	0	$\frac{q}{2}(\alpha-\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h_{ij}$
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=0}^{N}\sum_{j=0}^{M} h^2_{ij}$

Implantation de filtres digitaux récursifs à deux dimensions

FINITE REGISTER LENGTH EFFECTS IN THE IMPLEMENTATION OF TWO-DIMENSIONAL RECURSIVE DIGITAL FILTERS

TABLE III

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF PARALLEL REALIZED FILTER.

Representation		Rounding	Truncation
Sign-And Magnitude	$\bar{e}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$	$\frac{q^2}{3}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$
One's Complement	$\bar{e}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$	$\frac{q^2}{3}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$
Two's Complement	$\bar{e}$	0	$\frac{q}{2}(\alpha-\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$	$\frac{q^2}{12}(\alpha+\beta)\sum_{k=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^k_{ij})^2]$

TABLE IV

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A CASCADE REALIZED FILTER

Representation		Rounding	Truncation
Sign-And Magnitude	$\bar{e}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$	$\frac{q^2}{3}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$
One's Complement	$\bar{e}$	0	0
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$	$\frac{q^2}{3}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$
Two's Complement	$\bar{e}$	0	$\frac{q}{2}(\alpha-\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})]$
	σ^2_f	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$	$\frac{q^2}{12}(\alpha+\beta)\sum_{i=1}^{K}[\sum_{k=0}^{N}\sum_{\ell=0}^{M}(g^i_{k\ell})^2]$

B. ERRORS DUE TO COEFFICIENT QUANTIZATION

TABLE V

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A DIRECT FORM FILTER

Representation		Rounding	Truncation
Sign-And-Magnitude	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$	$\frac{q^2}{3}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$
One's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$	$\frac{q^2}{3}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$
Two's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$	$\frac{q^2}{12}(\sigma^2_x\alpha+\sigma^2_w\beta)/\Delta$

TABLE VI

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A PARALLEL REALIZED FILTER

Representation		Rounding	Truncation
Sign-And Magnitude	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$	$\frac{q^2}{3}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$
One's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$	$\frac{q^2}{3}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$
Two's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$	$\frac{q^2}{12}\sum_{k=1}^{K}[(\sigma^2_x\alpha_k+\sigma^2_{w_k}\beta_k)/\Delta_k]$

C. INPUT QUANTIZATION ERROR

TABLE VII

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A DIRECT FORM FILTER

Representation		Rounding	Truncation
Sign-And-Magnitude	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$	$\frac{q^2}{3}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$
One's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$	$\frac{q^2}{3}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$
Two's Complement	$\bar{f}$	0	$\frac{q}{2}\sum_{i=0}^{N}\sum_{j=0}^{M}h_{ij}$
	σ^2_f	$\frac{q^2}{12}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$	$\frac{q^2}{12}\sum_{i=0}^{N}\sum_{j=0}^{M}h^2_{ij}$

TABLE VIII

STATISTICS OF THE QUANTIZATION ERROR AT THE OUTPUT OF A PARALLEL REALIZED FILTER

		Rounding	Truncation
Sign-And-Magnitude	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$	$\frac{q^2}{3}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j-0}^{M}(h^r_{ij})^2]$
One's Complement	$\bar{f}$	0	0
	σ^2_f	$\frac{q^2}{12}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$	$\frac{q^2}{3}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$
Two's Complement	$\bar{f}$	0	$\frac{q}{2}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$
	σ^2_f	$\frac{q^2}{12}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$	$\frac{q^2}{12}\sum_{r=1}^{K}[\sum_{i=0}^{N}\sum_{j=0}^{M}(h^r_{ij})^2]$

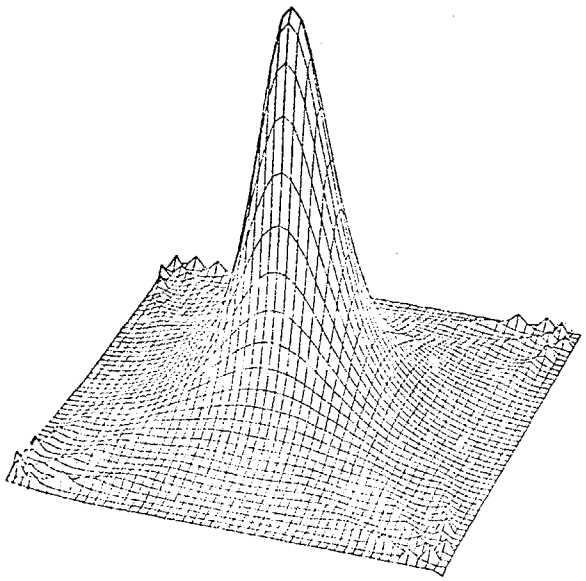

Figure 6 Frequency response of a two-dimensional digital filter with
1) Register Length = 8
2) Peak Magnitude = 1.028
3) Cutoff Frequency = 0.08