EFFICIENT QUANTIZATION OF PARALLEL FIR FILTER USING NUS AND MAD ALGORITHMS

MERLIN MATHEW, NISHA THOMAS, GEEVARGHESE TITUS

Abstract— Parallel **(or block) FIR digital filters can be used either for high-speed or low-power (with reduced supply voltage) applications. Traditional parallel filter implementations cause linear increase in the hardware cost with respect to the block size. Recently, an efficient parallel FIR filter implementation technique requiring a less-than linear increase in the hardware cost was proposed. The use of the appropriate fast FIR filter structures and the proposed quantization scheme can result in reduction in the number of binary adders. This paper makes two contributions for quantization. First, the coefficients are quantized using NUS (National University of Singapore) algorithm and then MAD (Maximum Absolute Difference) quantization algorithm. Quantization is done in a new method of SPT (Signed Power of Two) term allocation. The coefficient values are not allocated with the same number of SPT terms. Coefficient values with larger magnitudes are allocated with more SPT terms than those with smaller magnitudes. MAD quantization algorithm is considered more efficient for the implementation of parallel filters than NUS algorithm.**

I. INTRODUCTION

Low complexity FIR filter implementation has been given extensive consideration. However, most of the past efforts have been directed towards multiplier less implementation with coefficients expressed as sums of signed power-of-two terms (SPT)**.** Furthermore, very little work has been done that deals directly with reducing the hardware complexity of parallel FIR filters.

Traditionally, the application of parallel processing to an FIR filter involves the replication of the hardware units that exist in the original filter. If the area required by the original circuit is A, then the L-parallel circuit requires an area of L x A**.** Recently, an efficient parallel FIR filter implementation technique requiring a less-than linear increase in the hardware cost was proposed using FFAs (Fast FIR Algorithms) **.** The hardware cost can be reduced further by exploiting the frequency spectrum characteristics. This is achieved by selecting appropriate FFA structures out of many possible FFA structures all of whom have similar hardware complexity at the word-level. However, their complexity can differ significantly at the bit-level.

In order to implement a digital filter in hardware, the ideal filter coefficients must first be quantized. Traditionally, the quantization process consists of a direct binary conversion of the coefficients in conjunction with a truncation or rounding scheme. This quantization method is straightforward and time-efficient. However, the quantized filter coefficients contain large number of nonzero bits.

It is well known that if each coefficient value of a digital filter is a sum of signed power-of-two (SPT) terms, the filter can be implemented without using multipliers. In the past decade, several methods have been developed for the design of filters whose coefficient values are sums of SPT terms. Most of these methods are for the design of filters where all the coefficient values have the same number of SPT terms. It has also been demonstrated recently that significant advantage can be achieved if the coefficient values are allocated with different number of SPT terms while keeping the total number of SPT terms for the filter fixed. Some method allows the number of SPT terms for each coefficient to vary subject to the number of SPT terms for the entire filter. This provides the possibility of finding a better filter without increasing the number of adders, which determines the realization cost for a given filter length. The increase in the freedom of selecting the SPT terms will reduce the quantization error of the coefficients.

II. QUANTIZATION ALGORITHMS

Consider the general formulation of a length-N FIR filter,

$$Y_n = \sum_{i=0}^{N-1} \mathrm{h_i x_{n-i}}, \quad \text{n=0, 1, 2, ..., } \infty$$

where $\{x_i\}$ is an infinite length input sequence and $\{h_i\}$ are the length N FIR filter coefficients.

Quantization is done using these filter coefficients. Here we will discuss about NUS (National University of Singapore) and MAD (Maximum Absolute Difference) quantization algorithm.

A. *NUS Algorithm*

If the filter coefficients are first scaled before the quantization process is performed, the resulting filter will have much better frequency-space characteristics. The NUS algorithm employs a scalable quantization process. To begin the process, the ideal filter is normalized so that the largest coefficient has an absolute value of 1. The normalized ideal filter is then multiplied by a variable scale factor (VSF). The

VSF steps through the range of numbers from 0.4375 to 1.13 with a step size of 2^ (-W), where W is the coefficient word-length. Signed power-of-two (SPT) terms are then allocated to the quantized filter coefficient that represents the largest absolute difference between the scaled ideal filter and the quantized filter. The NUS algorithm iteratively allocates SPT terms until the desired number of SPT terms has been allocated or until the desired NPR specification is met. Once the allocation of terms has stopped, the normalized peak ripple (NPR) is calculated. The process is then repeated for a new scale factor. The quantized filter leading to the minimum NPR is chosen.

NUS algorithm is described by the following pseudocode:
Normalize the set of filter coefficients so that the magnitude of the largest coefficient is 1;
For VSF=Lower Scale: Step Size: Upper Scale,
{
Scale normalized filter coefficients with VSF;
Quantize the scaled coefficients using SPT term allocation scheme in NUS algorithm ;
Calculate NPR of the quantized filter;
}
Choose the quantized coefficient that leads to the minimum NPR.

Normalized Peak Ripple (NPR) can be calculated using the equation

$$NPR = \delta / g$$

where δ is the pass band ripple (in dB) and g is the pass band gain (in dB)

The parameter δ is chosen during the filter design, and g is calculated by:

$$g = (Pmax+Pmin)/2$$

where Pmax and Pmin are the maximum and minimum pass band frequencies.

SPT Term Allocation: If each coefficient value of a digital filter is a sum of signed power-of-two (SPT) terms, the filter can be implemented without using multipliers. Most of the methods are for the design of filters where all the coefficient values have the same number of SPT terms. It has been demonstrated recently that significant advantage can be achieved if the coefficient values are allocated with different number of SPT terms while keeping the total number of SPT terms for the filter fixed. The number of SPT terms allocated to a coefficient is determined by the statistical quantization step-size of that coefficient and the sensitivity of the frequency response of the filter to that coefficient. Coefficient values with larger magnitudes are allocated with more SPT terms than those with smaller magnitudes. Comparing with the allocation scheme where all the coefficient values are allocated with the same number of SPT terms, this new allocation scheme produces designs with significantly smaller normalized peak ripple magnitude.

DESIGN EXAMPLE: Consider an ideal filter section with the following parameters: pass band ripple=3dB, stop band ripple=60dB, pass band frequency=3000Hz, stop band frequency=3300Hz, sampling frequency=8000Hz.Frequency responses are shown below: From fig. 1 and 2, ripples of the NUS quantized filter increases compared to Ideal filter.

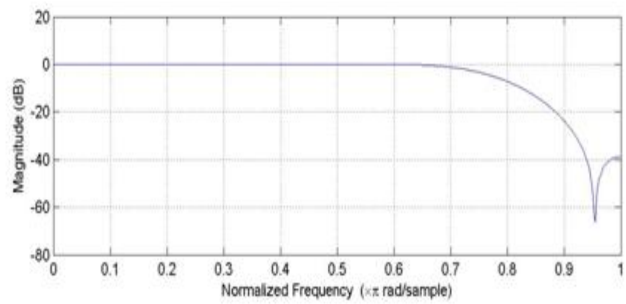

Fig. 1. Frequency response of Ideal filter

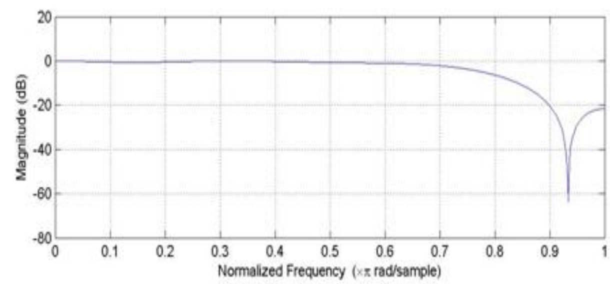

Fig. 2. Frequency response of NUS Quantized filter

B. Look-Ahead MAD Quantization

In parallel FIR filters, the NPR cannot be used as a selection criterion for choosing the best quantized filter since pass band/stop band ripples cannot be defined for the set of filters obtained by the application of FFAs. It is shown that the maximum absolute difference (MAD) between the frequency responses of the ideal filter and the quantized filter can be used as an efficient selection criterion for parallel filters.

The parameter used for Look-ahead MAD quantization algorithm instead of using Normalized Peak Ripple in NUS algorithm, is Maximum Absolute Difference between the frequency responses of ideal filter and quantized filter.

When the quantized filter is implemented, a post-processing scale factor (PPSF) is used to properly rescale the magnitude of the resulting data stream. The value of the PPSF is determined as follows:

PPSF=Max [Abs (Ideal Filter Coeffs.)] / VSF

Essentially, the PPSF reverses the normalization and scaling is introduced in the quantization process. While the scaling process changes the magnitude of the filter response, it should be noted that it does not change the functionality of the filter.

MAD algorithm is described by the following pseudocode:

For each filter section in the parallel FIR filter,
{Normalize the set of filter coefficients so that the magnitude of the largest coefficient is 1;
For VSF = Lower scale: Step size: Upper scale,
{Compute PPSF by (11);
Convert PPSF into Canonic Signed Digit form;
If (No. of nonzero bits in PPSF) < prespecified value,
{Scale normalized coefficients with VSF;
Quantize the scaled coefficients using SPT term allocation scheme in NUS algorithm;
Calculate MAD between the frequency responses

of the ideal and quantized filters;
}
}
Choose the quantized coefficient that leads to the minimum MAD;
}

CSD Algorithm:

An algorithm for computing the CSD format of a W bit number is represented here. Denote the two's complement representation of the number A as

$A=\hat{a}_{w-1}.\hat{a}_{w-2}..........\hat{a}_1\hat{a}_0$ and its CSD representation

$A= a_{w-1}.a_{w-2}........a_1a_0$. The conversion is illustrated using the following iterative algorithm:

$\hat{a}_{-1}=0$
$\gamma_{-1}=0$
$\hat{a}_{w}=\hat{a}_{w-1}$
for (i=0 to W-1)
{
$\Theta_i=(\hat{a}_i)\ \mathrm{xor}(\hat{a}_{i-1})$
$\gamma_i=\mathrm{inv}(\gamma_{i-1})*\Theta_i$
$a_i=(1-2\hat{a}_{i+1})*\gamma_i$
}

DESIGN EXAMPLE: Consider an ideal filter section with the following parameters: pass band ripple=3dB, stop band ripple=60dB, pass band frequency=3000Hz, stop band frequency=3300Hz, sampling frequency=8000Hz.Frequency responses are shown below:

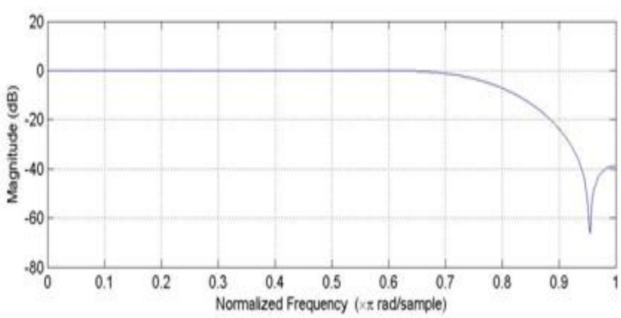

Fig. 3. Frequency response of Ideal filter

From fig. 3 and 4, ripples of the MAD quantized filter increases compared to Ideal filter.

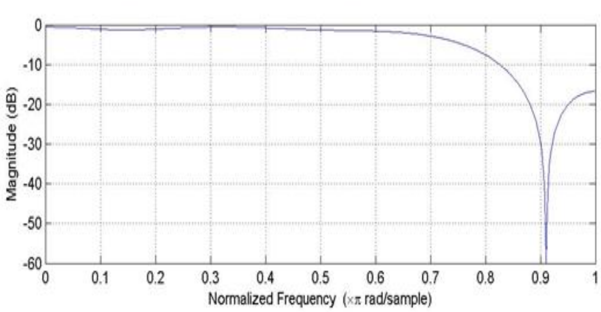

Fig. 3. Frequency response of MAD Quantized filter

III. CONCLUSION

It has been shown that the proposed NUS and look-ahead MAD quantization algorithms were shown to be very efficient for the implementation of parallel filters..

IV. REFERENCES

[1] Jin-Gyun Chung, Yong-Bae Kim, Hang- Geun Jeong, Keshab K. Parhi, Zhongfeng Wang, "Efficient parallel FIR filter implementations using frequency spectrum characteristics" in Proceedings of IEEE ISCAS, (Monterey, CA),vol.5, pp. 354-358, June 1998.

[2] D. Li, J. Song, and Y. C. Lim, "A polynomial-time algorithm for designing digital filters with power-of-two coefficients," in Proceedings of 1993 IEEE ISCAS, (Chicago, IL), pp. 84-87, May 1993.

[3] C. L. Chen, K. Y. Khoo, and A. N. Willson, Jr., "An improved polynomial- time algorithm for designing digital filters with power-of-two coefficients," in Proceedings of 1995 IEEE ISCAS, (Seattle, WA), pp. 223-226, May 1995.

[4] Yong Ching Lim, Rui Yang, Dongning Li, and Jianjian Song, "Signed Power-of-Two Term Allocation Scheme for the Design of Digital Filters" IEEE Trans. Circuits Syst., vol.46, NO. 5, MAY 1999.

[5] D. A. Parker and K. K. Parhi, "Low-area/power parallel FIR digital filter implementations," Journal of VLSI Signal Processing, vol. 17, pp. 75-92, Sept. 1997.

[6] J. H. Satyanarayana and K. K. Parhi, "HEAT : Hierarchical Energy Analysis Tool," in *33rd ACM/IEEE Design Automation Conf.*, (L a Vega, NA), pp. 9-14, June 1996.

[7] Y. C. Lim, "Design of discrete-coefficient-value linear phase FIR filters with optimum normalized peak ripple magnitude," *IEEE Trans. Circuits Syst.*, vol. 37, pp. 1480–1486, Dec. 1990.

[8] Y. C. Lim and S. R. Parker, "FIR filter design over a discrete power-oftwo coefficient space," *IEEE Trans. Acoust. Speech, Signal Processing,* vol. ASSP-31, pp. 583–591, June 1983.

[9] Y. C. Lim, S. R. Parker, and A. G. Constantinides, "Finite wordlength FIR filter design using integer programming over a discrete coefficient space," *IEEE Trans. Acoust. Speech, Signal Processing,* vol. ASSP-30, pp. 661–664, Aug. 1982.

[10] Q. Zhao and Y. Tadokoro, "A simple design of FIR filters with powerof- two coefficients," *IEEE Trans. Circuits Syst.*, vol. 35, pp. 566–570, May 1988.

[11] S. Sriranganathan, D. R. Bull, and D. W. Redmill, "Low complexity twodimensional digital filters using unconstrained SPT term allocation," in *Proc. 1996 Int. Symp. Circuits Syst.*, 1996, pp. 762–764.