EFFICIENT FIR FILTER DESIGN USING WALLACE TREE COMPRESSION

T. Sandhya Pridhini *, Jeena Maria Cherian, Diana Aloshius,  
Department of Electronics and Communication Engineering  
Hindustan University, Chennai, Tamilnadu.

Abstract -The major important metrics of the digital signal processing circuits are area, power and delay. In many signal processing system significant area, delay and power consumption is due to the multiplication. With advances in technology, many researchers have tried and are trying to design multipliers which offer either of the following design targets – high speed, low power consumption, regularity of layout and less area The Aim of this research is to design a low cost finite impulse response filter using the concept of faithfully rounded truncated multipliers. The optimization of bit width and the hardware resources are done with good accuracy. In direct FIR filter the multiple constant multiplication are implemented using the improved version of truncated multipliers.

Keywords: FIR filter; faithfully rounded; truncated multipliers; Wallace tree compression.

INTRODUCTION:

Low power design of digital integrated circuits has emerged as a very active and rapidly developing field of CMOS design. In the past the major concerns of the VLSI designer were speed, the area and cost, power consideration was typically a secondary importance. Applications like Digital Signal Processing involve a large number of multiplication using constants which lead to complex design and consumption of more area and power. FIR filters can be implemented with multiplier blocks through the use of multiple constant multiplication (MCM). MCM is an efficient way of implementing several constant multiplication with the same input data. The coefficients are expressed using shifts, adders and subtractors. The generation of multiplier block of set of constant is MCM block [3]. A finite impulse response (FIR) filter is usually implemented by using a series of delays, multipliers and adders to create the filter’s output as shown in fig 1. FIR filters can be easily designed to be “linear phase”. Linear phase has a constant group delay, where all the frequency components have equal delay time. So the filter does not cause any phase distortion or delay distortion. There are two types of FIR filter structure namely direct form and the transposed form.

Direct form: The delay of the adders can be reduced by extra pipeline registers between the adders. In transpose form there is no extra pipeline registers in order to reduce the delay. So the area of the delay will be high in transpose form when compare to the direct form. In direct form the MCM/A performs concurrent multiplication of delayed input signals and their respective coefficient followed by accumulation of all products. In order to decrease the area cost the hardware implementation of digital FIR filter is
classified into multiplierless and memory based FIR filter system. In multiplierless, multipliers can be replaced by the adders and shifter that leads to the reduction in silicon area. Using common subexpression elimination and Canonic Signed Digit can reduce the common suboperations in multiplierless technique. The memory based FIR filter can be implemented on LUT - look up table based FPGA. The transpose form FIR filter is best suitable for the memory based FIR filter [6].

Fig 2 Structure Of Multiple Constant Multiplication

In this research the FIR filters can be designed by using direct form FIR filter with multiple constant multiplication/accumulation using the concept of faithfully rounded truncated multiplier. The area of the delay elements is small in direct form when compared to the transpose form. The optimization of filter coefficients is the important design issue of the FIR filter implementation which has a direct impact on the silicon area cost of the arithmetic units and registers. After multiplication the bit widths grows and in many DSP application it does not require precision output. The accumulated partial products in MCM/A module are realized by removing the unnecessary PP bits (PPB) without affecting the final precision of the output. In order to achieve the optimized output, it is necessary to faithfully round the output where the maximum absolute error should not be more than 1 ulp.

USING IMPROVED VERSION

The generic flow of FIR filter design and its implementation show in fig 3, is divided into 3 stages namely (1) finding the order and the coefficients of the filter, (2) Quantizing the filter coefficients and (3) Optimizing the hardware resources of the filter. The first stage is to determine the filter order and filter coefficients with respect to the frequency response of the filter. The second stage is to quantize the obtained filter coefficients to obtain finite bit accuracy. The third stage describes the area cost can be optimized using the various methods.

Fig.3 Generic flow of FIR filter design and its implementation

Parks McClellan:

The most efficient method for designing optimum magnitude FIR filters with arbitrary specifications is the Remez exchange algorithm. Initially Parks McClellan ( ) is used to find the filter order M for the given frequency response. It is an iteration algorithm that accepts filter specifications in terms of passband and stopband frequencies, passband ripple, and stopband attenuation. Remez ( ) is to find the coefficients for the FIR filter of order M. Then, quantize the coefficients with enough bits and generate the set of uniformly quantized coefficients with equal bit width B.

PP Truncation and Compression:

Instead of accumulating individual multiplication for each product, it is more efficient to collect all the PPs into a single PPB matrix with carry-save addition shown in fig 4(a) and 4(b). In faithfully rounded FIR filter implementation, it is
required that the total error introduced during the arithmetic operations is no larger than one ulp.

Fig 4(a) Individual Compression of PPBs and (b) Combined PPBs Compression

**Truncated Multiplier**

Truncated $m \times n$ multipliers that produce results less than $m + n$ bits long [4]. Benefits of truncated multipliers include reduced area, delay, and power consumption. The correction constant, $C_r$, and the ‘1’ added for rounding are normally included in the reduction matrix as shown in fig.5. Truncation is the term for limiting the number of digits right of the decimal point by discarding the least significant ones.

Fig .5 Truncated two’s complement multiplier with constant correction

Fig .6 Improved version of the faithfully rounded truncated multiplier
Fig. 6 describes the improved version of the faithfully rounded truncated multiplier. In this method only the deletion and truncation takes place to eliminate the partial product bits. The range of the deletion error is two times larger than the previous results. The grey dot represents the deleted bits while the green dots represent the truncated bits. The Wallace tree multiplier is considered as faster than parallel multiplier and it is efficient implementation of a digital circuit which multiplies two integers. To reduce the latency a Wallace tree multiplier uses carry save addition algorithms. Basically Wallace tree multiplies two unsigned integers.

A multiplier is divided into three stages:- the first stage is partial product generation where the multiplicand and multiplier are multiplied by bit wise to generate partial products. The second stage is the partial product reduction which is more complicated and the final stage is the carry propagation stage using different compressors employed in high speed multipliers to reduce the latency of the accumulation stage.

**USING WALLACE TREE COMPRESSION**

Multiplication using improved version of the truncated multipliers consumes more area and power. In order to reduce power and area, in this paper 5:2 compression technique is used which gives good accuracy with less power consumption. During multiplication process the partial products are accumulated by the basic building block called compressor. The multiplier architecture consists of partial product generation, partial product reduction and partial product accumulation. By decreasing the number of adders the latency of the Wallace tree multiplier can be reduced in the partial product reduction stage. The full adders and half adders are replaced by the different compressors which speeds up the summation in general and multiplication in particular.

### 5:2 compressor

The basic idea of a n:2 compressor is that the n operands can be reduced by 2. The architecture of 5:2 compressors comprises of two serially connected 3:2 compressor as shown in the fig. 7. The A, B, C, Sum1 and Carry1 are the inputs and outputs of the first compressor. The sum1, D, E Sum2 and Carry2 are the inputs and outputs of the second compressor shown in fig 8.

![Fig 7. Structure of 5:2 compressor](image)

![Fig 8. Serially connected 3:2 compressor.](image)

Consider an example shown in fig. 6, in the given 8x8 multiplication the first 8 bit input is multiplied by 8 bit multiplier. If each bit is multiplied by another 8 bit totally 64 bits obtained. These bits are than added by using compressors which reduces the latency and increases the speed. The sum output of the intermediate compressor is fed as input to the next compressor of the same column and generated carry of the corresponding adders are propagated to the next column. Finally a 16 bit result will be obtained.

**SYNTHESIS REPORT:**

Synthesis report shown in fig. 9 gives the complete details of device and total power.
utilization for multiplication of two 8 bit data using 5:2 compression techniques. Using 5:2 compressor the number of gates is reduced to 2136 compared to the improved version of truncated multiplier. The total estimated power using 5:2 compressors is reduced to 48 mw.

Fig. 9 shows the synthesis report for the modified power

The table 1 gives the comparative study on area utilization and power consumption of the improved version of the truncated multiplier and the 5:2 compressors.

Table 1. Comparison of the area and power of the improved version of the faithfully rounded truncated multiplier and the 5:2 compressor

<table>
<thead>
<tr>
<th>PARAMETERS</th>
<th>IMPROVED VERSION OF THE TRUNCATED MULTIPLIER</th>
<th>5:2 COMPRESSOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Of Gate Counts (Area mm²)</td>
<td>2203</td>
<td>2136</td>
</tr>
<tr>
<td>Power (Mw)</td>
<td>56</td>
<td>48</td>
</tr>
</tbody>
</table>

CONCLUSION:

In this research, a linear phase direct form FIR filter is designed by considering the optimization of bit widths and the hardware resources. In the majority of these applications, multiplier has been a critical and obligatory component in dictating the overall circuit performance when constrained by power consumption and computation speed. Compressors are a critical component of the multiplier, which greatly influence the overall multiplier speed. The proposed method reduces the number of stages in partial product reduction which leads to reduction in silicon area and power consumption. The comparison results shows that a significant reduction in area and power. The results prove
that the proposed method is more conventional one in terms of area and power.

REFERENCES


