DESIGN OF A FAST AND AREA EFFICIENT FILTER

Pontus Åström, Peter Nilsson and Mats Torkelson
Dept. of Applied Electronics, University of Lund
Box 118, 22100 Lund, Sweden
Tel: +46 46 222 94 69; Fax: +46 46 12 99 48
e-mail: Pontus.Astrom@tde.lth.se

ABSTRACT

This paper shows how to optimize full custom, fixed coefficient filters to gain both in area and speed. The idea is to trade the filter order with the coefficient length and thus reduce the delay and the size of the multipliers. The result is a smaller design that needs fewer clock cycles per sample compared to a minimum order filter. Measurements on the manufactured chips verified a speed gain of 26% and a size reduction of 20%.

1 INTRODUCTION

Full custom designed chips have several advantages over standard DSP chips. They are often smaller and faster and consume less power. These features are achieved by stripping off unnecessary circuitry. However, in the core design of custom chips, the optimization methodology of standard processors is often used. In filter design this means that the main optimization parameter is the filter order. By optimizing filter coefficients instead, both size and speed can be improved. There are two major reasons why this gives better overall performance:

- In a multiplier, the delay is proportional to the length of the coefficient. The delay can be decreased by selecting short coefficients.
- Coefficients like 0, 0.5, 0.25 and so on are very attractive as they only result in a shift. This will reduce the area of the multiplier.

The idea behind the described optimization method is to focus on the filter coefficients rather than the filter order. By increasing the latter above the minimum order we get more freedom in the choice of coefficients. There might exist solutions with short and trivial coefficients. This principle is useful for both parallel and bit-serial arithmetic. Here, a bit-serial example is chosen to show the method.

2 BACKGROUND

The filter was designed to be part of a receiver employing wideband Orthogonal Frequency Division Multiplexing (OFDM). The main parts of the receiver are shown in figure 2. The incoming signal to the left consists of 40 separate channels, each 50 kHz wide. Thus the total bandwidth is 2 MHz, or rather ±1 MHz because it is a baseband signal.

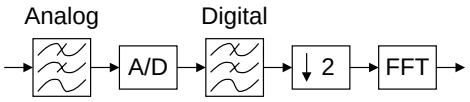

Figure 1: Block structure of the receiver.

Before this signal can be fed to the Fast Fourier Transform (FFT) block it has to be band-limited and A- to D-converted. The sample rate of the FFT chip is specified to 3.2 MHz. However, if this rate were to be used for A- to D-conversion, an analog anti-alias filter of very high order would be required. In order to reduce the filter order, the double sample frequency is chosen and a digital filter is inserted to attenuate the folding products. A decimator is inserted after the digital filter to divide the sample rate by two to obtain 3.2 MHz.

3 FILTER SPECIFICATION

The receiver demodulates Quadrature Phase Shift Keying (QPSK). This is a modulation method where the information is carried in the phase. Thus, an important parameter is group delay. Here it is specified to 200 ns, which equals 1% of the symbol duration.

The amplitude of each OFDM channel has to be flat. Therefore, the passband ripple must be low. Here it is specified to less than 0.5 dB. The stopband attenuation of the filter is specified to -40 dB and the system word length is 10 bits. In order to prevent arithmetic overflow, two extra bits are added internally in the filter.

4 FILTER ALGORITHM

The filter is implemented in a lattice wave digital filter (LWDF) algorithm [1]. If properly designed, this is an algorithm with low sensitivity to coefficient variations and guaranteed stability.

A LWDF is designed with z^{-1} elements, i.e. memories, and adaptors. The symbol and the internal struc-

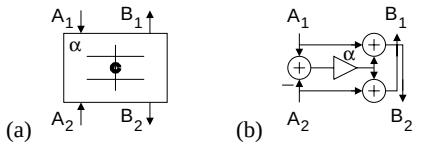

Figure 2: (a) The adaptor symbol. (b) The internal structure of the adaptor. A_i and B_i represent input respectively output signals. α is the adaptor coefficient.

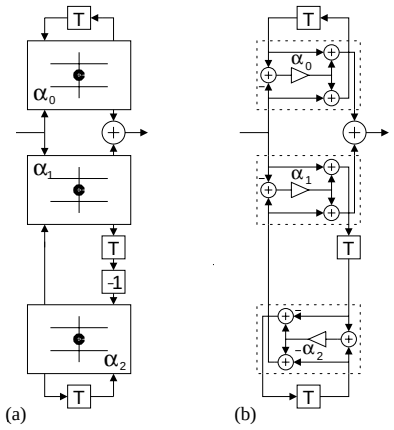

Figure 3: (a) Filter structure of third order LWDF. (b) The data path of the filter. Dashed lines indicate adaptor borders. Note that the block that takes the two's complement has been moved inside the adaptor α_2.

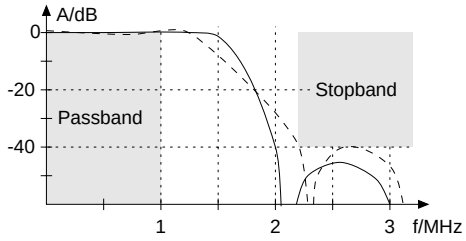

Figure 4: The amplitude response of both filters. Dotted line for the third order filter and solid line for the sixth order filter. The frequency axis is cut at 3.2 MHz, which equal half the sample rate.

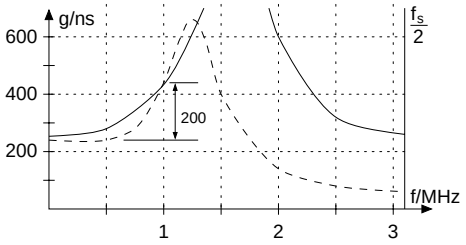

Figure 5: Groupdelay of the filters. As before, the dotted and solid lines belong to the third respectively the sixth order filter. The arrow indicates the largest difference of groupdelay within the passband.

ture of an adaptor are shown in figure 2.

4.1 Design of the Third Order Filter

A minimum third order filter is shown in figure 3a. The filter section consists of one first order all pass link at the top and one second order link at the bottom. In the figure the z^{-1} elements are represented by T-blocks. The coefficients are 7 bits long and shown in table 1. With this choice of coefficients, the filter response is given by the dashed line in figure 4 and 5. As shown in figure 5, the difference in groupdelay within the passband is 200 ns.

Table 1
Filter Coefficients

	3:rd order		6:th order	
α_0	0.375000	0011000	0.0	00
α_1	0.578125	0100101	0.5	01
α_2	-0.328125	1101011	0.0	00

4.2 Design of the Sixth Order Filter

The idea behind the choice of a higher filter order is illustrated in figure 6. As the filter order increases, the number of transfer functions that fit to the specification also increases. One drawback with an increased filter order is that the group delay also increases. It is desirable not to have to correct the phase with an all pass filter as this increases the complexity. Thus, groupdelay sets the upper limit of the possible filter order.

In the design of the coefficient optimized filter an exhaustive search algorithm was used. To limit the extent of the search, a maximum delay of the multiplier, i.e the length of the coefficients was specified. Due to the groupdelay, only filters with order less than three times the minimum order were searched for.

From all filters that were found to fit the specification a cost function was calculated. The aim of the cost function is to find the filter implementation with the lowest area cost. Important parameters in the cost function are the filter order and the number of ones in the coefficients.

The optimum performance was found with a sixth order filter built out of two identical cascaded third order links. With this solution the maximum delay of a multiplier is one clock cycle. The coefficients of the sixth order filter are given in table 1. The structure of the sixth order filter is shown in figure 7a. In figure 7b the data path is shown. Since both α_0 and α_2 equal zero, four adaptors consist of a feed-through only. The filter response is given by the solid lines in figure 4 and 5. Groupdelay within the passband varies from 263 ns to 430 ns. Thus, the maximum difference in group delay is 430-263=167 ns which is well below stipulated 200 ns.

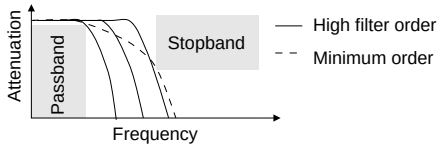

Figure 6: If the filter order is increased several sets of coefficients fulfill the specification. The set with the lowest area cost is chosen.

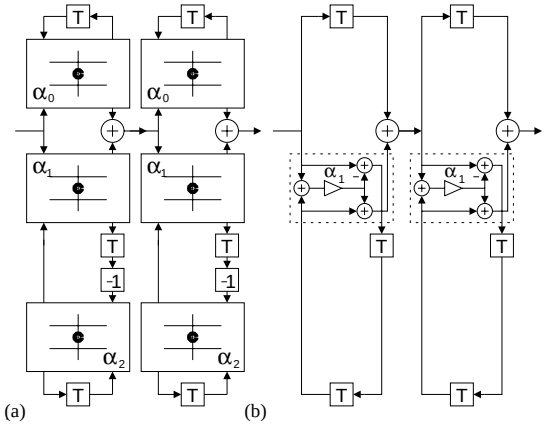

Figure 7: (a) Filter structure of a sixth order LWDF. (b) The data path of the filter where α_0 and α_1 equals 0.

5 COMPARISON OF THE FILTERS

In figure 8, the data path of the α_1 adaptor in the third order filter is shown including all shimming delays. The multiplier is realized in bit serial fixed coefficient arithmetic. The delay elements surrounded by dotted lines are inserted to shorten the longest delay between two delay elements to that of one adder. If size is more important than speed they can be removed.

Figure 9 shows the corresponding data path for the sixth order filter. Due to the simple coefficient the data path is significantly simplified. As for the third order filter, delay elements inside the dotted line can be removed.

5.1 Area

The data path consists of three types of cells: adders, subtracters and delay elements. From our own experience an adder and a subtracter need three times the area of a delay element. The area of one delay element is considered 1 area unit. Thus, the area of an adder or a subtracter will be 3 area units. To make the comparison complete the memory usage has to be included, i.e the T-blocks of figure 3 and 7. Each T-block stores one word. If the T-blocks are built with cascaded delay elements, as is the case here, the area cost in area units of a T-block will be equal to the word length. In table 2 the total area count for the third and sixth order filters are summed up. As seen in table 2 the area reduction is 28%.

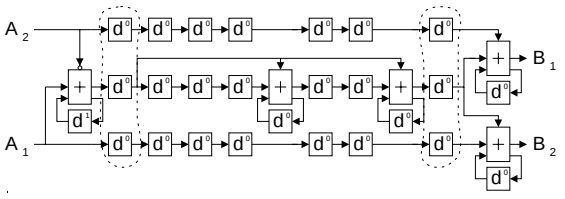

Figure 8: Data path including shimming delays of adaptor α_1 in the third order filter.

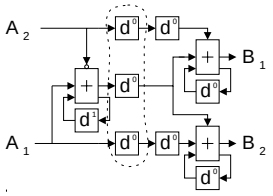

Figure 9: Data path of adaptor α_1 in the sixth order filter

Table 2:
Comparison of the two filters

Filter order	Area units	Clock cycles
three	138	19
six	100	14
Gain	28%	26%

5.2 Speed

In general the number of clock cycles needed for a sample is the word length plus twice the delay of a multiplier. For the third order filter this adds up to a total of 26 cycles. However, by a very simple rescheduling the delay can be decreased by that of a multiplier. Table 2 shows the number of clock cycles that is needed for both filters after rescheduling.

6 CHIP

To prove the result both the third and the sixth order filters were designed and fabricated. A 2 metal layers 0.8 μm CMOS process was used.

6.1 Layout

In order to get a stand alone chip, the following three blocks had to be added to the filter core.

I/O unit: Data is fed on and off the chip in a parallel mode. This is handled by the simple I/O unit. The basic building block of the I/O unit is the delay element.

Control unit: The control unit is used to reset the delay elements of the filter once every sample. It also keeps track of when a word is to be loaded on and off the chip. It is made up of delay elements coupled in a closed loop, along which a "1" is stepped forward. It makes one complete turn for every sample.

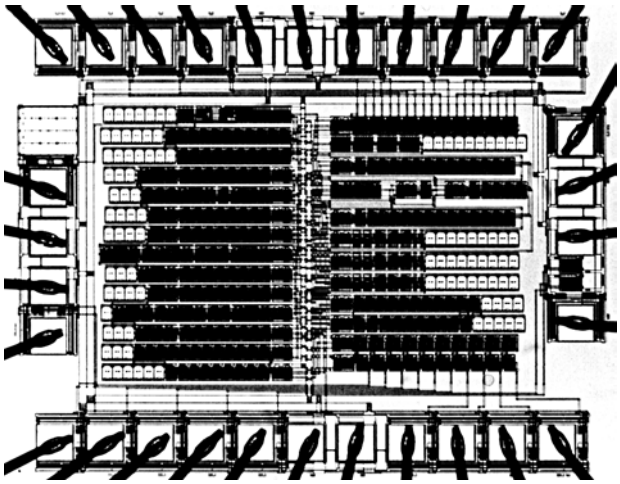

Figure 10: Die photo of third order filter.

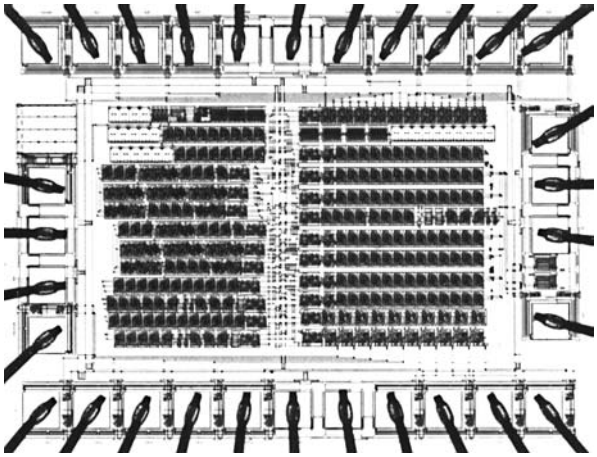

Figure 11: Die photo of sixth order filter.

On chip clock generator [3]: The high frequencies needed make it difficult to feed the clock signal from outside. Instead a trig signal is fed to the on-chip clock generator once every sample. The clock generator then generates a specified number of clock cycles.

6.2 Floor Plan

Die photos of the third and sixth order filters are shown in figure 10 and 11. Both filters have the same basic structure. The upper three rows to the left are the on-chip clock. The adaptors are situated below the clock generator. The control unit is placed at the bottom.

In the right column the I/O unit, the memory and the large buffers are placed. The I/O unit is split into two parts. Input signals are handled at the top of the column and output at the bottom. Memory and buffers are placed in between.

Total core area of the third and sixth order filter is 2.8 mm^2 and 2.2 mm^2 respectively. This means that the size reduction of the sixth order filter is 20%.

6.3 Measurements on the Fabricated Chips

At 5 volt supply both chips work properly. Total power consumption is 250mW for both chips. Due to the slower clock rate in the sixth order filter it is possible to reduce supply voltage to 3.7V, which reduces power consumption to 110mW.

7 SUMMARY

This paper has presented a method to find smaller and faster designs for custom designed chips. We have seen that an increase in filter order can be accepted if trivial or simple coefficients can be found.

The number of clock cycles that have to be carried out during a sample decreased from 19 to 14. This made it possible to decrease the on chip clock frequency from 122 MHz to 90 MHz. This simplified the design and made it possible to power the chip at 3.7V with power consumption reduced to 110mW.

The expected size reduction for the filter core was computed to 28%. However, after mapping to silicon the reduction decreased to 20%. This was mainly due to the extra circuitry that was needed for the off-chip interface.

References

[1] A. Fettweis. Wave digital filters: Theory and practice. *Proceedings of the IEEE*, 74, February 1986, pp. 270–327.

[2] P. Nilsson. *A CMOS VLSI Cell Library For Digital Signal Processing.* Lund University, Lund, Sweden, May 1992. Licentiate Thesis.

[3] P. Nilsson and M. Torkelson. A Monolithic Digital Clock-Generator for On-Chip Clocking of Custom DSPs. *IEEE Journal of Solid-State Circuits*, 31(5), May 1996.