AN OPTIMAL ENTROPY CODING SCHEME FOR EFFICIENT IMPLEMENTATION OF PULSE SHAPING FIR FILTERS IN DIGITAL RECEIVERS

A.P.Vinod, A.B.Premkumar, E.M-K.Lai

School of Computer Engineering, Nanyang Technological University
Nanyang Avenue, Singapore 639798

ABSTRACT

The most computationally intensive part of wide-band receivers is the IF processing block. Digital filtering is the main task in IF processing. Infinite precision filters require complicated digital circuits due to coefficient multiplication. This paper presents an efficient method to implement pulse shaping filters for a dual-mode GSM/W-CDMA receiver. We use an arithmetic scheme, known as pseudo floating-point (PFP) representation to encode the filter coefficients. By employing a span reduction technique, we show that the filters can be coded using an optimal entropy scheme employing PFP which requires only considerably fewer bits than conventional 24-bit and 16-bit fixed-point filters. Simulation results show that the magnitude responses of the filters coded in PFP meet the attenuation requirements of GSM/W-CDMA specifications.

1. INTRODUCTION

A first order estimate of the resources required to implement the wideband receiver of a software radio shows that the IF processing block to be the most computationally intensive part since it operates at the highest sampling rate [1]. Digital filtering is the core function in the IF processing block which is accomplished by FIR filters due to its linear phase characteristics. Hardware implementation of FIR filters generally requires large multipliers if they are realized with continuous-valued coefficients. For a multiplierless filter implementation, the hardware complexity increases with the number of the adders, which is proportional to the filter wordlength [2]. The fixed-point arithmetic implementation of channel filters in digital wideband receivers requires 24-bit wordlength to meet the channel specifications [3, 4]. It has been reported that the 16-bit implementation results in a significant degradation in stop-band attenuation preventing the required spectral mask requirements from being met [3]. The entropy of a coefficient is a measure of the information content that can be coded in minimum number of bits. The objective of efficient entropy coding scheme is to represent the filter coefficients using fewer bits while retaining its frequency response characteristics. Optimal coefficient entropy is a desirable feature in ASIC and FPGA implementation of pulse shaping FIR filters, where look-up tables are employed to store all the possible partial products formed when filter coefficients are convolved with the input signal [5, 6]. The memory required in look-up table implementation is a linear function of coefficient wordlength [6]. Efficient methods to implement fixed-point channel filters using fewer bits that meet the required spectral mask characteristics are hardly discussed in the literature. This paper presents the implementation of pulse shaping filters in digital receivers using an arithmetic scheme known as the Pseudo Floating-Point (PFP) representation. It has been shown that the PFP representation can be employed to implement the quadrature mirror filters (QMF) using the lower bound value of 11 bits in [7]. In this paper, we show that the coefficients of the pulse shaping filters can be coded using a wordlength which is considerably lower than conventional 24-bit and 16-bit implementations discussed in [3, 4]. The resulting filters can be implemented using fewer number of adders and is ideal for the look-up table implementation due to its low coefficient entropy.

This paper is organized as follows. Section II gives a brief review the PFP representation that is suitable for programmable Multiply-and-Accumulate (MAC) operations. We discuss the PFP coding scheme for implementation of pulse shaping filters for a dual-mode W-CDMA/GSM receiver in section III. A span reduction algorithm to obtain the optimal entropy PFP coefficients will also be described. In section IV, we illustrate the implementation of pulse shaping filters for W-CDMA/GSM receiver using several examples. Section V provides our conclusions.

2. THE PSEUDO FLOATING-POINT REPRESENTATION

The general representation of sum-of-powers-of-two (SOPOT) terms for the i^{th} filter coefficient is $h_i = \sum_{j=0}^{B-1} 2^{a_{ij}}$, where B is the number of digits in the power-of-two representation. The expression for h_i can be rewritten as, $h_i = 2^{a_{i0}}.\sum_{j=0}^{B-1} 2^{a_{ij}-a_{i0}} = 2^{a_{i0}}.\left[\sum_{j=0}^{B-1} 2^{c_{ij}}\right]$, where $c_{ij} = a_{ij} - a_{i0}$. The term a_{i0} is known as the *shift* and the upper limit value, $(a_{i(B-1)} - a_{i0})$, is known as the *span*. The bracketed term is known as the normalised value (n value). The shift and the normalised value are analogous to the exponent and mantissa in true floating-point representations. Instead of expressing the coefficients as a 16-bit integer, it can be expressed as a (*shift, n-value)* pair – this is the definition of the pseudo floating point representation. For a given coefficient set, let L and M be the number of bits needed to encode the shift and n-value respectively. Then,

$$L = \max_{0 \le i \le N-1} shift\,(h_i) \tag{1}$$

$$M = \max_{0 \le i \le N-1} span\,(h_i) \tag{2}$$

The following example illustrates this concept. Consider the coefficient $h(n)$, whose 16-bit SOPOT representation is given by $h(n) = 2^{-6} + 2^{-8} + 2^{-9} + 2^{-14}$. This can be written as $2^{-6}(2^0 + 2^{-2} + 2^{-3} + 2^{-8})$. In this expression, the term 2^{-6} is the *shift* part (implying 'right shift by 6'), and the bracketed term is the *span* part. The shifts are less complex since they can be hardwired. Therefore, only 3 bits are needed for storing the shift value and correspondingly, $L = 3$. The span value, $M = 8$, is obtained from the bracketed term. Hence the coefficient can be represented in PFP using $L + M = 11$ bits.

In the case of filter implementation in [3, 4], the *L* and *M* values of 24-bit fixed-point coefficients are 5 and 23 respectively. Hence, 28 bits are needed by the PFP for general coefficient sets. For the 16-bit coefficients, *L* and *M* are 4 and 15 respectively and thus require a total of 19 bits in PFP representation. It would seem that the PFP representation might not be an optimal representation. However, it would be interesting to investigate if the actual coefficient sets would require less than the 28 bits and 19 bits in these cases. It should be noted that the span contributes significantly more to the wordlength requirement than the shift. The shifts are well distributed across coefficients and so is not a parameter that could be optimized further. Therefore, it is beneficial to explore some efficient means of reducing the span without considerable implication on the magnitude response of the filter. In the following section, we show that by employing a span reduction technique, the wordlength requirement of the pulse shaping filters for a dual-mode GSM/W-CDMA receiver can be significantly reduced.

3. PFP CODING SCHEME FOR W-CDMA/GSM CHANNEL FILTERS

We consider the IF architecture for dual-mode GSM/W-CDMA receiver presented in [8] to show the filter implementation using PFP coefficients. The dual-mode architecture of IF processing for GSM/W-CDMA is shown in Fig. 1. The IF block performs frequency conversion for W-CDMA and channel extraction from a wide-band received signal for GSM. The input bandwidth of an IF signal covers one channel of 5 MHz in W-CDMA. In the W-CDMA mode, the filter $H_1(z)$, performs pulse shaping to achieve an attenuation of –40 dB at 5 MHz as in the W-CDMA specification. The output signal at 15.36 MHz, which is four times the W-CDMA chip-rate, is fed to base-band processing. The purpose of IF processing for GSM is to extract a single channel with a bandwidth of 200 kHz from the 5 MHz received signal. Therefore, in the GSM mode, $H_1(z)$ functions as a high decimation filter and $H_2(z)$ performs pulse shaping to attenuate the block signal at 200 kHz as per GSM specifications. Our focus is to realize the PFP implementation of the pulse shaping filters, $H_1(z)$, in W-CDMA mode and $H_2(z)$, in GSM mode.

In our attempt to achieve a minimum wordlength for any coefficient set, we fix the *shift* to the maximum value, *l,* corresponding to the worst-case coefficient set using (1). The span value is progressively reduced by throwing away the power-of-two terms and checking whether the resulting filter response meets the filter specifications at each stage. We can expect distortion in the frequency response characteristics when such a span reduction technique is employed to all the "offending coefficients". Our observation in employing the span reduction technique is that the pass-band response of the resulting filter does not change. It has also been noted that the effect of span reduction on stop-band attenuation and peak stop-band ripple is minimal in the case of filters having relatively few number of taps (filter length, $N < 40$). The reason for this behaviour can be explained as follows. Let the span value after performing the reduction be $\hat{m}$.

1. In the case of short length filters, the spans are more closely distributed around $\hat{m}$, whereas for long filters spans are sparsely distributed. Hence, the magnitudes of those terms whose span exceed $\hat{m}$, which are thrown away are considerably smaller for short length filters when compared to that of filters having large number of taps. As a result, the sensitivity of PFP coefficients to span reduction is very low. Sensitivity is a measure of the degree of influence on the frequency response of a digital filter when any one of the coefficients is quantized. The sensitivity can be computed by setting each coefficient, in turn, to its nearest power of two, yielding in each case a response $H_q(\omega_i)$.

$$s(n) = \frac{1}{M}\sum_{i=1}^{M}[H_q(\omega_i) - H(\omega_i)]^2 \qquad (3)$$

where $H(\omega_i)$ and $H_q(\omega_i)$ are the frequency responses of the infinite-precision and quantized coefficients respectively at *M* finite number of frequencies ω_i [9]. The equivalent time-domain expression for sensitivity is given by

$$s(n) = \frac{1}{M}\sum_{n=0}^{N-1}[h_q(n) - h(n)]^2 \qquad (4)$$

where $h_q(n)$ and $h(n)$ represent the impulse responses of the quantized and infinite-precision coefficients respectively. In the case of filters with relatively fewer number of taps, $[h_q(n) - h(n)]$ is considerably small due to the distribution of spans close to $\hat{m}$. Hence, the sensitivity, which is a square function is minimal. This will be illustrated in the examples provided in section IV.

2. The span deviation from $\hat{m}$ is relatively uniform across the coefficient grid in the case of filters with fewer taps when compared to that with larger taps. As a result, applying span reduction to the filter is similar to scaling the entire coefficient. Scaling the coefficient set will not affect the frequency response shape; instead it only changes the filter gain. In the case of PFP representation for short length filters, the deviation in gain is minimal since the span deviation from $\hat{m}$ is minimal. It is known that for fixed-point FIR filter coefficients, the worst-case quantization error usually occurs for the larger valued coefficient. However, in the design examples of raised cosine

filters illustrated in section IV it can be seen that the powers-of-two representation of larger valued coefficient has fewer number of non-zero digits. The span values of these coefficients are less than $\hat{m}$. In the case where the span of largest valued coefficient exceeds $\hat{m}$, the power of two is extremely small. Hence the response deterioration is within the filter stop-band attenuation specification limits when the PFP span reduction technique is employed.

It must be noted that the span reduction technique is limited to filters having fewer number of taps. The raised cosine filters employed as pulse shaping filters in digital receivers are short length filters whose taps is normally less than 40. Hence the proposed span reduction is suitable for this application.

3.1 PFP Filter Implementation Algorithm

The steps for PFP coding of filters using the span reduction approach are presented below.

Step 1: Design the raised cosine filter, $h(n)$, with infinite precision as in the specification for pulse shaping filter of W-CDMA/GSM. Determine the frequency response of the unquantized coefficients, $H_d(\omega) = \sum_{n=0}^{N-1} h(n)e^{-j\omega n}$.

Step 2: Set all the coefficients to their closest sum of power-of-two (SOPOT) coefficients using 16-bits and represent them as a (*shift, span)* pair in PFP. Fix the shift to the maximum value *l,* corresponding to the worst-case coefficient set. Determine the maximum span value, *M*. Set iteration *k=0.*

Step 3: Decrease the span to *M-1* by throwing away the power-of-two terms of offending coefficients and obtain the new set of coefficients, $h_q(n)$. Determine $h_q(n) - h(n)$.

Step 4: Determine the frequency response of the quantized filter whose span is reduced to *M-1* using:

$$H_q(\omega_i) = H_d(\omega_i) + H_e(\omega_i) = \sum_{n=0}^{N-1} [h(n) + \sum_{n=0}^{N-1} h_q(n) - h(n)]\, e^{-j\omega_i n}$$

where ω_i represents frequency samples in the stop-band.

Step 5: If $|H_{qs}(\omega_i)| \le |H_s(\omega_i)|$, where $H_{qs}(\omega_i)$ represents the stop-band response of the PFP filter and $H_s(\omega_i)$ as in the stop-band specification of the pulse shaping filter, set *k=k+1* and go to step 3. Otherwise, terminate the program and choose the PFP coefficients, $h_q(n)$, corresponding to the *'k'*th iteration.

4. DESIGN EXAMPLE

In this section, we implement the pulse shaping filters, $H_1(z)$, employed in W-CDMA mode and $H_2(z)$, employed in GSM mode. The infinite-precision filter, $h(n)$, is generated by the raised cosine FIR filter design program provided by the MATLAB® "firrcos" function. The roll-off factor is selected as 0.22 for bandwidth efficiency in 3G cellular applications.

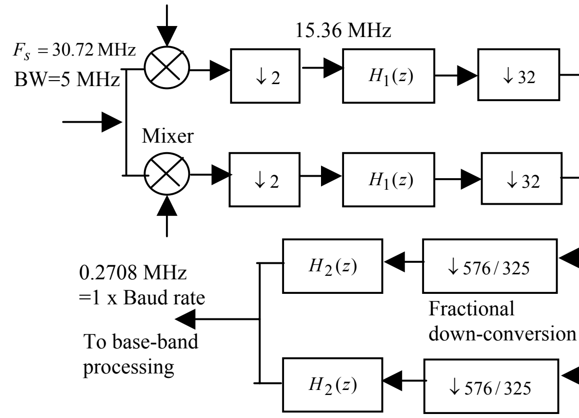

Fig. 1. IF architecture for dual-mode W-CDMA/GSM

Example 1: A raised cosine filter of length 19 is employed as the pulse shaping filter, $H_2(z)$, in GSM mode. The 16-bit SOPOT coefficients and 6-bit PFP coefficients obtained after span reduction are listed in Table 1. The sensitivity values of PFP coefficients are also shown in Table 1. The shift value is fixed at $l = 4$ bits based on the worst-case coefficients, $h(1)$ and $h(3)$. The lower bound of span obtained by employing the proposed algorithm is 2 bits. Hence the coefficient set is represented using 6 bits in PFP. The magnitude responses of the filters are shown in Figure 2. Response of the 8-bit PFP filter shows close resemblance to that of the 16-bit SOPOT filter. Though the attenuation of the PFP filter at 200 kHz (-37.7 dB) is slightly lower than the attenuation of the 16-bit SOPOT filter (-41.7 dB), it is still higher than the GSM specification of –20 dB. Both filters offer identical peak pass-band ripple (PPR) of 0.1 dB and peak stop-band ripple (PSR) of –35.80 dB. These comparisons show that there is practically no difference in the response of filters obtained using the proposed representation and the normal methods. The considerably low sensitivity values of PFP coefficients shown in Table 1 account for this achievement.

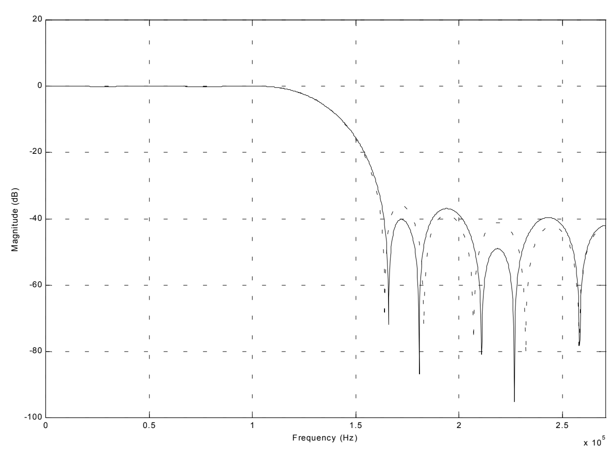

Fig. 2. Filter frequency responses in example 1.
Solid: 6-bit PFP, Dot: 16-bit SOPOT
(Frequency scale indicated is 0 - 270.833 kHz)

$h(n)$	SOPOT Coefficients (16-bit)	PFP Coefficients (6-bit)	Sensitivity of PFP coefficients (M=128)
$h(0)$	$2^{-7}+2^{-8}+2^{-12}+2^{-13}$	$2^{-7}+2^{-8}$	1.02e-009
$h(1)$	2^{-15}	2^{-15}	0
$h(2)$	$-2^{-6}-2^{-7}-2^{-10}-2^{-12}-2^{-13}-2^{-14}-2^{-16}$	$-2^{-6}-2^{-7}$	1.56e-008
$h(3)$	$-2^{-15}-2^{-16}$	$-2^{-15}-2^{-16}$	0
$h(4)$	$2^{-5}+2^{-6}+2^{-11}+2^{-15}+2^{-16}$	$2^{-5}+2^{-6}$	2.27e-009
$h(5)$	2^{-14}	2^{-14}	0
$h(6)$	$-2^{-4}-2^{-5}-2^{-10}-2^{-11}-2^{-12}-2^{-13}-2^{-14}-2^{-15}-2^{-16}$	$-2^{-4}-2^{-5}$	2.97e-008
$h(7)$	$-2^{-14}-2^{-16}$	$-2^{-14}-2^{-16}$	0
$h(8)$	$2^{-2}+2^{-4}+2^{-9}+2^{-12}+2^{-16}$	$2^{-2}+2^{-4}$	3.83e-008
$h(9)$	$2^{-1}+2^{-13}+2^{-16}$	2^{-1}	1.48e-010

Table 1. Coefficients of pulse shaping filter in example 1.

Example 2: For the W-CDMA mode, a raised cosine filter of length 33 is designed. The lower bound PFP obtained is 8-bit. The magnitude responses of the filters are shown in Figure 3. Both the filters meet the desired attenuation of –40 dB as in W-CDMA specifications. The 8-bit PFP filter response shows close resemblance to that of the 16-bit SOPOT filter.

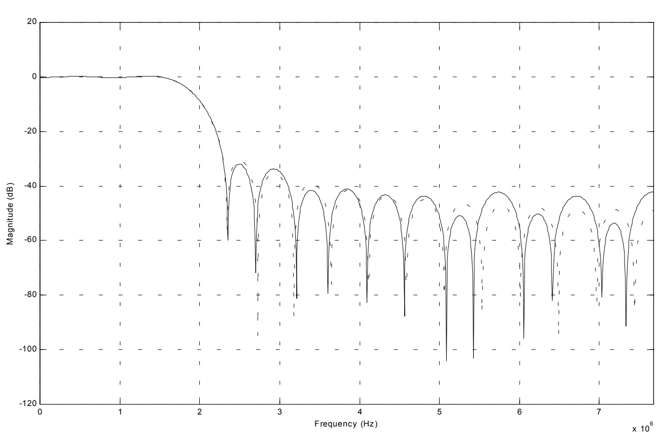

Fig. 3. Filter frequency responses in example 2.
Solid: 8-bit PFP, Dot: 16-bit SOPOT
(Frequency scale indicated is 0 – 7.68 MHz)

5. CONCLUSIONS

We have presented an efficient coefficient coding scheme using pseudo floating-point representation for implementation of pulse shaping filters in digital receivers. The span reduction algorithm reduces the PFP coefficient entropy to considerably lower bits when compared to the fixed-point realizations discussed in the literature. The computational complexity of the algorithm is relatively less since it is applied only to the filter stop-band response samples. Simulation results clearly show that the proposed method results in filters with good frequency response characteristics. The proposed method can be used to implement any FIR filters provided the number of taps is less than 40. It is also worth noting that the proposed PFP filter is hardware efficient due to its minimum wordlength requirement. The PFP representation requires less memory when look-up tables are employed in ASIC/FPGA implementation architecture for pulse shaping FIR filters in 3G mobile communications.

6. REFERENCES

[1] J.Mitola, *Software Radio Architecture.* New York: Wiley, 2000.

[2] Yong Ching Lim, Y.Sun, and Ya Jun Yu, "Design of discrete-coefficient FIR filters on loosely connected parallel machines," *IEEE Transactions on Signal processing,* vol. 50, no. 6, pp. 1409-1416, June 2002.

[3] H.R.Karimi, N.W.Anderson, and P.McAndrew, "Digital signal processing aspects of software definable radios," *IEE Colloquium, Adaptable and Multistandard Mobile Radio terminals,* March 1998.

[4] K.Kalbasi, "Performance tradeoffs in a dual-mode, W-CDMA/EDGE digital IF receiver," Tech. Rep., *Agilent Ees of Electronics Design Automation,* 2001.

[5] W.P. Zhu, M.O. Ahmad and M.N.S.Swamy, "ASIC implementation architecture for pulse shaping FIR filters in 3G mobile communications," *Proc. 2002 IEEE Int. Symp. on Circuits and Systems,* pp. 433-436, vol. 2, May 2002.

[6] Zheng Wu, Cheng Luo, Xin Su, and Xibin Xu, "Digital filter implementation for software radio," *53rd IEEE Vehicular Technology Conference,* pp. 1902-1906, vol. 3, 2001 (Spring).

[7] A.B.Premkumar, C.T.Lau, and A.P.Vinod, "High performance architectures for QMF banks," *NASA 9th Symposium on VLSI Architecture,* November 2000.

[8] M.Jian, W.H.Yung, and B.Songrong, "An efficient IF architecture for dual-mode GSM/W-CDMA receiver of a software radio," *IEEE International Workshop on Mobile Multimedia Communications,* pp. 21-24, November 1999.

[9] Xiaojuan Hu, L.S.DeBrunner, and Victor DeBrunner, "An efficient design for FIR filters with variable precision," *Proc. 2002 IEEE Int. Symp. on Circuits and Systems,* pp. 365-368, vol. 4, May 2002.