

# Design Algorithm of FIR Filter Based on Coefficient Compression

Qingcao Huang

College of Rail Transit Locomotive and Rolling Stock, Hunan Railway Professional Technology College, Zhuzhou, China

Email: 13974196270@163.com

**How to cite this paper:** Huang, H.C. (2025)

Design Algorithm of FIR Filter Based on Coefficient Compression. *Open Journal of Applied Sciences*, 15, 2128-2135.

<https://doi.org/10.4236/ojapps.2025.157140>

**Received:** June 24, 2025

**Accepted:** July 20, 2025

**Published:** July 23, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

This article proposes a new algorithm based on coefficient compression, which can be used for the design and implementation of various FIR filters, to address the problems caused by the limited quantization word length of traditional high-order filters. It addresses the performance degradation problem caused by quantization bit width in FIR, based on coefficient compression and modifications to the accumulator of FIR, which can improve the passband and stopband performance of FIR. And its serial implementation structure was proposed, and the algorithm's degree of simplification in circuit implementation was verified through experiments.

## Keywords

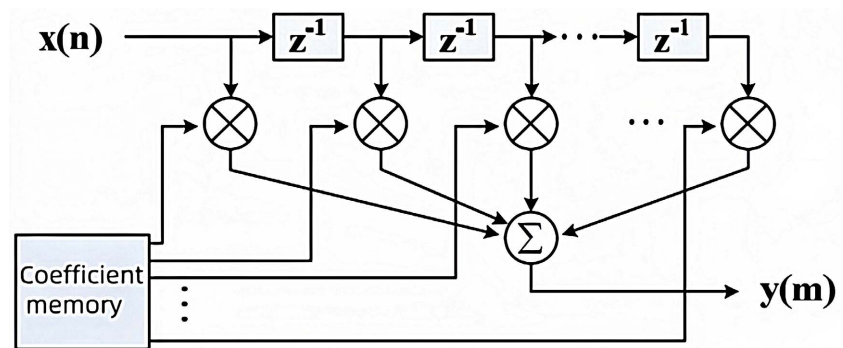
FIR Filter, Compensation Filter, Amplitude Response, Stop-Band Attenuation

## 1. Introduction

The application of Finite Impulse Response (FIR) filters in various digital systems has long been a challenging problem [1] [2]. The proposal of Parks-McClellan and other ripple design algorithms based on Remez exchange theory has made it possible to efficiently design high-order FIR filters through computer-aided design [2]-[5]. In the specific implementation process of FIR filters [6] [7], some problems will be encountered [8] [9]. The FIR filter structure of the most common coefficient pre storage architecture is shown in **Figure 1** [1] [2].

Many people have made efforts to design and improve high-performance FIR filters, such as decomposing the transfer function and implementing it with multi-stage filters; the problem of using floating-point filters to calculate coefficients; using a sharpening filter to obtain better passband stopband response, and utilizing a low sensitivity filter to reduce the impact of bit width on the passband; im-

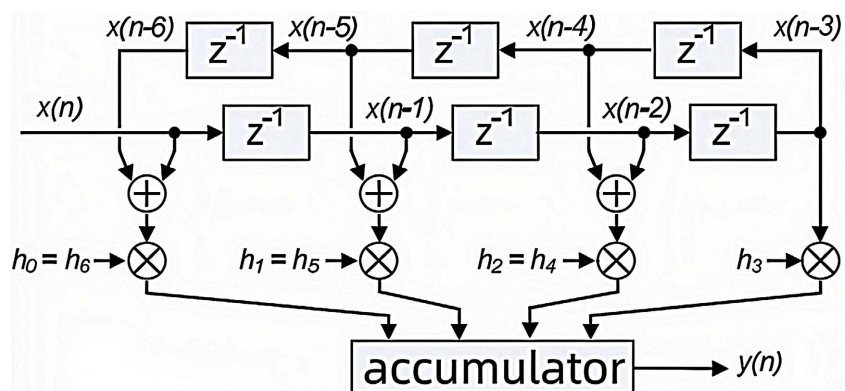
plementing a filter using a completely multiplier free approach.



**Figure 1.** Structure of direct FIR filter.

## 2. Methods and Filter Expression

Due to the linear phase characteristics of FIR filters, their coefficients are symmetric about the center. We can implement this using a folded filter structure, where the corresponding units of the lateral delay are added first and then multiplied by the coefficients. This can reduce the number of multiplications by half, as shown in **Figure 2**.



**Figure 2.** Folding structure of FIR filter.

For traditional quantification methods, their folded structure can be expressed as:

$$y[n] = \sum_{k=-\frac{N+1}{2}}^{\frac{N}{2}} [x[n-N+k] + x[n-k]] \cdot \text{round}[h[k] \cdot 2^{\text{BW}}] / 2^{\text{BW}} \quad (1)$$

$N$  is an odd number.

$$y[n] = x\left[n - \frac{N}{2}\right] \cdot \text{round}\left[2^{\text{BW}} \cdot h\left[\frac{N}{2}\right]\right] / 2^{\text{BW}} + \sum_{k=-\frac{N+1}{2}}^{\frac{N}{2}} [x[n-N+k]] \cdot \text{round}\left[2^{\text{BW}} \cdot h[k]\right] / 2^{\text{BW}} \quad (2)$$

$N$  is an even number.

where  $N$  is the order of the filter,  $h[k]$  is the ideal design coefficient value, and the  $\text{round}[]$  operation takes the closest integer, BW is the design coefficient quan-

tization bit width. When using non equal width quantization, the output is represented as:

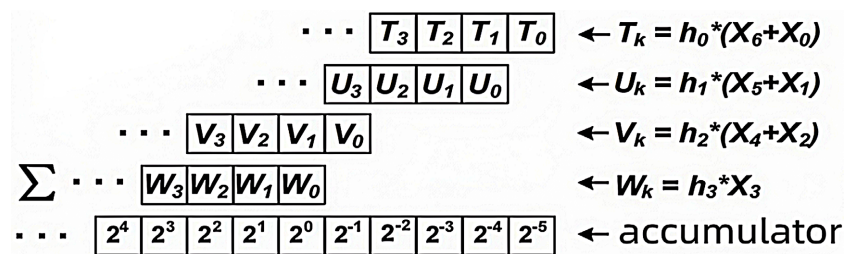
$$y[n] = \sum_{k=\frac{N+1}{2}}^N \left[ x[n-N+k] + x[n-k] \right] \cdot \text{round} \left[ 2^{\text{bw}[k]} \cdot h[k] \right] / 2^{\text{bw}[k]} \quad (3)$$

$$y[n] = x \left[ n - \frac{N}{2} \right] \cdot \text{round} \left[ 2^{\text{bw} \left[ \frac{N}{2} \right]} \cdot h \left[ \frac{N}{2} \right] \right] / 2^{\text{bw} \left[ \frac{N}{2} \right]} + \sum_{k=\frac{N+1}{2}}^N \left[ x[n-N+k] \right] \cdot \text{round} \left[ 2^{\text{bw}[k]} \cdot h[k] \right] / 2^{\text{bw}[k]} \quad (4)$$

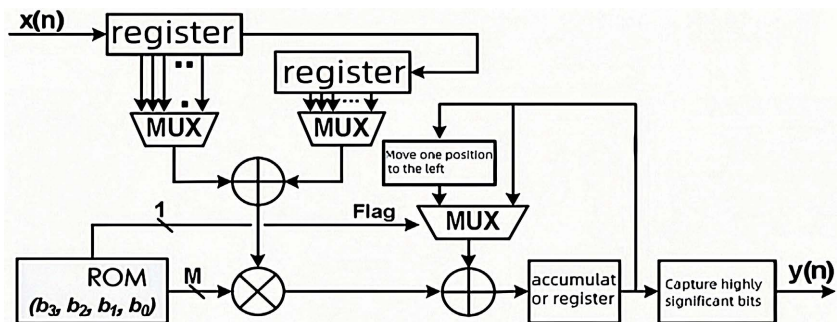
Among them,  $\text{bw}[k]$  is the equivalent quantization bit width of each coefficient. For  $\text{bw}[k]$ :

$$\forall i \in [k, N-1], \exists \text{bw}[i] \leq \text{bw}[i+1], \text{bw}[i] \in N \quad (5)$$

The accumulation method is shown in **Figure 3**. The serial implementation structure is shown in **Figure 4**.



**Figure 3.** Accumulator modification structure.



**Figure 4.** Serial implementation structure.

In this algorithm, due to the different equivalent quantization bit widths of the coefficients in FIR filters with coefficient compression, their quantization methods differ from traditional methods. The comparison of methods is also different. First, shift all coefficients to the power of  $2^M$ :

$$b'_i = b_i \cdot 2^M \quad (i \in [0, n], M \in \mathbb{Z}) \quad (6)$$

Make the coefficient with the maximum absolute value within the normalized range of 0.5 - 1, *i.e.*:

$$\max(|b'|) \in [0.5, 1) \quad (7)$$

Then, for the coefficient with the highest absolute value in the middle, allocate

the same fixed-point quantization bit width as the conventional method to it Quantify. The initial quantization bit width is set to BW based on the actual bit width of the final memory, and the ideal value of the middle coefficient is  $b'_n$ . The quantization operation that shifts  $b'_n$  by BW bit width is defined as  $\text{Qant}[b'_n, \text{BW}]$  and the quantized coefficient value is represented as  $B_n$  by binary complement. Therefore, the binary complement quantization range  $Q$  can be known as:

$$-2^{\text{BW}-1} \leq Q \leq 2^{\text{BW}-1} - 1, Q \in \mathbb{Z} \quad (8)$$

Its equivalent quantization bit width  $\text{bw}[n]$ :

$$\text{bw}[n] = \text{BW} \quad (9)$$

Its quantitative operation is as follows:

$$B_n = \text{Qant}[b'_n, \text{BW}] = \text{round}[b'_n \cdot 2^{\text{BW}}] \quad (10)$$

$$2^{\text{BW}-2} \leq |B_n| \leq 2^{\text{BW}-1} - 1 \quad (11)$$

The quantified range Temp\_Scale is:

$$\text{Temp\_Scale} = 1 \quad (12)$$

The quantization bit width Temp\_BW is:

$$\text{Temp\_BW} = \text{BW} \quad (13)$$

The cumulative left shift  $\text{Flag}_n$  is:

$$\text{Flag}_n = 0 \quad (14)$$

When the quantified range Temp\_Scale reaches half, it can be determined whether the following equation holds:

$$\forall i \in [0, n-1], \exists b'_i < \frac{1}{2} \cdot \text{Temp\_Scale} \quad (15)$$

$$\text{Temp\_BW} = \text{Temp\_BW} \quad (16)$$

$$\text{Flag}_{n-1} = 0 \quad (17)$$

To quantify it, the equivalent quantization bit width is:

$$B_{n-1} = \text{Qant}[b'_{n-1}, \text{Temp\_BW}] = \text{round}[b'_{n-1} \cdot 2^{\text{Temp\_BW}}] \quad (18)$$

$$\text{bw}[n-1] = \text{Temp\_BW} \quad (19)$$

When all unquantified coefficients meet the constraint requirements, there are:

$$\text{Temp\_BW} = \text{Temp\_BW} + 1 \quad (20)$$

$$\text{Flag}_{n-1} = 1 \quad (21)$$

Halving operation:

$$\text{Temp\_Scale} = \frac{1}{2} \text{Temp\_Scale} \quad (22)$$

We can obtain:

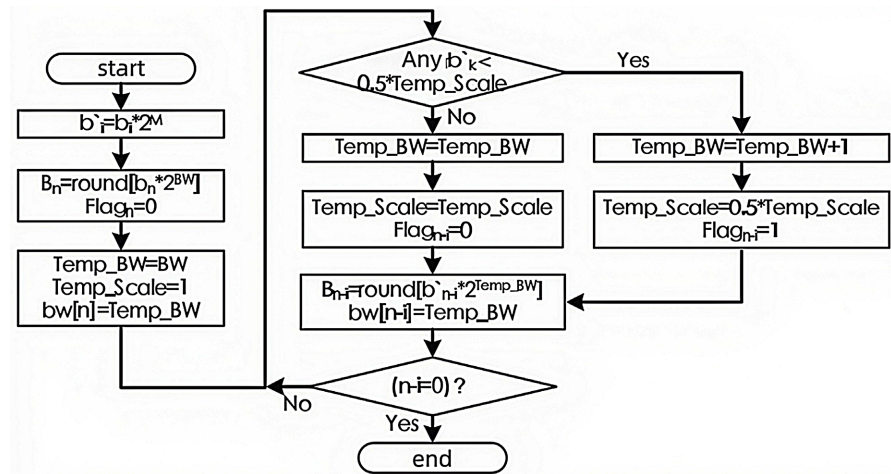
$$B_{n-1} = \text{Qant}[b'_{n-1}, \text{Temp\_BW}] = \text{round}[b'_{n-1} \cdot 2^{\text{Temp\_BW}}] \quad (23)$$

$$bw[n-1] = \text{Temp\_BW} \quad (24)$$

### 3. Example of Filter Design

The serial structure coefficient quantization flow chart is shown in **Figure 5**.

The serial structure quantization process is shown in the following **Table 1**:



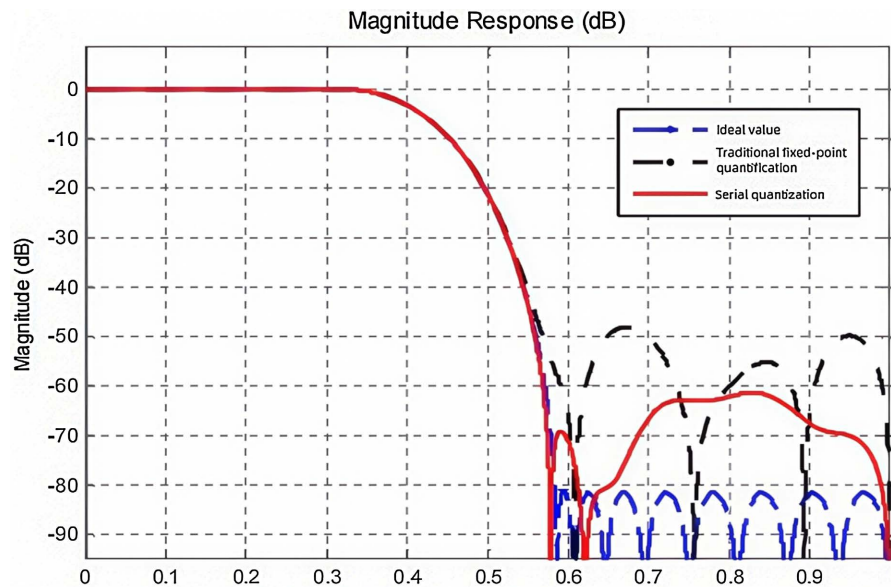
**Figure 5.** Serial structure coefficient quantization flow chart.

**Table 1.** Serial quantization process.

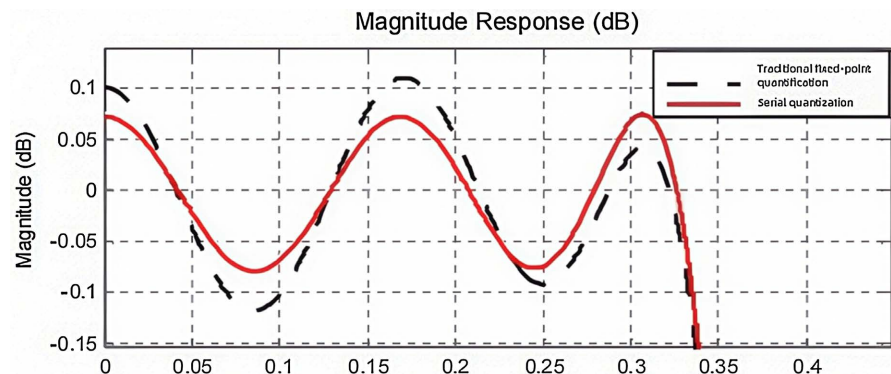
| Coefficient to be quantified $b_{n-1}$ | Unquantified coefficients all $< 0.5\text{Temp\_Scale}$ ? | After update Temp_BW | After update Temp_Scale | Equivalent quantization $n$ bit width $bw[n-1]$ | Shift flag Flag | Left shift and quantification           |
|----------------------------------------|-----------------------------------------------------------|----------------------|-------------------------|-------------------------------------------------|-----------------|-----------------------------------------|
| $b_4 = 0.87968$                        | No $ b_4  > 0.5$                                          | 8                    | 1                       | 8                                               | 0               | $B_4 = \text{round}[b_4 * 2^7] = 113$   |
| $b_3 = 0.37687$                        | Yes $ b_3 ,  b_2 ,  b_1 ,  b_0  < 0.5$                    | 9                    | 0.5                     | 9                                               | 1               | $B_3 = \text{round}[b_3 * 2^8] = 96$    |
| $b_2 = -0.26156$                       | No $ b_2  > 0.25$                                         | 9                    | 0.5                     | 9                                               | 0               | $B_2 = \text{round}[b_2 * 2^8] = -67$   |
| $b_1 = -0.05899$                       | Yes $ b_1 ,  b_0  < 0.25$                                 | 10                   | 0.25                    | 10                                              | 1               | $B_1 = \text{round}[b_1 * 2^9] = -30$   |
| $b_0 = 0.01751$                        | Yes $ b_1 ,  b_0  < 0.125$                                | 11                   | 0.125                   | 11                                              | 1               | $B_0 = \text{round}[b_0 * 2^{10}] = 18$ |

As shown in the above **Figure 6** and **Figure 7**, the new algorithm uses serial quantization, which improves the coefficient accuracy better than traditional fixed-point quantization methods. Specifically, it has more stop band attenuation, smaller transition band width, and smaller pass band ripple. The minimum stop band attenuation of traditional quantization methods is 48 dB, while the improved algorithm is 61.3 dB, and the effect is very significant. The comparison chart of the minimum stop band attenuation of its spectral response is as follows:

As shown in the above **Figure 8**, the new algorithm has more quantization word lengths on both sides of the coefficients, and its accuracy is significantly higher than traditional algorithms at lower quantization memory bit widths.



**Figure 6.** Comparison of amplitude frequency response of serial quantization coefficients.



**Figure 7.** Frequency response pass band details.

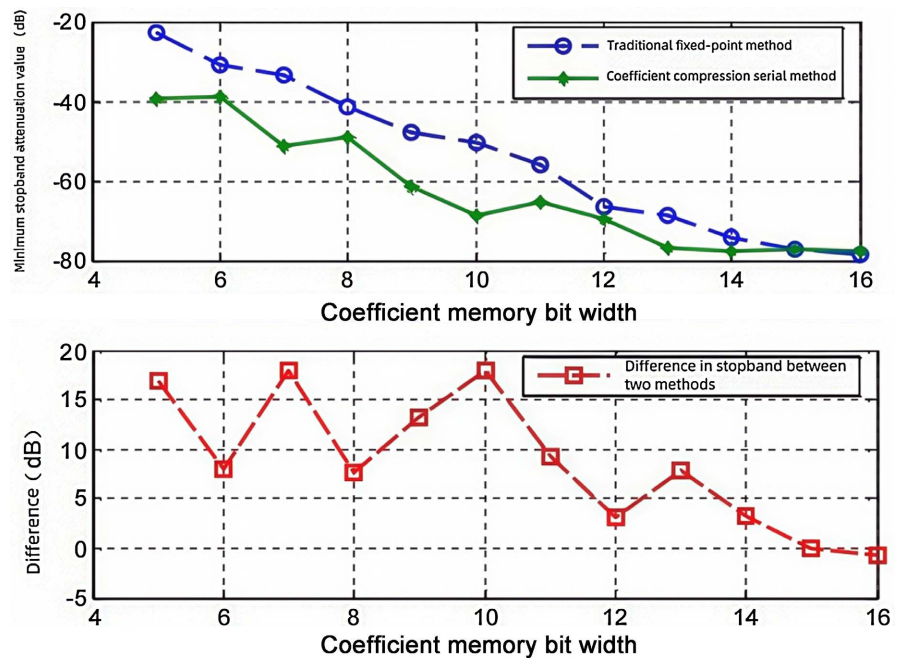
**Table 2.** Resource comparison of equal bit width serial implementation.

| DSP achieve                      | ALUT  | memory | DSP  | memory (bit) |
|----------------------------------|-------|--------|------|--------------|
| Traditional serial structure     | 118   | 240    | 1    | 126          |
| Serial structure of this article | 131   | 261    | 1    | 126          |
| Resource increment               | 11.0% | 8.8%   | 0.0% | 0.0%         |

As shown in the table above (**Table 2**), the implementation of the new algorithm includes a shift binary selector, which is the reason for the increased consumption of ALUT resources.

From the above table (**Table 3**), it can be seen that due to the increase in quantization bit width, ALUT and memory resources have increased by about 15%; The original  $9 \times 9$  multiplier could be implemented using only one DSP, but with the increase in the bit width of the multiplier, the system needs to use an  $18 \times 18$  multiplier, which is equivalent to two DSPs in effect. Therefore, the resources of the multiplier have increased, doubling by 100%.





**Figure 8.** Comparison of stop band attenuation under different position widths.

**Table 3.** Resource consumption of performance prerequisites such as serial implementation.

| DSP achieve                      | ALUT | memory | DSP  | memory (bit) |
|----------------------------------|------|--------|------|--------------|
| Traditional serial structure     | 562  | 1105   | 1    | 594          |
| Serial structure of this article | 611  | 1209   | 1    | 594          |
| Resource increment               | 8.7% | 9.4%   | 0.0% | 0.0%         |

## 4. Conclusion

Due to the use of serial quantization, the accuracy of coefficient quantization values has been significantly improved compared to traditional fixed-point quantization methods, provided that the bit width of the multiplier and coefficient memory is not increased, and the filter order is not increased. Next, parallel structures can be used to further improve its performance, or a series parallel hybrid approach can be adopted to further enhance the accuracy of coefficient quantization values.

## Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

## References

- [1] Shen, Z. (2010) Research on the Implementation of Key Technologies for MC-QAM Modulator Signal Processing. Doctoral Dissertation, Huazhong University of Science and Technology.
- [2] Jovanovic Dolecek, G. (2009) Simple Wideband CIC Compensator. *Electronics Letters*, **45**, 1270-1272. <https://doi.org/10.1049/el.2009.1860>

- [3] Kim, S., Lee, W., Ahn, S. and Choi, S. (2006) Design of CIC Roll-Off Compensation Filter in a W-CDMA Digital IF Receiver. *Digital Signal Processing*, **16**, 846-854. <https://doi.org/10.1016/j.dsp.2006.06.003>
- [4] Torres, F.J.T. and Dolecek, G.J. (2007) Compensated CIC-Cosine Decimation Filter. *Proceedings International Symposium on Communications and Information Technologies, ISCIT'07*, Shanghai, 25-27 September 2007, 256-259.
- [5] Dolecek, G.J. and Mitra, S.K. (2008) Simple Method for Compensation of CIC Decimation Filter. *Electronics Letters*, **44**, 1162-1163. <https://doi.org/10.1049/el:20081603>
- [6] Willson, A.N. (2007) Desensitized Half-Band Interpolation Filters. 2007 50th *Midwest Symposium on Circuits and Systems*, Montreal, 5-8 August 2007, 1034-1037. <https://doi.org/10.1109/mwscas.2007.4488739>
- [7] Fernandez-Vazquez, A. and Dolecek, G.J. (2009) A General Method to Design GCF Compensation Filter. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **56**, 409-413. <https://doi.org/10.1109/tcsii.2009.2019337>
- [8] McClellan, J.H. and Parks, T.W. (2005) A Personal History of the Parks-McClellan Algorithm. *IEEE Signal Processing Magazine*, **22**, 82-86. <https://doi.org/10.1109/msp.2005.1406492>
- [9] Boonyanant, P. and Tantaratana, S. (2005) Design and Hybrid Realization of FIR Nyquist Filters with Quantized Coefficients and Low Sensitivity to Timing Jitter. *IEEE Transactions on Signal Processing*, **53**, 208-221. <https://doi.org/10.1109/tsp.2004.838968>