A Novel Multi-Stage Interpolation Filter Design Technique for High-Resolution Σ-Δ DAC*

Chen Run[1,†], Liu Liyuan[2], and Li Dongmei[2]

(1 *Institute of Microelectronics, Tsinghua University, Beijing* 100084, *China*)
(2 *Department of Electronic Engineering, Tsinghua University, Beijing* 100084, *China*)

Abstract: This paper presents an efficient way to implement an interpolation filter in a 20bit Σ-Δ DAC with an oversampling ratio of 128. A multistage structure is used to reduce the complexity of filter coefficients and the finite word length effect. A novel method based on mixed-radix number representation is proposed to realize a polyphase multiplier-free half-band subfilter with a high resolution. This approach reduces the complexity of the control system and saves chip area dramatically. The IC is realized in a standard 0.13μm CMOS process and the interpolation filter occupies less than 0.63mm^2. This realization has desirable properties of regularity with simple hardware devices which are suitable for VLSI and can be applied to many other high resolution data converters.

Key words: interpolation filter; mixed-radix; multistage; Σ-Δ DAC
EEACC: 1265H
CLC number: TN43 **Document code**: A **Article ID**: 0253-4177(2007)11-1735-07

1 Introduction

Σ-Δ digital-to-analog converters (DACs) are widely used in high quality multi-media processing chips because of their high resolution and their ability to be easily integrated within the digital system due to the advances in VLSI technology. For the digital-to-analog conversion functions, noise shaping and oversampling principles play an important role in inhibiting noise. A general system diagram of a Σ-Δ DAC is depicted in Fig. 1[1]. An interpolation filter changes the input data rate f_N of x to an oversampled value M_{fN}, where M is the oversampling ratio (OSR). The data then enters into a Σ-Δ noise shaping loop (NSL) and a coarse DAC, which carries most of the quantization noise power outside of the baseband and changes the word length to a single bit. The following analog lowpass filter (LPF) suppresses the noise outside of the baseband and generates the analog output y.

This paper primarily focuses on the design method and implementation of the high-performance interpolation filter, which suppresses the image spectrum introduced by oversampling. In order to realize an overall 20bit DAC, the noise has to be attenuated to under a −120dB level. In many cases, a linear-phase finite impulse response (FIR) filter is used since it has symmetric coefficients. In order to avoid using a hardware multiplier, which occupies a great amount of chip area, many multiplier-free realizations are proposed and most of them use sum-of-powers-of-two (SOPOT) to approximate coefficients[2,3], i.e.,

$$a = \sum_{i=1}^{N} c_i 2^{-i} \tag{1}$$

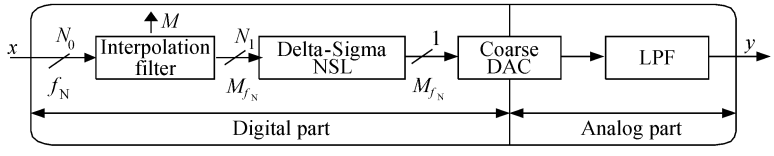

Fig. 1 Block diagram of a Σ-Δ DAC

* Project supported by the National High Technology Research and Development Program of China (No. 2004AA1Z1100)
† Corresponding author. Email: chen-r02@mails.tsinghua.edu.cn
Received 5 April 2007, revised manuscript received 17 July 2007

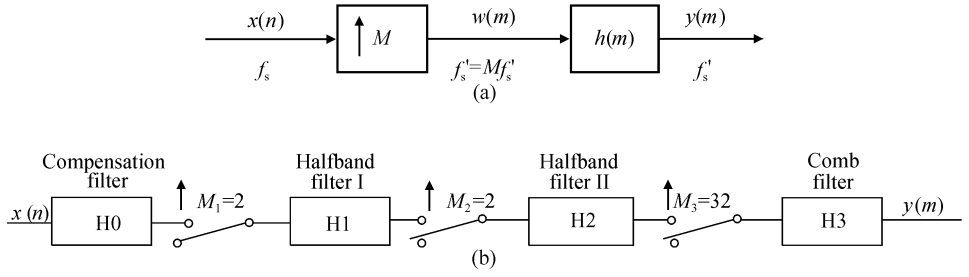

Fig. 2 Block diagram for an interpolating filter (a) General structure; (b) Four-stage structure

where a is a filter coefficient, $c_i \in \{-1, 0, 1\}$, and N determines the approach precision. In practice, N should be very large to achieve high data resolution, whereas the resulting coefficients based on SOPOT are not uniform since the powers of two can be quite different from each other. This results in irregularity of VLSI implementation and greater hardware expenses because especially long registers are needed to store small numbers. Therefore, a SOPOT method leads to a trade-off between data precision and chip area; thus, more design effort is required. Compared to the uni-radix method such as SOPOT, another approach is to approximate the coefficients by mixed-radix number representation (MRNR)[4]. This method provides much higher bit precision and can be regularly designed. Moreover, it reduces hardware resources dramatically. However, it needs a comparatively complicated control system, thus a higher clock speed is required. This paper combines the merits of MRNR and the half-band FIR filter technique to obtain high resolution while using relatively less hardware resources and a less complicated control system. In addition, a multistage structure and polyphase architecture is used to further reduce the algorithmic complexity of the interpolation filter.

2 Architecture and algorithm

The interpolation filter is composed of an interpolator and a digital filter, as shown in Fig. 2 (a). The interpolator inserts zero-valued samples between each sample of input data $x(n)$; $\uparrow M$ means to insert $M-1$ zeros between adjacent samples. The intermediate signal $w(m)$ has $M-1$ imaged replications of spectrum of $x(n)$, which should be attenuated by the filter $h(m)$. High performance of the filter requires a very narrow transition bandwidth, so the filter must have high orders. A multistage structure is preferred because it is easy to design, and it reduces the complexity of filter coefficients and finite word length effect. This paper describes the design of a four-stage interpolation filter as shown in Fig. 2 (b). The first subfilter H0 compensates the passband amplitude distortion caused by the following subfilters. The overall passband response stays well within the limits of ± 0.003dB. The second subfilter H1 is a half-band filter, which provides at least 120dB attenuation on stopband and the normalized (by sampling rate) transition bandwidth is less than 0.1. Its performance determines the precision of the overall interpolation filter. H2 is also a half-band filter which attenuates the image spectrum caused by the second oversampling unit. After subfilters H1 and H2, the following subfilter is less stringently required. Since the comb filter has a simple structure and is easy to implement in hardware, it constitutes the main part of H3.

2.1 Half-band filter design

A half-band filter is a special linear-phase FIR filter with the order of $N_f - 1$ (N_f is an odd number). The impulse response $h(n)$ satisfies

$$h(n) = h(N_f - 1 - n) \tag{2}$$

Using another two important characteristics: (1) Passband ripple δ_p equals stopband ripple δ_s; (2) Band edges are related to $\omega_p + \omega_s = \pi$, we find $h(n)$ satisfies

$$h(n) = \begin{cases} 0.5, & n = (N_f - 1)/2 \\ 0, & n = \text{other odd numbers} \end{cases} \tag{3}$$

Nearly half of the coefficients are zero. Take H2 for example. The stopband attenuation is designed to be above 100dB, and the normalized transition bandwidth is less than 0.55. The filter coefficients are listed in Table 1 by using the equiripple design method[5]. It has the order of 26, 15 of which are nonzero numbers. Actually, there are only 8 unique numbers when taking account of symmetric

Table 1　Coefficients of half-band filter H2

n	$h(n)$	n	$h(n)$
0,26	0.00018903727821	8,18	0.03640237667686
2,24	−0.00128751925222	10,16	−0.08707739259022
4,22	0.00503747813522	12,14	0.31145585454288
6,20	−0.01471705807905	13	0.50000000000000
others	0		

characteristics. Table 2 gives the coefficients of H1.

2.2　Structure of half-band filter based on MRNR

Mixed-radix signed-digit number representations (MRNR) with periodically time varying (PTV) coefficients can provide much higher coefficient approach precision than the SOPOT method. Given the filter coefficients $h(n)$, an NK-digit MRNR can be written as[6]

$$h(n) = C\sum_{k=0}^{K-1}\sum_{i=1}^{N} c_n^i(k) r_1^{-i} r_2^{-(K-1-k)} = Cr_1^{-1}\sum_{k=0}^{K-1}\sum_{i=0}^{N-1} c_n^i(k) r_1^{-i} r_2^{-(K-1-k)},\ n = 0,1,\cdots,M \tag{4}$$

where r_1 and r_2 are the radices, and M is the number of coefficients. When $r_1 = 4$, PTV coefficients $c_n^i(k)$ are signed digits belonging to the set $\{0, \pm 1, \pm 2\}$, and the optimum choice of r_2 is

$$(r_2^{-1})_{\text{opt}} = \frac{3}{4^{N+1}-1} = 2^{-p} + 2^{-q} \tag{5}$$

The value of C that normalizes the range of representation $h(n)$ to $[-1,1]$ is in the form of $1 \pm 2^{-s} \pm 2^{-t}$. Here p, q, s, and t are the natural numbers. According to Eq. (4), the half-band filter can be implemented by only shifters and adders. Since $c_n^i(k)$ belongs to the set $\{0, \pm 1, \pm 2\}$, the shift operation is simply one or two bits, which can be hardwired. This method avoids long bit shift operations, thus improving the coefficient approach precision deduced from Eq. (4) as

$$P = -\log_2(Cr_1^{-N} r_2^{-(K-1)}) + 1 \tag{6}$$

where P is the approach precision. From Eq. (6), a larger N or K yields higher precision. This paper selects $N = 3$ and $K = 5$, so the bit precision can be up to 32bit, which can be hardly achieved by the SOPOT method. According to Eq. (4), $c_n^i(k)$ make up an $N \times K$ matrix, written as

$$MR_{h(n)} = \begin{bmatrix} c_n^0(0) & c_n^0(1) & c_n^0(2) & c_n^0(3) & c_n^0(4) \\ c_n^1(0) & c_n^1(1) & c_n^1(2) & c_n^1(3) & c_n^1(4) \\ c_n^2(0) & c_n^2(1) & c_n^2(2) & c_n^2(3) & c_n^2(4) \end{bmatrix} \tag{7}$$

The matrix actually completes a PTV data process, which is depicted in Fig. 3. The switch connects from 1 to 5 periodically, and the function f is determined by

$$f(c_n^0(k), c_n^1(k), c_n^2(k)) = \sum_{i=0}^{2} MR_{h(n)}(i+1, k+1) \cdot r_1^{-i},\quad k = 0,1,\cdots,4 \tag{8}$$

Based on Eq. (4) and Fig. 3, a half-band filter prototype structure is proposed in Fig. 4 (a). The block D is a D flip-flop (DFF) as a delay unit. The block DSH is a digital sample and hold unit, which keeps the input data K times, so the signal rate of DSH output is K times the input data rate.

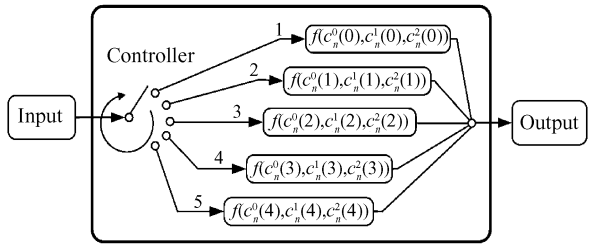

Fig. 3　PTV data process by $MR_{h(n)}$

Table 2　Coefficient of half-band filter H1

n	$h(n)$	n	$h(n)$	n	$h(n)$	n	$h(n)$
0,162	0.000000490846	22,140	−0.000146720483	44,118	0.002380922775	66,96	−0.017333390917
2,160	−0.000001139478	24,138	0.000201412111	46,116	−0.002899948773	68,94	0.021039274037
4,158	0.000002423614	26,136	−0.000272043421	48,114	0.003510540995	70,92	−0.025963893248
6,156	−0.000004618544	28,134	0.000362071671	50,112	−0.004226762620	72,90	0.032894964121
8,154	0.000008165154	30,132	−0.000475454602	52,110	0.005065520799	74,88	−0.043524175418
10,152	−0.000013650549	32,130	0.000616687385	54,108	−0.006047669323	76,86	0.062256284634
12,150	0.000021838608	34,128	−0.000790843562	56,106	0.007199681826	78,84	−0.105254308062
14,148	−0.000033702877	36,126	0.001003622789	58,104	−0.008556241936	80,82	0.318025956122
16,146	0.000050461492	38,124	−0.001261411343	60,102	0.010164361887	81	0.500000000000
18,144	−0.000073613426	40,122	0.001571362382	62,100	−0.012090146949	others	0
20,142	0.000104975695	42,120	−0.001941508845	64,98	0.014430376956		

Each DSH output goes through M branches of PTV coefficients with scaling by $r_1^{-i}(i=0,1,\cdots,N-1)$. Here $M=(N_f+5)/4$, so there are 8 branches for subfilter H2. Consider that the block in Fig. 3 is reused K times per sample. It requires NM shift operations and $NM-1$ addition operations within $1/K$ time of the input signal period.

Close scrutiny of Fig. 4(a) reveals that further simplification can be achieved. In Fig. 4 (a), a signal at point a has a zero between every two samples due to oversampling. These zeros will be processed in the same way as nonzero numbers. Thus, this wastes time and hardware resources. A modified structure is proposed based on a simple linear transformation[4], and the PTV coefficients are changed to be

$$d_n^i(\mu K+\eta)=c_{\mu+2n}^i(\eta),\quad \eta=0,1,\cdots,K-1 \text{ and } \mu=0,1 \tag{9}$$

This realization saves half of the DFF delay units to store data. However, it increases the burden of timing control because the MR matrix is extended to

$$MR^*_{h'(n^*)}=\begin{bmatrix} d^0_{n^*}(0) & d^0_{n^*}(1) & d^0_{n^*}(2)\cdots d^0_{n^*}(9)\\ d^1_{n^*}(0) & d^1_{n^*}(1) & d^1_{n^*}(2)\cdots d^1_{n^*}(9)\\ d^2_{n^*}(0) & d^2_{n^*}(1) & d^2_{n^*}(2)\cdots d^2_{n^*}(9)\end{bmatrix},\quad n^*=0,1,\cdots,M-2 \tag{10}$$

It requires $N(M-1)$ shift operations and $N(M-1)-1$ addition operations within $1/2K$ time of the input signal period. By comparing Eqs. (7, 9, 10), we find

$$MR^*_{h'(n^*)}=(MR_{h(2n^*)},MR_{h(2n^*+1)}) \tag{11}$$

Since about half of the half-band filter coefficients are zero and it is common that the MR matrix of 0 is zero matrix $\Theta_{N\times K}$, we have

$$MR_{h(2n^*+1)}=\begin{cases} MR_{0.5}, & n^*=M-2\\ \Theta_{N\times K}, & n^*=\text{others}\end{cases} \tag{12}$$

Equation (12) shows that there is only one special coefficient 0.5 that makes the pattern of $MR_{h'(2M-3)}$ different from the other MR^* matrixes, i.e.,

$$MR^*_{h'(n^*)}=\begin{cases}(MR_{h(2n^*)},MR_{0.5}), & n^*=2M-3\\ (MR_{h(2n^*)},\Theta_{N\times K}), & n^*=\text{others}\end{cases} \tag{13}$$

From Eq. (13), the modified structure can be effectively simplified by the following representation:

$$(MR_{h(2n^*)},MR_{0.5})=(MR_{h(2n^*)},\Theta_{N\times K})+(\Theta_{N\times K},MR_{0.5}) \tag{14}$$

Polyphase architecture helps to realize the representation in Eq. (14). Actually, the coefficient 0.5 does not need any transformation and can be easily implemented by a 1bit shifter. A double-phase half-band filter is designed using this method and the coefficient $g_m(n)$ satisfies:

$$g_m(n)=\begin{cases}g_0(n)=h(n)-0.5\delta(n-2M+3)\\ g_1(n)=0.5\delta(n-2M+3)\end{cases} \tag{15}$$

From Eq. (15), we see half of the elements of $MR^*[g'_0(n^*)]$ are zero, which saves half the time of operations, thus the control complexity is reduced. This novel structure is depicted in Fig. 4 (b). It completes $N(M-1)$ shift operations and $N(M-1)-1$ addition operations within $1/K$ time of the input signal period. It uses only half of the DFFs as the prototype and it maintains high coefficient precision. A comparison of structures discussed here is listed in Table 3. The performance advantage of the proposed structure is apparent.

2.3 Comb filter and compensation filter design

The comb subfilter is designed to accomplish OSR = 32 oversampling and to filter the image spectrum. We adopt a 4-order 5-stage multistage structure. The transfer function of each stage is:

$$H_{\text{each\_stage}}=(1+z^{-1})^4=(1+z^{-4})+4(z^{-1}+z^{-3})+6z^{-2} \tag{16}$$

The poly-phase method, as stated above, is used to reduce the control complexity. The coefficients of the comb filter are integers so it is easy to implement the structure using simple shifters and adders. The amplitude response of the comb filter reveals that there is distortion at about 1dB in the passband. In order to compensate for this effect, the frequency sampling method[7] is used to design a pre-subfilter H0, which has the order of 10 and is realized also based on MRNR.

3 Realization and experiment results

According to section 2, the basic operations are shift and addition. The shifters and adders can be reused several times as multiplexing cells. A general multiplexing cell is designed as shown in Fig. 5 (a). It comprises two 16-1 multiplexers (Mux1 and Mux2), two shifters, two 2-1 multi plexers (Mux3 and Mux4), a full adder, and two

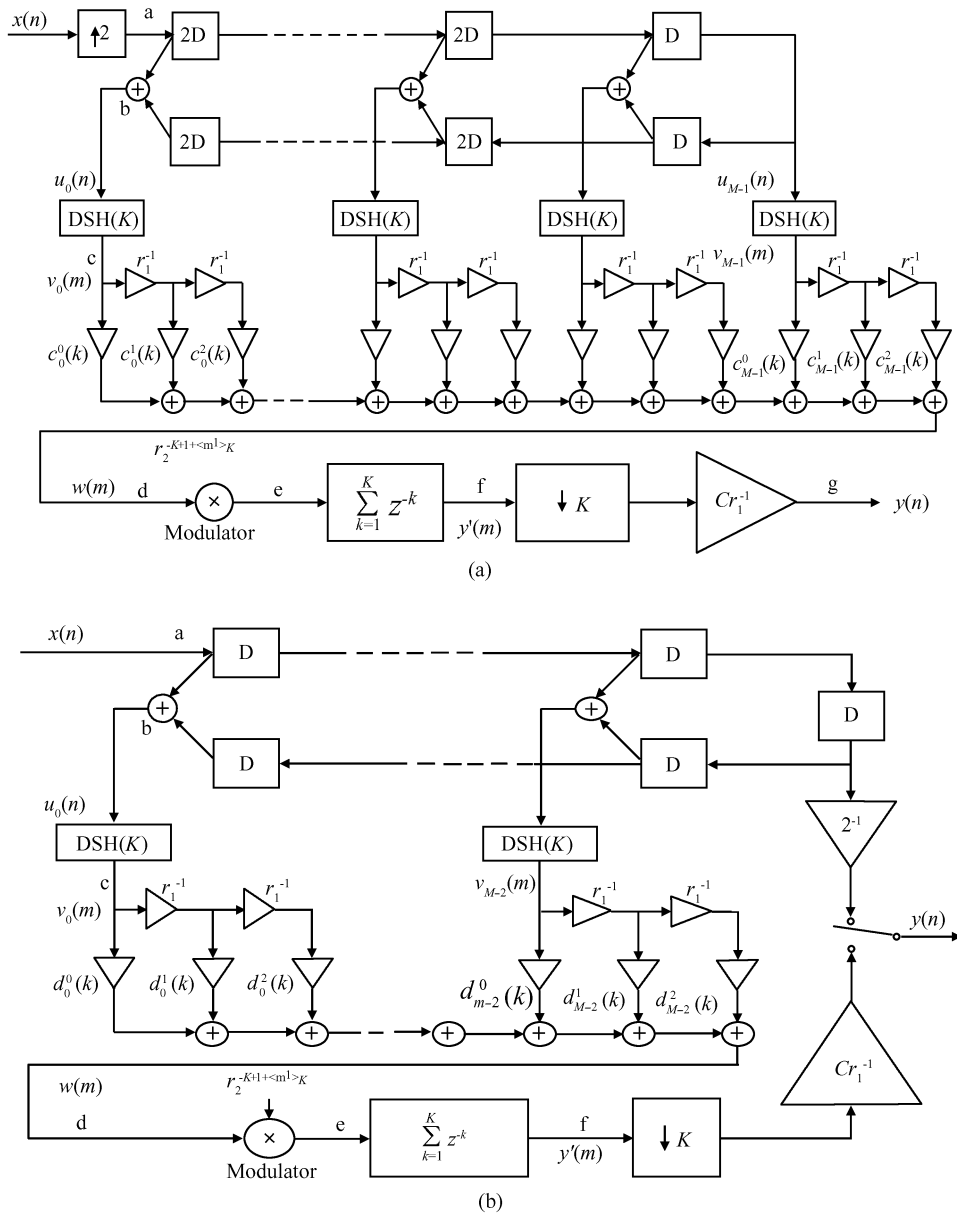

Fig. 4　Half-band filter with multifree realization based on MRNR　(a) Prototype structure (N = 3, K = 5); (b) Double-phase structure

DFFs. The control code of such shifters and multiplexers constitute the instruction of the cell, as shown in Fig. 5 (b). The instruction is 18bits long to complete the following functions: shift operation, direct addition, direct subtraction, addition with shift operation, subtraction with shift operation, and accumulation. The number of multiplexing cells depends on the complexity of the filter itself and the signal rate. For instance, H1 needs 5 cells while H2 needs 2.

Table 3　Comparison of different structures

Structure	Registers needed in the delay chain	Least operation times during a signal period		Coefficient approach precision
		shift	addition	
SOPOT method[Eq. (1)]	$(4M-6)\times$ word length	–	–	hard to achieve 20bit
Prototype structure [Fig. 4(a)]	$(4M-6)\times$ word length	NMK	$(NM-1)K$	32bit (N = 3 and K = 5)
Modified structure [Eq. (9)]	$(2M-1)\times$ word length	$2N(M-1)K$	$2[N(M-1)-1]K$	32bit (N = 3 and K = 5)
Proposed structure [Fig. 4(b), this work]	$(2M-1)\times$ word length	$N(M-1)K$	$[N(M-1)-1]K$	32bit (N = 3 and K = 5)

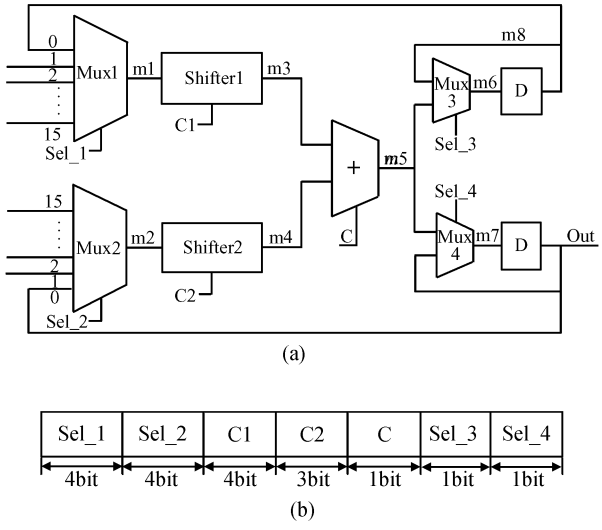

Fig. 5 (a) A general multiplexing cell; (b) Instruction format for a multiplexing cell

The interpolation filter is realized by using VHDL, and synthesized by DC tools. Figure 6 gives the microphotograph of the overall 20bit DAC, which is mainly comprised of two parts: the digital part and the analog part. The DAC is fabricated in a standard 0.13μm CMOS process. The digital part is composed of an interpolation filter and a noise shaping loop. The filter takes up about 90% of total digital area, which is 0.9mm × 0.7mm = 0.63mm^2.

Figure 7 shows the measured output spectrum of the interpolation filter by impulse input. The subfigures have the same dimension with the main figure. The input word length is 21bit (including a sign bit). We use an Agilent 16702B logic analyzer to capture the output data of the chip. The captured data are then analyzed using Matlab. Results from the analysis show that the normalized (by sampling rate) passband edge is 0.0075 and the stopband attenuation is −121dB. The first sidelobe is suppressed to −113dB; the second sidelobe

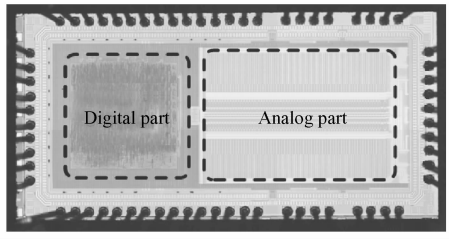

Fig. 6 Microphotograph of 20bit Σ-Δ DAC

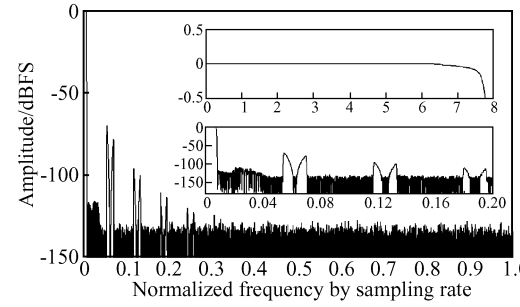

Fig. 7 Frequency response of the interpolation filter

is under −70dB. The passband ripple is with in ± 0.003dB. Figure 7 also shows that the noise floor is about −130dB.

4 Conclusions

This paper presented an efficient way to implement a linear-phase FIR interpolation filter in a 20bit Σ-Δ DAC. A multistage structure was used to reduce the complexity of filter coefficients and the finite word length effect. A novel method based on MRNR was used to design a double-phase half-band subfilter. It reduces the complexity of the control system, obtains a much higher coefficient approach precision, and reduces chip area dramatically compared to conventional methods.

References

[1] Norsworthy S R, Schreier R, Temes G C. Delta-sigma data converters-theory, design and simulation. IEEE Press, 1996: 309

[2] Zhao Q, Tadokoro Y. A simple design of FIR filters with powers-of-two coefficients. IEEE Trans Circuit Syst, 1988, 35: 566

[3] Horng B R, Samueli H, Willson A N Jr. The design of two-channel lattice structure perfect-reconstruction filter banks using powers of two coefficients. IEEE Trans Circuits Syst I, 1993, 40: 497

[4] Li J L, Tantaratana S. Multiplier-free realizations for FIR multirate converts based on mixed-radix number representation. IEEE Trans Signal Process, 1997, 45: 880

[5] Oppenheim A V, Schafer R W, Buke J R. Discrete-time signal processing. Upper Saddle River, NJ: Prentice Hall, 1999

[6] Ghanekar S, Tantaratana S. Signal-digit based multiplier-free realizations for multirate converters. IEEE Trans Signal Process, 1995, 43: 628

[7] Rabiner L R, Gold B, McGonegal C A. An approach to the approximation problem for nonrecursive digital filters. IEEE Trans Audio and Electroacoustics, 1970, AU-18: 83

用于高精度 Σ-Δ 数模转换的多级插值滤波器的设计技术*

陈　润[1,†]　刘力源[2]　李冬梅[2]

（1 清华大学微电子学研究所，北京　100084）
（2 清华大学电子工程系，北京　100084）

摘要：提出了一种用于 20bit Σ-Δ 数模转换器中的内插滤波器的有效实现方法，内插滤波器的过采样率为 128. 该方法使用多级结构以降低滤波器系数的复杂度和有限字长效应. 同时提出了基于系数混合基分解的多相半带滤波器的无乘法器实现方法，它降低了控制逻辑的复杂程度，并大大节省了芯片面积. 芯片采用 0.13μm CMOS 工艺实现，整个插值滤波器面积小于 0.63mm^2. 整个电路系统仅用简单的硬件单元实现，且结构规整，这有利于大规模集成电路制造，并可应用于高精度数据转换电路中.

关键词：插值滤波器；混合基；多级；Σ-Δ 数模转换器
EEACC：1265H
中图分类号：TN43　　**文献标识码**：A　　**文章编号**：0253-4177(2007)11-1735-07

* 国家高技术研究发展计划资助项目（批准号：2004AA1Z1100）
† 通信作者. Email：chen-r02@mails.tsinghua.edu.cn
2007-04-05 收到，2007-07-17 定稿