doi:10.3772/j.issn.1002-0470.2023.06.003

基于一种新的三模集的低通 FIR 数字滤波器设计^①

赵海军② 陈毅红 贺春林 蒲 斌

(西华师范大学计算机学院 南充 637009)

(物联网感知与大数据分析南充市重点实验室 南充 637009)

摘 要 本文对低通有限脉冲响应(FIR)数字滤波器的设计进行了研究;在对常用三模集和传统的具有脉冲响应系数的 FIR 数字滤波器分析的基础上,提出将一种新的三模集应用于低通 FIR 数字滤波器设计的新方法。该 FIR 滤波器由多个乘法-累加器单元构成,实现正向转换器、并行算术通道和逆向转换器的功能。正向转换器将一个二进制数编码成一个关于模集的余数表示的数,每个算术通道对模集的每个模进行模乘和累加,逆向转换器将一个余数表示的数解码为其等效的二进制数,算术通道以完全并行的架构工作,从而实现轻量化的计算和简单的结构;最后的仿真实验结果不仅验证了本文提出的设计方法的有效性,而且相比于传统的单向转换设计,本文提出的设计方法有更小的滤波器系数偏差和更好的频率响应等波纹特性。

关键词 低通有限脉冲响应(FIR)数字滤波器;余数系统(RNS);三模集;乘法-累加器单元:脉冲响应:量化误差

0 引言

余数系统(residue number system, RNS)是可用于数字信号处理计算算法高速硬件实现的一种有效替代数字系统^[1-2]。在 RNS 中,把一个较大字长的整数值用一个特定的模集分割成几个相对较小的整数副本,然后对 RNS 整数副本并行地执行加法和乘法运算,每个副本称为一个通道,且 RNS 通道执行模算术。这样,RNS 算术就不会受到通道间传播延迟的影响。系统性能可以通过选择较小的字长通道和较短的内部进位延迟来提高。因此,RNS 成为实现高速有限脉冲响应(finite impulse response,FIR)滤波器的一种有效方法,其中主要的运算是加法和乘法。

由于这一特性,人们基于 RNS 提出了多种数字信号处理架构。文献[3]提出了一种基于余数系统

和有限域置换多项式的混沌序列生成方法,该方法可以降低硬件实现中迭代环路的计算位宽,并提高生成速率。文献[4]提出了一种基于余数系统和蒙哥马利模乘算法的抗攻击算法。采用余数系统把大数运算转变成小数运算,具有一定的理论研究意义与实际应用意义,这为数字信号处理计算算法高速硬件实现提供了新的思路。文献[5]提出了基于RNS的奇偶结构误差检测与校正方案。采用180 nmCMOS标准单元库的综合结果表明,该方案在延迟、功耗和面积开销方面比传统的RNS有所提高,而不失去误差检测与校正能力。

每个 RNS 的基本要素是一个模集,它由一组成对的质数构成^[6]。针对 RNS 已提出了多种模集的应用,其中模集 {2ⁿ - 1, 2ⁿ, 2ⁿ + 1} 最为常见。尽管这种模集可以简化正/逆转换器的设计,但 RNS 算术单元的性能受到模 2ⁿ + 1 通道的时间性能限

① 国家自然科学基金(61871330)资助项目。

② 男,1966 年生,教授;研究方向:网络数据通信及电子信息;E-mail: zhaohai _ jun@163.com。(收稿日期:2022-05-07)

制。文献[7]采用一类复合形式 $\{2^k, 2^p - 1\}$ 的 RNS 模集提出了一种新的方法来设计具有符号输 出的逆向转换器。对所采用的模加法器结构进行改 进,以重用内部电路产生有符号输出。实验结果表 明,对于一个4模转换器,该设计优于传统方法。文 献[8]以 $\{2^n-1,2^n,2^n+1,2^{n-1}-1,2^{n+1}-1\}$ 为 余数基,设计了一种 128 抽头 FIR 滤波器,利用基于 华莱士(Wallace)树结构的纯组合逻辑电路,实现了 二进制到余数的转换。文献[9,10]提出了一种采 用混合-基转换(mixed-radix conversion, MRC)技术, 分别将三模集 $\{2^{n-1}, 2^n - 1, 2^{n+k}\}$ 和 $\{2^m - 1, 2^m,$ 2" +1 的 RNS 转换到二进制转换器。并设计了 2 个逆向转换器,从而获得硬件资源和转换时间的权 衡。但这些利用 RNS 模集要么是二进制到余数的 正向转换,要么是从 RNS 转换到二进制转换器的逆 向转换,这种单向转换存在不足之处。

对此,本文提出采用三模集 {2ⁿ-1, 2ⁿ, 2ⁿ⁺¹-1} 用于低通 FIR 数字滤波器的设计。具体而言,从 FIR 结构的角度出发,将基于单个 RNS 的乘法-累加器 (multiplication-accumulator, MAC) 结构应用于每个通道 FIR 滤波器设计,实现正向转换器、并行算术通道和逆向转换器的功能,从而实现轻量化的计算和简单的结构;最后的仿真实验结果不仅验证了本文提出的设计方法的有效性,而且证明了相比于传统的单向转换设计本文设计有更小的滤波器系数偏差和更好的频率响应等波纹特性。

1 预备知识

RNS 是由 k 个互质的整数定义的,k 个互质的整数即 $m = \{m_1, m_2, \cdots, m_k\}$, 其中对于 i, j (i, j = 1, \cdots ,k) 且 i < j, 有 $gcd(m_i, m_j) = 1$, $gcd(\cdot)$ 表示求最大公约数[11]。任意整数 $X(0 \le X < M$, 其中 $M = \prod_{i=1}^k m_i$)在 RNS 中表示为 $X = \{x_1, x_2, \cdots, x_k\}$,其中 $x_i = \langle X \rangle_{m_i}$, $\langle X \rangle_{m_i}$ 表示 X 模 m_i 运算,即求余数。

RNS 是一种非加权数字系统,它通过将算术运算分割成更小的运算来加速算术运算。由于每个模通道的算术运算独立于其他模通道,在它们之间不

存在进位传播,因此 RNS 得到的是无进位加法和无借位减法。

为了实现余数到二进制转换,也就是将余数 $\{x_1, x_2, \dots, x_k\}$ 转换为整数 X, 一般采用中国余数 定理(Chinese remainder theorem, CRT) $^{[12]}$ 和混合-基转换(MRC) $^{[9-10,13]}$ 技术。

CRT 的计算公式为[14]

$$X = \langle \sum_{i=1}^{k} \langle x_i N_i \rangle_{m_i} M_i \rangle_M \tag{1}$$

式中 $M_i = M/m_i$ 且 $N_i = \langle M_i^{-1} \rangle_{m_i}$ 是 M_i 的逆模 m_i 运算。这种方法的主要缺点是它需要 M_i (通常为很大的数)的逆模 m_i 运算和模 M 运算。

MRC 将余数 $\{x_1, x_2, \dots, x_k\}$ 转换为整数 X 的 计算公式为

$$X = a_k \prod_{i=1}^{k-1} M_i + \dots + a_3 m_1 m_2 + a_2 m_1 + a_1$$
(2)

式中 a_i 为混合-基数字(mixed-radix digit,MRD),可以从余数中得到:

$$a_{i} = \left\langle \left(\cdots \left(\left(x_{i} - a_{1} \right) c_{1,i} - a_{2} \right) c_{2,i} - \cdots \right. \\ \left. - a_{i-1} \right) c_{i-1,i} \right\rangle_{m}$$
(3)

式中 $c_{i,j}$ 对于 $1 \le i \le j < k$ 为 m_i 模 m_j 的逆,或者 $\langle c_{ij} \times m_i \rangle_{m_i} = 1$ 对于k > 1且 $a_1 = x_1 \circ$

2 基于 RNS 的三模集 FIR 滤波器设计

具有脉冲响应系数 b_k 的 N 抽头 FIR 数字滤波器一般描述为

$$y(n) = \sum_{k=0}^{N-1} b_k x(n-k)$$
 (4)

这种 FIR 滤波器的可能实现^[15] 如图 1 所示,可以看出,需要执行 N 次乘法和 N-1 次加法。乘法在硬件和计算时间方面是非常昂贵的,因为算术单元要对以 2 的补码形式表示的数字执行定点计算,而且算术单元由专用硬件乘法器和与累加器连接的

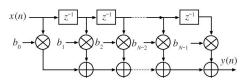


图 1 传统的 FIR 滤波器结构

加法器构成,以便能够有效执行乘法-累加操作。为此,本文提出了基于 RNS 的三模集 $\{m_1, m_2, m_3\}$ 的 MAC 结构来实现它。

图 2 所示为本文提出的基于 RNS 的三模集 $\{m_1, m_2, m_3\}$ 的 MAC 结构。RNS 的主要组成部分是正向转换器、并行算术通道和逆向转换器。正向转换器将一个二进制数编码成一个关于模集的余数表示的数,每个算术通道需要对模集的每个模进行模乘和累加;逆向转换器将一个余数表示的数解码为其等效的二进制数,算术通道以完全并行的架构工作,没有任何关联,这使得运算速度大幅提高。

该结构有 2 个独立的存储空间,可以同时访问。 其中一个存储器用于存储系数,另一个存储输入数 据样本,它们都以余数的形式。通过采用三模 FIR 滤波器模块在 RNS 域实现 FIR 滤波。在每个模 m_1 , m_2 和 m_3 上,FIR 滤波执行一系列模 MAC 运算。

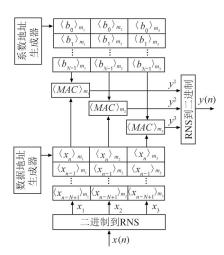


图 2 基于 RNS 的三模集 FIR 滤波器结构

2.1 乘法-累加单元

对于本文提出的 RNS FIR 滤波器来说,式(4) 可表示为

$$\langle y(n) \rangle_{m_i} = \langle \sum_{k=0}^{N} \langle b_k \rangle_{m_i} \langle x(n-k) \rangle_{m_i} \rangle_{m_i}$$

$$i = 1, 2, 3 \quad (5)$$

事实表明, FIR 滤波器模 m_i 的设计式(5)实际上是一个乘积算法的总和,需要一个模 m_i MAC 单元。

图 3 所示为模 m_i MAC 单元,它将 2 个数 $\langle b_k \rangle_{m_i}$ 和 $\langle x(n-k) \rangle_{m_i}$ 相乘,并对结果求和。乘法和累加

操作可由 MAC 单元在单个周期中执行。设 $\langle b_k \rangle_{m_i}$ 和 $\langle x(n-k) \rangle_{m_i}$ 为 2 个输入序列,在图 3 所示的 MAC 单元中,输入相乘并添加 0,最初结果存储在内存(寄存器)中,然后将和存储在内存单元中。在下一个时钟,接下来的输入相乘并加上先前存储在内存中的数据。

例如,最先计算乘法 $z_0 = \langle b_0 \rangle_{m_i} \langle x(n) \rangle_{m_i}$,然后将模 m_i 结果 z_0 加到余数积 $\langle b_1 \rangle_{m_i} \langle x(n-1) \rangle_{m_i}$ 中,得到中间量 $z_1 = z_0 + \langle b_1 \rangle_{m_i} \langle x(n-1) \rangle_{m_i}$ 。在 N次加法和乘法后,递归地得到结果 $\langle y(n) \rangle_{m_i}$ 。因此,最终 结果 y(n) 通过 RNS 结果 $\{\langle y(n) \rangle_{m_1}$, $\langle y(n) \rangle_{m_2}$, $\langle y(n) \rangle_{m_3}$ }的余数到二进制转换得到。

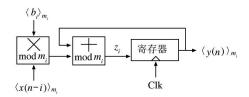


图 3 模 m_i MAC 单元

以下提出可用于 MAC 单元实现的模加法器和模乘法器。

2.1.1 模加法

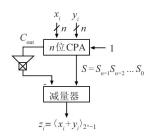
避免零的双重表示的模 (2^n-1) (模 $2^{n+1}-1$) 加法定义为

 $\langle x_i + y_i \rangle_{2^{n-1}} = \langle x_i + y_i + 1 - \overline{C}_{out} \rangle_{2^n}$ (6) 式中 \overline{C}_{out} 是加法 x + y 的输出。图 4(a) 所示为相应运算的硬件结构,它包括一个进位传播加法器(carry propagate adder, CPA)、一个 NOR 门和一个减量器,n 位 CPA 实现来自两路数据 x_i 和 y_i 的进位 C_{out} 计算,NOR 门实现进位 C_{out} 和先前进位 1 的运算,减量器获得输出结果。图 4(b) 所示为基于传统的 n 位纹波借半减法器的减量器的内部逻辑电路原理图。

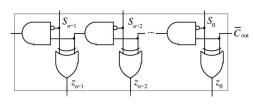
模 2^n 加法器可以直接用忽略溢出的 n 位加法器实现。

2.1.2 模乘法

模
$$(2^{n}-1)$$
 (模 $2^{n+1}-1$) 乘法可以表示为
$$\langle x_{i} \times y_{i} \rangle_{2^{n}-1} = \langle \langle x_{i} \times y_{i} \rangle_{2^{n}} + x_{i} \times y_{i} \operatorname{div} 2^{n} \rangle_{2^{n}-1}$$
 (7)



(a) 模 (2^n-1) 硬件结构的模加



(b) 纹波借减量逻辑结构

图 4 模加法的硬件结构

式中 $\langle x_i \times y_i \rangle_{2^n}$ 对应于乘法 $x_i \times y_i$ 的低输出字, $x_i \times y_i$ div2ⁿ 对应于高输出字。其中 y_i div2ⁿ 的值是 $y_i/2^n$ 的值,按 0 的方向四舍五入到最接近的整数。因此,模 (2^n-1) 乘法可以通过n位无符号乘法然后是n位模 (2^n-1) 加法来完成。

式(7)可以改写为部分积的和:

$$\langle x_i \times y_i \rangle_{2^{n-1}} = \sum_{k=0}^{n-1} \langle pp_{i,k} \rangle_{2^{n-1}}$$
 (8)

式中, $x_i = x_{i,n-1}x_{i,n-2}$, …, $x_{i,0}$; $y_i = y_{i,n-1}y_{i,n-2}$, …, $y_{i,0}$, 且 $pp_{i,k} = x_k \times (y_{i,n-k-1} \dots y_{i,0}y_{i,n-1} \dots y_{i,n-k})$ (用AND 门实现) 为第 k 个部分积模 $(2^n - 1)_{\circ}$

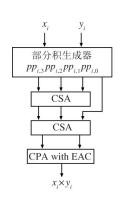
注意,全部 n 位部分积 $pp_{i,k}$ 有相同的大小(相对于普通乘法),即对于全部位位置,要加入的部分积位数是相同的。表 1 所示为 4 位宽输入的部分积生成实例。

表 1 模(24-1)的部分积生成

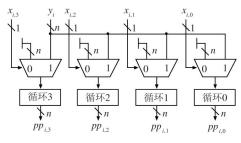
2 ⁶	2 ⁵	24	2 ³	2^2	21	2°
			$x_{i,0}y_{i,3}$	$x_{i,0}y_{i,2}$	$x_{i,0}y_{i,3}$	$x_{i,0}y_{i,0} = pp_{i,0}$
		$x_{i,3}y_{i,3}$	$x_{i,1}y_{i,2}$	$x_{i,1}y_{i,1}$	$x_{i,1}y_{i,0}$	$x_{i,1}y_{i,3} = pp_{i,1}$
	$x_{i,2}y_{i,3}$	$x_{i,2}y_{i,2}$	$x_{i,2}y_{i,1}$	$x_{i,2}y_{i,0}$	$x_{i,2}y_{i,3}$	$x_{i,2}y_{i,2} = pp_{i,2}$
$x_{i,3}y_{i,3}$	$x_{i,3}y_{i,2}$	$x_{i,3}y_{i,1}$	$x_{i,3}y_{i,0}$	$x_{i,3}y_{i,3}$	$x_{i,3}y_{i,2}$	$x_{i,3}y_{i,1} = pp_{i,3}$

在这个乘法中,权重大于 2³ 的位(位于竖直直线左边)被重新放到竖直直线的右边。

在传统的基于存储的技术中,只读存储器 (read-only memory,ROM)存储全部可能值的乘法结果。本文将其扩展以得到基于无存储的实现。提出的模 (2^n-1) 乘法实现结构如图 5(a) 所示。假设系数字长和输入样本字长均为 4 位,图 5(a) 所示为 4×4 Wallace 树逻辑的分层分解。对于 4×4 位,生成 4 个部分积,且并行相加。部分和是通过采用 2 个进位保存加法器 (carry save adders,CSA)和 1 个带末端循环进位的进位传播加法器 (CPA with end around carry,CPA with EAC)相加;提出的基于无存储的部分积生成器实现结构如图 5(b) 所示。它由n 个复用器 (multiplexer,MUX)构成,n 为输入样本字长。部分积是通过将 0 和系数值连接到 MUX 数据输入、将输入数据位连接到选择输入、将 MUX 剩下的 $s-1(1 \le s \le n)$ 位进行循环移位输出得到的。



(a) 模(24-1)乘法器结构



(b) 部分积生成器

图 5 模乘法的硬件结构

假设乘法器 Y 和被乘数 $x_i = x_{i,3}x_{i,2}x_{i,1}x_{i,0}$ 为 n = 4 位的字长,则部分积将有 4 个值。注意,每个 MUX 由并行工作的 n 个 MUX 单元构成(对于 2^n – 1 通道),或者并行工作的 n+1 个 MUX 单元(对于 2^{n+1} – 1 通道)构成。

模 2" 乘法可表示为

$$\langle x_i \times y_i \rangle_{2^n} = \sum_{k=0}^{n-1} p p_{i,k} \tag{9}$$

式中 $pp_{i,k} = x_i \cdot (y_{i,n-k-1}y_{i,n-k-2}\cdots y_0 \underbrace{00\cdots 0}_{k})$ 。部分积 采用 CSA 相加,并丢弃进位输出。

2.2 二进制到 RNS 的转换

在[0,M)范围内的整数 X,用 2^n 记号表示为

$$X = \sum_{i=0}^{3n-1} X_i 2^i = N_2 2^{2n} + N_1 2^n + N_0$$
 (10)

在 RNS 中,可以用模集 $\{m_1, m_2, m_3\}$ 的集合 (x_1, x_2, x_3) 表示。为了得到整数的 RNS 表示,需要 3 个转换器,每个元素一个。

最简单的就是 $m_1 = 2^n$ 通道的转换器。值 x_1 可以通过 X 除以 2^n 的余数得到,这可以通过截短 X 值来实现,是因为:

$$x_1 = \langle X \rangle_{2^n} = X_{n-1} X_{n-2} \cdots X_0 \tag{11}$$

对于 $(2^n - 1)$ 通道和 $(2^{n+1} - 1)$ 通道,相应的余数计算更复杂,是因为转换的最终结果依赖于全部 X_i 位的值。 $(2^n - 1)$ 余数的计算比较复杂,而且在芯片面积和速度上都很昂贵,因而本文不采用除法运算,而是通过一系列加法进行,如下所示。

$$x_2 = \langle X \rangle_{2^{n-1}} = \langle N_2 2^{2^n} + N_1 2^n + N_0 \rangle_{2^{n-1}}$$
 (12)
通过取方程:

$$\langle 2^n \rangle_{2^{n-1}} = 1 \tag{13}$$

式(12)可以改写为

$$x_2 = \langle N_2 + N_1 + N_0 \rangle_{2^{n-1}} \tag{14}$$

从而 X 到模 $(2^n - 1)$ 的转换可以简单地通过相加 X 的模 $(2^n - 1)$ 的 N_i 个部分来执行。

以相同的方式,可计算 $(2^{n+1}-1)$ 的余数为 $x_3 = \langle X \rangle_{2^{n+1}-1} = \langle N'_2 2^{2(n+1)} + N'_1 2^{n+1} + N'_0 \rangle_{2^{n+1}-1}$ (15a)

式(15a)也可以改写为

$$x_{3} = \langle N'_{2} + N'_{1} + N'_{0} \rangle_{2^{n+1}-1}$$
式中 $N'_{0} N'_{1}$ 和 N'_{2} 为 $n+1$ 位数。

2.3 RNS 到二进制的转换

给定关于模集 $\{2^n, 2^n - 1, 2^{n+1} - 1\}$ 的 RNS 数 (y_1, y_2, y_3) ,则本文算法采用 MRC 技术来计算 RNS 数的等效二进制数。对于本文提出的模集 (k=3),则式(2)简化为

$$y = a_3 2^n (2^n - 1) + a_2 2^n + a_1$$

$$= (a_2 + a_3 2^n - a_3) 2^n + a_1$$

$$= (a_4 - a_3) 2^n + a_1$$

$$= a_5 2^n + a_1$$
(16)

对于本文所提出的模集,各种乘法逆为: $c_{12} = 1$, $c_{13} = 2$ 和 $c_{23} = -2$ 。用式(3)计算 MRD。

$$\begin{cases} a_1 = y_1 \\ a_2 = \langle y_2 - y_1 \rangle_{2^{n-1}} \\ a_3 = \langle (a_2 - 2(y_3 - y_1)) 2 \rangle_{2^{n+1}-1} \end{cases}$$
 (17)

二进制数 a_4 仅是 MRD a_2 和 a_3 的级联,其中 a_3 是最高有效位(most significant bit, MSB), a_2 是最低有效位(least significant bit, LSB):

$$a_4 = a_2 + a_3 2^n (18)$$

 a_5 是 MDR a_4 和 a_3 的传统减法:

$$a_5 = a_4 - a_3 \tag{19}$$

所提出的 RNS 到二进制数转换的结构如图 6(a) 所示。它包含 $2 \land$ 模 $(2^{n+1}-1)$ 减法器, $1 \land$ 模 (2^n-1) 减法器和 $1 \land$ 传统的借位传播减法器 (borrow propagation subtractor, BPS)。其中 n 位输入 y_1 和 y_2 经过模 (2^n-1) 减法器运算得到 n 位的 a_2 , n 位的 y_1 和 n+1 位的 y_3 经过模 $(2^{n+1}-1)$ 减法器运算、循环,再与 a_2 经过模 $(2^{n+1}-1)$ 减法器运算和循环得到 n+1 位的 a_3 , a_3 和 a_2 连接得到 2n+1 位的 a_4 , 然后与 a_3 经过 BPS 运算得到 2n+1 位的 a_5 , 最后和 y_1 连接得到 3n+1 位的输出 y_0 .

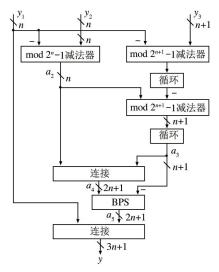
模
$$(2^n - 1)$$
 减法可以表示为式 (20) 。

$$\langle x - y \rangle_{2^{n-1}} = \langle x - y - B_{\text{out}} \rangle_{2^{n}}$$
 (20)

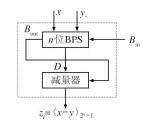
借出信号(B_{out})可用于模 ($2^n - 1$) 减法的计算过程,这是由于:

$$B_{\text{out}} = \begin{cases} 1 & x < y \\ 0 & x \ge \gamma \end{cases} \tag{21}$$

则模 2ⁿ 减法器与借出反馈回借输入(以实现末端循环借)可用于实现模 (2ⁿ - 1) 减法器式(20)。 这种类型的减法器也称为带末端循环借位的借位传 播减法器(borrow propagate subtractor with end around borrow, BPS with EAB)。提出的模 (2^n-1) 减法算法避免了零的双重表示,覆盖了整个动态范围,而基于 CPA with EAC 的模减法器不能实现这个功能,因此本文将模减法器实现为带末端循环借位 (end around borrow, EAB) 的普通减法器,图 6(b) 所示是基于 EAB 的 BPS 的模减法器结构,其中考虑了 n 位 BPS 的延迟 D,通过减量器后会得到大大缓解。



(a) 模集{2ⁿ, 2ⁿ⁻¹, 2ⁿ⁺¹-1}的 MRC 结构



(b) 基于 BPS 和 EAB 的模 2ⁿ⁻¹减法器

图 6 RNS 到二进制数的转换结构

由于 $c_{12} = 1$,所以在模减 $y_2 - y_1$ 后,混合-基数字 a_2 是可用的。接下来,在 4 个运算后得到 a_3 :第1 个运算是模减 $y_3 - y_1$;第2 个运算是将得到的差值左循环移动 1 位;第3 个运算是模减法,将 a_2 作为被减数,循环移位的结果作为减数;第4 个运算是对结果进行左循环移动 1 位。

假设 MRD a_1, a_2, a_3 有如下二进制表示:

$$\begin{cases}
a_1 = a_{1,n-1}a_{1,n-2} \cdots a_{1,0} \\
a_2 = a_{2,n-1}a_{2,n-2} \cdots a_{2,0} \\
a_3 = a_{3,n-1}a_{3,n-2} \cdots a_{3,0}
\end{cases}$$
(22)

如式(18)和(19)所示,要加入3个操作数来得到 a_5 ,这3个操作数简化为2个(2n+1)位的字 a_4 和 a_5 ,是因为 a_2 和 a_32^n 一起在全部(2n+1)位上都是0。

$$\begin{cases}
 a_2 = \underbrace{00\cdots0}_{n+1} \underbrace{a_{2,n-1}a_{2,n-2}\cdots a_{2,0}}_{n} \\
 a_32^n = \underbrace{a_{3,n}a_{3,n-1}\cdots a_{3,0}}_{n+1} \underbrace{00\cdots0}_{n}
\end{cases}$$
(23)

因此:

$$a_4 = a_{3,n} a_{3,n-1} \cdots a_{3,0} a_{2,n-1} a_{2,n-2} \cdots a_{2,0}$$
 (24)

减法 $a_5 = a_4 - a_3$ 可以通过传统的 BPS 实现。最后,在式(16)的基础上,经过比特组织(拼接)后,得到的最终结果为二进制数 $y = a_5 2^n + a_1$:

$$y = \underbrace{a_{5,2n} a_{5,2n-1} \cdots a_{5,0}}_{2n+1} \underbrace{a_{1,n-1} a_{1,n-2} \cdots a_{1,0}}_{n}$$
 (25)

3 算法仿真实验及结果

3.1 仿真环境及参数设置

为了验证本文提出的算法,本文采用 Parks-Mc-Clellan 方法^[16]和 Matlab 对三模集 $(2^{n-1}, 2^n, 2^{n+1} - 1)$ 、n = 6的 27 阶 FIR 低通滤波器进行设计、数值计算和仿真,分 2 步进行。

第 1 步用 firpmord 命令估计出最佳 Parks-Mc-Clellan FIR 滤波器的阶数,以满足设计规格。命令语法为:[N,fo,mo,w] = firpmord(f,m,dev),向量 $f=(0.4\ 0.5)$ 为频带频率向量,向量 $m=[1\ 0]$ 包含滤波器通带和阻带处期望的幅值响应值,向量 $dev=[0.01\ 0.1]$ 具有滤波器幅值响应与期望幅值响应的最大允许偏差。

第 2 步是滤波器的实际设计,用 firpm 命令 b = firpm(N, fo, mo)找到设计的 Parks-McClellan FIR 滤波器的脉冲响应系数 b_i 。

对于双精度和 9 位精度(包括符号位)的整数 表示法的滤波器系数如表 2 所示。

表 2 中第 3 列的整数值是通过 2 个步骤从浮点值(第 2 列)转换得到。第 1 步是用 2 个 Matlab 函数 Q_1 = quantizer('round', Format)和 b_1 binary = num2bin(Q_1 , b)对二进制串 b 二进制中的浮点滤波器系数 b 进行转换;量化器中的值格式 Matlab 函

数创建二进制数参数 [wordlength, fractionlength]用于符号定点模式,对于9位精度格式为 wordlength = 9和 fractionlength = 8。第2步是用2个新的 Matlab 函数 q_1 = quantizer ('round', Format)和 b_1 int = bin2num(q_1 , b_1 binary)将二进制串 b_1 进制转换为整数值。在这种情况下,值格式 fractionlength = 0即 Format = [9,0];最后,将滤波系数的整数值转换为 RNS 数。对于系数的正向转换,用 Matlab 函数

mod o

仿真实验研究模集 $\{64,63,127\}$ 的二进制到余数转换器。实例描述了系数 b_1 的定点到余数系统的转换。滤 波 器 系 数 b_1 的 双 精 度 为 0.023540422135223,将其转换为二进制数 b_1 binary = 1111111010,再转换为整数数 b_1 int = -6,最后转换为 RNS 数 b_1 RNS = (58,57,121)。

表 2	三模集(2"-1.	$2^{n} \cdot 2^{n+1}$	$n^{-1} - 1$, $n = 6$	的 27 阶 FI	R 低通滤波器系数

	双	整数表示 -	RNS 数		
	双精度表示		$b_{i,1}$	$b_{i,2}$	$b_{i,3}$
$b_0 = b_{27}$	0.004493088227562	1	1	1	1
$b_1 = b_{26}$	- 0. 023540422135223	-6	58	57	121
$b_2 = b_{25}$	-0.010537273615351	-3	61	60	124
$b_3 = b_{24}$	0.014434554627302	4	4	4	4
$b_4 = b_{23}$	0.017852276297252	5	5	5	5
$b_5 = b_{22}$	-0.014709152081930	-4	60	59	123
$b_6 = b_{21}$	-0.031608802439920	-8	56	55	119
$b_7 = b_{20}$	0.009659673837214	2	2	2	2
$b_8 = b_{19}$	0.051468586283075	13	13	13	13
$b_9 = b_{18}$	0.005218868131868	1	1	1	1
$b_{10} = b_{17}$	- 0. 084444425211876	- 22	42	41	105
$b_{11} = b_{16}$	-0.047670664338718	- 12	52	51	115
$b_{12} = b_{15}$	0.179378544446062	46	46	46	46
$b_{13} = b_{14}$	0.413118538653651	106	42	43	106

3.2 本文设计方法的验证

采用 Matlab 进行的仿真验证本文的设计方法。图 7 所示为得到的理想滤波器输出曲线(虚线)和实际输出曲线(实线)。可见,基于余数的三模集FIR 低通滤波器具有良好的衰减性能。

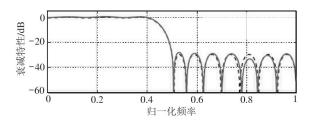


图 7 衰减响应

假设数据序列量化为 10 位(包括符号),且滤波器必须在没有舍入误差的情况下实现,则滤波器

响应的绝对上界 | y(n) | 由式(26)给出。

$$| y(n) | \le \max\{| x(n) |\} \sum_{k=1}^{30} | b_k | = 476718$$

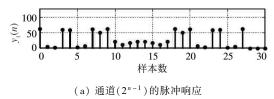
 $\approx 18.86 \text{ bit}$ (26)

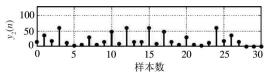
模集{63,64,127}提供了 18.96 位的动态范围,这对于大多数实际情况是足够的,因为式(26)给出的 18.86 位是上界。

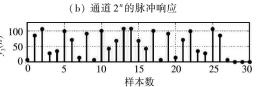
图 8 所示为 RNS 通道的脉冲响应。第 1 个和第 2 个子滤波器的系数值必须存储在 6 位字中,第 3 个子滤波器的系数值必须存储在 7 位字中。在 {64,63,127}余数系统中,单位样本为

$$\delta(n) = \begin{cases} (63,15,7) & n = 0\\ (0,0,0) & n \neq 0 \end{cases}$$
 (27)

对于整个滤波器的脉冲响应,可以用 MRC 技术 来转换一个在传统的数字系统中用余数系统给出 的数(如图 6),得到的数字滤波器的脉冲响应如图 9(a)所示;将系数量化为 9位后,脉冲响应的量化误差如图 9(b)所示。可以看出,9位就足以保持误差小于 0.02。

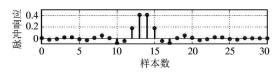




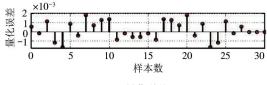


(c) 通道(2ⁿ⁺¹-1)的脉冲响应

图 8 RNS 通道的脉冲响应



(a) 脉冲响应



(b) 量化误差

图 9 RNS 低通滤波器的脉冲响应及量化误差

3.3 与其他设计方法的比较

表 3 所示为不同设计方法得到的 27 阶 FIR 低通滤波器系数 b_i 的偏差的比较。可以看到,由于本文设计方法采用的基于 RNS 的三模集 MAC 结构实现了正向转换器、并行算术通道和逆向转换器的功能,所以最终获得的滤波器系数 b_i 的偏差更小,而且是近似不变的,而文献[8]和文献[9]的设计算法分别仅实现了二进制到余数的正向转换器和从 RNS 到二进制的逆向转换器,所以获得的滤波器系数 b_i 的偏差不仅较大,而且在不同阶时还存在正负方向的偏差,从而会导致滤波器输出的纹波特性变差。

表 3 不同设计方法的滤波器系数偏差比较

系数 <i>b</i> _i	本文算法	文献[8]算法	文献[9]算法
	偏差	偏差	偏差
$b_0 = b_{27}$	0.0177	0.0321	-0.0330
$b_1 = b_{26}$	0.0176	0.0325	0.0320
$b_2 = b_{25}$	0.0174	-0.0324	0.0340
$b_3 = b_{24}$	0.0175	-0.0319	-0.0322
$b_4 = b_{23}$	0.0178	0.0327	-0.0340
$b_5 = b_{22}$	0.0172	-0.0331	-0.0335
$b_6 = b_{21}$	0.0174	-0.0319	0.0310
$b_7 = b_{20}$	0.0175	0.0311	-0.0323
$b_8 = b_{19}$	0.0177	-0.0341	-0.0340
$b_9 = b_{18}$	0.0175	-0.0313	0.0319
$b_{10} = b_{17}$	0.0176	0.0330	-0.0314

图 10 所示为采用 Parks-McClellan 方法基于 3 种设计算法设计的最优长度 15 的 FIR 线性相位滤 波器得到的频率响应的等波纹特性。长度 15 的线 性相位 FIR 滤波器的频率响应可以写成 8 个余弦函 数 $\{1,\cos\omega,\cdots,\cos7\omega\}$ (即 M=7),根据交替定理, 有M+2=9个极值点(见频率响应上的点表示)。 两条带边即 ω=0 和 ω=π 都是极值点;可以看到,3 种设计算法得到的最佳频率响应都在理想响应值上 下波动,理想响应在通带[0,0.375π]为1,在阻带 $[0.5\pi,\pi]$ 为0。但本文设计算法的通带偏差 δ_{pass} 和 阻带偏差 δ_{stop} 分别为 0.0883 和 0.0884,都要低于文 献[8]和文献[9]设计算法的 0.1211 和0.1210,以 及 0.1280 和 0.1282,这主要是由于本文设计算法 的 MAC 结构的轻量化计算和简单结构,而文献[8] 和文献[9]的设计算法的单向转换器的结构和复杂 计算造成量化误差和延迟变大。

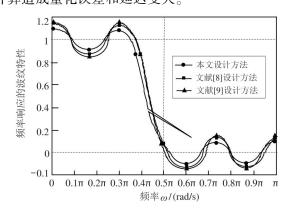


图 10 不同设计方法的滤波器的频率响应的 等波纹特性比较

4 结论

本文针对大动态范围 RNS 的三模集 {2ⁿ - 1, 2ⁿ,2ⁿ⁺¹ - 1} 的数字信号处理技术进行了研究。应用于常系数 FIR 滤波器的 RNS 尚未见文献进行深入研究,基于初步研究,本文提出了新的 RNS 的三模集,可平衡通道间松弛,从而最大限度地利用时钟周期;利用松弛平衡可实现低功耗的余数通道结构;描述了算法实现的原型结构;最后对设计方法进行了仿真实验来验证其有效性,并与其他典型设计方法进行方法进行了比较。

在某种意义上,对于FIR 滤波器设计,涉及到理 论算法和实际实现。但通常要着重考虑 2 个方面, 以使设计方法更高效:(1)在通带和阻带之间分配 极值频率(对于低通情况);(2)在算法迭代时使极 值在频带之间移动。因此,通带和阻带的极值个数 可以通过带宽大小的比值来分配。此外,通带边缘 和阻带边缘也要具有较好的特性;这些新的设计思 想是将来要着重考虑和进一步深入研究的内容。

参考文献

- [1] 卢有亮,赵成,滕云龙,等. 基于随机计算与余数系统相结合的 FIR 滤波系统:110620566 A [P]. 2021-07-02.
- [2] VALUEVA M V, NAGORNOV N N, LYAKHOV P A, et al. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation[J]. Mathematics and Computers in Simulation (MATCOM), 2020,177(C);232-243.
- [3] 刘剑锋. 基于余数系统的混沌序列发生器设计与实现 [D]. 成都:电子科技大学, 2017.
- [4] 刘莺迎. 基于余数系统抗 MES D 攻击的 RS A 算法 [J]. 计算机应用与软件,2021,38(4):324-327,333.
- [5] ARMAND A, TIMARCHI S, MAHDAVI H. Optimized parity-based error detection and correction methods for residue number system [J]. Journal of Circuits, Systems and Computers, 2019,28(1):195002-195016.
- [6] XIAO L, XIA X G, HUO H Y. Towards robustness in

- residue number systems [J]. IEEE Transactions on Signal Processing, 2017,65(6):1497-1510.
- [7] ZARANDI A A E, MOLAHOSSEINI A S, SOUSA L, et al. An efficient component for designing signed reverse converters for a class of RNS moduli sets of composite form {2k, 2P - 1} [J]. IEEE Transactions on Very Large Scale Integration Systems, 2017, 25(1):48-59.
- [8] 王巍,李双巧,徐媛媛,等. 基于 RNS 算法的高阶 FIR 滤波器设计[J]. 微电子学,2017,47(6):788-792.
- [9] MADHAVI LATHA M V N, RACHH R R, ANANDA MOHAN P V. RNS-to-binary converters for a three-moduli set {2ⁿ⁻¹ - 1,2ⁿ - 1,2^{n+k}} [J]. IETE Journal of Education, 2017, 58(1);20-28.
- [10] PHALGUNA P S, KAMAT DATTAGURU V, ANANDA MOHAN P V. Novel RNS-to-binary converters for the three-moduli set {2^m 1,2^m,2^m + 1} [J]. Sadhana of Indian Academy of Science,2019,44(99):1-10.
- [11] KROEMER P, PLATO J, NOWAKOVA J, et al. An acceleration of quasi group operations by residue arithmetic
 [J]. Concurrency and Computation, 2018,30(2):1-14.
- [12] 曹成虎,赵永波,庞晓娇,等. 基于中国余数定理的目标距离估计算法[J]. 系统工程与电子技术, 2019,41 (12): 2717-2722.
- [13] HIASAT A. A scaler design for the RNS three-moduli set $\{2^{n+1}-1,2^n,2^n-1\}$ based on mixed-radix conversion [J]. Journal of Circuits ,Systems and Computers , 2019 , 29(3);1-8.
- [14] CHERVYAKOV N I, MOLAHOSSEINI A S, LYAKHOV P A, et al. Residue-to-binary conversion for general moduli sets based on approximate Chinese remainder theorem [J]. International Journal of Computer Mathematics, 2017,94(9):1833-1849.
- [15] 林跃杉,林郁,尹韬,等. FIR 基于 FPGA 的高并行度 DA 结构[J]. 太赫兹科学与电子信息学报,2018,16 (1):170-175.
- [16] TSENG C C, LEE S L. Minimax design of graph filter using Chebyshev polynomial approximation [J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2021,68(5):1630-1634.

Design of low-pass FIR digital filter based on a new three moduli set

ZHAO Haijun, CHEN Yihong, HE Chunlin, PU Bin

(School of Computer China-West Normal University, Nanchong 637009)

(Internet of Things Perception and Big Data Analysis Key Laboratory of Nanchong, Nanchong 637009)

Abstract

In this paper, the design of low-pass finite impulse response (FIR) digital filter is studied. Based on the analysis of common three moduli set and traditional FIR digital filter with impulse response coefficients, a new three moduli set is proposed for the design of low-pass FIR digital filter. The FIR filter is composed of multiple multiplication-accumulator units, which can realize the functions of a forward converter, parallel arithmetic channels and a reverse converter. The forward converter encodes a binary number into a residue represented number, with regard to the moduli set, each arithmetic channel requires modular multiplication and accumulation for each modulo of set, the reverse converter decodes a residue represented number into its equivalent binary number, the arithmetic channels are working in a completely parallel architecture, so as to realize the calculation of lightweight and simple structure. Finally, the simulation results not only verify the effectiveness of the proposed design method, but also show the proposed design method has smaller filter coefficient deviation and better frequency response ripple characteristics compared with the traditional one-way conversion design.

Key words: low-pass finite impulse response (FIR) digital filter, residue number system (RNS), three moduli set, multiplication-accumulator unit, impulse response, quantization error