DESIGNING EFFICIENT RESIDUE ARITHMETIC BASED VLSI CORRELATORS

A. A. Deodhar, Manish Bhardwaj
Microelectronics Design Center
Telecommunication System ICs
Siemens Components Private Limited
166 Kallang Way, Singapore 349249.

C. T. Clarke
Submetrics Limited
Minerva Clos
Lower Bristol Road, Bath
United Kingdom BA29ER.

T. Srikanthan
School of Applied Science
Nanyang Technological University
Nanyang Avenue
Singapore 639798.

ABSTRACT

The most important reason for the lack of commercial residue arithmetic (RA) based systems is not the "slow" and area consuming reverse conversion, but the absence of research that explores the *system-level* trade-offs of such arithmetic in *actual* VLSI implementations. Such system-level issues are - choice of the moduli set, effect of moduli imbalance on resulting VLSI implementation, choice of the reverse and forward converters, use of lookup versus computation for modular operations, system characteristics that indicate RA suitability and finally, typical VLSI area and performance figures. This paper explains these concerns by presenting novel RA architectures for VLSI correlators employed in radio-astronomy and ultrasonic blood flow measurement. A state-of-the-art, high performance (80-100 MHz), RA-based correlator ASIC was successfully fabricated as a result of this research[1].

1. INTRODUCTION

1.1 Overview of Residue Arithmetic (RA)

The Residue Number System (RNS) is a "carry-free" non-weighted number system that exploits parallelism by representing numbers as sets of remainders or *residues*. [9,12].

The base of RNS is a set $\{m_1, m_2, \ldots, m_p\}$, where each member of the set is called a modulus and the set elements are pair-wise relatively prime. For any given moduli set, the residue representation of an integer X is a p-tuple, $(x_1, x_2, \ldots, x_p)$, where $\{x_i\}$ are least positive residues of X modulo m_i, and are defined by [3]

$$x_i = |X|_{m(i)}, \qquad i = 1, 2, \ldots, p.$$

Beginning with zero, all numbers up to and excluding M given by,

$$M = \prod_{i=1}^{p} m_i$$

may be unambiguously encoded in RNS. This is often called the dynamic range of the given system.

The advantage of arithmetic using RNS i.e. residue arithmetic is that arithmetic operations of addition, subtraction and multiplication can be performed for each residue independent of all other residues. The individual arithmetic operations can thus be done in parallel and the resulting speedup is close to the number of moduli.

This tremendous speedup rarely translates to system speed-up due to the need to convert weighted binary numbers into and out of the residue domain. These *forward* and *reverse* converters have to be carefully crafted to retain the advantages of parallel arithmetic.

1.2 Issues in RA-based VLSI implementations

Traditionally, published RA research either focuses on forward and reverse converters [4,11] or efficient residue operators [1]. Due to this, two extreme facets of RA are highlighted. If one looks at converters alone, the speed and area disadvantages of RA look insurmountable. A look limited to residue operators, on the other hand, leads to impractical expectations of system speed-ups. As we shall see in the following sections, the key to achieving an overall increase in the system's cost-performance is a domain-specific integration of these two perspectives.

The strategic architectural issues in RA-based designs are -

- **Choice of a moduli set.** As explained above, more moduli imply higher speedup, *as long as they are well balanced* (this means that the highest residues these moduli leave should be of the same width). The downside to a higher set cardinality is the increased cost of forward and reverse conversion. Apart from the cardinality, an architect must decide whether the popular $\{2^n-1, 2^n, 2^n+1\}$ set be used or a more general one. Again, the 2^n set offers fast conversion, but limits RA speedup to about 3. This might not be enough to meet the computational bandwidth requirement.
- **Design of forward and reverse converters.** Previously, a good reason to stick to 2^n sets was their tremendous advantage in ease of conversion. This is changing fast with highly area-time efficient converters for general moduli sets being reported [2, 10, 11]. One such converter, called the CMAC reverse converter, has been used in a radio astronomy correlator described in the following section.
- **Design of residue operators.** The main issue here is lookup versus computation. While, there is no consensus on this, efficient implementations result when lookups are restricted to small tables (less than 32 locations). Larger tables should be broken into computational circuitry followed by multiplexers (and many a times, small lookups).

[1] Funded by Advanced Research Project Grant - ACRFRG-22/95

In the following sections, we present two out of the six correlator architectures we developed [6]. We have chosen a radio astronomy example to illustrate an extremely successful RA-based VLSI implementations. From this example, it is clear that by selectively applying RA to only certain parts of the system, we can minimize conversion and modulo correction penalties. The other system example, an ultrasonic blood flow meter, illustrates how we can improve upon pre-VLSI RA implementations. Issues like replacing lookup by computation and increasing moduli cardinality are highlighted in this example.

2. SYSTEM EXAMPLES

2.1. Radio Astronomy

Radio interferometers and synthesis arrays are ensembles of two-element interferometers used to make measurements of the fine angular detail in the radio emission from the sky [13]. Two antennas situated on the surface of the earth receive radio signals which are multiply-accumulated to give the desired output. The combination of the multiplier circuit and the time-averaging circuit is referred to as a correlator. The most important issues for a Radio Astronomy Correlator design are -

- *High throughput* : At least 250 Ms/s in a multi-channel correlator.
- *Large number of lags per chip* : Each *lag* corresponding to a distinct correlation.
- *High rate of correlation but a lower rate of "observation"* : Allows a slower reverse converter to co-exist with very fast RA-based MACs.
- *Small input wordlength (2 bits) but a large dynamic range (32 bits)* : This is due to the large number of computations performed per second per chip.

A correlator architecture that employs RA in integration and co-exists with a binary MAC is illustrated in Figure 1. Figure 2 shows the magnified photo of the chip (VAM96b).

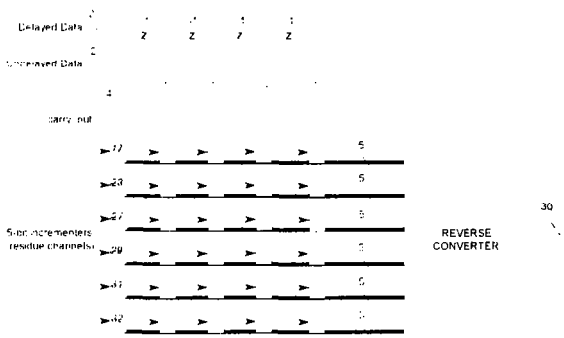

Figure 1: RA based correlator for Radio Astronomy

The implementation above brings out several points --

- *Smart Use of RA* : The architecture is a perfect example of how RA can be used only where it is needed. The 32-bit accumulation is done in RA, while the multiply-accumulate is left in the binary domain. Since, the MAC being only 4 bits wide, RA cannot be used there. The accumulation, however, is 32 bits long and offers a wide dynamic range for RA to exploit.
- *No Forward Converters* - since the accumulators merely perform an incrementing function (as in binary).
- *No modulo correction* - the incrementing operation was implemented as a set of synchronous down counters.
- *Speed* : Six 5-bit increments are done *in parallel* leading to high computational speed. Due to the low output rate requirement we can perform MAC at a high rate while reverse converting the output at a lower rate.
- *Reverse Converter*. The converter used in this design incorporates a technique known as Compressed Multiply ACcumulate (CMAC) [11] It involves the merging of partial sum generation and partial sum addition into a single step i.e. the CMAC evaluation. This approach results in significantly better area-time performance than existing solutions. The same reverse converter (area of $85FA_{width}$ by 12 FA_{height} and a pipelined delay of 20 ns) is assumed throughout all applications in this paper.

Comparison between RA and an existing non-RA designs

Herzen et al. [8] describe a circuit consisting of 12 lags in a chip 2.2mm x 2.2mm. We have managed to fit in 32 lags *and* a reverse converter in a chip of dimension 2.4 x 4.8 mm.

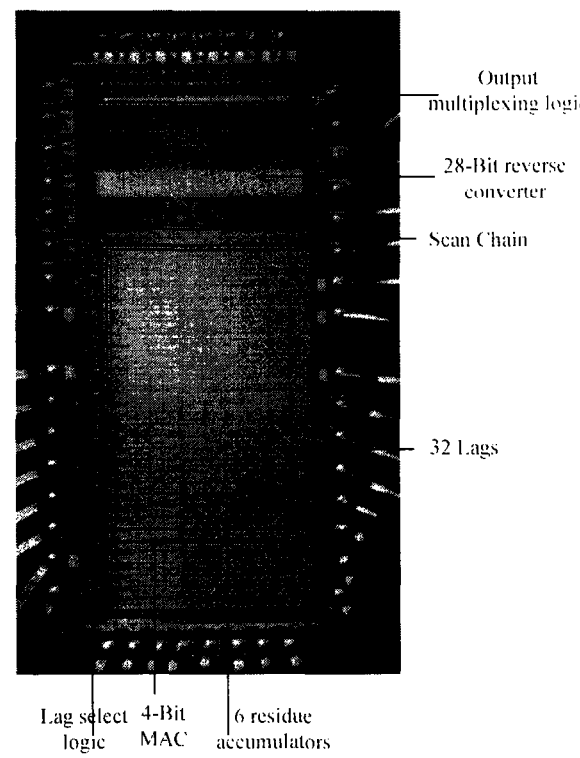

Figure 2: Correlator Chip for Radio Astronomy (0.8μ)

The following table compares the salient features of *VAM96b* (our development chip) with those of

Herzen91. The significant increase in performance is clear from the table.

	Herzen 91	VAM96b	Markup
Area	2.2mm x 2.2mm ($4.84\ mm^2$)	2.4mm x 4.8mm ($11.52\ mm^2$)	2.3↑
Speed	50 MHz.	250 MHz.	5↑
Process	2μ	0.8μ	↑
# of Lags	12	32-36	3↑
Multiplication	2 bit x 2 bit → 3 bit (truncated)	2 bit x 2 bit → 4 bits (full)	↑
Accumulation	16-bit	28-bit	2↑
Lag selection	No	Yes	↑
Latching MAC	No	Yes	↑
μ-pipelining	Yes	No	↑

Table 1: VAM96b versus Herzen91

Comparison between RA and "perfect" non-RA designs

Since the VLSI process and the chip size in Herzen91 are different from that in VAM96b, such comparisons become difficult. Next, we compare the two circuits with identical variables like the process, chip area etc. --

	RA-based	Conventional	Markup
Area (grids)	8mm x 9mm		--
Process	0.8 μ		--
Correlator lag area	483 x 27 sq. grids		--
Total # of lags	500	512	1.02↓
Area / Lag	0.144 mm^2	0.14 mm^2	1.03↓
MAC delay (ns)	4	4	--
Acc. delay (ns)	3	20	6.7↑
Throughput	142 MHz. (= 7 ns)	41 MHz. (= 24 ns)	3.5↑
Lags / chip / second	7.1×10^{10}	2.1×10^{10}	3.4↑

Table 2: Residue Arithmetic versus weighted Arithmetic

The above analysis shows that a RA-based correlator results in a speed up of 3.5 times at the expense of the area equaling a mere 12 lags !

2.2 Ultrasonic blood flow meter

Volumetric blood flow measurement is an important diagnostic aid in many diseases. Due to its non-invasive nature, ultrasonic measurement has been fast replacing traditional Doppler based means. Jenkins et. al. have shown that Ultrasonic Time Domain Correlation techniques can result in 20% higher accuracy. The purpose of UTDC hardware is to perform a correlation function using the samples from two different echoes [5, 7].

$$F(a,b) = \sum_{i=0}^{n} x(a-i) * y(b+i)$$

where x and y represent echoes while *a* and *b* represent the initial offsets into the respective echoes.

Each echo is composed of 1024 8-bit samples divided into 25 forty-sample ranges which represent different depths across the blood vessel. The samples from a particular range are correlated across different echoes to derive a flow rate for that range. With an 8-bit by 8-bit multiplication (giving a product of 16 bits) followed by 40 correlations (6-bits), the dynamic range becomes 22 bits. Our proposed architecture (*ULTRA96b* - Figure 3) resembles the one proposed by Jenkins et al., but is modified for a different moduli set and eliminates all lookup.

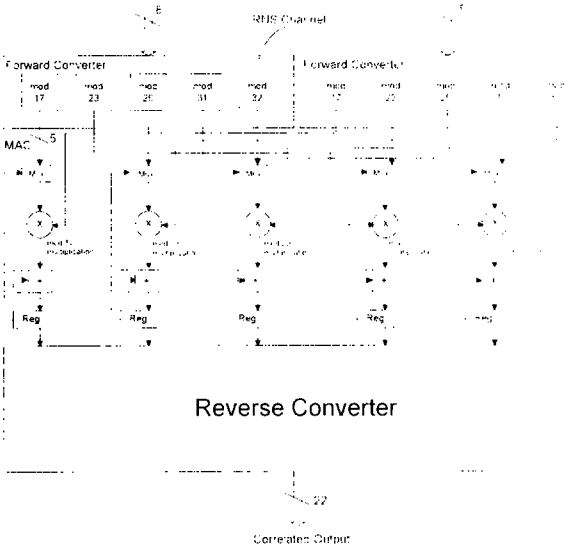

Figure 3: Correlator for Blood Flow Measurement

The reason for choosing this application is to illustrate that cost and performance improvements can result *even within RA-based solutions*.

Comparison between existing and proposed RA designs

	Jenkins90	ULTRA96	Mark up
Area	-- not applicable -- (wire-wrap only)	9mm^2 (expected)	-
Speed (output)	0.4 MHz	3.5 MHz	8.5↑
Memory	- 256K x 8 (echo memory) - 4K x 6 (mod multipliers) - 4K x 6 (mod adders)	Negligible	↑
Moduli used	Four 6-bit moduli (64, 63, 61 and 59)	Five 5-bit moduli (17, 23, 29, 31 and 32)	↑
DR	23 to 24 bits	23 to 24 bits	-

Table 3: ULTRA96b versus Jenkins90

Thus a 8.5 times faster RA correlator could be implemented in 0.8μ process technology in a chip of area of $9mm^2$.

Effect of changing the moduli set in hypothetical design

The following table compares an implementation that uses a *Jenkins90* moduli set versus one that uses our moduli set --

	Jenkins90	ULTRA-96	Markup
Area	85FA X 25FA	85FA X 24FA	Same
MAC Delay	18 FA = 12.6 ns	15 FA = 10.5 ns	2.1ns↑
Delay for 40-point correlation	40 * 12.6 = 504 ns	40 * 10.5 = 420 ns	84 ns↑
Speed	2 MHz	2.5 MHz	25%↑

***Table 4: Jenkins90** versus ULTRA96*

We see that the area complexity does not change much with a change in the moduli set as long as the dynamic range remains the same (23/24 bits). There is an advantage in using the smaller moduli set (as in ULTRA96b) as far as speed is concerned. Knowing that area is not critical, this is indeed an important argument in favor of choosing the 5-bit moduli set. A speed up of 1.25 is achieved, by merely choosing the smaller balanced moduli set.

Comparison between an RA and a "perfect" non-RA design

The following table illustrates an interesting system trade-off. By using RA, our area goes up roughly 3 times, but so does the performance.

	Jenkins-90 on 0.8μ	ULTRA-96	Markup
Area	700 grid x 900 grid	950 grid x 950 grid	2.7↓
MAC Delay	40 FA = 28 ns	15 FA = 10.5 ns	2.7↑
Delay for 40-point correlation	40 * 28 = 1120 ns	40 * 10.5 = 420 ns	2.7↑
Speed	0.89 MHz	2.5 MHz	2.7↑

***Table 5: Jenkins90 on 0.8μ** versus ULTRA96*

It is up to the architect to decide if such a trade-off is required. Also note that in absolute terms, cutting down delay by 2.7 times is very tough to achieve by traditional means like super-scalarity or pipelining.

3. SUMMARY

In this paper, we described two different correlation architectures highlighting various system level issues encountered in RA-based VLSI correlators. From these case studies, we can list the characteristics of systems best suited for RA deployment --

- High Speed/Real-time computations
- High Input Rate with slower output rate
- Large Dynamic Range/Precision (16 bits or greater).
- Inherently simple but otherwise time-consuming computations. In the radio astronomy architecture, RA was used for something as simple as incrementing a stored 28-bit value.

A radio astronomy correlator illustrated a hybrid architecture in which the forward conversion and residue correction problems were eliminated by selective application of RA. Details of an actual, high-speed (80-100 MHz, 32-lag, 32-bit precision) radio astronomy chip were also presented. The second correlator architecture (ultrasonic blood-flow measurement) highlighted how lookup can be replaced by computation and the moduli cardinality increased to reach at a more efficient system.

4. REFERENCES

[1] G. Alia and E. Martinelli, "A VLSI modulo-*m* multiplier", IEEE Trans. Computers, vol. 40, July 1991.

[2] F. Barsi and M.C. Pinotti, "A Fully Parallel Algorithm for Residue to Binary Conversion," Information Processing Letters, vol. 50, pp. 1-8, 1994.

[3] M. A. Bayoumi, G. A. Julien and W. C. Miller, "A Look-up Table VLSI Design Methodology for RNS Structures used in DSP Applications," IEEE Trans. CAS, vol. CAS-34, pp. 604-615, June 1987.

[4] R. M. Capocelli and R. G. Carlo, "Efficient VLSI Networks for converting an integer from Binary System to Residue Number System and vice-versa," IEEE Trans. CAS, vol. 35, pp. 1425 - 30, Nov. 1988.

[5] J. T. Chen, W. K. Jenkins, I.A. Hein and W. D. O'Brien Jr.,"Design and Implementation of a High Speed Residue Number System Correlator for Ultrasonic Time Domain Blood Flow Measurement", IEEE *ISCAS*, vol. 4, pp. 2893-2896, 1990.

[6] A. Deodhar, "Residue Arithmetic Based Correlator System Design", Honors Year Thesis, School of Applied Science, Nanyang Technological University, 1997.

[7] I. A. Hein, J. T. Chen, W. K. Jenkins and W. D. O'Brien Jr., "A Real-Time Ultrasound Time-Domain Correlation Blood Flowmeter: Part I - Theory and Design", IEEE *Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 40, pp. 768-775, Nov. 1993.

[8] B. Herzen, "VLSI Partitioning of a 2-Gs/s Digital Spectrometer", IEEE *JSSC*, vol. 26, pp. 768-772, May 1991.

[9] S. Oddvar, Y. Lundh and J.K. Hagene, "Very High Performance Signal Processing using the Residue Number System", *Proc. 6 MIT Conference - Advanced Research in VLSI*, ed. W. J. Dally, MIT Press, pp. 18-32, 1990.

[10] S. J. Piestrak, "Design of Residue Generators and Multi-Operand Modulo Adders Using Carry Save Adders," IEEE Trans. Computers, vol. 423, pp. 68-77, Jan. 1994.

[11] T. Srikanthan, M. Bhardwaj and C. T. Clarke, "Implementing Area-Time efficient VLSI Residue to Binary Converters," To be presented at the IEEE Intl. Conference on Signal Processing Systems : Design and Implementation (SIPS'97), Leicester, Nov. 3 - 5, 1997.

[12] N. S. Szabo and R. I. Tanaka, "Residue Arithmetic and its applications to Computer Technology", McGraw-Hill, 1967.

[13] A. R. Thompson, J. M. Moran and G. W. Swenson Jr., "Interferometry and Synthesis in Radio Astronomy", John Wiley and Sons Inc., 1986.