Residue Number System for Low-Power DSP Applications

Gian Carlo Cardarilli[(1)], Alberto Nannarelli[(2)] , Marco Re[(1)]

[(1)]Dept. of Electronic Engineering
University of Rome "Tor Vergata"
[(2)] Department of Informatics & Math Modelling
Technical University of Denmark

***Abstract*— In previous works ([1]-[8]) we performed different experiments implementing FIR filtering structures. Each filter was implemented using both the Two's Complement System (TCS) and the Residue Number System (RNS) number representations. The comparison of these two implementations allows to conclude that, for these applications, the RNS uses less power than the TCS counterpart. The aim of the present paper is to highlight the reasons of this power consumption reduction.**

I. Introduction

Power consumption is a main concern for circuit designers. Different techniques can be used for the mitigation of this problem.

In previous works ([1]-[8]) the authors proposed the use of the RNS number representation as a method for power reduction in DSP architecture implementation. To validate the approach, different experiments implementing FIR filtering structures have been developed. Each filter was implemented using both Two's Complement System (TCS) and Residue Number System (RNS) number representations. These experiments showed a large power reduction for RNS representation.

The analysis of the different contributions of power consumption allows us to develop a simple model able to categorize the experimental results. This model expresses the dependence of power consumption on the different implementation related factors (such as area occupancy, length of the interconnects, and input gate capacitance).

The aim of this work is to explain the reasons of power reduction.

In the next section we will recall the fundamentals of RNS, underlining those characteristics that make the RNS very suitable for high speed and low power applications. In the following sections we we give a very simple model for analyzing ASIC and FPGA power consumptions. Finally, we apply this model to ASIC and FPGA implementations of FIR filters, using both Two's Complement (TCS) and Residue Number Systems, drawing some conclusions on the properties of the different representations.

II. RNS Fundamentals

The **integer number representation** based on Residue Number System (**RNS**) is defined by a set of P relatively prime integers $\{m_1, m_2, \ldots, m_P\}$, quoted as the RNS base.

The dynamic range for this base is given by $M = m_1 \cdot m_2 \cdot \ldots \cdot m_P$.

Any integer $X \in \{0, 1, 2, \ldots M-1\}$ has an unique RNS representation:

$$X \overset{RNS}{\rightarrow} (\ \langle X \rangle_{m_1}, \langle X \rangle_{m_2}, \ldots, \langle X \rangle_{m_P}\ )$$

where $\langle X \rangle_{m_i}$ means $X \ mod\ m_i$

If op represents a generic arithmetic operation (a sum, a difference, a product, a modular division), the most interesting property of RNS is the possibility to translate this operation among integers in a set of modular operations on the different moduli m_i. We can express this property as:

$$Z = X\ op\ Y \quad \overset{RNS}{\rightarrow} \quad \begin{cases} Z_{m_1} = \langle X_{m_1}\ op\ Y_{m_1} \rangle_{m_1} \\ Z_{m_2} = \langle X_{m_2}\ op\ Y_{m_2} \rangle_{m_2} \\ \ldots \quad \ldots \quad \ldots \\ Z_{m_P} = \langle X_{m_P}\ op\ Y_{m_P} \rangle_{m_P} \end{cases} \tag{1}$$

The resulting structure of a RNS processor is shown in Fig. 1. The modular operations are implemented with very simple circuits and can be performed in parallel.

Since RNS processors normally work in an environment on TCS representation, conversions TCS-RNS and vice versa must be performed. The modular processors are interfaced with the external world through an input (TCS to RNS) and an output (RNS to TCS) converter.

978-1-4244-2110-7/08/$25.00 ©2007 IEEE

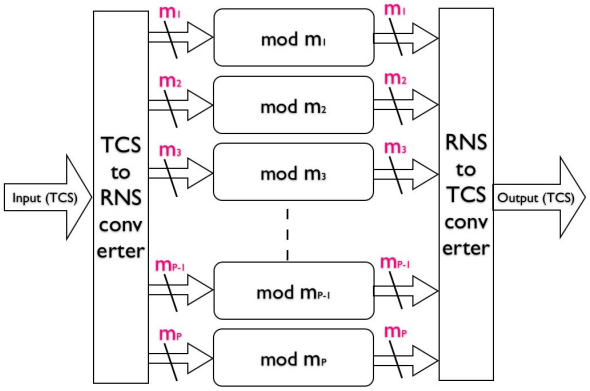


Fig. 1. Structure of RNS processor

Due to the absence of any carry propagation among the different modular processors, the architecture implementation is very efficient. The propagation on the whole wordlength is only present in the input and output converters.
In the following sections, we show as this signal locality can be exploited to increase the speed and to reduce the power consumption.

III. Reduction of Power Consumption

It is well known that power consumption is one of main concerns for digital designers.
Different levels of power optimization are possible:

- Algorithm level
- Architecture level
- Arithmetic level
- Implementation level
- Technology level

In our work, we focus our attention on the arithmetic level optimization, using the Residue Number System representation.
We also analyze the performance (in terms of area, speed and power) of DSP algorithms implemented with different technologies.
In particular, we study implementations based on ASIC-SC (Standard Cells) and FPGA technologies.
The RNS representation offers great advantages in the implementation of DSP systems. These advantages are mainly related to the absence of carry propagation among modular blocks. This property allows to obtain two fundamental characteristics for RNS implementation

1) simpler structures (in particular if high-speed processing is required),
2) signals are more local (in fact signals are bounded inside each modular block)

The main objective of this paper is to evaluate the effects of RNS on power consumptions. To this purpose, in our work we develop a model for characterizing the power consumption of DSP algorithms implementation.

As a general remark, we can observe that different technologies have different characteristics with respect to power consumption model. In particular:

- ASIC-SC (Standard Cells) are characterized by very variable logic and interconnect structures.
- FPGAs have, on the other hand, fixed structure with CLBs, clock and interconnects.

Consequently, the general evaluation of power consumption contributions is more complex for ASIC-SC.

In previous works ([1]-[8]), we compared different FIR filtering structures.
In our analysis, A_x and P_x represent the area and the power consumption of filter, and x identifies the number system representation used in the actual implementation. In our case, x can take the two values RNS or TCS. In general, A_x and P_x grow linearly with the number of taps (N_{TAP}), following the laws (2).

$$A_x = k_1^{A_x} + k_2^{A_x} N_{TAP} \qquad (2)$$
$$P_x = k_1^{P_x} + k_2^{P_x} N_{TAP}$$

The constants k_1 represent the offset of the plots (two different values apply for area and power) while k_2 are the growing rate values. An example of the above trends is shown in the Fig. 2. From these plots, we observe that the RNS has larger values of offset. This is related to the presence of input and output converters. On the other hand, RNS slopes are less steep than TCS ones.

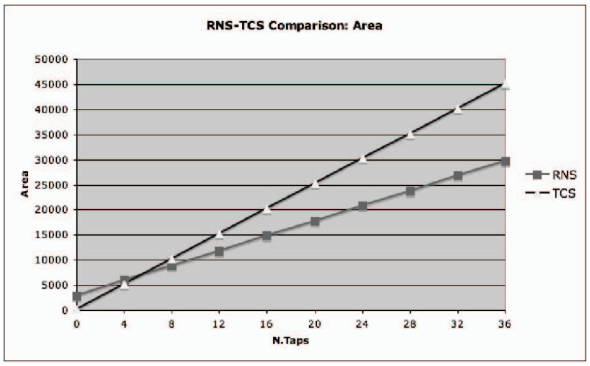
IV. ASIC-SC: Power Consumption Contributions

Considering the contributions to the dynamic power consumption in an ASIC-SC implementation, as remarked above, the contribution breakdown is highly dependent on the implemented structure. Here we consider the breakdown for a microprocessor structure [10].

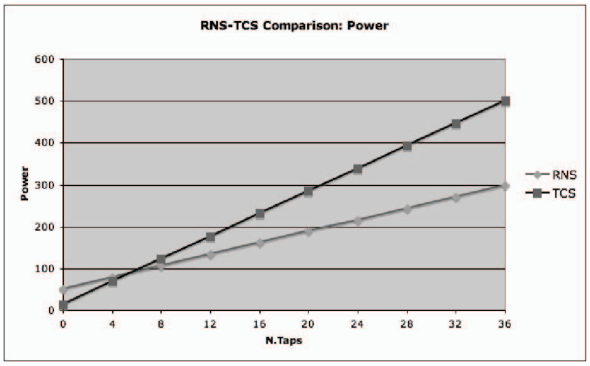
Power consumption contributions are divided between local and global interconnects. While local interconnects route signals inside a functional block, global interconnects route signals among different blocks.

In Fig. 3 the breakdown for the selected system is shown. Fig. 3.a shows that local interconnects play a fundamental role in power consumption, followed by local and global clock distributions.

The breakdown of the local interconnect power consumption is shown in Fig. 3.b. Contributions are divided in three sets: 1) power required for gate charging (requiring about 50% of the overall local power), 2)

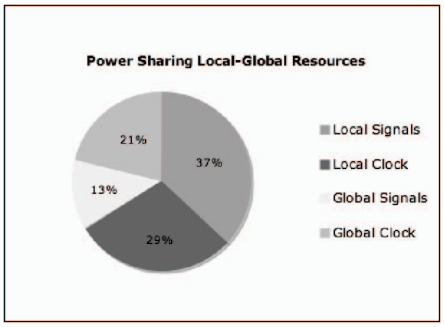


a) RNS vs. TCS: Area comparison

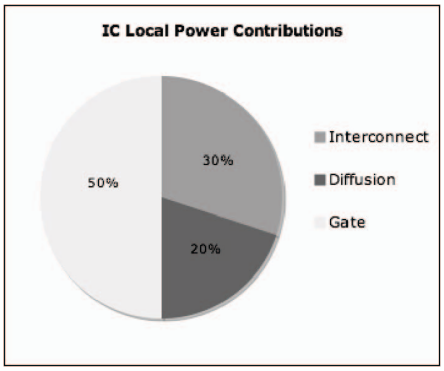


b) RNS vs. TCS: Power comparison

Fig. 2. Comparisons of TCS and RNS implementations of FIR flters



a) IC power contributions



b) IC local contribution breakdown

Fig. 3. ASIC-SC power contributions

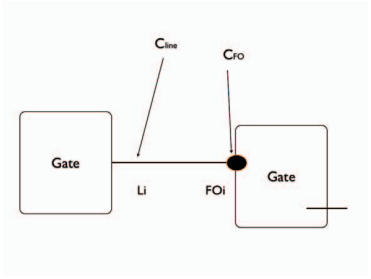


Fig. 4. Contributions to the node capacitance.

power required for interconnect capacitance charging (about 30%), and 3) power required for charging diffusion capacitances (about 20%).

In ASIC-SC technology, carry propagation properties of RNS potentially give the following area-power benefits:

1) reduction of complexity (number of gates - area),
2) reduction of interconnection capacitances.

For evaluating these effects, we develop a model linking power consumption to circuit complexity (area) and interconnects *locality*.

In our model we express ASIC area A in terms of number of (NAND2) equivalent gates. The term N_L represents the number of nodes. A capacitance $C(i)$ is connected at each node i. Considering a constant switching activity factor α, the power consumption is expressed by

$$P_{tot} = \Sigma_{i=1}^{N_L} \alpha(i) C(i) F V_{DD}^2 = \alpha C_{tot} F V_{DD}^2 \quad (3)$$

If α, F, and V_{DD}^2 can be considered constant values, we obtain that the total power P_{tot} is proportional to the total capacitance C_{tot}.

Above considerations allow to derive a model for studying RNS properties.

Fig. 4 shows the contributions to the capacitance of the node i.

For this capacitance, we can derive the following expression

$$C_{node} = C_{line} + C_{FO} = \alpha_C L_{line} + \beta_C FO_{node} \quad (4)$$

where C_{node} represents the node capacitance obtained combining line length L_{line} and node fan-out FO_{node} with two coefficients (α_C and β_C).

The mean value of this capacitance is

$$C_{mean} = \alpha_C L_{mean} + \beta_C FO_{mean} \quad (5)$$

Consequently, the total capacitance is

$$C_{tot} = N_L C_{mean} \quad (6)$$

where N_L is the number of nodes of the circuit.

From above expressions, we obtain the following value of power consumption

$$P_{tot} = \alpha N_L C_{mean} F V_{DD}^2 \quad (7)$$

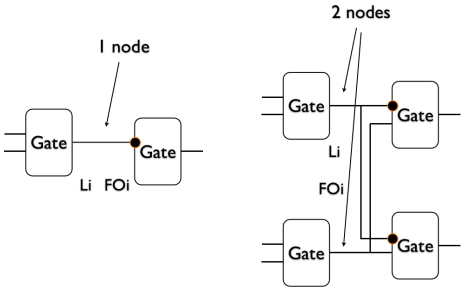


Fig. 5. Increase of node capacitances with complexity

If the number of gates increases, the global interconnect capacitance is modified for two reasons:

1) increase of the number of nodes (corresponding to the number of outputs);
2) increase of the wire length and Fan Out (corresponding to an increase of signal globality).

A rough model for the capacitance value can assume that

1) N_L is proportional to the circuit complexity (number of gates or area A): $N_L = \gamma A$
2) L_{mean} & FO_{mean} are related to the signal locality (through Globality Index or GI)
   $L_{mean} = \phi_1 GI; FO_{mean} = \phi_2 GI$

Consequently, we obtain
$C_{tot} = N_L C_{mean} = \gamma A(\alpha_C \phi_1 + \beta_C \phi_2) GI = K \cdot A \cdot GI$ with $K = \gamma(\alpha_C \phi_1 + \beta_C \phi_2)$
and, finally,

$$P_{tot} = \alpha C_{tot} F V_{DD}^2 = \alpha \cdot K \cdot A \cdot GI \cdot F \cdot V_{DD}^2 \qquad (8)$$

expresses the dependency of the power dissipation on the globality index GI.

A. Analysis of Experimental Results: ASIC-SC

In this section, we compare different results obtained for the RNS and the TCS implementation of FIR filters based on ASIC-SC technology.

From expression (8), we obtain the power ratio

$$\frac{P_{RNS}}{P_{TCS}} = \frac{\alpha K F V_{DD}^2 A_{RNS} GI_{RNS}}{\alpha K F V_{DD}^2 A_{TCS} GI_{TCS}} \qquad (9)$$

In our experiments, area (A_R) and power (P_R) ratios are measured. From these ratios we can derive the GI ratio (GI_R) as follows

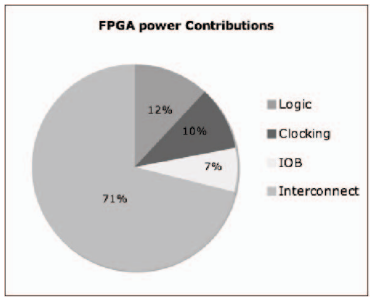
$$GI_R = \frac{GI_{RNS}}{GI_{TCS}} = \frac{P_{RNS}}{P_{TCS}} / \frac{A_{RNS}}{A_{TCS}} = \frac{P_R}{A_R} \qquad (10)$$

Using data from our previous work and equation (10) we obtain the ratios shown in Table I.

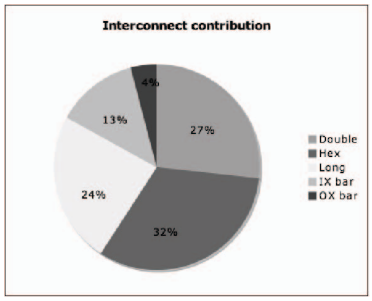
With the exception of filter 3 in Table I, the results show that the locality (GI) does not play a significant role in power savings for the ASIC-SC implementation

	Filter	A_R	P_R	GI_R
1	64-taps transp. FIR	0.630	0.560	0.890
2	8-taps direct FIR	0.994	0.990	0.995
3	64-taps complex FIR	0.578	0.340	0.589
4	Polyphase Filter	0.747	0.624	0.850

TABLE I
AREA, POWER AND GLOBALITY RATIOS FOR ASIC-SC.



a) FPGA power contributions



b) FPGA power interconnect

Fig. 6. Power consumption breakdown for FPGA implementation

(GI_R is close to 1). For the complex FIR filter (entry 3 in Table I), the difference is related to the activity factor that is significantly smaller for the Quadratic RNS (QRNS) as highligthed in [9].

V. FPGA: POWER CONSUMPTION CONTRIBUTIONS

In FPGA implementations factor weights are quite different. An analysis of the FPGA power consumption identifies three main contributions ([11]):

1) power consumption in logic and IOB;
2) power consumption in clocking structure;
3) power consumption in interconnects.

The breakdown of these different contributions in general FPGA implementations is shown in Fig. 6. From Fig. 6.a it is possible to observe that, for this technology, interconnect power consumption plays a fundamental role.

Moreover, the use of a hierarchical structure increases power consumption for global interconnecting. These different types of interconnects are summarized in Tab. V. Different interconnects are used for routing signals characterized by a different degree of locality. Capacitances are noticeably different for local or global

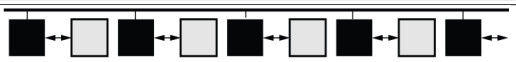
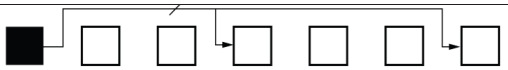
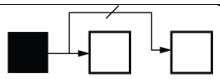
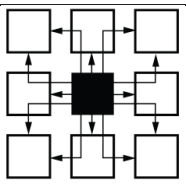
Name	Description	Architecture	Capac.
Long lines	span the full height & width		26.10 pF
Hex lines	route signal to every 3rd or 6th block		18.40 pF
Double lines	route signal to every 1st or 2nd block		13.20 pF
Direct Conn.	route signal to neighboring blocks		v.low

TABLE II
INTERCONNECTS IN FPGA.

	Description	$\frac{A_{RNS}}{A_{TCS}}$	$\frac{P_{RNS}}{P_{TCS}}$	$\frac{GI_{RNS}}{GI_{TCS}}$
1	8-taps FIR	1.1	0.612	0.554
2	16-taps FIR	0.947	0.51	0.539
3	8-taps CS FIR	1.095	0.533	0.487
4	16-taps CS FIR	0.93	0.416	0.447

TABLE III
AREA, POWER AND GLOBALITY RATIOS FOR FPGA.

lines of interconnect. As shown in Fig. 6.b, global interconnects give a great contribution to the overall power consumption. From this observation, we can expect a greater advantage in exploiting RNS locality in the FPGA technology.

A. Analysis of Experimental Results: FPGA

For the analysis of FPGA consumption we apply the same model used for ASIC-SC.
In this case, the area corresponds to the number of slices used in the implementation.

The effects of RNS locality are evident in Tab. III. In fact, for the FPGA technology the GI ratio (GI_R) is about 0.5.

VI. CONCLUSIONS

In this paper we compared power consumption of RNS and TCS for DSP applications. The analysis was performed on ASIC-SC and FPGA implementations.
Different contributions were analyzed using area (A), power (P) and global index (GI) ratios
Our analysis gave the following results.

1) The RNS allows to reduce power, both in ASIC-SC and FPGA implementations.
2) The FPGA implementation exploits both the complexity reduction and the locality of the RNS representation.

These favorable properties of RNS allow to extend the use of FPGA technology to power constrained DSP systems.

REFERENCES

[1] G.C. Cardarilli, A. Nannarelli and M. Re. "Reducing Power Dissipation in FIR Filters using the Residue Number System", *Proc. of 43rd IEEE Midwest Symposium on Circuits and Systems*, p. 320-323, Lansing, USA, Aug. 2000.

[2] A. D'Amora, A. Nannarelli, M. Re and G.C. Cardarilli. "Reducing Power Dissipation in Complex Digital Filters by using the Quadratic Residue Number System", *Proc. of 34th Asilomar Conference on Signals, Systems, and Computers*, p. 879-883, Asilomar Conference Grounds, California (USA), Nov. 2000.

[3] A. Nannarelli, M. Re and G.C. Cardarilli. "Tradeoffs between Residue Number System and Traditional FIR Filters", *Proc. of IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. I I, p. 305-308, May 2001.

[4] A. Del Re, A. Nannarelli, and M. Re. "Implementation of Digital Filters in Carry-Save Residue Number System",*Proc. of 35th Asilomar Conference on Signals, Systems, and Computers*, p. 1309- 1313, Asilomar Hotel and Conference Grounds, Pacific Grove, California (USA), Nov. 2001.

[5] G.C. Cardarilli, A. Del Re, A. Nannarelli and M. Re. "Power Characterization of Digital Filters Implemented on FPGA", *Proc. of 2002 IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. V, p. 801-804, May 2002.

[6] A. Nannarelli, G.C. Cardarilli, and M. Re. "Power-delay tradeoff in residue number system", *Proc. of 2003 IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. V, p. 413-416, May 2003.

[7] G.C. Cardarilli, A. Del Re, A. Nannarelli and M. Re. "Low-Power Implementation of Polyphase Filters in Quadratic Residue Number System", *Proc. of 2004 IEEE International Symposium on Circuits and Systems (ISCAS)*, Vol. I I, p. 725-728, May 2004.

[8] G.C. Cardarilli, A. Del Re, A. Nannarelli and M. Re. "Low Power and Low Leakage Implementation of RNS FIR Filters", *Proc. of 39th Asilomar Conference on Signals, Systems, and Computers*, Asilomar Hotel and Conference Grounds, Pacific Grove, CA, USA. October 30 - November 2, 2005.

[9] T. Stouraitis and V. Paliouras, "Considering the alternatives in Low-Power Design", in *IEEE Circuits and Devices*, July 2001.

[10] N. Magen, A. Kolodny, U. Weiser and N. Shamir "Interconnect-power Dissipation in a Microprocessor", *Proc. of SLIP'04*, Paris, France, February 2004.

[11] L. Shang, A.S. Kaviani and K. Bathala " Dynamic Power Consumption in Virtex-II FPGA Family", *Proc. of FPGA'02*, Monterey, California, USA, February, 2002.