Evaluating the Impact of Data Encoding Techniques on the Power Consumption in Networks-on-Chip

José C. S. Palma [1], Leandro Soares Indrusiak [2], Fernando G. Moraes [3],
Alberto Garcia Ortiz [2], Manfred Glesner [2], Ricardo A. L. Reis [1]

[jcspalma, reis]@inf.ufrgs.br, [lsi, agarcia, glesner]@mes.tu-darmstadt.de, moraes@inf.pucrs.br

1. *PPGC - II - UFRGS - Av. Bento Gonçalves, 9500, Porto Alegre, RS – Brazil*

2. *MES – TU Darmstadt – Karlstr. 15, 64283 Darmstadt - Germany*

3. *PPGCC - FACIN – PUCRS - Av. Ipiranga, 6681, Porto Alegre, RS – Brazil*

Abstract

This work addresses the problem of power consumption in networks-on-chip (NoCs). It investigates the reduction of dynamic power consumption through the reduction of transition activity using data coding techniques. Power macromodels for various NoC modules were built, allowing the estimation of the power consumption as a function of the transition activity at each module's inputs. Such macromodels were embedded in a system model and a series of simulations were performed, aiming to analyze the trade-off between the power savings due to coding techniques versus the power consumption overhead due to the encoding and decoding modules.

A network-on-chip (NoC) is an infrastructure essentially composed of routers interconnected by communication channels. It is suitable to support the GALS paradigm, since it provides asynchronous communication, high scalability, reusability and reliability [1][2].

The growing market for portable battery-powered devices adds a third dimension (power) to the previously two-dimensional (speed, area) VLSI design space [3]. NoCs are not particularly efficient in power consumption. They solve the problem of high capacitances in wires and, consequently, the power consumption in long lines, but on the other hand the power consumed by the router modules is not negligible.

The dynamic power consumption in a NoC grows linearly with the amount of bit transitions in subsequent data packets sent through the interconnect architecture. Using the Hermes NoC architecture [4] as a case study we could show that bit transitions affect the dynamic power consumption by as much as 6400% for interconnect lines, 180% for router input buffers and 20% for router control logic. Based on such results, this paper addresses the hypothesis that the power dissipation on a NoC-based system can be reduced by encoding the data sent through the network with coding schemes that reduce the average number of signal transitions [5]. Several such schemes were proposed in the late 90's, all of them addressing bus-based communication architectures. The contribution of this work is on the evaluation of such schemes in the context of NoC-based systems and on the trade-off analysis of the power savings obtained by the application of such coding schemes versus the power consumption overhead due to the additional encoding and decoding circuitry.

In Networks-on-chip, the data is transmitted in packets which are sent through routers, from one source to one target core. Encoding and decoding operations must be done in the source and target cores only, so to convert the original data to the encoded (and transmitted) one and vice-versa. In this work the encoder and decoder modules were inserted in the local ports of the routers, that is, between the integrated cores and the NoC interconnect structure.

For the latest CMOS technologies, static power accounts for the smallest part of the overall consumption. Accordingly, this work focuses only on NoC dynamic power consumption, using it as an objective function to evaluate the quality of data encoding schemes.

Dynamic power consumption is proportional to the switching activity arising from packets moving across the network. Interconnect wires and routers dissipate dynamic power. Several authors have proposed to estimate NoC power consumption by evaluating the effect of bits/packets traffic on each NoC component.

Average power per hop[1] (*APH*) is used here to denote the average dynamic power consumption in a single hop of a packet transmitted over the NoC.

Proceedings of the 2006 Emerging VLSI Technologies and Architectures (ISVLSI'06)
0-7695-2533-4/06 $20.00 © 2006 IEEE

Table 2 – Experimental results.

Stream	Bit transition reduction (reported in [5])	Bit transition reduction (simulated within the scope of this work)	Average Power consumption without encoding	Average Power consumption with encoding	Encoding power consumption	# of hops
HTML	9,3 %	10,5 %	18,58 mW	18 mW	23 mW	39
GZIP	16,3 %	8%	20,21 mW	19,7 mW	23,65 mW	46
GCC	15,6 %	4 %	19,8 mW	19,7 mW	23,5 mW	235
Bytecode	-	13 %	23,26 mW	22,24 mW	24,64 mW	24
Synthetic 1	-	75 %	31,39 mW	21,23 mW	25,9 mW	3
Synthetic 2	-	- 11 %	20,9 mW	21,92 mW	26,02 mW	-

APH can be split into three components: average power consumed by a router comprised of buffers, router wires and logic gates for switching (*APR*); average power consumed on a link between routers (*APL*); and average power consumed on a link between the router and the system core attached directly to it (*APC*). Equation (1) gives the average power consumption of a packet transmitted through a router, a local link and a link between routers.

(1) $APH = APR + APL + APC$

Moreover, our analysis showed that a better understanding of the average power consumption in the router (*APR*) can be achieved by dividing it into its buffer (*APB*) and control (*APS*) components. This is because the bit transition effect on power consumption at the router control is much smaller than its effect on the power consumption at the router buffer.

In regular tile-based architectures, tile dimension is close to the average core dimension, and the core inputs/outputs are placed near the router local channel. Therefore, *APC* is much smaller than *APL* and may be safely neglected without significant errors in total power dissipation. Therefore, Equation (2) computes the average router-to-router communication power dissipation, from tile τi to tile τj, where η corresponds to the number of routers through which the packet passes.

(2) $RRPij = \eta \times (APB + APS) + (\eta - 1) \times APL$

Considering now an approach with data encoding in the NoC local ports, two new parameters can be introduced to Equation (2): *APE* and *APD* (encoder and decoder average power consumption, respectively), producing the Equation (3).

(3) $CRRPij = APE + \eta \times (APB + APS) + (\eta - 1) \times APL + APD$

The parameters of the macromodel were acquired after the SPICE simulation of the communication infrastructure and encoding modules with different traffic patterns. Table 2 shows the experimental results obtained by system level simulation within Ptolemy II, using the macromodels described above. The first column describes the type of traffic. The second column reproduces the reduction of transition activity reported by Benini et al. in [7] (when available) and the third column presents the reduction of found on our experiments by reproducing the coding techniques proposed in [5]. Both cases report the results in terms of reduction in the number of transitions with respect to the original data streams.

The fourth column shows the power consumption (*APH*) without use of data coding techniques, while the fifth column shows the same measurement when coding techniques are used. Finally, the sixth and seventh columns presents the power consumption overhead due to the encoder and decoder modules (*APE* + *APD*) and the number of hops which are needed to amortize this overhead.

Experiments with the transmission of synthetically generated streams showed that the use of coding techniques can be advantageous if the average hop count for packages is larger than 3 (so the overhead of encoding and decoding is amortized), or larger than 24 hops for a java bytecode (power reduction of 1,02 mW/switch). By synthesizing a variety of traffic patterns, we could see that the power savings due to coding range from -1 to 18 mW/switch, thus pointing the direction for further research addressing the use of multiple encoding schemes to better match the transition activity patterns, and the use of NoC topology and routing information to help the decision whether a given stream should be encoded or not.

References

[1] A. Iyer and D. Marculescu. "Power and performance evaluation of globally asynchronous locally synchronous processors". 29th Annual International Symposium on Computer Architecture (ISCA), pp. 158-168, May 2002.

[2] W. Dally and B. Towles. "Route packets, not wires: on-chip interconnection networks". Design Automation Conference (DAC), pp. 684–689, June 2001.

[3] D. Singh, J. M. Rabaey, M. Pedram, F. Catthoor, S. Rajgopal, N. Sehgal and T. J. Mozdzen. "Power conscious cad tool and methodologies: A perspective". Proc. IEEE, vol.83, pp. 570-594, Apr. 1995.

[4] F. Moraes, N. Calazans, A. Mello, L. Möller and L. Ost. "HERMES: an infrastructure for low area overhead packet-switching networks on chip". The VLSI Journal Integration (VJI), vol. 38, issue 1, pp. 69-93, Oct 2004.

[5] L. Benini, A. Macii, E. Macii, M. Poncino, R. Scarsi. "Synthesis of low-overhead interfaces for power-efficient communication over wide buses". DAC, 1999. Proceedings. 36th, 21-25 June 1999 Page(s):128 – 133.

0-7695-2533-4/06 $20.00 © 2006 IEEE