2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). Mexico City, Mexico. November 10-12, 2021

Reduction Of Energy Consumption in NoC Through The Application Of Novel Encoding Techniques

Dr Ramakrishna Vakulabharanam
Department of ECE
School of Engineering & Technology, Sri Padmavati Mahila Visvavidyalayam
Tirupati, Andhra Pradesh, India
ramakrishna.bharan@gmail.com

***Abstract*— With the advancement of technology, the power consumed by network on chip connections (NOCs), particularly routers and network interfaces, comes into the picture to compete with the power consumed by other network communication subsystems. This paper provides a set of data coding techniques for reducing NoC connection power consumption. The proposed techniques are generic and transparent in terms of the NoC fabric (i.e. no changes to routers or connection design). Experiments on simulated and real-world traffic demonstrate that the suggested methods can save more than 50% of power dissipation and less than 15% reduction in energy consumption without compromising performance with reduced percentage area overhead in network interfaces.**

Keywords*— Network On-Chip, Encoding, Decoding, Power Consumption, VLSI

I. INTRODUCTION

Moving to the next node results in faster and more powerful gates, but the wires are slower. In actuality, nearly half of the dynamic power in connectivity processors is wasted [1], and this figure is expected to rise to 65 percent to 80 percent over the next several years. The global link is not lengthened, and the local cables are not lengthened. Because of the continual rise in chip function, chip size remains practically constant, whereas RC latency increases exponentially. The RC-wire delay, for example, is at least 1 mm global wire at 32/28 nm [2], which is at least 5 times larger than the intrinsic delay. If the raw computational power looks infinite, because more and more cores are installed in one single silicone, scaling problems are the actual challenge because of the requirement to communicate effectively and reliably with an ever rising number of cores. The on-chip network (NoC) architecture paradigm is largely regarded as the most feasible solution to the ultra-deep metre's scalability and predictability problems. In today's world, on-chip communication concerns are just as important as computational issues, if not more so in some cases. In practise, the communication subsystem influences traditional design objectives such as costs (i.e. the silicon area), performance, energy dissipation, consumption of energy and dependability. As technology improves, the communication subsystem consumes a larger portion of chip's overall power budget.

Network connections are as energy-efficient as routers and network (NIs)[3] and are anticipated to increase with progress in technology. This work propose a set of data encoding methods that function at the flit and end-to-end levels, allowing us to reduce switching and coupling switching activity on packet routing route connections. At both the algorithmic and architectural levels, the suggested encoding techniques are given and discussed, as well as evaluated using simulations in both synthetic and real-world traffic scenarios. In the coming years, chips with 1000 cores are projected to be accessible. These chips' interconnection networks use a considerable part of the system's total power budget. As a result, most of the research in the literature on NoC designs has concentrated on the creation of power-efficient connection networks. Various interconnection network components like as routers, NIs, and connections are the subject of these research. Because the objective of this study is to cut down on how much energy is lost through connections, Shielding, expanding line-to-line spacing, and inserting repeaters are all instances of these techniques. They're all meant to increase the surface area of the chip. The adoption of a data encoding scheme was another method for lowering connection power usage [4].

There are two different sorts of data encoding techniques. The first group of encoding techniques focuses on decreasing power dissipation due by individual bus lines' self-switching activity while ignoring power dissipation.

II. NETWORK-ON-CHIP

As the contemporary VLSI technology enables such comprehensive integration with transistors, the amount of resources for computation on a single chip has risen rapidly in order to address the rising demands of high-performance computing applications and low-power systems. There are many extra computer resources, including CPUs, DSPs, IPs and much more[5]. In a System-on-Chip, the interconnection of the systems becomes another tough problem to address. However, because only one master may utilise the bus at a time, the scalability of such shared bus connection is restricted, necessitating the arbitrator to serialise all bus accesses. Traditional large-scale multiprocessors and distributed computing networks inspired the fundamental concept. The adoption of NoC-based systems has grown because to its scalability and flexibility, as well as their assistance with effective on-chip connectivity. As a result, if the current trend continues, one of the most important issues to consider is wiring.

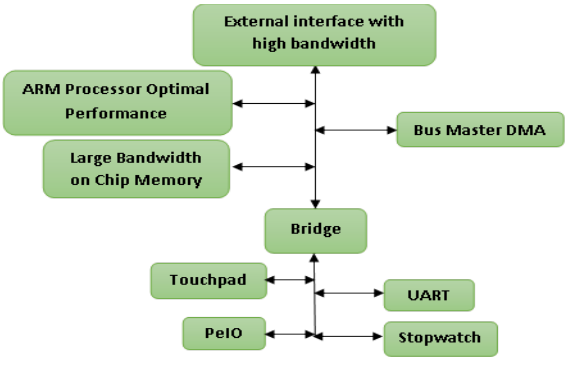

Fig 1. Typical system for ARM & AMBA

978-1-6654-0029-9/21/$31.00 ©2021 IEEE

As illustrated in figure 1 basic construction of ARM-AMBA[6] the majority of sophisticated bus architectures employ a hierarchical structure to provide a scalable transmission and separate communication domains into different bandwidth requirement groups like high performance and low performance and so on. In a System-on-Chip (SoC) environment, numerous bus-based designs and point-to-point communication techniques are typical ways to communicate across several cores. The bus systems are the most prevalent because to their simplicity and user-friendliness. On the other hand, the bus-based design has a basic bandwidth restriction, as the number of components associated with a bus grows, so does the true capacity of the bus lines and the cabling delay.

Another possibility in future technologies is to employ network connections, commonly known as network-on-chip (NoC) architecture, to avoid such a large cable delay. The basic idea of such links may be derived from the creation of the contemporary computer network, as mentioned before. The use of network communication between a few routers and each communication object can reduce the needed wire. As a result, the switch-based connection method offers good scalability and eliminates the need for complicated wiring. When economic data demonstrates that NoC decreases SoC production costs, SoC buses are replaced with NoCs[7]. In terms of system performance, the NoC method provides a number of benefits over conventional buses. In terms of performance and complexity, cross-bar hierarchies or multi-level buses fall midway between regular buses and NoC, although they are still far behind NoC. The study of the interfaces between NoC processing components and connected structures is critical to the effectiveness of the NoC design. The connectivity of a SoC produced process has several problems in terms of slow bus reaction time, energy restriction, scalability issues, and bandwidth limitation. Due to the influence of bus sharing, the connection between a bus and a network interface with a high number of components may result in a slow interface time. Furthermore, the connectivity has a fault in that all communication devices consume an excessive amount of power. Furthermore, because to the bandwidth limitations of a bus, the number of connections to the components cannot be expanded indefinitely. As a result, the NoC architecture's performance is heavily dependent on the connection paradigm. Although network technology is well established in the computer network, achieving chip-level intercommunication without any adjustments or reductions is basically impossible. As a result, a number of researchers are working on network topologies that are suited for communication on-chip.

A. NoC Design:

Figure 2 depicts the fundamental NoC architecture.

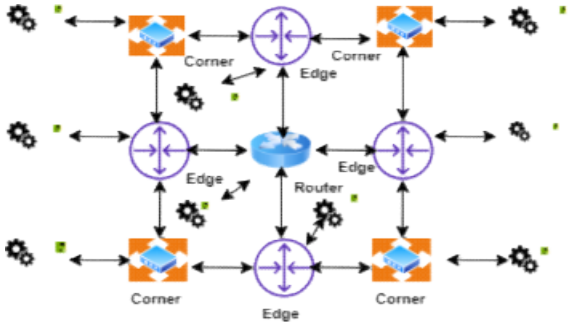

Fig.2. NoC Architecture

The structure of the NoC design is shown in Figure 2. Data is sent from the Processing Elements to Network Interfaces[8], which bundles it into flits and transmits it to the routers. Corner routers, edge routers, and routers are the three types of routers. To choose the majority of body flits for transmission into crossbar, switch arbiter uses a two-stage arbitration method. The arbiter will begin working as soon as the item arrives. The switch arbiter process begins when the flit arrives. The router will then unregister the transmission channel after the last flit, which is known as the tail flit. There is mesh, star, and fat tree router topologies to choose from.

III. Overview of Proposed Scheme

The recommended approach encrypts and reduces the switching activity on the links via which the flit pass before it is injected into the network. The automatic dissipation of the link power is the cause of automatic switching and coupling activity. This paper describes the design of decoders and encoders, as well as how they may be used to reduce network interface power dissipation.

A. Encoder Architectue

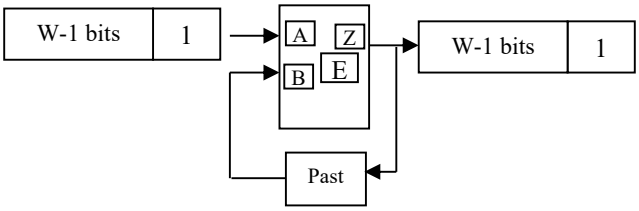

Fig.3.Encoder Architecture.

Figure 3 shows the encoder architecture which has two major components. Encoder block and storage element (including encoding logic) (which consists of previously encoded flit). The block encoder is given as an input by w bits of data. In this scenario, w-1 bits of data and 1 bit inversion data are sent to the encoder[9]. The input data is strangely flipped, even reversed or even inverted depending on the encoding system if the reversal bit is set to 1.

B. Decoder Architectue

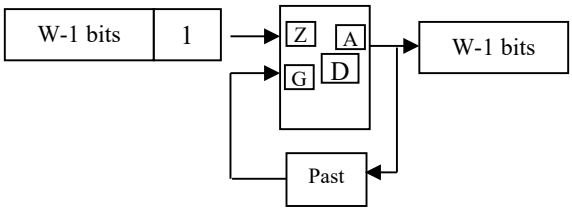

Fig.4. Decoder Architecture.

In the decoder architecture which is shown in figure 4, there are two major components. Decoder block and storage element (which contains logic decoding) (which consists of previously decoded flit)[10]. The decoder block is provided with w bits of data. The encoder generates w-1 bits of data and one bit is a bit of reversal. If the inversion bit is set to 1, the input information in the decoder block depending on the method of decoding is either odd or inverted. If the reverse bit is set to 0, the input of the decoder will not be reversed. The encoder and decoder block format are common to all three techniques. The basic encoding and logic of each scheme will vary. Comparing the efficiency and power dissipation of connections using various data coding

methods[11]. Since all connections on the route have the same flit sequence, the choice of encoding on the NI can save the same power on all connections[12]. An encoder and a decoder block will be added to the NI for the suggested method. By encoding the output flows of the packet, the encoder decreases the energy required for the inter-router connection point-by- point.

C. Proposed Encoding Schemes

This paper, provided two data encoding strategies for dynamic power dissipation, as well as the potential use of decoding to minimise network connection.

Scheme 1: The first scheme focuses on lowering Type I (converted to Type III and Type IV) and Type 2 numbers (by converting them to Type I transition). The approach compares current data to past data to see if an unusual reversal of current data or no reversal at all, might result in a loss of power. A section of the link is utilised to demonstrate whether or not the flash travelling across the connection is inverted. The NI connects the body to W-1 bits in particular. The NI's encoding logic E must determine whether or not the reversal is to take place and, if so, reverse it.

Scheme 2: The odd and complete investment in the proposed encoding scheme II are both used, as previously indicated. Scheme II focuses on minimising Type I to Type III and IV transitions and Type II to Type I and Type II to Type IV transitions. In order to assess if an unknown or full reverse in the existing data might cause a connection power loss, the system compares current data to previous ones.

The second stage consists of a single block collection with one input block. The blocks' output has a width of log2 w. When the pair bits are inverted, the output of Top 1s influences the number of transitions that cut power. The amount of transitions with full inversion pairs decreases the connection power in the central 1s block. Finally, the number of transitions is represented by the 1s at the bottom of the block; the entire pair of bits reverses to give additional connecting power. Using the quantity of 1s for each transition type to determine if a damage reversal or complete inversion operation should be performed to save energy.

IV. RESULTS AND DISCUSSIONS

The RTL Schematic of scheme1 shown in figure 5(a) &(b) shows that Data Out is created based on the clock reset and data input. Depending on the amount of power consumed by data encoding, the data is either odd inverted or not inverted.

Scheme1

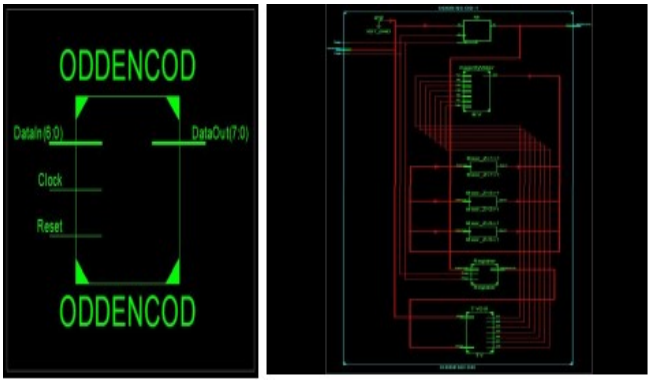

Fig.5(a) RTL Schematic of Scheme 1. 5(b) Internal view

bIf given Clock at positive edge and Reset = 1 then whatever be the value of Data Input, the Data Out will be always = 0.

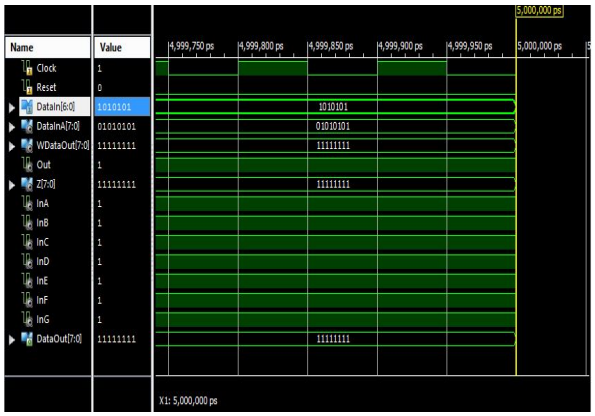

Fig.6 (a) Scheme1 when Reset = 0 but Not-Inverted

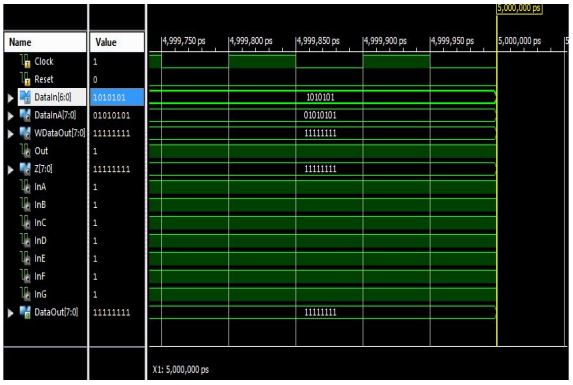

Fig.6 (b) Scheme1 Result when Reset = 0 but Inverted

Figure 6(a) and 6(b) shows if given Clock at positive edge and Reset = 0 then depending on the value of Data Input, the Data Out will be either odd inverted or not inverted. The inversion is based on the power calculation with inversion of data. Odd inversion based on Odd inversion Table.

Scheme2

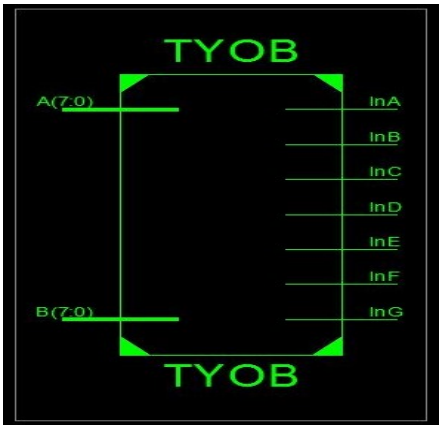

Fig.7(a) RTL Schematic of Scheme2.

Here in the RTL Schematic, it is shown that Depending on clock reset and Data input, Data Out is generated. The Data in either odd inverted full inverted or not inverted, depending on the amount of power consumption by the encoding of data. An extra block is added to the scheme I, to check full inversion and is shown in Fig.7(a) and 7(b) .

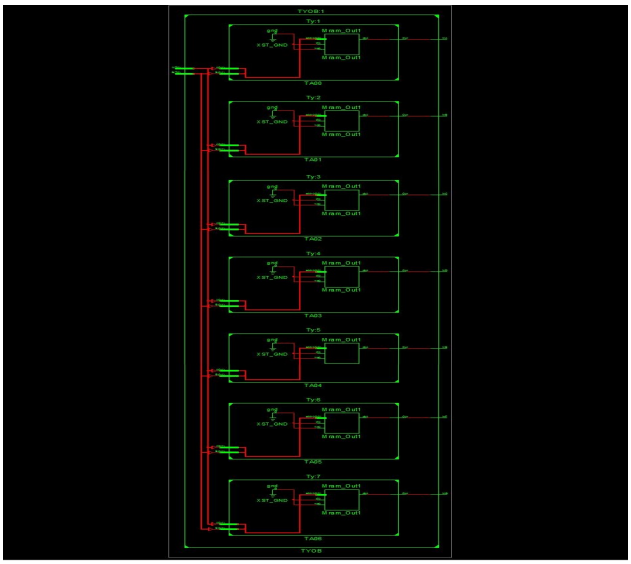

Fig.7(b) Internal View of RTL schematic TYOB.

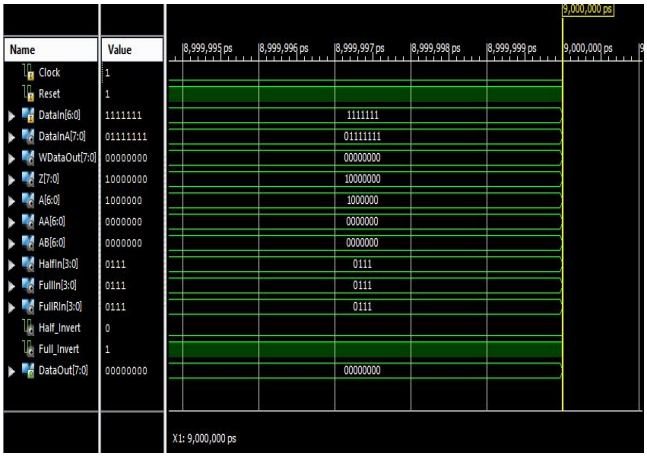

Fig.7(c) Scheme2 Result when Reset = 1

Figure 7(c) shows that If given Clock at positive edge and Reset = 1 then whatever be the value of Data Input, the Data Out will be always = 0.

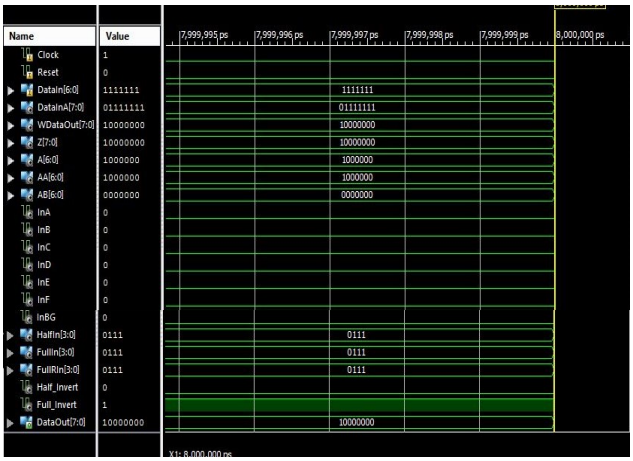

Fig.7(d). Scheme2 Result when Odd Inverted

If given Clock at positive edge and Reset = 0 then depending on the value of Data Input, the Data Out will be either odd inverted, even inverted or not inverted. The inversion is based on the power calculation with inversion of data is shown in figure 7(d).

CONCLUSION

In this work, it is suggested to minimise the power wasted by a NoC a number of new approaches for data coding. Connections really represent a large part of the overall power consumption of the communication system. Moreover, at future technological nodes its relevance will surely rise. Compared with early encoder schemes, which are proposed in the literature and which are mainly responsible for dissipating the power of the connections under the deep sub-micrometer technology regime, the proposed scheme seeks to minimise not only change but also (and in particular) coupling activity. In the sense that using the described encoding schemes, no modifications to the router or the underlying NoC architecture are required. The logical implications of the NI encoder and decoder were thoroughly studied. An study of the encoder's dissipation and silicone area using the techniques given. The efficiency, power, and energy measurement implications were assessed in synthetic and actual traffic situations using the cyclic and bit-precise NoC simulator. The suggested encoding approaches may often cut power dissipation by up to 51% and NI energy use by 14% without sacrificing performance.

REFERENCES

[1] Pelley S, Meisner D, Zandevakili P, Wenisch TF, Underwood J. Power routing: dynamic power provisioning in the data center. ACM SIGARCH Computer Architecture News. 2010 Mar 13;38(1):231-42.

[2] Qian W, Babecki C, Karam R, Paul S, Bhunia S. Enfire: A spatio-temporal fine-grained reconfigurable hardware. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2016 Jun 24;25(1):177-88.

[3] Li CL, Yoo JC, Han TH. Energy-efficient custom topology generation for link-failure-aware network-on-chip in voltage-frequency island regime. JSTS: Journal of Semiconductor Technology and Science. 2016;16(6):832-41.

[4] Jafarzadeh N, Palesi M, Eskandari S, Hessabi S, Afzali-Kusha A. Low energy yet reliable data communication scheme for network-on-chip. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2015 Jun 2;34(12):1892-904.

[5] Bahn JH, Lee SE, Yang YS, Yang J, Bagherzadeh N. On design and application mapping of a Network-on-Chip (NoC) architecture. Parallel Processing Letters. 2008 Jun;18(02):239-55.

[6] Deeksha L, Shivakumar BR. Effective design and implementation of AMBA AHB bus protocol using Verilog. In2019 IEEE International Conference on Intelligent Sustainable Systems (ICISS) 2019 Feb 21 (pp. 1-5).

[7] Hammami O, Jabbar MH. NOC-based MPSoC design and implementation on FPGA: DCT application. In2012 4th IEEE Asia Symposium on Quality Electronic Design (ASQED) 2012 Jul 10 (pp. 277-285).

[8] Silva MR, Zeferino CA. Confidentiality and Authenticity in a Platform Based on Network-on-Chip. In2017 VII IEEE Brazilian Symposium on Computing Systems Engineering (SBESC) 2017 Nov 6 (pp. 225-230).

[9] Sarkar S, Biswas A, Dhar AS, Rao RM. Adaptive bus encoding for transition reduction on off-chip buses with dynamically varying switching characteristics. IEEE Transactions on Very Large Scale Integration (VLSI) Systems. 2017 Aug 3;25(11):3057-66.

[10] Ercan F, Tonnellier T, Gross WJ. Energy-efficient hardware architectures for fast polar decoders. IEEE Transactions on Circuits and Systems I: Regular Papers. 2019 Oct 2;67(1):322-35.

[11] Rajendran B, Sebastian A, Schmuker M, Srinivasa N, Eleftheriou E. Low-power neuromorphic hardware for signal processing applications: A review of architectural and system-level design approaches. IEEE Signal Processing Magazine. 2019 Oct 31;36(6):97-110.

[12] Prasad N, Chakrabarti I, Chattopadhyay S. An energy-efficient network-on-chip-based reconfigurable Viterbi decoder architecture. IEEE Transactions on Circuits and Systems I: Regular Papers. 2018 Apr 24;65(10):3543-5