RIBiT: Reduced Intra-flit Bit Transitions for Bufferless NoC

Akshay Sarman*, Alwin Shaju*, Rose George Kunthara*, Neethu K*, Rekha K James* and John Jose[†]

* Division of Electronics Engineering, School of Engineering, CUSAT, Cochin, India

[‡] Dept. of Computer Science and Engineering, Indian Institute of Technology Guwahati, Guwahati, India

akshay4me@ug.cusat.ac.in, alwin.shaju@ug.cusat.ac.in, rosegeorgekunthara@cusat.ac.in, neethukuriyedam@cusat.ac.in, rekhajames@cusat.ac.in, johnjose@iitg.ac.in

Abstract—In modern Tiled Chip Multicore Processor (TCMP) systems, Network on Chip (NoC) is the preferred interconnect solution to overcome scalability and performance bottleneck issues that conventional bus-based architectures face. For low to medium NoC traffic, the energy and area efficient bufferless router is a better design choice compared to buffered structures. Dynamic power contributes to the majority of total power dissipation during data transmission whereas only a fraction of it is due to leakage power. Self-switching and cross-coupling activities across NoC links are responsible for total dynamic power, of which latter is the prime contributor. In any NoC system, data encoding techniques are generally employed at Network Interface (NI) level to minimize power dissipation across NoC links. We propose a data encoding mechanism for bufferless NoCs to minimize bit transitions within the flit which will result in reduced dynamic link power. Our suggested approach leverages a modified version of Delta encoding technique where the flit is encoded into data differences by a configurable module placed inside NI of each core. No additional control lines and hence no changes to the network are required for our proposed encoding scheme. Experimental analysis done using Xilinx Vivado shows that our proposed design approach has significant reduction in intra-flit bit transitions in comparison to the baseline designs.

Keywords—Network on Chip, cross-coupling, bufferless, delta encoding, deflection router

I. INTRODUCTION

Advancements in transistor technology facilitates the integration of multiple processors onto a single chip known as System on Chip (SoC). Network on Chip (NoC) is a scalable interconnect solution compared to traditional bus-based and point-to-point intercommunication frameworks for SoC [1], [2], [3]. Early designs in NoC show a strong inclination for buffered systems. However, with increase in core count, power and area consumed by NoC will also become significant. Even though buffered architectures endure a higher load capacity and more straightforward routing techniques, buffers take in nearly 30% of total power used by the chip [4], [5]. Thus bufferless NoC is a better choice for area efficient and lowpower NoC designs [6]. Experiments reaffirm that bufferless routers surpass buffered structures for low to medium network workloads [7].

Another major concern in modern Tiled Chip Multi Processor (TCMP) design is the power dissipation that occurs during data transmission across NoC links. This arises from the

978-1-6654-9005-4/22/\$31.00 ©2022 IEEE

switching within a link and cross-coupling between adjacent links. Although numerous methods exist to reduce the coupling capacitance between these links, they all cost additional physical area [8]. Encoding the flits at NI before injecting them into network is found to be helpful in minimizing the above-said effects [9], [10], [11]. Hence, we propose a configurable multistage encoding approach at NI for bufferless NoC to decrease number of intra-flit bit transitions. We employ Delta encoding in first stage of our design to transform original data into a representation with fewer 1s, resulting in fewer bit transitions within the flit.

Remainder of this paper is structured as follows. Section II gives an overview about relevant works related to crosstalk, dynamic link power reduction and data encoding techniques for the same. Section III discusses background & motivation for our design approach and Section IV explains our proposed design in detail. Experimental results are discussed in Section V and finally, Section VI concludes the paper.

II. RELATED WORK

The power dissipation along NoC links is generally reduced by using techniques such as shielding [12], increasing lineto-line spacing [13], and repeater insertion [14]. However, they incur extra chip area. This necessitates the adoption of suitable encoding schemes to minimize power dissipation across network links. Stan et al. proposes bus invert technique to decrease power consumption across network lines [15]. Their model first calculates Hamming distance between current bus value and subsequent data value. This is followed by bus inversion if Hamming distance is greater than $\frac{n}{2}$, where n is the bus width. INC-XOR [16] is another attempt to reduce the switching activity across interconnects. They provide seven encoding techniques and minimize number of switching transitions by assigning code words with lower transitions to the original signal that occurs more frequently.

To minimize power consumption and crosstalk, Yan et al. follows a technique where the data is first both odd and even inverted. Then transmission is carried out using a suitable type of inversion, which is selected conditionally, resulting in less coupling [17]. Fan et al. details how the coupling and switching activity is reduced up to 39% in buffered NoCs [18]. They propose a coding technique by taking advantage of end-to-end encoding for wormhole switching, as has been suggested by Palesi et al. [9]. This lowers the dynamic link power by eliminating only odd inverted transitions. Dehyadegari suggests an encoding mechanism to lessen dynamic energy consumption of NoC packets for a 16-core processor setup [19]. They propose Sig-NoC to predict energy consumption of each packet in the source node. Their model is able to reduce the number of 1s within every flit.

Shen et al. puts forth a configurable NoC with four encoding approaches to provide reliability and power efficiency with minimal impact on performance [20]. Firstly, they employ two encoding techniques to reduce frequency of two nearby transitions. The other two encoding techniques are designed primarily to eliminate cross-talk interference by transmitting one or two additional flits, respectively. Ascia et al. propounds a data encoding technique to cut back the power dissipation across NoC [21]. Their approach inverts bits of the flit to be transmitted if it results in reduction of both switching activity and coupling activity along NoC links.

Chen et al. modifies Smart NoC [22] to lessen the large interconnection overhead [23]. By shortening the existing wires and integrating switches to remove overlapping, their model achieves 63% and 15% reduction in area and dynamic energy respectively. The hybrid coding scheme designed by Behnam et al. employs a slow-transition fast-level (STFL) coding technique to overcome performance impact of low power links [24]. Dual Binary-Weighted Code (DBWC) [25] limits cross talk fault by obviating triplet opposite direction transition in the entire network. This is achieved by generating Forbidden Pattern Free (FPF) codes and thus DBWC minimizes number of NoC links.

Delta encoding is an effective coding scheme based on sending data in the form of differences instead of sending it as such [26], [27], [28]. To the best of our knowledge, majority of work on different encoding schemes to improve network performance of NoC systems are done for buffered architectures. Here we propose a novel configurable multilevel encoding design approach for a bufferless NoC system to minimize bit transitions inside a flit which leads to reduced cross-coupling.

III. BACKGROUND & MOTIVATION

Velayudham et al. employs various coding schemes such as Gray and odd-bit inversion to encode data in buffered NoC architectures [29]. We evaluate suitability of Gray encoding (GR), Odd-first Even-last encoding (OE) and a combination of Gray and Odd-first Even-last encoding (GR&OE) techniques in bufferless NoC to analyse reduction in intra-flit bit transitions. We assume a 64-core system arranged as an 8×8 mesh NoC with both link bandwidth and flit size as 128 bits (16 bytes). CHIPPER [7], a popular bufferless deflection router with reduced router complexity is considered for our experimental evaluation. Each and every flit is independently routed in a bufferless deflection router. To avoid livelock, CHIPPER employs a golden flit concept. Flit which acquires



Fig. 1: Normalized comparison of Delta encoding for different Base values in CHIPPER NoC

golden status is considered as the maximum priority flit in the network, such that it gets required port in every router.

For encoding, initial bits of the flit are not encoded as it contains destination address and golden flit status bit. We consider first byte of each flit to contain the 6-bit destination address for an 8×8 mesh NoC and 1 bit to indicate golden status. Remaining bits are encoded for minimizing bit transitions within the flit.

Experimental evaluations for a bufferless NoC done on GR, OE and GR&OE by realizing them at RTL level gives negligible improvement compared to scheme without any encoding approach, which is considered as baseline design technique. Functional simulations using Verilog test benches with 32KB random data gives intra-flit bit transition reduction of 2.12%, 1.06% and 2.87% respectively for GR, OE and GR&OE models.

To get more reduction in bit transitions, we employ Delta encoding approach where data is subdivided into chunks and are encoded using a Base, which refers to the reference value taken. The differences from Base, known as Deltas (Δ s) are found using following expression:

$$\Delta_i = Base - Chunk_i \tag{1}$$

where 'i' denotes index of each chunk in the flit. Thus the Base and Deltas together constitute encoded flit. Following are the 3 cases which we have considered to calculate Base value for Delta encoding in CHIPPER based NoC:

- Δ _B1: First byte of the flit is considered as Base
- Δ _B2: Largest valued byte within the flit is considered as Base
- Δ_B3: Mean of largest and smallest byte within the flit is considered as Base

Experimental evaluations are done on 64-core set up arranged as 8×8 mesh CHIPPER NoC for a 128-bit flit width with 1-byte chunk size. Analysis of Δ_B1 , Δ_B2 and Δ_B3 in comparison to baseline design gives 17.96%, 14.15% and 25.2% reduction in bit transitions within the flit respectively. Figure 1 shows comparison of the above 3 cases for a CHIPPER based NoC.



Fig. 2: Block diagram of Encoder



Fig. 3: Block diagram of Decoder

127 - 122	121	120	119 - 0
DEST. ADDRESS		ENCODING FLAG	DATA

Fig. 4: Type I flit format

The better results attained for Δ _B3 makes it a fair design option. We refer Δ _B3 as Model 1 in rest of this paper. Multilevel encoding along with Delta encoding further minimizes bit transitions in a flit.

In this paper, we propose an encoding scheme at NI that incorporates configurable encoding approach to bring down bit transitions within the flit that reduces cross coupling which in turn decreases dynamic link power dissipation.

IV. PROPOSED DESIGN

A three level encoding approach with Delta, Gray, Oddfirst Even-last encoding schemes is proposed. Based on the configuration selected, flit gets encoded in these 3 levels. Figures 2 and 3 portray the workflow of encoder and decoder.

A. Level 1: Delta Encoding

The functioning of encoder is given in Algorithm 1. Base for Delta encoding is taken as mean of largest (C_{large}) and smallest (C_{small}) chunk within the flit. Delta values from this base for each chunk (C_i) is calculated. A priority encoder is used which takes input as highest Delta and provides encoding bits (001-111) at the output. Value of these encoding bits indicate minimum number of bits needed to represent each Delta along with an extra bit for denoting the sign. Any flit with encoding bits higher than value 6 are named as Type I flits and original flit is sent as such without encoding as shown in Algorithm 1. This is indicated by 1-bit encoding flag in the

Algorithm 1 Base algorithm for Encoding

Input: 128-bit flit data $C_{large} \leftarrow Max[C_1, C_2...C_i]$ $C_{small} \leftarrow Min[C_1, C_2...C_i]$ $Base \leftarrow \frac{C_{large} + C_{small}}{2}$ $r \leftarrow C_{large} - Base$ $n \leftarrow (log_2(r) + 1) + 1$ // Number of bits needed // additional 1-bit to indicate sign if Number of bits, n < 6 then Encoding $Flag \leftarrow 0$ for each chunk C_i in the flit do $\Delta_i \leftarrow Base - C_i$ // Find the Deltas $P_i \leftarrow Pack[\Delta_i]$ // Type II $G_i \leftarrow Gray[P_i]$ // Gray coding $OE_i \leftarrow OddEven[G_i]$ // Odd-first Even-last end for else Encoding $Flag \leftarrow 1$ Original flit is sent as such // Type I

Type I flit structure as shown in Figure 4. Deltas of flits with encoding bits less than or equal to 6, named as Type II, are to be packed.

B. Packing

end if

After the Delta encoding stage, obtained Delta values may contain trailing zeroes. We pack them together by stripping off these unwanted zeroes. The number of zeros stripped off is according to highest Delta value. For doing this, the encoding bits are utilized. To preserve flit size as 128-bit, stripped off zeroes are appended as least significant bits (LSB) of the flit. Thus the resulting flit has a series of zeroes at its LSB positions, which will reduce number of intra-flit bit transitions, resulting in minimal cross-coupling activity across NoC links.

C. Level 2: Gray Coding

To minimize intra-flit bit transitions further, multi-level encodings are incorporated which is configured by the enabler. At Level 2, packed Delta-encoded flit is further Gray coded if it results in lesser number of bit transitions. Configuration status bits are set accordingly as shown in Table II to indicate whether Gray coding is performed or not.

D. Level 3: Odd-first Even-last Encoding

Odd-first Even-last encoding is performed at Level 3 to reduce number of bit transitions within the flit. The grouping of



ODD-FIRST EVEN-LAST ENCODING

Fig. 5: Block level representation of Type II encoding architecture



Fig. 6: Type II flit format

Odd-first Even-last flit is done if it results in lesser number of intra-flit bit transitions. This is indicated by the configuration status bits as shown in Table II.

E. Formatting

After multi-level encoding process, the final encoded flit is formatted as follows:

1) Type I: Since data is not encoded for Type I, flit format remains the same as shown in Figure 4.

2) Type II: Figure 5 depicts block level representation of Type II encoding architecture. The Type II flit undergoes different levels of encoding. This configurable multi-level encoding is indicated by 8-bit Base for Delta encoding, 3-bit encoding bits for the length of highest Δ , 2-bit configuration

TABLE I: Indication of encoding bits

Encoding bits	Flit type	Number of bits of highest $ \Delta $
000	Type II	0
001	Type II	1
010	Type II	2
011	Type II	3
100	Type II	4
101	Type II	5
110	Type I	6 (not encoded)
111	Type I	7 (not encoded)

TABLE II: Encoding type used at each NI for multi-stage encoding

Configuration Status bits	Encoding type
00	Δ only
01	Δ & GR
10	Δ & OE
11	Δ & OE & GR

status to determine levels of encoding as shown in Figure 6. Table I and II indicate encoding bits corresponding to length of highest Δ and encoding type employed in multistage encoding.



Fig. 7: Comparison of Encoding Models

As shown in Table I, Type II flit formatting is adopted only if the number of bits in absolute value of highest delta is less than or equal to 5. So, the total number of bits with sign will be 6 to represent difference. This gives an indication that at least two bits of every Delta are zeroes. 15 Deltas will be formed for 128-bit flit (16 byte chunks) format. While packing these 15 Deltas, 30 bits will be zeroes. 13-bits of these trailing zero bits are utilized to place meta data. The meta data includes 8bit Base, three encoding bits and two configuration status bits as shown in Figure 6. Thus our proposed design approach does not require extra control lines (metadata related to encoding technique) for flit transmission and no modifications to the router architecture.

Our proposed encoding scheme works well for networks with higher linkwidth. Also, when network is scaled, atmost 1 more byte in the flit will be required to denote destination address. Thus first 2 bytes of the flit may not be encoded for larger network sizes and remaining bytes inside the flit can undergo multi-stage encoding.

F. Decoder

The block level workflow of decoder is depicted in Figure 3. This includes the extraction of metadata like Base byte, encoding flag bit, encoding bits and configuration status bits followed by the three levels of decoding. Flit undergoes Odd-first Even-last decoding at first stage, followed by Gray to Binary decoding and finally Delta decoding depending on the value of configuration status bits.

V. RESULTS AND ANALYSIS

A. Experimental Setup

For experimental evaluations, we compare the baseline design (without any encoding scheme) against following four variants of our proposed design approach:

- Model 1: Only Delta Encoding
- Model 2: Delta with Gray coding
- Model 3: Delta with Odd-first Even-last Encoding
- Model 4: Delta with Gray and Odd-first Even-last Encoding

Encoder-decoder pair corresponding to the above designs are realized at RTL level using Verilog HDL. The functional simulation of all models are done in Xilinx Vivado Design Suite 2020.3 by generating a random 32KB data, using Verilog test benches. All the considered models are synthesized using Vivado targeted to Zynq Ultra Scale+ ZCU106 board to obtain hardware overhead incurred [30]. Parameters like count of bit transitions within the flit, LUT utilization and static power are considered for comparison of all four models.

B. Result Analysis

Figure 7a shows percentage reduction in bit transitions within flits for all the four variants compared against baseline design. Figure 7b and Figure 7c depict normalized hardware utilization and static power comparison of the models under consideration.

It is clear that Model 1 shows a 26.6% reduction in bit transitions whereas with addition of Gray coding stage, Model 2 show a reduction of almost 30% when compared to baseline model. Models 3 has only around 27% reduction in intra-flit bit transitions while Model 4 exhibits the highest reduction of 31.6%. Model 4 which incorporates all levels of encoding, shows promising results in intra-flit bit transition reduction at the expense of hardware overhead compared to other models. Though M2 and M3 have almost same hardware utilization, M2 shows better result in terms of bit transition reduction and static power.

VI. CONCLUSION

Our paper proposes a configurable multi-stage data encoding scheme in bufferless NoCs, primarily based on Delta encoding. All four variants of design are implemented in NI and data is encoded before injecting into the network. Our proposed encoding mechanism reduces number of bit transitions within a flit leading to reduced cross-coupling and dynamic power dissipation across NoC links. Unlike other encoding schemes modeled to minimize power dissipation across network links, our design eliminates the need for additional control lines to indicate encoding technique used. Thus, our multi-level encoding approach gives clear-cut advantage of including meta data for encoding in the 128-bit link itself. Our experimental analysis shows that Type II formatting delivered a fair reduction in bit transitions with little hardware overheads, achieving a maximum reduction of 31.6% in Model 4. Future improvements include exploring viability of converting Delta encoding mechanism into an efficient compression technique to further increase the system's data throughput for better network performance. Possibility of integrating a layer of encryption to this encoding scheme can be explored, which can add an extra degree of security to systems that process vital data.

REFERENCES

- W. J DALLY, "Route packets, not wires: On-chip interconnection networks," in *Proc. of IEEE Design Automation Conference*, 2001, pp. 684–689.
- [2] W. J. Dally and B. P. Towles, *Principles and Practices of Interconnection Networks*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2004.
- [3] L. Benini and G. De Micheli, "Networks on chips: a new soc paradigm," *Computer*, vol. 35, no. 1, pp. 70–78, 2002.
- [4] M. B. Taylor, W. Lee, J. Miller, D. Wentzlaff, I. Bratt, B. Greenwald, H. Hoffmann, P. Johnson, J. Kim, J. Psota *et al.*, "Evaluation of the raw microprocessor: An exposed-wire-delay architecture for ilp and streams," *ACM SIGARCH Computer Architecture News*, vol. 32, no. 2, p. 2, 2004.
- [5] Y. Hoskote, S. Vangal, A. Singh, N. Borkar, and S. Borkar, "A 5-ghz mesh interconnect for a teraflops processor," *IEEE micro*, vol. 27, no. 5, pp. 51–61, 2007.
- [6] T. Moscibroda and O. Mutlu, "A case for bufferless routing in on-chip networks," in *Proceedings of the 36th annual international symposium* on Computer architecture, 2009, pp. 196–207.
- [7] C. Fallin, C. Craik, and O. Mutlu, "Chipper: A low-complexity bufferless deflection router," in 2011 IEEE 17th International Symposium on High Performance Computer Architecture. IEEE, 2011, pp. 144–155.
- [8] S. Mittal and S. Nag, "A survey of encoding techniques for reducing data-movement energy," *Journal of Systems Architecture*, vol. 97, pp. 373–396, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1383762118304521
- [9] M. Palesi, G. Ascia, F. Fazzino, and V. Catania, "Data encoding schemes in networks on chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, no. 5, pp. 774–786, 2011.
- [10] N. Jafarzadeh, M. Palesi, A. Khademzadeh, and A. Afzali-Kusha, "Data encoding techniques for reducing energy consumption in networkon-chip," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 3, pp. 675–685, 2013.
- [11] S. S. Dev, S. M. Krishna, S. S. Archana, R. G. Kunthara, K. Neethu, and R. K. James, "Dual stage encoding technique to minimize cross coupling across noc links," in 2021 25th International Symposium on VLSI Design and Test (VDAT), 2021, pp. 1–6.
- [12] M. Ghoneima, Y. Ismail, M. Khellah, J. Tschanz, and V. De, "Formal derivation of optimal active shielding for low-power on-chip buses," *IEEE Transactions on Computer-Aided Design of Integrated Circuits* and Systems, vol. 25, no. 5, pp. 821–836, 2006.
- [13] R. Ayoub and A. Orailoglu, "A unified transformational approach for reductions in fault vulnerability, power, and crosstalk noise &; delay on processor buses." New York, NY, USA: Association for Computing Machinery, 2005. [Online]. Available: https://doi.org/10.1145/1120725.1121004
- [14] K. Banerjee and A. Mehrotra, "A power-optimal repeater insertion methodology for global interconnects in nanometer designs," *IEEE Transactions on Electron Devices*, vol. 49, no. 11, pp. 2001–2007, 2002.
- [15] M. R. Stan and W. P. Burleson, "Bus-invert coding for low-power i/o," *IEEE Transactions on very large scale integration (VLSI) systems*, vol. 3, no. 1, pp. 49–58, 1995.
- [16] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "A coding framework for low-power address and data busses," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 7, no. 2, pp. 212–221, 1999.
- [17] Y. Zhang, J. Lach, K. Skadron, and M. R. Stan, "Odd/even bus invert with two-phase transfer for buses with coupling," in *Proceedings of the* 2002 international symposium on Low power electronics and design, 2002, pp. 80–83.
- [18] C.-P. Fan and C.-H. Fang, "Efficient rc low-power bus encoding methods for crosstalk reduction," *Integration*, vol. 44, no. 1, pp. 75–86, 2011.
- [19] M. Dehyadegari, "Signature codes for energy-efficient data movement in on-chip networks," *Journal of Computing and Security*, vol. 7, no. 2, pp. 95–101, 2020.

- [20] J.-S. Shen, C.-H. Huang, and P.-A. Hsiung, "Learning-based adaptation to applications and environments in a reconfigurable network-on-chip," in 2010 Design, Automation & Test in Europe Conference & Exhibition (DATE 2010). IEEE, 2010, pp. 381–386.
- [21] G. Ascia, V. Catania, F. Fazzino, and M. Palesi, "An encoding scheme to reduce power consumption in networks-on-chip," in 2009 International Conference on Computer Engineering & Systems. IEEE, 2009, pp. 15–20.
- [22] C.-H. O. Chen, S. Park, T. Krishna, S. Subramanian, A. P. Chandrakasan, and L.-S. Peh, "Smart: A single-cycle reconfigurable noc for soc applications," in 2013 Design, Automation Test in Europe Conference Exhibition (DATE), 2013, pp. 338–343.
- [23] X. Chen and N. K. Jha, "Reducing wire and energy overheads of the smart noc using a setup request network," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 10, pp. 3013–3026, 2016.
- [24] P. Behnam and M. N. Bojnordi, "Stfl: Energy-efficient data movement with slow transition fast level signaling," in 2019 56th ACM/IEEE Design Automation Conference (DAC). IEEE, 2019, pp. 1–6.
- [25] B. Subramaniam, S. Muthusamy, and G. Gengavel, "Crosstalk minimization in network on chip (noc) links with dual binary weighted code codec," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 5, pp. 4603–4608, 2021.
- [26] J. J. Hunt, K.-P. Vo, and W. F. Tichy, "Delta algorithms: An empirical analysis," ACM Transactions on Software Engineering and Methodology (TOSEM), vol. 7, no. 2, pp. 192–214, 1998.
- [27] J. Zhan, M. Poremba, Y. Xu, and Y. Xie, "Noδ: Leveraging delta compression for end-to-end memory access in noc based multicores," in 2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC). IEEE, 2014, pp. 586–591.
- [28] D. Deb, M. Rohith, and J. Jose, "Flitzip: Effective packet compression for noc in multiprocessor system-on-chip," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 1, pp. 117–128, 2021.
- [29] S. Velayudham, S. Rajagopal, and S.-B. Ko, "An improved low-power coding for serial network-on-chip links," *Circuits, Systems, and Signal Processing*, vol. 39, no. 4, pp. 1896–1919, 2020.
- [30] "https://www.xilinx.com/products/boards-and-kits/zcu106.html."