An Efficient Low Power LDPC Method for Network-on-Chip

Ms. Keerthana Padmanabhan[1]
PG Scholar, Applied Electronics
Dept .Electronics And Communication Engineering
R.M.K Engineering College
Chennai, India keerthana071093@gmail.com

Mrs. V. Sumitra M.E (Ph.D)[2]
Research Scholar, Anna University
Asst.Professor
Dept .Electronics And Communication Engineering
R.M.K Engineering College
Chennai, India rush2sumi@gmail.com

Abstract—Network on- chip structure makes a fitting replacement for system on – chip designs incorporating large number of processing cores. In NoC, the links are the main element for power dissipation. The packets transmitted on the links are multiplexed and transmitted in serial form which dissipates more energy due to bit multiplexing and also consumes more power due to transition switching and coupling switching. A novel encoding method is proposed to reduce the transmission energy of serial communication. This encoding method uses LDPC encoding for reducing the transition activity in Noc link. Therefore the transition activity reduced bits are transmitted to reduce power consumption in the link of the NoC.

Keywords-switching activity;power dissipation; LDPC

I. INTRODUCTION AND MOTIVATION

A. INTRODUCTION:

To meet the growing computation-intensive applications and the needs of low-power and high-performance systems, the number of computing resources in a single-chip hasincreasedenormously since theVLSI technology can support an extensive integration of transistors. By adding many computing resources like CPU, DSP, specific IPs, etc to build a system in System-on-Chip, its interconnection between each other becomes a challenging issue.

The most widely used interconnect systems in such design are BUS based system Buses significantly impacted by the DSM effects. The network-on-chip will be helpful for future SoC due to several advantages and applications. These are regular geometry that is scalable, flexible QoS, higher bandwidth, reusable components, reliable and predictable electrical and physical properties and no longer global wires.NoC technology allows designers to optimize SoCs for often-competing three metrics: Multi-gigahertz frequencies, controllable latency for all NoC IP connections, and scalable bandwidth by individual trace. NoC interconnect IP offers a nnumber of advantages for SoC power, performance and area. There is no cheating Nature, but NoC interconnect exploits its leniencies. With effective design, for the entire SoC,NoC interconnect IP enables efficient optimizations between power, performance, and area. Traditionally, ICs have been designed with point-to-point connections, with the wire specified to each signal. For large designs, in particular, this has several limitations from a physical design point of view. The wires in a chip occupy much of the area, and in nanometer both performance and dynamic power dissipation of CMOS technology, interconnects dominate, as signal propagation in wires across the chip requires multiple clock cycles. As the technology scales down, the power consumed by the building blocks of NoC will be raised. Similarly, there are several issues related to NoC, such as the nature of the NoC link length, serial Vs Parallel links, bus Vs Packet-based switching and leakage currents. Of these, the power consumption in the links reaches high as compared to other specified building blocks. Therefore, we proposed a data encoding scheme which operates to reduce the switching activity in the links of NoC. The encoder and decoder designed in this method consume less power due to the reduction in transition activity of binary bits.

B.RELATED RESEARCH:

Traditionally, ICs have been designed with dedicated point-to-point connections, with one wire dedicated to each signal. For large designs, in particular, this has several limitations from a physical design point of view. The wires occupy much of the area of the chip, and in nanometerboth performance and dynamic power dissipation of CMOS technology,the interconnects dominate, since the signal propagation in wires across the chip requires multiple clock cycles.

As the technology node scales down to 65nm there won't be much increase in dynamic power dissipation. However the static or leakage power is same or exceeds the dynamic power beyond 65nm technology node. It is known fact that power is directly proportional to voltage and the frequency of the clock. Reduction of power consumption makes a device more reliable. Behind the development of CMOS technologies the need for devices that consume a minimum amount of power was a major force. As a result, CMOS devices are best known for low power consumption.

However, in order to minimize the power requirements of a board or a system, simply knowing that CMOS devices may use limited power than equivalent devices from other technologies does not help much. It is important to know not only how to calculate the power consumption, but also to understand how such factors of input voltage level, input rise time, power-dissipation capacitance, and output loading affect the power consumption of a device. The different methods have been proposed in the existing architecture to reduce the power consumed in the NoC architecture.

Reduction of the transmission energy of the serial communication by minimizing the number of transitions on the serial wire. It demonstrates the significant energy saving in a multimedia application, 3D graphics. It also applies the coding technique to a CMOS SoC implementation which integrates various processing units with packet switched on-chip networks[1].

A coding scheme is proposed for the data in the serial links to decrease the number of self-transitions to reduce the power during data transmission. It accounts the correlations in the data and hence is more effective for real-life applications. The system architecture with the encoding and decoding schemes have been implemented to explore the proposed algorithm applicable for any CMP or NoCarchitectures [2].

The data encoding schemes proposed, operating at flit level and on an end-to-end basis, which results in minimizing both the transition switching activity and the coupling switching activity on links of Noc. The encoding method undergoes even invert, odd invert or full invert based on the Type 1 and 2 transitions[3].

The encoding scheme comprises of the wormhole switching techniques and it works on an end-to-end basis. That is, flits are encoded by the network interface (NI) before they are injected into the network and it is decoded by the destination NI. This method makes the scheme transparent to the underlying network since the encoder and decoder logic is integrated in the NI and no modification of the routers architecture is needed[4].

It involves three schemes in order to reduce the dynamic power of the NoC data path through a decrease in the number of bit transitions. The transitions are of Type (I to IV) and correspondingly. Here, LDPC encoder is replaced with data encoding schemes in control to reduce the dynamic power consumption in Low Density Parity Check Techniques (LDPC)[5].

The idea in this paper is based on encoding the packets before they are injected into the network in such a way as to minimize both the switching activity and the coupling switching activity in the NoC's links which represent the main factors of power dissipation [8].

The encoder in LDPC is replaced with a data encoding schemes in order to reduce the power consumption in Low Density Parity Check Techniques. Experiments carried out on both synthetic and real time traffic scenarios shows the effectiveness of the scheme, which allow saving up to 27% of power dissipation[9].

The above papers discussed shows the reduction in transition activity of data bits in the link, but not accurate for the increase in transition activity after encoding the data bits. But the proposed encoding method reduces the transition activity.

The power dissipation and latency of the NoC link discussed throughout all the research papers are also reduced in case of proposed Low Density Parity Check encoding

C. PAPER ORGANIZATION:

This paper is organized as follows. Section II focuses on the contribution of the proposed method. The encoding and decoding technique of the proposed system are described in section III. Section IV discusses the proposed encoder and decoder architecture. Section V explains the flow diagram of a proposed coding method. Section VI deals with the experimental results and analysis of results. Finally, the conclusion and future scope of the proposed method is discussed in section VII.

II. CONTRIBUTION OF THE PAPER

The novel contributions of this paper to the state-of-art are the following:

- First, the proposed technique reduces the number of transitions significantly as well as guarantees a low-power dissipation during on chip communication through NoC links.
- Second, proposed method of data encoding/decoding is completely independent of data types.
- Third, proposed encoder and decoder parts are modeled in VHDL and Synthesized using Xilinx 14.5 ISE.
- Extensive experiments have been carried out over different types of binary data values under communication to show average power reduction for various binary values.
- Comparative discussions with related research results using similar existing techniques.

III. PROPOSED ENCODING AND DECODING METHOD

A. PROBLEM DEFINITION:

In an information theory, LDPC code is a linear error correcting code, which is a method of transmitting a message over a noisy transmission channel. LDPC codes are capacity-approaching codes, which means that practical construction exists that allow the noise threshold to be set very close to the theoretical maximum which is known as the Shannon limit for a symmetric memory less channel. The noise threshold defines an upper bound for the channel noise, until the probability of lost information is made as small as desired. Using iterative belief propagation techniques, Low Density Parity Check codes can be decoded in time linear to their block length.

LDPC codes has an increasinglyhigh applications which requires reliable and highly efficient information which is transfered over bandwidth or return channel-constrained links in the presence of the noise which is corrupted. Implementation of LDPC codes has lagged behind the other codes, which is called as turbo codes. The patent for Turbo Codes expired on August 29, 2013.

LDPC codes are also known as Gallager codes.

B. PROPOSED ENCODING METHOD:

Encode binary low-density parity-check code specified by parity-check matrix.The previous encoding algorithm suffers from extensive usage of back-substitution can be replaced by a few matrix-vector multiplications.

C. PROPOSED DECODING METHOD:

Decode binary low-density parity-check code specified by parity-check matrix. It implements the message-passing algorithm in the decoding step of low-density parity-check (LDPC) codes, which is a linear error control codes with sparse parity-check matrices and has long block lengths which can attain performance near the Shannon limit.

D. LDPC CODE CONSTRUCTION:

In order to achieve good performance, the LDPCcode comprisesof the following properties:

(a) ***Large code length*****:** The performance of this method improves as the code length increases, and the code length cannot be too small (at least 1K).

(b) ***Not too manysmall cycles*****:**The error-correcting performancewill seriously degradeif there are too many small cycles in the code bipartite graph.

(c) ***Irregular node degree distribution*****:** It has been demonstrated thatcarefully designed LDPC codes with irregular node degree distributions will remarkably outperform regular ones.

Code construction parameters:

Each block matrix has the size ofis $p \times p$, the size of parity check matrix is $(m \cdot p) \times (n \cdot p)$, and $g = \gamma \cdot p$. The row and column weight distributions are $\{w_1^{(r)}, w_2^{(r)}, \cdots, w_m^{(r)}\}$ and $\{w_1^{(c)}, w_2^{(c)}, \cdots, w_n^{(c)}\}$, where$w_i^{(r)}$ and $w_j^{(c)}$ represent the weight of i-th block rows and j-th block columns, respectively.

Output:

$(m \cdot p) \times (n \cdot p)$isparity check matrix **H** with the structure shown in Fig. 1 , in which each $p \times p$ block matrix $\mathbf{H}_{i,j}$ is either a zero matrix or a right cyclic shift of an identity matrix.

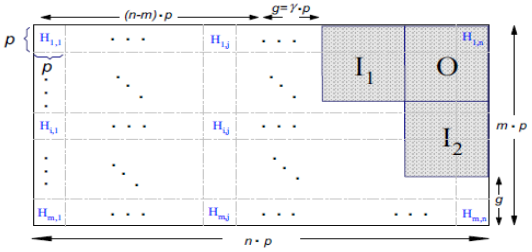

Fig 1. The parity check matrix H

Procedure:

1. Generate an $(m \cdot p) \times (n \cdot p)$ matrix with the structure as shown in Fig. 1, where $\mathbf{I}_1$ and $\mathbf{I}_2$ are identity matrices with roughly the same size and **O** is a zero matrix. Theun-shaded region blocks are initially set as NULL blocks.

2. According to the column weight distribution, generate a set $\{a_1, a_2, \cdots, a_n\}$, in which $a_j = w_j^{(c)}$ if $\mathbf{1} \le j \le n - m + \gamma$, and $a_j = w_j^{(c)} - \mathbf{1}$ if $n - m + \gamma + \mathbf{1} \le j \le n$.

3.According to the row weight distribution, generate a set $\{b_1, b_2, \cdots, b_m\}$, in which $b_i = w_i^{(r)} - \mathbf{1}$ if $\mathbf{1} \le i \le m - \gamma$, and $b_i = w_i^{(r)}$ if $m - \gamma + \mathbf{1} \le i \le m$.

4. Initialize the cycle degree constraint $d = d_{init}$.

5. For j = 1 to n, replace aj NULL blocks on the j-th block column with aj right cyclic shifted identity matrices:

- Randomly pick $i \in \{1, 2, \cdots, m\}$ such that bi >0 cyclic shift of a p × p identity matrix with randomly generated shift value.
- Let f(**H**) denote the minimum cycle degree in the bipartite graph which corresponds to the current matrix **H**. If f(**H**) <doris the bipartite graph that contains 4-cycles, then reject the replacement and go back to (a). If f(**H**) remains less than d after a certain number of iterations, decrease d by one before go back to (a).
- bi = bi – 1
- Terminate and restart the procedure if d <dmin, Where,

dmin is the minimum allowable cycle degree.

6. Replace the remaining NULL blocks with zero matrices and output the matrix **H** and **H**i,j is a NULL block. Replace **H**i,j.

IV. PROPOSED ENCODER AND DECODER ARCHITECTURE

It Encodes the binary low-density parity-check code which is specified by parity-check matrix.The block given below Fig 2.supports encoding of low-density parity-check (LDPC) codes, which are linear error control codes with the sparse parity-check matrices and long block lengths which can attain performance near the Shannon limit. Both the input and the output of the encoder are discrete-time signals. The ratio of the output sample to the input sample time is k/n. The input must be a real k×1 column vector signal. The output signal inherits the data type from the input signal, and the input must be binary-valued (0 or 1).

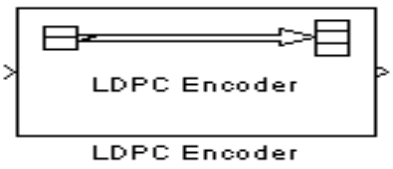

Fig 2. LDPC Encoder

The straightforward encoding process uses thegenerator matrix which results in prohibitive VLSI implementation complexity. Richardson and Urbanke demonstrated that, if the parity check matrix is approximate upper triangular, then the encoding complexity can be significantly reduced. However, the encoding algorithm suffers from extensive usage of back-substitution operations which in turn increases the encoding latency and makethe hardware implementation problematic. The authors shows that all the back-substitution operations will be replaced by a few matrix-vector multiplications where $\mathbf{I}_1$and $\mathbf{I}_2$are identity matrices and **O** is a zero matrix which is as shown in Fig 3.

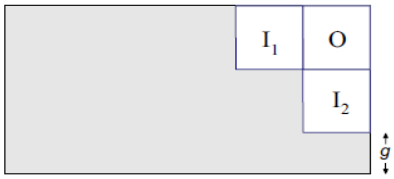

Fig 3. The encoder aware parity check matrix structure.

It Decodes the binary low-density parity-check code specified by parity-check matrix.The block given below Fig 4.accepts a real-valued, *n*×1 column vector input signal. Each element is the log-likelihood ratio for a received bit. The first *k* elements which are corresponding to the information part of a code word. Both the input and the output of the decoder are discrete-time signals. The ratio of the output sample to the input sample time is *n*/*k* if only the information part is decoded, and 1 if the entire code word is decoded.

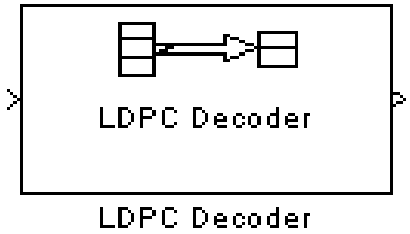

Fig 4. LDPC Decoder

The LDPC Decoder block is designed:

- In order to Decode generic binary LDPC codes where no patterns in the parity-check matrix are assumed.
- In order to Execute a number of iterations you specify or run until all parity-checks are satisfied.
- To get Output hard decisions or soft decisions (log-likelihood ratios) for decoded bits.

Message Passing Decoding Algorithm

LDPC codes are defined by an M × N binary matrix which is called the parity check matrix H. The number of columns, N, defines the code length. The number of rows in H, M, which defines the constraints of number of parity check for the code. The information length K is K = N − M for full-rank matrices, otherwise K = N − rank. Column weight Wc is the number of ones per column and row weight Wr is the number of ones per row.

LDPC codes which can also be described by a bipartite graph or Tanner graph. The parity check matrix and the corresponding Tanner graph of a (Wc = 2, Wr = 4) (N = 12, K = 7) LDPC code are shown in Fig 5. The rank of the matrix is 5, therefore the information length K is 7. In the graph, there are two sets of nodes: they are check nodes and variable nodes. Each and every column of the parity check matrix corresponds to a variable node in the graph which is denoted by V . Each and every row of the parity check matrix corresponds to a check node in the graph which is represented by C. There is an edge between a variable node Vjand check node Ci, if the position (i, j) in the parity check matrix is 1, or H(i, j) = 1. For example,in the Tanner graph the first row of the matrix corresponds to C1 which is connected to V3, V5, V8 and V10 variable nodes. A neighbor variable node is a node were a variable node is connected to a check node. Similarly, a check node that is connected to a variable node is called the neighbor check node.

Total number of edges (connections) representation of a 12 column (N), 6 row (M), column weight 2 (Wc), row weight 4 (Wr), LDPC code with information length 7 (K). between variable node and check nodes is N × Wc or M × Wr. For clearer explanations, in this work, we examine cases where H is regular and thus Wr and Wc are constants. For VLSI implementation examples, we use a (6,32)-regular (2048,1723) RS-LDPC code which has been adopted for the forward error correction in the IEEE 802.3an 10GBASE-T standard [4].

The code has a Wr = 32 and Wc = 6 of H matrix 384×2048. There are M = 384 check nodes and N = 2048 variable nodes in its graph representation. The most widely used method for practical decoding isthe iterative message-passing algorithm. After receiving the corrupted information from an AWGN channel (λ),by processing it the algorithm begins and then iteratively corrects the received data.

First, inputs of all check node are initialized to “0” and then a check node update step (i.e. row processing) is done in order to produce α messages. Second, the variable node receives the new α messages, and then the variable node update step (i.e. column processing) is done in order to produce β messages. This process will be repeated for another iteration by passing the previous iteration’s β messages to the check nodes. Finally the algorithm terminates when it reaches a maximum number of decoding iterations (Imax).

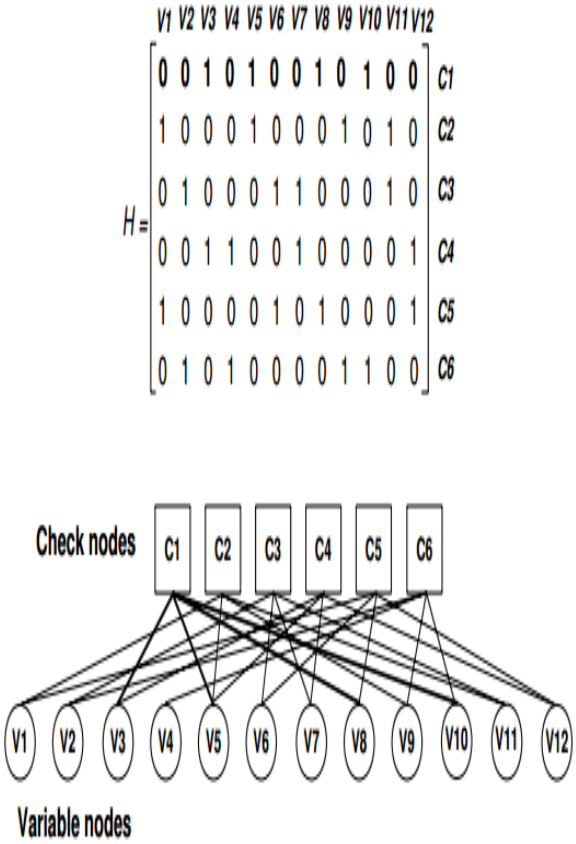

Fig 5. Parity check matrix (upper) and Tanner graph (lower)

V FLOW DIAGRAM OF THE PROPOSED ENCODINGMETHOD

The encoder and decoder employed in the network interface, convert the data bits in the link with reduced number of transitions and in turn retrieve the original data bits from the link respectively. The process starts by counting the number of transitions in the original data bits followed by the number of transitions in encoded bits is calculated.

Thedecoder compares and retrieves the original binary data values.

The Flow of this method is shown in Fig. 6.

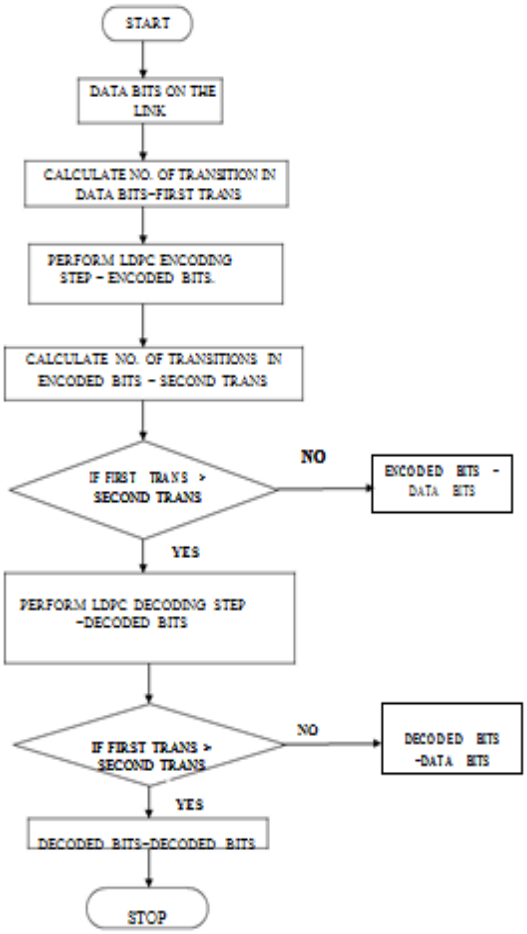

Fig 6. Flow diagram of the proposed coding method

VI EXPERIMENTAL RESULTS

A. COMPARISON OF PROPOSED TECHNIQUE WITH DATA CORRELATION ENCODING:

The number of transitions in the data correlation method is more than that of the proposed encoding. In data correlation encoding, the bit transitions are checked and inverter to reduce the number of transitions. In proposed model, the number of transitions in input data is calculated and encoded by LDPC encoding technique and the number of transitions is reduced. Therefore the proposed LDPC coding model is compared with the data correlation aware serial encoding for low switching power on-chip communication which is shown below with the help of comparison table.

TABLE I Comparison of proposed LDPC coding with data correlation

Input Data stream	# Trans w/o Encoding	Data correlation	# Trans of data correlation	Proposed Encoding	# Trans with encoding
11111111	0	11111111	0	11111111	0
01010101	7	00111001	3	11100001	2
01111011	3	01111011	3	11100001	2
01110111	3	01111011	3	10011110	3
TOTAL	13		09		07

B. EXPERIMENTAL SET UP:

The proposed method was modeled in VHDL and Synthesized using Xilinx 14.5 ISE. The simulation result of the proposed encoding method shows the correct functionality of the LDPC technique. The number of transitions is reduced in the proposed method is higher than the previous methods which results in the reduction of power consumption.

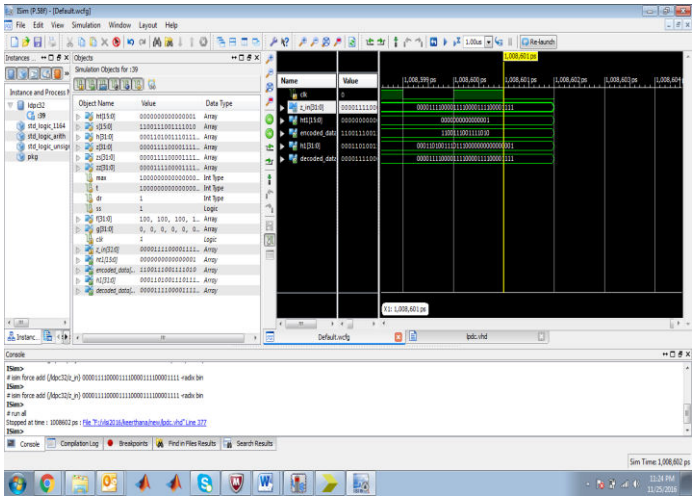

Fig 7. Simulation result of proposed LDPC coding technique

The proposed method obtains the overall power about 0.041mW which is shown below.

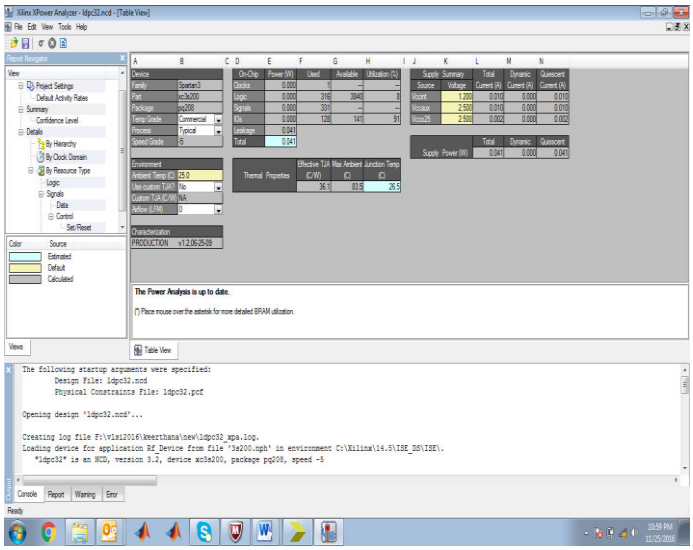

Fig 8. Power Analysis of proposed LDPC coding technique

C. DISCUSSIONS:

It is evident from experimental results that the proposed scheme obtains on an average of 15.39% reduction in number of transitions in case of binary bits, 51.89% reduction in power consumption.

In comparison to data correlation aware serial encoding, the proposed system is much more effective where the binary data bits are highly encoded, decoded and power reduction is achieved. Unlike data correlation aware serial encoding, the proposed method always shows improvement in case of switching activity and thereby power reduction.

VII CONCLUSION AND FUTURE WORK

The transition activity of the data bits is minimized to reduce the power consumption in the link of NoC by using a novel LDPC (Low Density Parity Check) data encoding method. This results in the reduction of power consumption in the link of NoC.

The future scope of this project makes use of end-to-end encoding scheme with the overall implementation and further reduction of power and area.

REFERENCES

[1] Lee K. and Yoo H. J. (2004), "SILENT: Serialized Low Energy Transmission Coding for On-chip Interconnection Networks", Proceedings of the IEEE/ACM International Conference on Computer-aided design, pp. 448-451.

[2] Ghosh S and Mohanty P. (2014), "Data Correlation Aware Serial Encoding for Low Switching Power On-Chip Communication", IEEE Computer Society Annual Symposium on VLSI, pp. 124-129.

[3] Jafarzadeh N and Palesi M. (2014), "Data Encoding Techniques for reducing Energy Consumption in Network-on-Chip", IEEE Transactions on Very Large Scale Integration Systems, Vol. 22, No. 3.

[4] Giuseppe Ascia and Maurizio Palesi, (2011), "Data Encoding Schemes in Networks on Chip", IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, Vol. 30, No. 5

[5] Suresh Dannan (2014) "Data Driven Data Encoding for Low Power LDPC Applications", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.8, No.10, pp.375-388.

[6] M. M. Jadhav, C. N. Kharkar, and A. M. Sapkal, (2013), "FPGA Implementation of linear LDPC encoder", vol. 2, no. 11.

[7] T. S. Bharath Kumar and M. Z. Kurian (2014), "Architecture for Low Density Parity Check Encoder", (IJAIEM), vol. 3, no. 3, pp. 414-417.

[8] S. R. Mullainathan and S. Ramkumar, (2014), "Switching Reduction Through Data Encoding Techniques in NoC", IEEE, vol. 23, no.76, pp. 38-41.

[9] V HanumanthGoud, (2015) "Low Power Data Encoding Schemes in LDPC Applications", International Journal of Science, Engineering and Technology Research (IJSETR), Volume 4, Issue 4.