Optimized Recurrent based Training Accelerator for Network-On-Chip Communication System

Sumana Achar, Jayadevappa D

***Abstract*—The chip application has become the trending element in all digital applications because of its easy and flexible use. However, chip-based communication's chief demerits are power consumption and delay. The huge data broadcasting often raises these two problems in Network-on-Chip (NoC) architecture. So, the present research work has aimed to create a novel Strawberry-based Recurrent neural (SbRN) framework for minimizing the NoC buffer length and reducing the latency and power consumption. Here, the fitness of Strawberry was utilized to sense the large data size; once the large data size was identified, the compression process was started to compress the data. Moreover, the NoC architecture was designed with an optimized buffer with strawberry fitness. Once the compressed data was present in the network medium, it allowed broadcasting to the other end. If the data is not compressed or not in the minimum compression range, the compression process was again to optimize the data buffer. Finally, the proposed architecture was validated with other conventional schemes and has gained the best outcome by reducing 5% of device utilization and 0.5% of power consumption than other methods.**

***Index Terms*—Router, communication channel, network-on-chip, neural layer, optimized buffer, power consumption**

I. Introduction

RECURRENT-NEURAL-NETWORK(RNN)[1] are subset of deep neural networks. It is used for some applications, including inputs, and time series such as control of the dynamic system and speech recognition[1]. Moreover, RNN is a fully connected network, and it will take inputs as a one-dimensional vector and output as a product vector[2]. RNN is also connected to the high dimensional input and is useful for regression tasks and sequential classification. Thus, the RNN are connected among nodes, forming the directed graph next to the temporal sequence [4], [5]. The RNN is used for processing inputs of variable sequential length and to execute practical tasks such as handwriting recognition, unsegmented and speech recognition; it exhibits temporal dynamic behavior[5].

Network on Chip (NoC) is the Integrated Circuit (IC) network-based communication system among modules. The key components of NoC communication system are illustrated in Fig.1[6]. Modular networks design the various functions of computer subsystems, and the NoC depends on router-based packets among switching network[7]

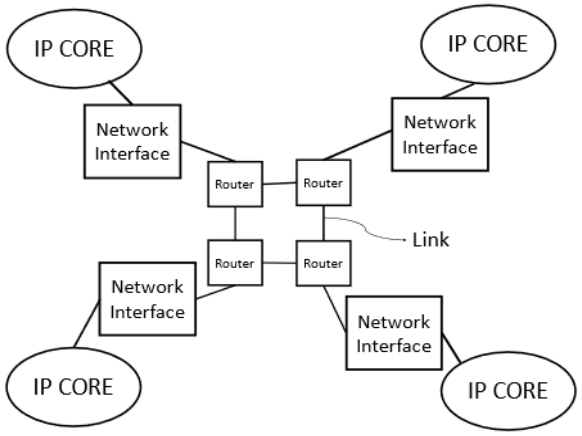


Fig.1. Key components of NoC

System on Chip (SoC) design contains a group of heterogeneous mechanisms with varying block sizes[8]. It has various components like storage elements, processor core, analogue peripheral devices and embedded hardware[9]. The main aim of the NoC is to provide scalability and performance in applications such as multimedia and communication[10].

Furthermore, NoC enables communication, distributed computation, and synchronization between system components[11]. The architectural parameters and quality of communication play a role in the determination of executing the NoC. In some studies, it is found by deducting the memory-accompanied materials of the network. The most achieved technique is the neural network. Also, in the software study, the neural synthesizer is included for transforming Suspended particulate matter (SPM) into the approximate Hem for the precise drop constraint.

But the main critical task in NoC is serviceability, packaging, power consumption and programmability constraints[12]. Also, the main important role of NoC is to improve scalability and power efficiency[13]. Many existing techniques are introduced to enhance scalability, power consumption, and cost[14]. Some other methods are neural neuro stim framework[15], memory architecture[16], asynchronous spiking neural system[17], recurrent neural system accelerator [18]and so on but still have some problems in NoC performance[19]. An optimized recurrent accelerator technique is introduced in this paper to enhance the NoC communication system[20]. The main aim of the

Manuscript received April 05, 2023; revised October 02, 2023.

Sumana Achar is the Research Scholar at Department of Electronics and Instrumentation Engineering, JSS Academy of Technical Education Bengaluru, VTU, Belagavi, India Phone: 8971605150; e-mail: sumanavtuphd@gmail.com

Jayadevappa D. is a Professor at Department of Electronics and Instrumentation Engineering, JSS Academy of Technical Education, Bengaluru, VTU Belagavi, India, Phone: 9986134424; e-mail: djayadevappa@jssateb.ac.in

developed framework is to improve scalability and power usage[21].

Attention mechanism and adaptive attention fusion module work contributes to the high performance of any communication systems[22]. Time delays can affect the property of the neural network[23]. Caching strategy on contents at the network router will increase the efficiency of the network and hence the users are served better [24].

II. Related Work

Many researchers have claimed their views on Recurrent based training accelerator. Shanshi huang *et al.* [15] have proposed a deep neural neurosim framework of an integrated network for accelerating benchmark computer memory. Thus, the proposed framework offers an automatic algorithm for hardware mapping, estimated chip-level places, throughput, accuracy, and energy efficiency. An optimized open-source technique is introduced to enhance the performance of reliability, but neural network training has been more difficult. Memory architecture is commonly used for speech applications and data time series. It requires large complexity and weight storage. Deepak *et al.* [16]have developed an energy-efficient memory architecture-based recurrent neural system accelerator to achieve less error rate in degradation. Moreover, developed techniques attain low error rates in speech recognition but have taken more time to execute tasks.

Chang gao *et al.* [20] have introduced a lightweight gated recurrent component-based recurrent neural system accelerator to enhance less latency and low power portable. Also, it is called EdgeDRNN; it will adopt a neural network and inspire the data using a network algorithm. The main aim of the technique is to improve the latency and accuracy of neural network systems, but it has the issue of exploding. To improve the life quality of leg prostheses and reduce energy, Rachel *et al.*[25] have proposed learning complex regulators to assist dynamic robotics devices. For controlling power, a recurrent system of recurrent system is introduced and runs under a hardware accelerator; it develops temporal sparsity. Thus, the technique executes more complex tasks but has been obtaining errors during computation time because of large data. Edith Beigne *et al.* have developed a globally asynchronous spiking neural system for scalability. It will allow the specified neutrons to be inhibitory, and the pruning distance will cut communication and memory. Thus, the developed technique limits the latency of the hop and reduces traffic. Moreover, input channels are parallelized and achieve good performance in latency, but it has gradient vanishing.

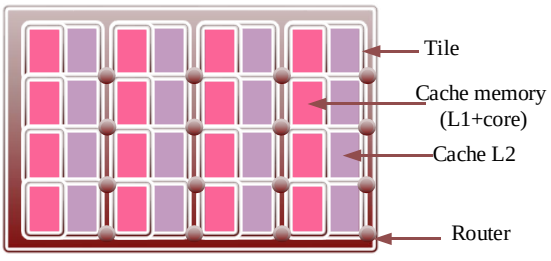


Fig.2. Basic Noc Architecture

The key steps of this planned design are detailed as follows

- Initially, an optimized deep neural model named a novel SbRN model is designed with the required weights and parameters
- Moreover, NoC is structured with required processing elements, registers, and memory
- Consequently, the designed SbRN is updated in the NoC random-access memory (RAM) to optimize the buffer size
- Hereafter, the key metrics are calculated and compared with other models in terms of latency, power, Flipflop, LUT, throughput, etc

III. System Model and Problem Statement

All the memory must pass through the huge-energy-cost off-chip Dynamic Random Access Memory (DRAM) in evidence with an impact on both throughput and energy efficiency. The data movement is exclusively deduced using data reuse through the multi-level memory evaluation.

Moreover, the reprocessed local data has maximized power, less throughput and low-level memory. Hence, the usual NoC architecture is exposed in Fig.2. The chip-to-chip interconnection highly influences an interconnection's latency, reliability, cost, and throughput. All these issues have motivated this research towards the network-on-chip communication area.

IV. Proposed Methodology

Training accelerator is the trending topic in the digital industry to advance digital chip-based applications. Hence, to enrich the resources of NoC, the current research work has aimed to focus on a novel optimized deep learning-based accelerator was introduced, which is named as Strawberry based Recurrent Neural (SbRN) model**.** Moreover, the efficiency of the proposed design is validated by calculating the key metrics. The proposed design is detailed in Fig.3.

In the strawberry algorithm, the threshold of minimum data size was fixed. Once the NoC was designed in the MATLAB platform, the data-sharing process began. Before transferring data, the fitness of the Strawberry is utilized to check the data size. If the data has not met the minimum optimized data, then the compression procedure was activated to compress the data.

A. Design of SbRN Model

The presented work is a combination of dual models that are recurrent models and strawberry algorithms[26]. This hybrid model in NoC model is used to improve training rapidity and minimize the data broadcasting measure and power consumption.

To develop the SbRN model, initially, three points are taken in an unstructured format. Consequently, the process is initiated by separating the points into three overlapped blocks, which are processed independently.

$$t_i = \sigma(p_i A_{pt} + G_{i-1} A_{ht} + b_{i-1} A_{bt} + c_t) \quad (1)$$

$$f_i = \sigma(p_i A_{pf} + G_{i-1} A_{hf} + b_{i-1} A_{bf} + c_f) \quad (2)$$

$$T^*_{i,j}, G_{i,j} = f(r_{i,j}, G_{i-,j1}) \quad (3)$$

Where updated output is represented by $T^*_{i,j}, G_{i,j}$, and the current input, which is determined as $r_{i,j}, G_{i-,j1}$. In this Strawberry based recurrent model, the dense layer of the recurrent neural layer is enhanced with strawberry fitness.

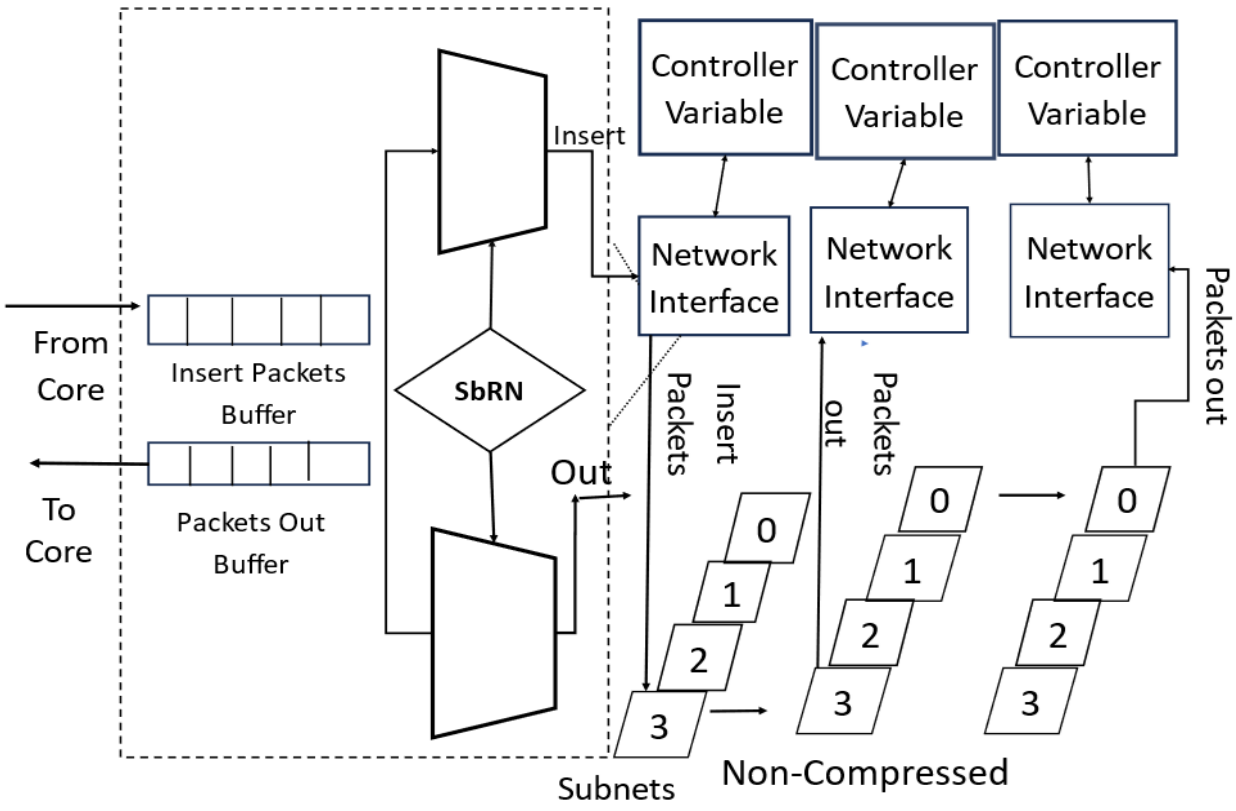


Fig.3 Proposed NoC Architecture.

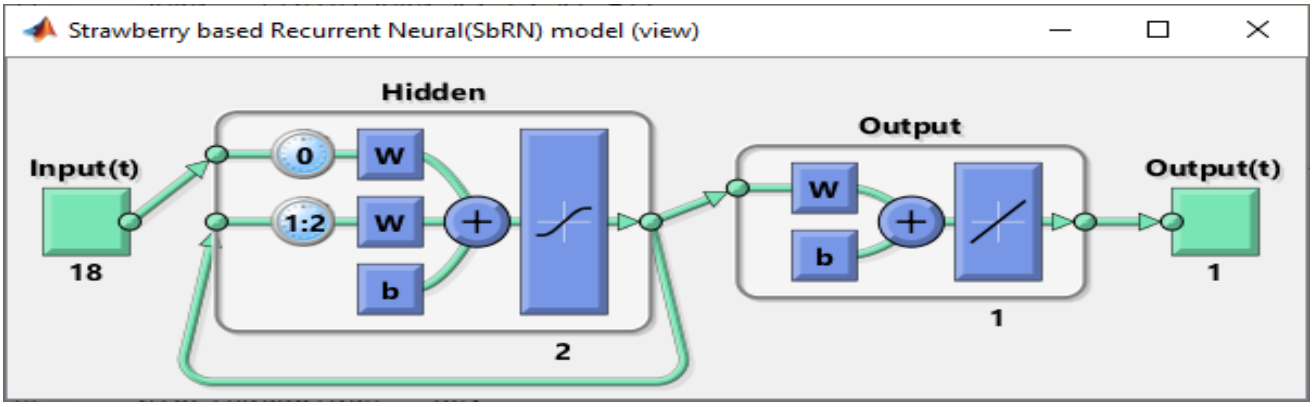


Fig.4 Layers of Proposed SbRN

The purpose of this model is to reduce energy consumption and enhance the speed of application. Moreover, the presented SbRN layers are depicted in Fig. 4. Here, b is the bias and W determines the neuron weights.

The NoC[27] has designed three models: links, routers and network adapters or network interfaces. Here, some wires and router connections link the communications. In addition, the links might contain more physical and logical channels, which are composed of a set of wires. On the other hand, the router was designed with a switching function, virtual channel (VC), flow control and tiny buffer cost.

Compressing the data is an important process to maximize the acceleration of the NoC performance. So, if the data compression process is processed, then automatically decompression process also takes place. The internal module of the projected NoC model is depicted in fig.5. Here, VC (b) is the highly compressed data, so it gets access to transfer. In addition, the reason for this compression module is to optimize the buffer size. Hence, the validation of buffer size is processed by (4)

$$\min f(k), B_k \le B \le B_s \tag{4}$$

Here, the parameter $\min f(k)$ represents the optimal size of data and B determines the neuron, S denotes a total number of data and k is the data in the specific neuron.

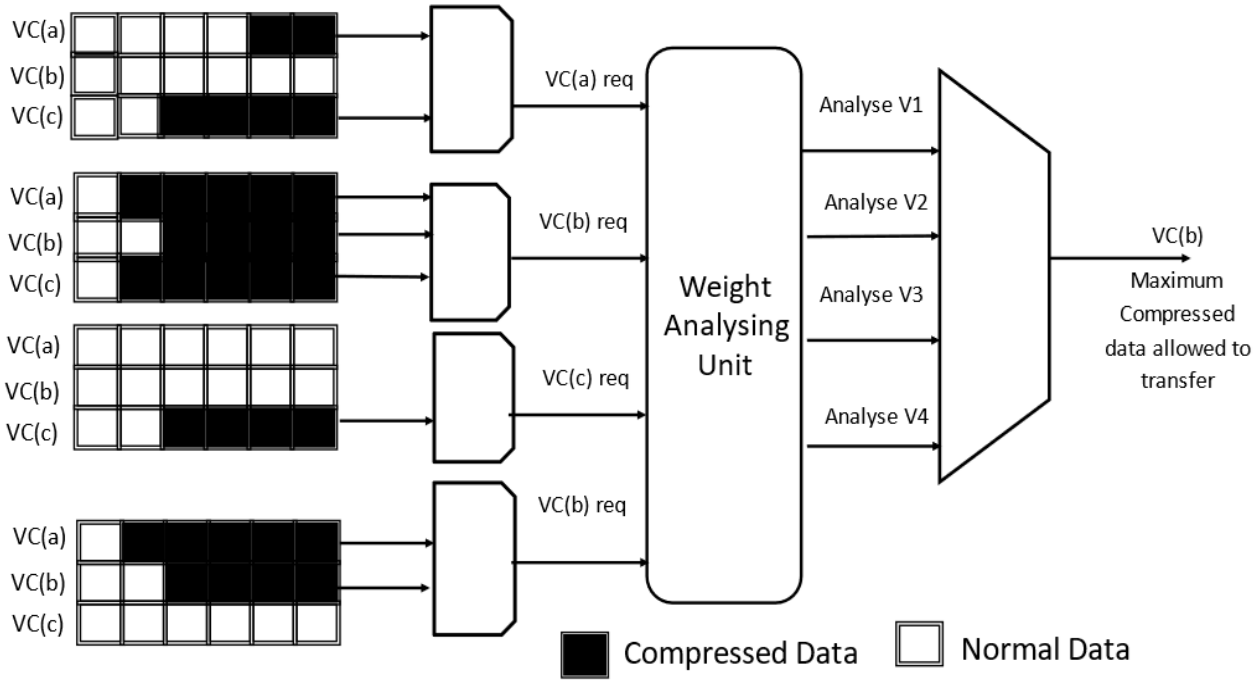


Fig.5 Internal Architecture of Proposed NOC

There are 18 layers in the designed network, which is defined as β.

$$B(i) = [B_p(i)B_k(i)] \tag{5}$$

Here, $B_p(i)$ represents the activation parameter of the data compression and decompression process, and $B_k(i)$ denotes each input data. Moreover, the input data of each process is detailed as $B(i)$. The input data was measured by using (5)

$$B_k = \begin{cases} \dfrac{1}{B + f(B_p.B_f)} & f(B_f(i)) = 0 \\ B + f(B_p.B_f) & f((B_f(i)) = 1 \end{cases}, else(hc) \tag{6}$$

Here, hc is the high data cache, optimized buffer and data are determined as B_f, also, the optimized buffer is represented as 0, and optimized data is denoted as 1, which is done using (6) The working flow is exposed in Algorihm.1.

Usually, the compression of the bit plane functions by compressing the non-zero membrane. Here, in the proposed approach, initially, the bit was compressed using a conventional procedure then that compressed data was trained to the proposed model.

The bit compression function is processed based on the following steps; the reason for compression in deep neural networks is to enhance the performance of hardware implementation.

Here the compressed recurrent model is denoted, and $\vec{v}_\theta$ finally, the compression rate is estimated by comparing the performance of the original deep neural network with the compressed network.

The quality and performance of compression procedure are estimated in terms of $Q(\vec{v}_\theta)$ and $M(\vec{v}_\theta, v^*)$ is for fidelity validation.

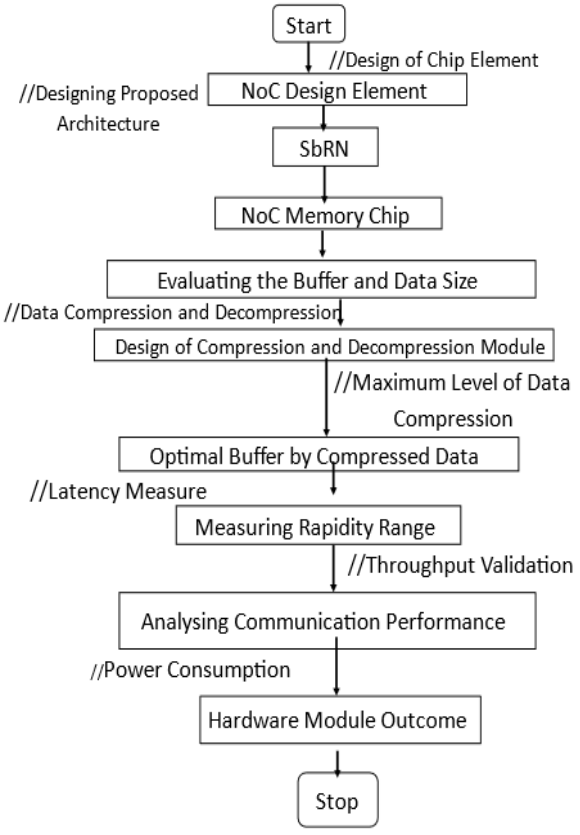
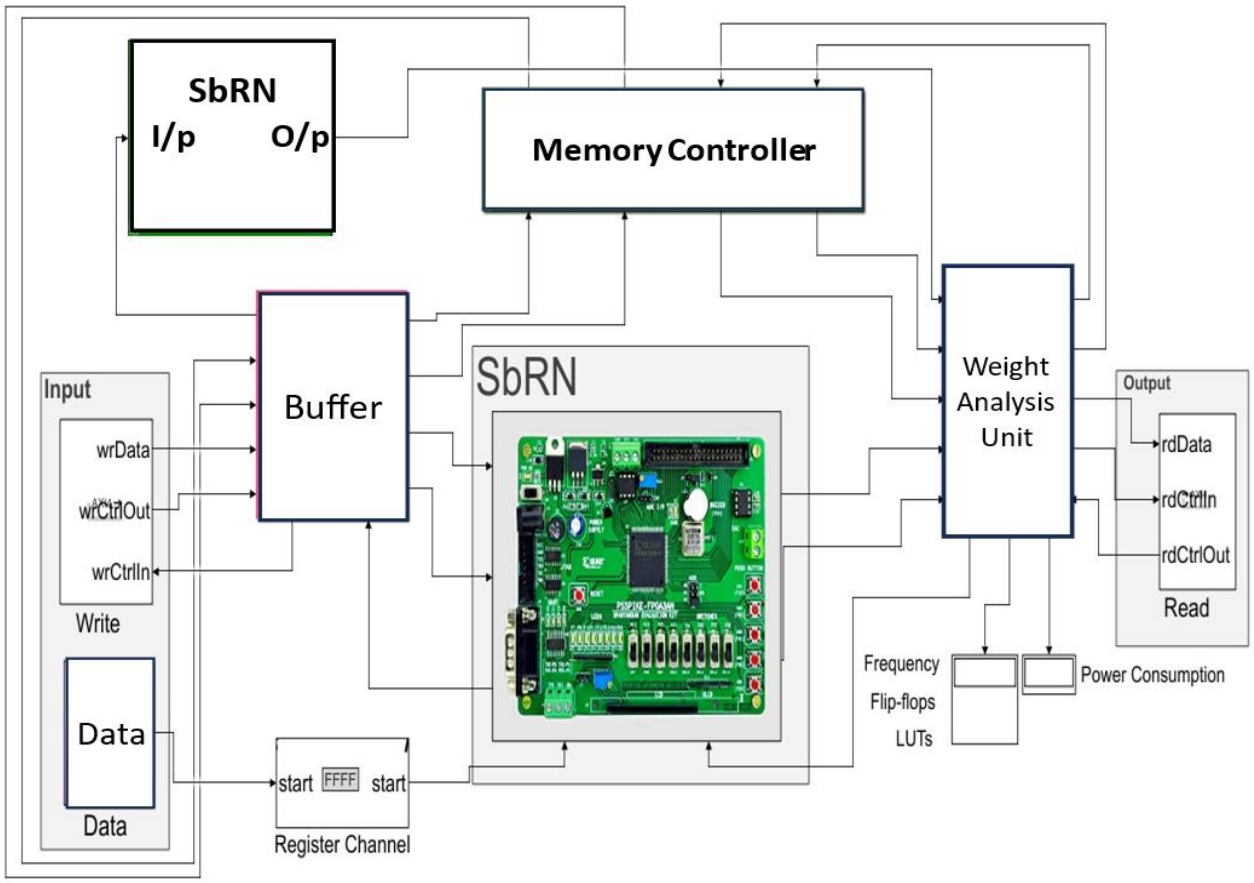
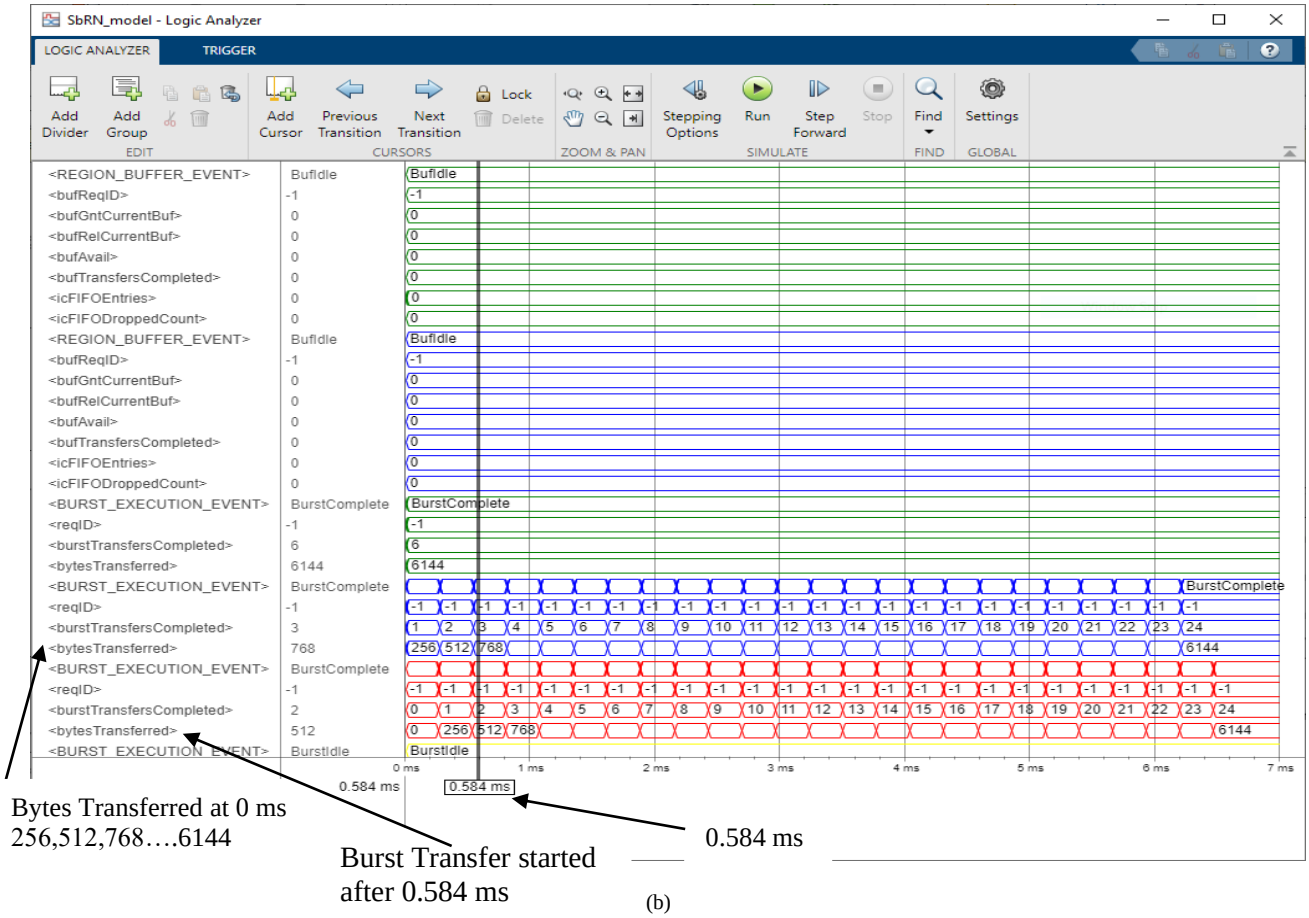


Fig.6. Proposed SbRN Process Flow



(a)



(b)

Fig.7 SbRN Architecture (a) Board Diagram (b) Logic Analyzer

Algorithm 1: Pseudocode for application-based SbRN

Activation weight initialization $T^*(i = 1, 2, ...n)$
Begin with i, j *and* K *(number of neurons)*
Estimate the fitness of all NoC elements

α *is to estimate the NoC feature weights of each layer* G
If G *is the map of each neuron layer, then* $\beta = 14$,
Here, 14 is the maximum pixel of feature maps
Consequently, $\min f(k)$ *is the threshold range of trained data.*

if $(B_k \leq \min f(k))$
{
Buffer is in Compressed State
}
else
Compress Data
end if
while $(T^* < B_k)$
int B_f *;*
$\sigma = B_k \rightarrow B_f(0)$ *// optimized data*
$B_k \rightarrow B_f(1)$ *// optimized buffer*
// B_f *Buffer and data size estimator (after compression process)*
end while
Update $T^* = T^* + 1$
return B_f

Moreover, $N(\vec{v}_\theta, v^*)$ determines the ratio of bit compression, defined as the ratio of parameters count divided by the parameters of the original network. Based on the application type, the compression rate is maximized with the support of strawberry fitness.

The developed compression strategy is utilized to minimize the energy usage by an external RAM for interrupting communication. The proposed SbRN function flow is depicted in Fig.6. Here, the small chip is sufficient in the hardware platform when the data is compressed in a neural layer. To perform the hardware process module, the compressor and decompress frame are integrated with a hardware accelerator that has achieved better energy efficiency for specific neural weights. In addition, if the number of compression units increases, the cost of I/O gets reduced.

V. Results and Discussion

The present research has implemented MATLAB R2021a/ Simulink running in a windows 10 environment. The successive measure of the designed NoC router is analyzed with parameters like throughput, latency, frequency, LUT, power consumption and flipflop.

The chip-based memory has attracted all digital applications with its advancement; still, in some case, this chip-based technology has consumed more power because of large datasets. To overcome this problem, the current research has projected a novel SbRN to maximize the running speed and minimize the running time. Hence, to attempt the rapidity of the process, the dataset should be compressed in an appropriate format.

Here, the data was compressed in recurrent layers with the help of strawberry fitness. Once the data is compressed, it makes the data transaction an easy case. The designed model in the MATLAB platform is depicted in Fig. 7 a) and the log analyzer is detailed in Fig.7 b).

A. Case Study

To explain the designed approach, a network data transmission application is taken. Here, the communications among dual users are elaborately explained by using the proposed approach. Let us imagine I and J is the end user broadcasts the data in the wireless environment. Moreover, for example, image data is taken as the input. Initially, MSB of the images is framed in row and column form. Simultaneously, to compress the data, the usual compression mechanism was processed.

To detail the bit compression and decompression procedure, an example was taken in the form of an image MSB shown in Fig.8. Compression will help reduce the buffer size of Router and hence Optimized buffer can provide the best resource utilization in Network.

178	1	78	32	200	27
119	225	160	32	67	143
225	160	178	11	133	143

Fig..8 Image Pixel Value

In the next steps, the arithmetic values are converted into binary bits, explained in Fig.9. From the binary values, the MSBs are extracted, which is elaborated in Fig.10.

10110010	00000001	01001110	00100000	11001000	00011011
01110111	11100001	10100000	00100000	01000011	10001111
11100001	10100000	10110010	00001011	10000101	10001111

Fig.9. Binary Representation of Entered Input Signal

1	0	0	0	1	0
0	1	1	0	0	1
1	1	1	0	1	1

Fig.10. MSB Extraction

For 8-bit compression, 8-bit is needed, but in the previous two stages, only 6 bits. So, the padding function is performed to attain 8 bits. Thus, padding the bit is shown in Fig.11.

1	0	0	0	1	0	0	1
1	1	0	0	1	1	1	1
0	1	1	0	0	0	0	0

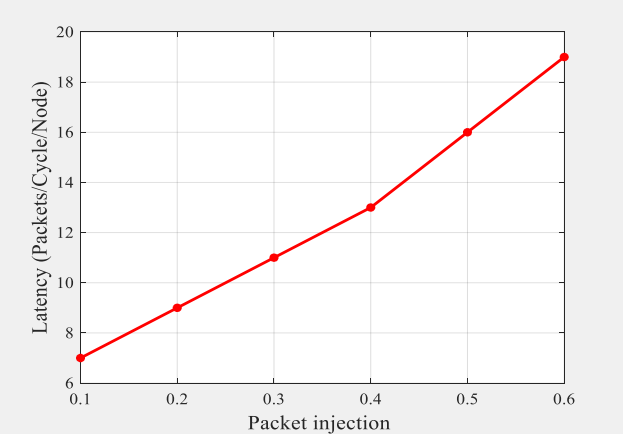
Fig.11. Bit Padding

Consequently, the decimal value of the previous binary data is shown in Fig.12, which is the attained binary data from the usual compression.

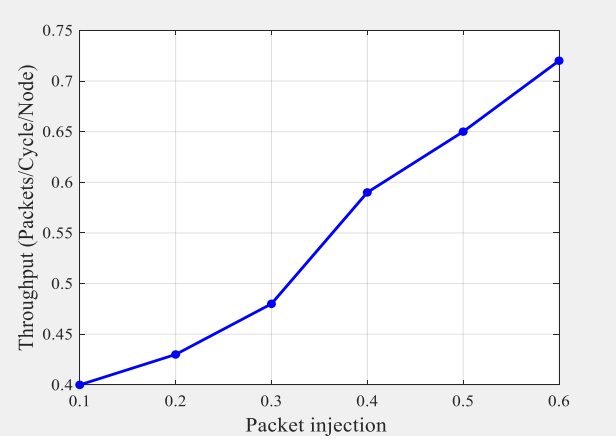
137
158
192

Fig.12. Compressed Data

Once, the bits or feature maps are compressed, then that compressed data is trained to the chip applications, and the device or chip performance is analyzed and compared with other models. Hence, internal performance parameters like latency and throughput were considered to measure the optimal score of the NoC router. Also, its performance was validated with different packet injection rates, as defined in Fig.13.



(a)



(b)

Fig.13.Performance (a) Latency (b) Throughput

The recorded maximum latency range is 19 cycles for the 0.6 packet injection rate, and the reported elevated throughput level is 0.75 packets/cycle, sufficient for the NOC to enrich the performance.

B. Performance Metrics

To Validate the robustness of the model, a few existing approaches obtained, Learning Enabled High Performance (LEHP)[28], and NeuronLink[29], were taken and compared with dissimilar metrics like throughput, latency, power consumption, LUT, flipflop and frequency.

Throughput Ratio

In all digital appliances, the rate of communication broadcasting is based on the throughput ratio. Also, throughput is validated as packets/cycles/nodes. In addition, the throughput ratio is evaluated with dissimilar packet injection rates.

The proposed model has attained the finest throughput value by calculating throughput measures while considering other techniques

TABLE I
THROUGHPUT ASSESSMENT

Packet injection	Throughput (Packets/Cycle/Node)		
	LEHP	NeuronLink	Proposed (SbRN)
0.1	0.12	0.13	0.4
0.2	0.22	0.23	0.43
0.3	0.29	0.29	0.48
0.4	0.39	0.32	0.59
0.5	0.4	0.34	0.65
0.6	0.42	0.38	0.72

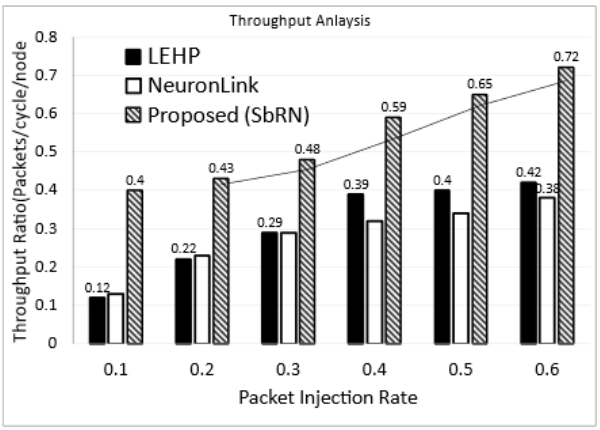


Fig.14. Throughput Analysis

For the 0.1 packet injection measure, the proposed model has attained the maximum throughout ratio of 0.4 cycles. Also, the existing scheme Neuronlink has achieved 0.13 throughput measure, and the LEHP model has earned 0.12 throughput measure for o.1 packet injection, as depicted in Fig.14 and Table I.

Latency

After importing the input, the time required to create the proposed design output is termed latency. If the latency was high, it reduced the system's rapidity. Here, the parameter latency was estimated based on the utilized clock cycles. Reducing the latency makes the Router access rapid.

TABLE II
LATENCY COMPARISON

Packet injection	Latency (Packets/Cycle/Node)		
	LEHP	NeuronLink	Proposed (SbRN)
0.1	10	25	7
0.2	12	25	9
0.3	14	50	11
0.4	18	80	13
0.5	20	84	16
0.6	23	96	19

Here, the latency is estimated based on the data sharing rate, the model that has reported a low delay score is considered the robust approach. The presented design has attained much less latency as 7 packets/cycle/node in that case. Simultaneously, the existing approaches LEHP has recorded the delay period as 10 packets/cycle/node and Neuronlink.

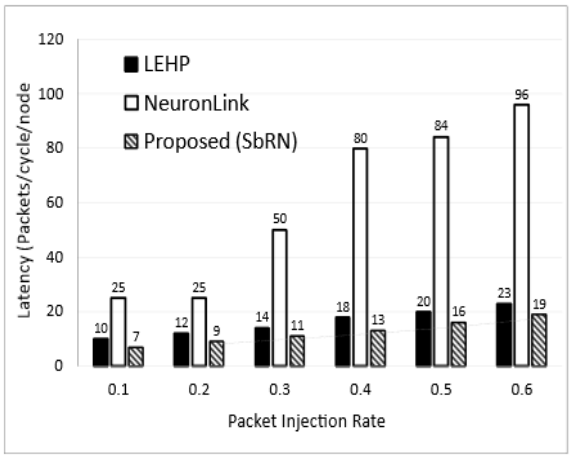


Fig.15. Latency Comparison

The duration required to broadcast the data is depicted in Table II and Fig.15. From this comparison, the successive measurements of the projected scheme were estimated.

Frequency

In data transmission applications measuring the frequency is the chief factor in estimating the data broadcasting rate. Hence, the frequency comparison is detailed in fig.16. The proposed design has been executed for the high frequency of 350MHz. Here, the hardware parameters were compared with the other techniques, Reconfigurable NoC [30]and NeuronLink [29].

Moreover, conventional models like NeuronLink have processed in 300MHz frequency, and the reconfigurable NoC has been executed in 250 MHz platforms, which is exposed in Fig.16.

Moreover, conventional models like NeuronLink have processed in 300MHz frequency, and the reconfigurable NoC has been executed in 250 MHz platforms, which is exposed in Fig.16.

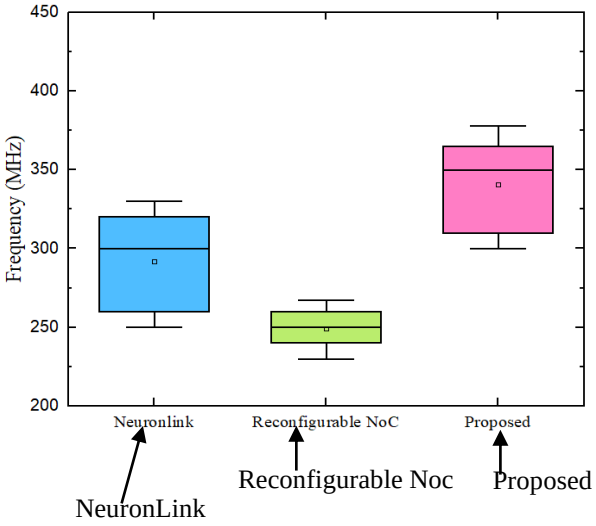


Fig.16. Frequency Assessment

Lookup-Table (LUT)

In VLSI, the memory cache size is an important parameter in improving the device's performance. So, to reduce the memory size, the weight of each neuron activation layer should be minimized. Because changing the program for each function process is t a complicated task and requires more computational resources. For that, compressions of feature maps are developed. The need of cache line compression is necessary while the data is printed back from level one cache to level two caches. Moreover, a simple circuit is processed to compress the cache line. Also, each word is arranged in parallel form. Hence, the LUT is described in Fig.17.

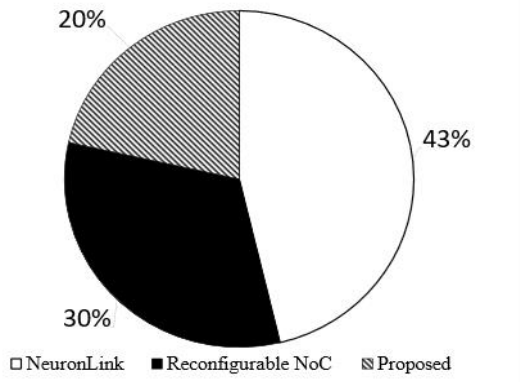


Fig.17. Validation of LUT

The LUT looks like a crossbar, which is utilized to make the output table from trained inputs. The LUT's have cluster interconnections that are completely joined in the form of a crossbar. To organize logical chunks, the LUTs are accumulated together then the LUT is designed with the resistive elements.

Flipflop

To fix an appropriate register line size, the projected model has optimized the neuron activation layers to store more data and improve the data broadcasting rate.

Moreover, layer optimization has reduced energy consumption and helps to quickly process a huge quantity of data. Usually, more instructions or storage modules were added to the register to process the large data. But it needs an additional energy source to perform the function. To avoid this problem, feature map optimization was introduced for the compressed data.

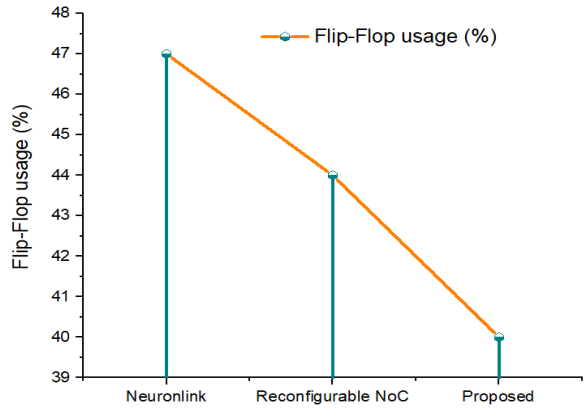


Fig.18. Comparison of Flipflop Usage

The usage of Flip-Flop is to save the data bits, also, it has dual static states. In addition, the Flip-flop is considered the sequential circuit model, which utilizes memory elements that depend on input and output. Therefore, the Flip-flop estimation is exposed in Fig.18.

Power consumption

After switching off the transistor, the emitting unwanted current id is called leakage current. The power consumption rate was recorded based on the leakage current measure. The power consumption is estimated by Watt (W).

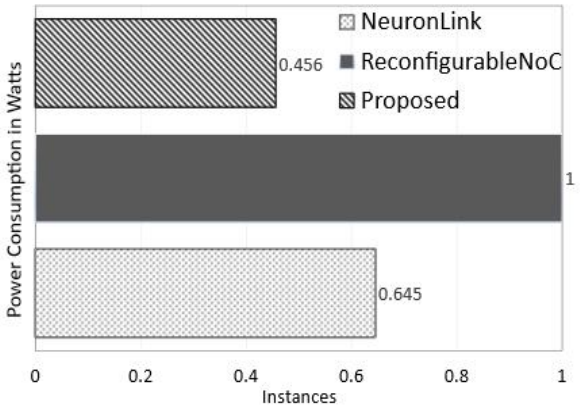


Fig.19. Power Analysis

The recorded power of the Neuronlink model is 0.645W and reconfigurable NoC 1W. Comparing the existing model, the proposed scheme has reported 0.456W of power. Hence, the robustness of the presented model was verified in Fig.19.

Table III shows the hardware outcome of the proposed model in comparison with the LEHP and Neuron Link devices. It has been observed that the proposed model gives good efficiency comparatively.

TABLE III
OUTCOME OF THE HARDWARE MODULE

Parameter	Hardware module outcome		
	Neuronlink	NeuronLink	Neuronlink
Power consumption	0.645W	1	0.456
Frequency	300MHz	250 MHz	350MHz
Flipflop	47%	44%	40%
LUTs	2644	2000	1900

VI. CONCLUSION

Nowadays, NoC model is an advanced and compactable design structure in many digital applications. However, high energy consumption and power leakage have made the chip performance difficult because of the wide range of data. For this reason, the current research has planned to optimize the feature maps of deep neural networks to optimize the chip performance. Henceforth, the projected novel approach is SbRN, which is used for weight compressing the chip activation layer. Then that designed chip router is used to transfer the data. Subsequently, the data transmission applications were adopted to validate the proposed model one of the chip memory routers, and its performance was validated. Finally, the parameters of the proposed metrics were compared with other approaches and have gained better results as 0.46W power consumption, 40% flipflop, 1900 LUT usage and 350 MHz frequency. Hence, the depicted results are quite better than other compared models. The NOC performance was improved by 20% compared to the previous model.

REFERENCES

[1] Savadkoohi, Marzieh, Timothy Oladunni, and Lara A. Thompson, "Deep Neural Networks for Human's Fall-risk Prediction using Force-Plate Time Series Signal," *Expert Systems with Applications*, 2021.

[2] Zhang, Zufan, et al., "Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions.," *Neurocomputing*, vol. 410, pp. 304–316, 2020.

[3] Li, Xiang, Qian Ding, and Jian-Qiao Sun, "Xiang, Qian Ding, and Jian-Qiao Sun. "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering & System Safety*, vol. 172, pp. 1–11, 2018.

[4] Ma and Xiaolei, "Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20.6, pp. 2278–2288, 2018.

[5] Singh Sukhdeep, Anuj Sharma, and Vinod Kumar Chauhan, "Online handwritten Gurmukhi word recognition using fine-tuned Deep Convolutional Neural Network on offline features (2021): 100037.," *Machine Learning with Applications*, vol. 5, no. 100037, 2021.

[6] Reinbrecht, Cezar, Bruno Forlin, and Johanna Sepúlveda, "Cache timing attacks on NoC-based MPSoCs," *Microprocessors and Microsystems*, vol. 66, pp. 1–9, 2019.

[7] Perez and Ivan, "BST: A BookSim-based toolset to simulate NoCs with single-and multi-hop bypass," *IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS) IEEE*, 2020.

[8] Davidson and Scott, "The Celerity open-source 511-core RISC-V-tiered accelerator fabric: Fast architectures and design methodologies for fast chips," *IEEE Micro 38.2*, pp. 30–41, 2018.

[9] Ray, Partha Pratim, Dinesh Dash, and Debashis De, "Edge computing for Internet of Things: A survey, e-healthcare case study and future direction," *Journal of Network and Computer Applications*, vol. 140, pp. 1–22, 2019.

[10] Kumar, Aruru Sai, and TVK Hanumantha Rao, "Scalable benchmark synthesis for performance evaluation of NoC core mapping," *Microprocessors and Microsystems*, vol. 79, p. 103272, 2020.
[11] Penna and Pedro Henrique et al., "Inter-kernel communication facility of a distributed operating system for NoC-based lightweight manycores," *Journal of Parallel and Distributed Computing*, vol. 154, pp. 1–15, 2021.
[12] Achballah, Ahmed Ben, Slim Ben Othman, and Slim Ben Saoud, "Problems and challenges of emerging technology networks− on−chip: A review," *Microprocessors and Microsystems*, no. 53, pp. 1–20, 2017.
[13] Yang and Pengfei, et al., "Partially shared cache and adaptive replacement algorithm for NoC-based many-core systems," *Journal of Systems Architecture*, vol. 98, pp. 424–433, 2019.
[14] Hamzei, Marzieh, and Nima Jafari Navimipour, "Toward efficient service composition techniques in the Internet of Things," *IEEE Internet of Things Journal*, vol. 5.5, pp. 3774–3787, 2018.
[15] Peng and Xiaochen, et al., "DNN+ NeuroSim V2. 0: An end-to-end benchmarking framework for compute-in-memory accelerators for on-chip training.," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2020.
[16] Kadetotad and Deepak, et al., "An 8.93 TOPS/W LSTM recurrent neural network accelerator featuring hierarchical coarse-grain sparsity for on-device speech recognition," *IEEE Journal of Solid-State Circuits*, vol. 55.7, pp. 1877–1887, 2020.
[17] Cho, Sung-Gun, Edith Beigne, and Zhengya Zhang, "A 2048-neuron spiking neural network accelerator with neuro-inspired pruning and asynchronous network on chip in 40nm CMOS," *2019 IEEE Custom Integrated Circuits Conference (CICC), IEEE*, 2019.
[18] Gao and Chang, et al., "EdgeDRNN: Recurrent neural network accelerator for edge inference," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10.4, pp. 419–432, 2020.
[19] Gao, Wei, Zhiliang Qian, and Pingqiang Zhou, "Reliability-and performance-driven mapping for regular 3D NoCs using a novel latency model and Simulated Allocation," *Integration*, vol. 65, no. 351–361, 2019.
[20] Mandal and Sumit K., et al, "A Latency-Optimized Reconfigurable NoC for In-Memory Acceleration of DNNs," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10.3, pp. 362–375, 2020.
[21] Noel and Adam B, et al., "Structural health monitoring using wireless sensor networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 19.3, pp. 1403–1423, 2017.
[22] Xinmiao Yu, Zhengpeng Li, Mingao Liu, and Jiansheng Wu, "Multi-module Fusion Relevance Attention Network for Multi-label Text Classification," *Engineering Letters*, vol. 30, no. 4, pp. 1237–1245, 2022.
[23] Yongzhi Liao and Qilin Tang, "Multiple Periodic Solutions for Cohen-Grossberg BAM Neural Networks with Mixed Delays and Impulses," *Engineering Letters*, vol. 30, no. 4, pp1185-1198, 2022.
[24] Mohammad Alkhazaleh, S.A. Aljunid, and Naseer Sabri, "An Efficacious Content Caching and Eviction Priorities (CCEP) for In-network Caching High Performance in Information-centric Networking," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp169-182, 2023.
[25] Gao and Chang, et al., "Recurrent neural network control of a hybrid dynamical transfemoral prosthesis with EdgeDRNN accelerator," *IEEE International Conference on Robotics and Automation (ICRA). IEEE*, 2020.
[26] Merrikh-Bayat and Farshad, "A numerical optimization algorithm inspired by the strawberry plant," no. arXiv preprint arXiv:1407.7399, 2014.
[27] Baharloo and Mohammad et al., "ChangeSUB: A power efficient multiple network-on-chip architecture," *Computers & Electrical Engineering*, no. 83, p. 106578, 2020.
[28] Li, Yuan, and Ahmed Louri, "ALPHA: A learning-enabled high-performance Network-on-Chip router design for heterogeneous manycore architectures," *IEEE Transactions on Sustainable Computing*, 2020.
[29] Xiao and Shanlin, et al., "NeuronLink: An Efficient Chip-to-Chip Interconnect for Large-Scale Neural Network Accelerators," *IEEE Transactions on Very Large-Scale Integration (VLSI) Systems*, vol. 28.9, pp. 1966–1978, 2020.
[30] Ramy, Ahmed, Hassan Mostafa, and Ahmed H. Khalil, "Design of a reconfigurable network-on-chip for next generation FPGAs using Dynamic Partial Reconfiguration," *Microelectronics Journal*, vol. 108, no. 104964, 2021.

Sumana Achar born in Ambikanagar, Uttara Kannada District of Karnataka, India in 2nd July 1980. She is the Research Scholar at JSS Academy of Technical Education, Bengaluru and has received M.Tech. degree in VLSI Design and Embedded Systems from BVB College of Engineering, Hubli, Karnataka, India in 2009. Her major field of study is VLSI Design and Embedded Systems. She has 18 years of experience in teaching at Engineering Institutions and 2 years of Industrial Experience. Her last service was for APS College of Engineering, Bengaluru, India. Her current research interest is Networking. Ms. Achar is the member of ISTE, IFERP and currently applied for IAENG membership

D. Jayadevappa is the member of IAENG since 2008. He was born in Kabbur, Karnataka, India in 1970. He holds a Ph.D. degree from JNTU college of Engineering, Kakinada, Andhra Pradesh., India, M.Tech. degree from SJCE, Mysore, VTU, in 2000 with specialization in Bio-Medical Instrumentation. He received his B.E. degree in Instrumentation Technology from SIT, Tumkur, Bangalore University in 1994. He is currently working as Professor and Head of the department of Electronics and Instrumentation Engineering, JSS Academy of Technical Education, Bengaluru, Karnataka, India. He has 27 years of teaching and industrial experience and is the Member of IAENG, IETE, ISTE & BMESI. He has published more than 115 papers in International Journals and conferences. One paper published in IAENG-IJCS Journal in the year 2009. His areas of interests are Digital Image Processing and Bio-Medical Signal Processing.