IEICE Electronics Express, Vol.7, No.11, 791–796

A high performance network-on-chip scheme using lossless data compression

Hong-Sik Kim[1], Youngha Jung[2], Hyunjin Kim[1], Jin-Ho Ahn[3], Woo-Chan Park[4], and Sungho Kang[1a)]

[1] *Department of Electrical & Electronic Engineering, Yonsei University, Seoul, Korea*

[2] *Samgsung Electronics Co. Ltd., Suwon, Kyun-gi Do, Korea*

[3] *Department of Electronic Engineering, Hoseo University, Chungchung Do, Korea*

[4] *Department of Computer Science, Sejong University, Seoul, Korea*

a) *shkang@yonsei.ac.kr*

Abstract: A new NoC (network on chip) architecture using lossless data compression and decompression to improve the performance and power efficiency of the on-chip interconnect is proposed. In the proposed NoC scheme, the sender compresses the data to be transferred in order to reduce the number of data packets and the receiver decompresses the encoded data to restore the original data. For the lossless compression and decompression, we have implemented a hardware CODEC based on a Golomb-Rice algorithm. According to the experimental results using a cycle-accurate NoC simulator, the proposed scheme could significantly improve the performance and power efficiency of the conventional NoC architecture.

Keywords: NoC, lossless data compression, Golomb-Rice algorithm

Classification: Integrated circuits

References

[1] D. Park, C. Nicopoulos, J. Kim, N. Vijaykrishnan, and C. R. Das, "A Distributed Multi-Point Network Interface for Low-Latency, Deadlock-Free On-Chip Interconnects," *Proc. Int. Conf. Nano-Networks and Workshops*, pp. 1–6, 2006.

[2] P. Rosinger, B. M. Al-Hashimi, and N. Nicolici, "Scan Architecture with Mutually Exclusive Scan Segment Activation for Shift and Capture Power Reduction," *IEEE Trans. Computer Aided Design Integr. Circuits Syst.*, vol. 23, no. 7, pp. 1142–1153, 2004.

[3] A. Jalabert, S. Murali, L. Benini, and G. D. Micheli, "Xpipes Compiler: A Tool for Instantiating Application Specific Networks on Chip," *Proc. Design and Test in Europe*, pp. 884–889, 2004.

[4] R. Mullins, A. West, and S. Moore, "Low-latency Virtual Channel Routers for On-Chip Networks," *Procs. Int. Symp. Computer Architecture*, pp. 188–197, 2004.

[5] M. Palesi, R. Holsmark, S. Kumar, and V. Catania, "Application Specific

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010

Routing Algorithms for Networks on Chip," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 3, pp. 316–330, 2009.
[6] R. F. Rice "Some Practical Universal Noiseless Coding Techniques," *Technical Report 79-22*, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, 1979.

1 Introduction

The increasing complexity of SoC (system on chip) products and the decreasing feature sizes in the DSM (deep sub-micron) technologies have made the on-chip global data communication a critical bottleneck for the high performance design [1]. Recently, in order to guarantee the performance and reliability of the on-chip interconnect, a new design methodology called as an NoC (network on chip), has been introduced [2]. Contrary to the macro-networks, the NoC design inherently encounters high resource limitation since all of the NoC hardware should be implemented on a chip. Therefore, lots of researches have been performed to improve the NoC performance under high resource limitation, such as a pipelined channeling [3], a virtual channel router [4] and a deadlock-free routing [5].

In this paper, a new NoC architectural approach using lossless data compression and decompression to improve the NoC performance is proposed. The original data to be transferred are encoded into a reduced number of data packets at the sender. After the data packets arrive at the receiver, the original data are restored by decompressing the encoded data. We have developed hardware encoder and decoder architectures based on the Golomb-Rice algorithm [6] for the lossless compression and decompression. Without loss of generality, the proposed NoC scheme can be applied to most of the existing NoC architectures with a slight modification regardless of the topologies and the routing algorithms to improve the performance of the NoC system. For the performance evaluation, a cycle-accurate NoC simulator has been developed with a *C* language. In addition, NoC router with the proposed Golomb-Rice encoder and decoder were design with Verilog HDL and synthesized with *Design Vision* from *Synopsys*. According to experimental results for mesh and torus NoC's, the proposed NoC scheme could improve the performance and power consumption efficiency.

2 Proposed NoC scheme

Fig. 1 (a) illustrates the proposed NoC scheme based on a simple 4x4 mesh architecture. Note that the proposed NoC scheme can be applied to not only meshes but also various NoC's. In the proposed NoC scheme, when a data transaction is issued at a sender, the data to be sent are compressed by an on-chip hardware encoder and packetized by the NI (network interface). Then, the data packets will traverse the on-chip network and finally will be delivered to the receiver. At the receiver, the packets are de-packetized

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010

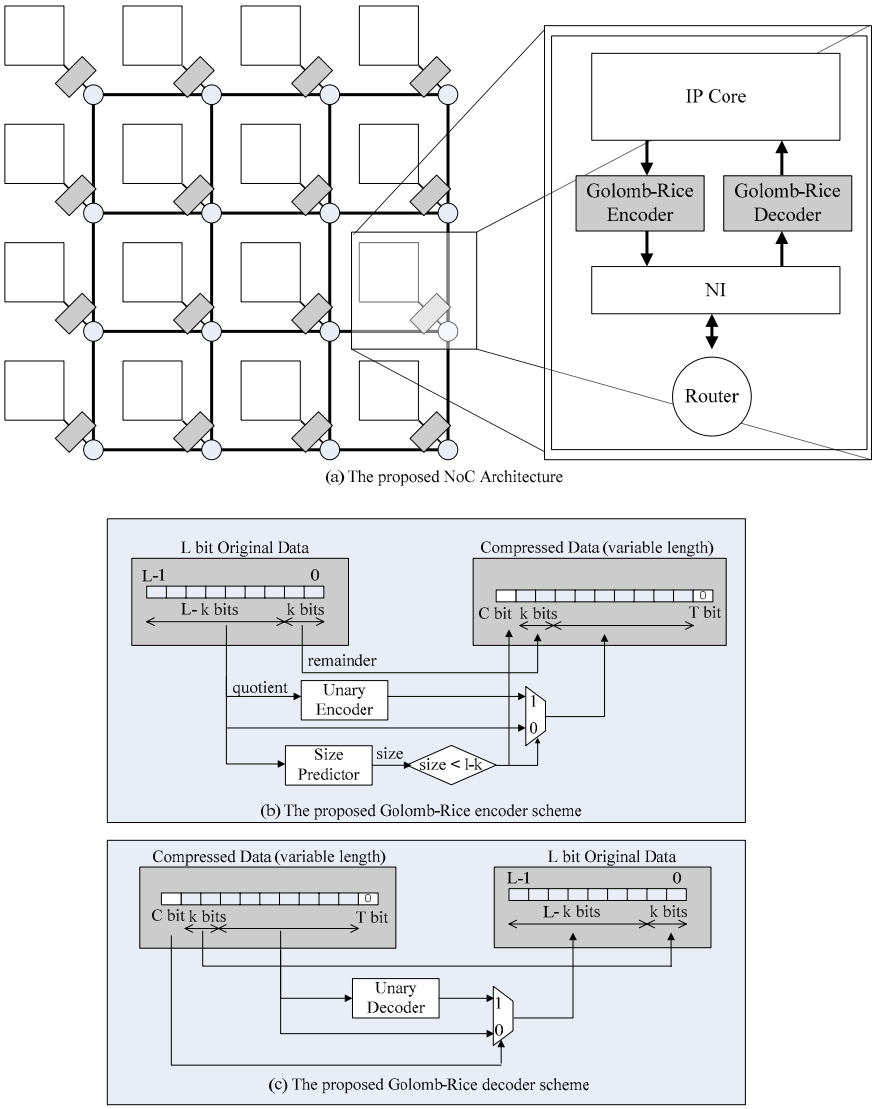

Fig. 1. Proposed NoC Architecture

by the NI and decompressed to restore the original data by the hardware decoder. Since the original data has been compressed by the encoder, the number of the data packets produced by the NI will be decreased, which will contribute to the improvement of the network throughput. For the lossless data compression, we have adopted the Golomb-Rice algorithm for its low complexity. According to the Golomb-Rice algorithm, for an input data, n, and the dividend of 2^k, the quotient value is *unary* encoded and the residual value is *binary* encoded. Since, in the Golom-Rice algorithm, the dividend value is always the power of 2, the quotient and the residual values can be easily calculated so that the hardware complexity of the Golomb-Rice encoder and decoder is very low compared to the other lossless compression algorithms.

Fig. 1 (b) and (c) depict the hardware architectures for the proposed Golomb-Rice CODEC. We assumed an L bit original data and a 2^k dividend. In the Golomb-Rice algorithm, the quotient value is *unary* encoded into a variable length data and the residual value is *binary* encoded into a fixed

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010

length data so that, in the worst case, the length of the encoded data may be larger than the original data. Thus, the size prediction logic pre-calculates the length of the encoded data and if the predicted length is larger than the original data length, then the original data is used without any compression. The 'C' bit signifies whether the encoded data are compressed or not. In addition, since the length of the quotient encode will be variable, in order to identify the end of the data, the 'T' bit (the terminating zero) is attached to the quotient encode.

In the Golomb-Rice decoding stage, the 'C' bit is first checked to identify whether the input data are compressed or not. If the 'C' bit is '1' or the input data is compressed one, the earlier k bits are *binary* decoded and the remaining data before the 'T' bit are *unary* decoded. Then the two decoded data are concatenated into the original data. If the 'C' bit is '0', the k bit data after the 'C' bit and the '$L-k$' bit data before the 'T' bit are directly concatenated into the original data. Since the compression ratio is much larger than '1', it is highly probable that multiple 'T' bits exists in the current input data. In addition, the current encode data may be placed to the successive two input data (the current data and the next data). Therefore, the hardware encoder and decoder include small registers and control flow logic besides the combinational blocks in the Fig. 2.

3 Experimental results

In order to evaluate the proposed NoC scheme, we have implemented a cycle-accurate NoC simulator with a C-language. The developed simulator involves the Golomb-Rice encoder/decoder, NI's, routers and interconnect channels. Several performance enhancement techniques such as the pipelined channeling, the virtual channels and the adaptive routing algorithm [5] are involved for the target NoC platform. For the simulation scenario, internal IP cores generate random data according to the data generation ratio. The size of the data is selected randomly from 4 bytes to 1024 bytes. The generated random data are compressed by the Golomb-Rice encoder and packetized by the NI. The data packet will travel through the on-chip network and arrive at the destination node or the receiver. Then the data packets are de-packetized by the NI and are decompressed to restore the original data by the Golomb-Rice decoder. For the evaluation of the network throughput, the cycle times between the moment of data generation at the sender and the moment of the data restoration at the receiver are calculated. This process is repeated until 5,000,000 random data are generated and diminished. Simulation is performed for 3 physical channel bandwidths cases such as 8, 16, and 32 bits in order to evaluate the performance impact of the proposed scheme on diverse channel bandwidth.

The proposed NoC router including NI has been designed with Verilog HDL and synthesized with *Design Vision* from *Synopsys* under 0.15 um CMOS technology. The NoC configurations for the simulation are used for the router design. In case of a conventional NoC without any compression

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010

method, the hardware area (gate count), the dynamic power consumption, and the critical path delay were estimated such as 10,114, 689.61 uW, and 3.56 ns, respectively. In case of the proposed scheme with encoder and decoder, the hardware area, the dynamic power consumption, and the critical path delay were estimated such as 11,602, 703.21 uW, and 3.59 ns, respectively. These data for the power consumption and the critical path delay were used by the simulator in order to calculate the performance and power efficiency of the NoC system.

Table I and Fig. 2 show the comparison between the proposed NoC system and two conventional NoC systems (mesh and torus topologies) without any encoder/decoder in terms of the throughput and the power efficiency, respectively. DR (data generation ratio) stands for the number of IP cores generating random data divided by the total number of the IP cores at one cycle. For example, if DR is 0.4, then 25.6 ($8 \times 8 \times 0.4$) cores generate random data at every clock cycle. CB means the size of the physical channel bandwidth. TP stands for the throughput of the NoC system, or the number of millions of packets processed per second (Mpps). For the evaluation of the power efficiency, the energy consumption of the proposed scheme and the convectional NoC schemes to process the whole data set are calculated and the energy reduction introduced by the proposed scheme over the conventional mesh and torus NoC architecture are illustrated in Figure 2.

According to Table I and Figure 2, the proposed scheme improved both the network throughput and power efficiency in all cases. The average improvements of throughput and power efficiency were estimated by about 15.53% and 9.20%, respectively. In addition, the proposed scheme shows more improvement of throughput and power efficiency over the conventional NoC system without any compression as the size of channel bandwidth decreases, which means that the compression methodology could contribute more to narrow bandwidth case.

Even with the slight increase of area and power consumption of the router

Table I. Performance evaluations

DR	CB (bits)	8x8 mesh NoC			8x8 torus NoC		
		TP (Mpps)		ⓑ−ⓐ / ⓐ	TP (Mpps)		ⓓ−ⓒ / ⓒ
		Without ENC/DEC ⓐ	With ENC/DEC ⓑ		Without ENC/DEC ⓒ	With ENC/DEC ⓓ	
0.1	8	3.38	4.12	18.89%	3.92	4.65	18.72%
	16	4.22	4.90	14.11%	4.63	5.33	15.12%
	32	7.96	9.16	13.08%	8.21	9.26	12.78%
0.2	8	3.32	4.08	19.72%	3.55	4.26	20.11%
	16	4.02	4.77	15.66%	4.31	5.02	16.54%
	32	7.55	8.89	14.54%	7.84	8.88	13.21%
0.3	8	3.02	3.72	19.82%	3.39	4.05	19.33%
	16	3.96	4.75	15.67%	4.10	4.72	15.01%
	32	7.06	8.34	14.62%	7.46	8.50	13.92%
0.4	8	1.22	1.51	19.80%	1.89	2.26	19.42%
	16	2.01	2.33	15.70%	2.49	2.89	16.02%
	32	2.72	3.22	14.65%	3.05	3.51	15.33%
Average		4.20	4.86	15.59%	4.57	5.28	15.48%

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010

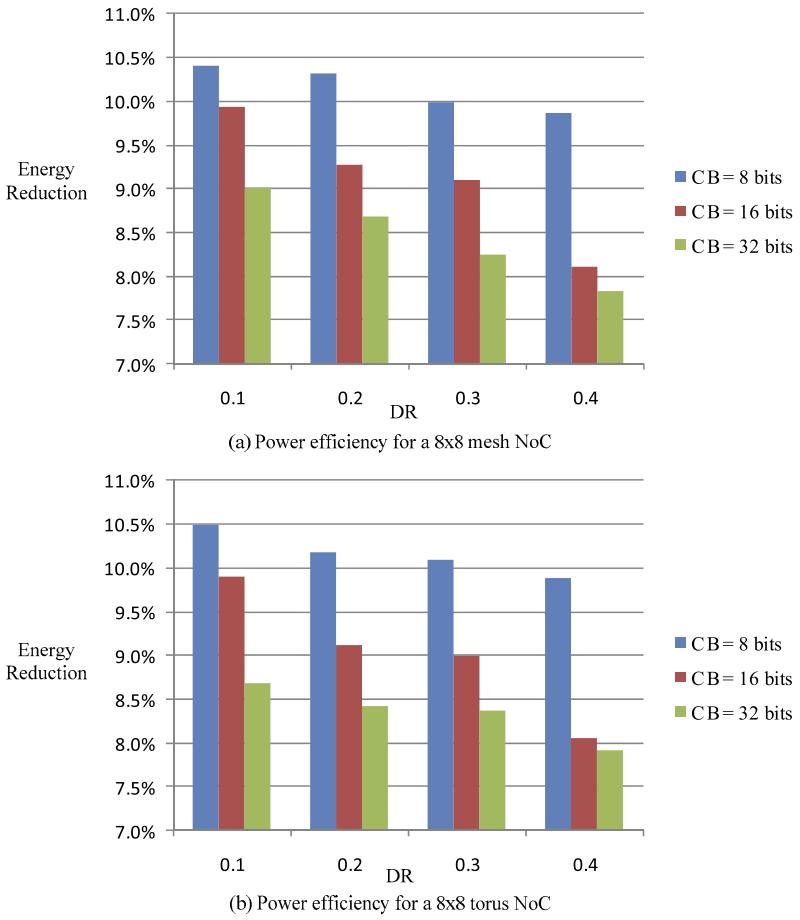

Fig. 2. Power Efficiency Evaluation

due to the overhead of the decoder and encoder, the proposed scheme could reduce the number of the network data transaction so that the throughput and the power efficiency of the NoC systems can be significantly improved.

4 Conclusion

A new NoC design scheme using Golomb-Rice based lossless data compression and decompression is proposed to improve the on-chip network performance. The proposed scheme can be used to enhance the performance of any existing NoC architectures regardless of their topologies and routing algorithms. According to the experimental results, the proposed scheme could considerably improve both the network throughput and power efficiency with a little hardware area overhead, compared to convectional NoC's without any decoder/encoder.

DOI: 10.1587/elex.7.791
Received April 25, 2010
Accepted May 15, 2010
Published June 10, 2010