다중 프로세서를 갖는 SoC 를 위한 CDMA 기술에 기반한 통신망 설계

천익재*, 김보관**
충남대학교 공과대학 전자공학과

A CDMA-Based Communication Network for a Multiprocessor SoC

Ik-Jae Chun*, Bo-Gwan Kim**
Dept. of Electronics Engineering
Chungnam National University
E-mail : *s_amity@cnu.ac.kr, **bgkim@cnu.ac.kr

Abstract

In this paper, we propose a new communication network for on-chip communication. The network is based on a direct sequence code division multiple access (DS-CDMA) technique. The new communication network is suitable for a parallel processing system and also drastically reduces the I/O pin count. Our network architecture is mainly divided into a CDMA-based network interface (CNI), a communication channel, a synchronizer. The network includes a reverse communication channel for reducing latency. The network decouples computation task from communication task by the CNI. An extreme truncation is considered to simplify the communication link. For the scalability of the network, we use a PN-code reuse method and a hierarchical structure. The network elements have a modular architecture. The communication network is done using fully synthesizable Verilog HDL to enhance the portability between process technologies.

I. Introduction

The most frequently used communication architecture in SoC is a bus-based interconnection. However, the bus architecture is inadequate for future SoC design. In particular, in the implementation of a multiprocessor SoC, the bus architecture comes to the forefront because the performance of system is dependent not only on the CPU speed but also on the bus architecture. Hence, we believe that a new communication network addressing future issues of the bus architectures is needed. Many systems would like to maintain the serialization and broadcast properties of the bus, but cannot use a shared medium because of electrical constraints. For example, the parasitic resistance and capacitance of on-chip interconnects requires repeaters every few millimeters to maintain signaling speed. Thus, a large on-chip bus cannot be realized as a single electrical node driven and sensed by many modules.

NoC has been proposed as a solution to the interconnection problems of existing on-chip buses for highly complex chips. NoC architecture is a scalable communication infrastructure. Its basic idea is borrowed from traditional large-scale multi-processors and wide-area networks domain, and envisions on-chip router (or switch)-based net-works on which packetized communication takes place. Therefore, the NoC demands communication network to be scalable, flexible, and efficient.

Most of the researches for NoC concentrate on packet

This work was supported in part by the ITRC (Information Technology Research Center) and the IDEC (IC Design Education Center) at KAIST and IT-SoC (Information Technology System on Chip).

switched network. However, there are rare researches for on-chip communication network [1], [2], [3]. The researches are based on code division multiple access (CDMA) technique [4]. [1] proposed using Kasami sequences to rout packets among processors in a multiprocessor network. However, it used only one large central switching element to perform all of the routing and did not consider network issues such as buffering and packet contention. [2] proposed a hierarchical star network based on CDMA switching element. However, it did not consider a flow control for packet transmission and has a performance degradation due to arbitration which is used to overcome a destination contention. [3] considered multi-valued signaling with CDMA to increase bus bandwidth, but these did not use a network architecture and relied on non-traditional signaling methods.

In this paper, we propose a new on-chip communication network for a NoC. Our approach decouples the functional task from the communication task by well-defined network elements. It can be also optimized for area-efficient interconnection according to an application by configurable and modular design. We propose a latency-insensitive flow control scheme. It is done at the transport level by the implementation of a reverse channel. The network is easily scalable and it can be applied to various interconnection types.

II. Subnetwork Architecture

The architecture of our communication network is shown in Fig. 1. It is largely composed of 3 functional blocks: CDMA-based network interface (CNI), communication channel, and synchronizer. The resources are attached to the communication channel through the CNI. The resource can be computational IPs, such as a computation or storage unit or an integrated processing unit, in which memory, DSP, controller, and peripherals may be integrated together. Information is transferred from a source to a destination by a packet structure. Our communication network elements are modular, configurable and reusable. Therefore it can be used in various applications. Our network uses an extreme binary truncation for small I/O pin count. Due to the truncation, we can connect up to 7 resources. The CDMA network connecting 7 resources is called subnetwork on which scalable CDMA network is based.

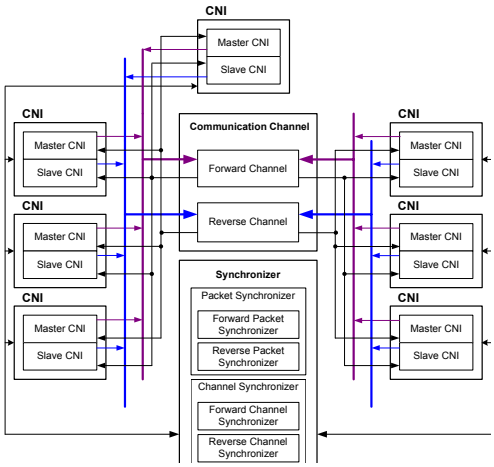

Figure 1. On-chip communication network architecture.

2.1 CDMA-based Network Interface (CNI)

The CNI decouples functional task and communication task and interconnects resource and communication channel. The resource dumps the message on the buffer in the CNI without a direct control for the communication channel. The CNI consists of master CNI module (MNI) and slave CNI module (SNI). In the implementation, the CNI may have one of MNI and SNI or both of them. The MNI is for a data packet transmission and the SNI is for the reception of the data packet. Each of MNIs and SNIs has both TX channel and RX channel. The TX channel is a modulator and the RX channel is a demodulator. The CNI has an assigned 7-bit PN codeword for TX channel and RX channel can concurrently receive all the packets from other CNIs. The modulator and demodulator operation is depicted in Fig. 2.

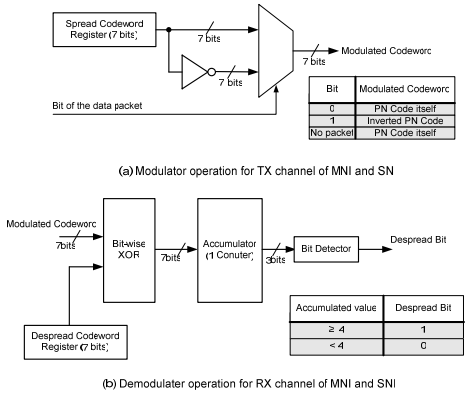

Figure 2. Modulator and demodulator for the CNI.

2.2 Communication Channel

The communication channel is a physical link to interconnect among the CNIs. The communication channel consists of forward channel and reverse channel. The forward channel is for data packet transmission and the reverse channel is for a packet flow control. The forward and the reverse channel have the same architecture. Each of the channels is equipped with summer module and binary truncation module for mixing codeword transferred from the CNIs in a parallel fashion. The separated communication channels reduce the latency of packet transmission by the fast flow control through the reverse channel.

2.3 Synchronizer

Each of forward and reverse channels requires bit sync and packet sync for sending and receiving packet. The synchronizer consists of a channel synchronizer and a packet synchronizer. The packet synchronizer supports a packet start timing sync for the acquisition of valid data in order to remove a full-time correlator raising the problem of power consumption. The channel synchronizer provides a channel sync clock to the CNIs in order to maintain the orthogonality of channels.

2.4 Packet Structure

The type of packets is two: data packet and ACK packet. The data packet is for transmitting message data from the MNI to the SNI. The data packet is divided into three fields. For 7 resources, 3-bit destination ID, 3-bit source ID, and payload. The ACK packet is for controlling the transmission of the data packet through the reverse channel. It sends packet reception information of the receiver to the transmitter. The ACK packet is divided into two fields: A subnet ID for the source CNI waiting for ACK packet and ACK field containing ACK/NACK information about whether the slave CNI has received or is receiving the data packet

2.5 Packet Flow Control

The start of transmission is from the MNI's TX channel. Resource sends the modulated packet through the MNI (source CNI) to the forward channel. The forward channel adds the modulated codewords from the MNIs, truncates for each bits of the summed codewords, and contributes the truncated data to communication line. The SNI (destination CNI) demodulates packet through the SNI's RX channel. The destination CNI extracts the destination ID from the packet, and then propagates ACK/NACK information back to the source CNI according to the RX buffer level of itself.

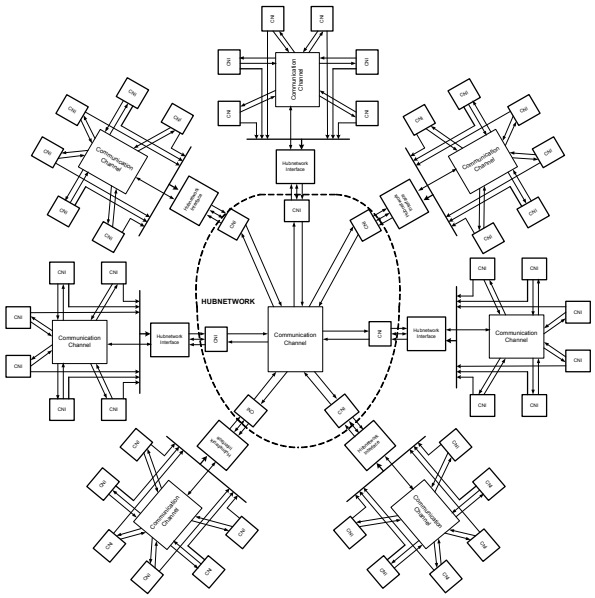

Figure 3. Example of scaled communication network.

III. Scaled Network Design

Our on-chip network is based on the subnetwork which consists of 7 resources. For more than 7 resources, we propose a hierarchical network architecture by a hubnetwork. The hubnetwork interconnects among the subnetworks. The architecture of the hubnetwork is exactly the same as the subnetwork. Two networks are connected by a hubnetwork interface (HI). The HI is basically an interface module for sub-to-subnetwork interconnection. We replace a CNI of the subnetwork with the HI, and then directly connect a CNI of the hubnetwork to the HI. We can support various scalable networks by reusing the HI and the subnetwork. Figure 3 shows a scaled network for 42 resources. Each of the hubnetwork and the subnetwork uses the same 7 PN codes for 7 nodes. The 7

PN codes are reused by the hubnetwork and the subnetwork.

The overall flow control from source to destination in the scaled network is performed on the basis of the subnetwork. This means that the overall flow control is achieved by the flow control of each of the subnetworks belonging with source to destination path.

IV. Experimental Results

We have simulated the communication network with a Verilog simulator. The throughput and latency are considered for the subnetwork and the scaled network for 42 resources. We have used the most common benchmark, which is the uniformly random distribution of packet destinations.

We randomly generate the packets at a set of fixed communication ratios and compute the average time spent on transferring one packet from source to destination nodes. In here all paths of the network are equally loaded.

Figure 4 shows the throughput and latency in the subnetwork. The offered traffic is the average proportion of cycles for packet injection. Communication channel for one resource is contention-free in the subnetwork.

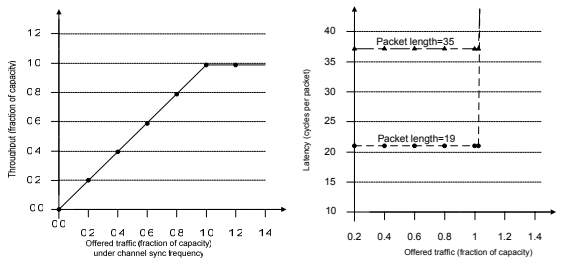

Figure 4. Throughput and latency of the subnetwork.

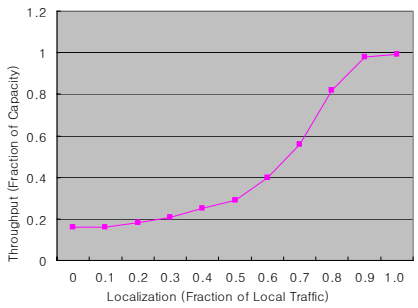

Figure 5. Throughput with localization factor.

In the scaled CDMA communication network connecting 42 resources, the benchmark with uniformly random distribution is a pessimistic estimation of the practical performance of a network [5]. Hence, we consider the locality of network traffic. In the Fig. 5, we show the effect of traffic localization. Our network has very high throughput for the uniformly distributed traffic with a high locality.

V. Conclusion

In this paper, we have presented a new on-chip communication network based on CDMA technique. The communication network uses the simple flow control scheme insensitive to latency. It also uses parallel codeword transmission method for reducing the spreading overhead. In spite of the parallel transmission, the complexity of interconnection is very small due to the extreme binary truncation. In our future work, buffer optimization could be considered for more area reduction. In addition, we plan to model an application to determine its performance.

References

[1] R.H. Bell,Jr., Chang Yong Kang, L. John, and E.E. Swartzlander,Jr., “CDMA as a Multiprocessor Inter-connect Stratagy,” Conference on Signals, Systems and Computers, Vol. 2, pp. 1246-1250, Nov. 2001.

[2] Daewook Kim, Manho Kim, and Gerald E. Sobelman, “CDMA-BASED NETWORK-ON-CHIP ARCHITEC-TURE,” 2004 IEEE Asia-Pacific Conference on Circuits and Systems, pp. 137-140, Dec. 2004.

[3] Ryuji Yoshimura et al., “DS-CDMA Wired Bus with Simple Interconnection Topology for Parallel Processing system LSIs,” IEEE International Solid-State Circuits Conference, pp. 370–371, Feb. 2000.

[4] Samuel C. Yang, CDMA RF System Engineering, ARTECH HOUSE INC., 1998

[5] Partha Pratim Pande, Cristian Grecu, Michael Jones, André Ivanov, and Resve Saleh, “Performance Evaluation and Design Trade-Offs for Network-on-Chip Interconnect Architectures,” IEEE Trans. on Computers, vol. 54, no. 8, pp. 1025-1040, Aug. 2005.