

DCP 환경에서의 고속 Serial 네트워크 환경구현

High Speed Serial Network Environment on DCP

전기만*, 박창원, 정하중

(Ki-Man Jeon, Chang-Won Park and Ha-Joong Chung)

Abstract : Nowadays, we can enjoy access to high speed network and advanced services of convergence between broadcasting and communication anywhere and anytime through a ubiquitous computing. So, now digital convergence devices come out constantly. These devices are required faster network environment for high speed data processing than before. In this paper, we describe the design of InfiniBand network adapter, which is included two FPGA chipsets. When this adapter is installed to Digital Convergence Platform (DCP), network performance will be improved. The adapter includes 12channel serial ports for external communication and internally, uses PCI-Express bus. We have finished the test of high speed serial based network adapter through composing complete InfiniBand network and applied fabric management software. So, we have verified that it can be applied on DCP environment.

Keywords: Digital Convergence Platform, High Speed Serial Network, InfiniBand, PCI-Express

I. 서론

시간과 장소, 정보기기, 콘텐츠에 관계되지 않고 통신, 방송, 인터넷서비스 등 융합서비스를 언제 어디서나 이용할 수 있는 환경이 도래함에 따라 능동화, 융복합화, 광대역화가 광범위하게 생활화 되고 있고, 유비쿼터스 및 지능형 홈네트워크 환경의 발전은 멀티미디어 콘텐츠의 질적 향상은 물론 단순형 정보 서비스에서 지식처리와 같은 정보창조형의 지식형 서비스로 발전하고 있다. 또한 데이터 전송기술의 발달로 네트워크에 가중되는 데이터의 양은 수직적으로 증가하고 있는데, 취급되는 데이터의 형식이 단순한 문자나 그림이 아닌 비디오/오디오 와 같은 멀티미디어인 점을 감안한다면 네트워크다뤄야 할 데이터량은 폭발적이라 할 수 있다. 그러한 방대한 데이터를 처리하기 위해, 해당 기기들의 내부 프로세서의 성능향상은 물론, 외부 네트워크환경 역시 획기적인 향상이 요구된다.

인터넷 서비스 시장의 급성장과 더불어 가정내의 디지털 정보기반 기술의 발전은 고품질 멀티미디어 서비스와 양방향 디지털 아이템 거래와 같은 신산업 서비스가 미래의 정보 서비스로 급부상함으로 인해, 이를 지원할 수 있는 신개념의 종합 정보서비스 솔루션의 개발이 필요하다. 차세대 디지털 컨버전스 플랫폼(Digital Convergence Platform, DCP) 기술은 이러한 요구를 만족시키는 토털 정보서비스 솔루션으로서 방송, 통신, 컴퓨터, 가전 등을 포함한 다양한 미디어 및 디지털 데이터들을 융합하여 서비스하는 기술로서 전자산업 분야의 중추적인 역할을 담당할 것으로 예상된다.

현재 서버 및 스토리지 시스템에서 사용되는 주된 입출력 방식은 PCI 버스를 통해 연결망에 접속하는 다중 계층구조이다. 그러므로 각 노드의 전송능력은 PCI 버스 성능에 의해 제한되고 있다. PCI 버스는 공유버스 구조를 갖는 입출력의

구조적 단점으로 확장성 및 성능향상에 한계를 벗어나지 못한다. 이러한 한계를 극복하기 위해 모든 연결을 스위치 기반형 단일 연결망으로 구성하여, 가용성 및 확장성을 보장하고, 점대점 연결을 통한 고속 직렬 통신방법을 사용 전송속도를 향상시키는 새로운 시스템 연결방식이 대두되기 시작했다. 현재 PCI-Express, Hyper-Transport, Rapid I/O 규격 제품들이 등장되고 있으며, 전체 네트워크의 상태파악 및 효율적인 관리와 운영에 장점을 가진 인피니밴드 표준에 따른 제품들 또한 등장하고 있다 [2,4,5].

인텔, IBM 등 강력한 시장 지배력을 지닌 IBTA 운영위원회 회원사를 주축으로 현재 다양한 InfiniBand 관련 제품들이 개발 및 상용화되고 있다. 관련 제품군으로는, H/W 분야로 Host & Target Channel Adapter, Switch, Router 등이 있으며, 네트워크 환경설정 및 운영을 위한, S/W 및 Server, Storage 등의 제품들이 있다. 또한, SI 형태의 통합 Solution 을 제공하는 업체들이 등장하여, Network Storage 및 Server 분야에 새로운 시장을 형성하고 있다. 그 중 Mellanox사 “InfiniScale”, “InfiniHost”, “Infini-Bridge” 등의 칩셋을 적용한 다양한 제품들을 선보이며, 시장을 주도하고 있으며, 앞서 언급한 여러 H/W 업체 또한, Mellanox사의 칩셋을 채용하여, 다양한 제품군을 출시하고 있다. 또한, 최근 국내에서도 스토리지 시스템 관련 업체를 중심으로 InfiniBand 관련 제품에 많은 관심을 가지고 있으며, 제품개발을 진행하고 있어, 가까운 미래에 새로운 시장이 형성될 것으로 기대된다 [1-3].

본 논문의 구조는 다음과 같다. 2장에서는 InfiniBand 네트워크 구조 및 구성요소들의 설명하고, 기존에 상용화되어 Infini-Band 네트워크에 활용되고 있는 네트워크 어댑터의 종류 및 구조에 대한 설명한다. 3장에서는 고속 네트워크 환경의 DCP 를 위해 설계된 12x 채널기반의 테트릭 어댑터의 구조 및 특징에 대해 언급한다. 마지막으로 4장에서는 어댑터를 통한 S/W 적용 및 테스트 결과를 바탕으로, 추후 SoC 형태의 ASIC 칩셋 개발 및 해당 칩셋이 적용된 네트워크 어댑터 설계계획으로 본 논문을 정리한다.

* 책임저자(Corresponding Author)

전기만, 박창원, 정하중 : 전자부품연구원 지능형정보시스템연구센터
(kmjeon@keti.re.kr, parkcw@keti.re.kr, chungji@keti.re.kr)

※ 본 연구는 산자부 중기적용개발사업비의 지원을 받아 연구되었음.

II. InfiniBand Network Adapter

InfiniBand 네트워크 어댑터는 지역 또는 원격 DMA 기능을 수행하는 Programmable DMA 엔진으로 정의될 수 있으며, 가상 메모리 보호 메커니즘을 포함하고 있다. 전송 요구를 해독하여 해당 전송 요구를 처리하기 위한 패킷을 발생시키고 수신하는 기능을 수행하며, 기능에 따라 HCA와 TCA로 구분된다. HCA는 스위치와 TCA모두에 직접 연결할 수 있으므로 서버 구성시 유연하게 적용할 수 있으며, TCA는 입출력 장치와 연결망을 이어주는 정지로 활용된다 [6-8].

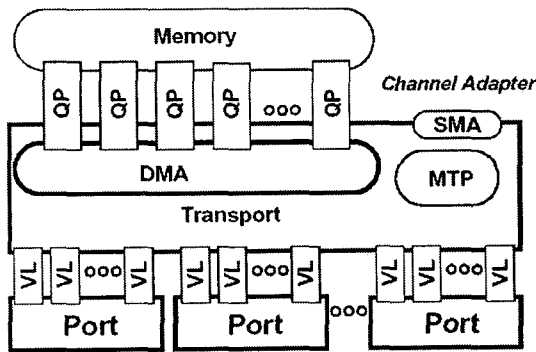


그림 1. 채널 어댑터 개념도.

1. InfiniBand 네트워크 특징

차세대 서버 시스템구조 기술개발을 위해, Compaq, Dell, IBM, MS, Sun, Microsystems 등이 주축이 되어 1999년 8월 IBTA (InfiniBand Trade Association)가 결성되었고, 2000년 9월 발표된 InfiniBand Architecture 표준규격 1.0이후 2002년 11월 표준규격 1.1까지 발표 되고있다. InfiniBand는 상호 독립적인 프로세서 플랫폼, 입출력 처리 그리고 입출력 장치를 연결하는 시스템 연결망으로 통신 및 관리 메커니즘을 포함하며, 소규모 서버 시스템에서 대규모 인터넷 서버 및 슈퍼컴퓨터에 이르기까지 다양한 영역에서 사용가능하다. 더욱이 인터넷 접속, 인터넷 접속, 원격서버 접속이 쉽게 구현될 수 있다 [1,4-6].

InfiniBand 연결망은 스위치 기반의 비점형 연결망으로 종단에 프로세서 노드, 입출력 노드가 연결되는 여러개의 서브넷 (Subnet)으로 구성된다. 서브넷은 종단노드, 스위치, 라우터로 구성되며 서브넷간의 연결은 라우터를 통한다. 하나의 서브넷에는 최대 65,536개의 종단 노드가 연결가능하며, 각 종단 노드는 InfiniBand 연결망 접속을 위한 채널 어댑터를 가지며, 프로세서 노드쪽은 HCA(Host Channel Adapter), 입출력 처리노드 및 장치 쪽에서는 TCA(Target Channel Adapter)가 사용된다. InfiniBand 링크는 양방향 점대점 통신 채널로 되어있어, 1x, 4x, 12x 의 독립적인 양방향 직렬연결 포트를 가지며, 2.5Gbps에서 최대 30Gbps 의 데이터 전송 속도를 지원한다 [3,9-11].

표 1. 고속 입출력 규격비교.

Feature	InfiniBand™	PCI-X	1Gb & 10Gb Ethernet	Hyper-Transport™	Rapid I/O
Bus/Link Bandwidth	2.5, 10, 30Gb/s*	8.51 Gb/s	1Gb, 10Gb	12.8, 25.6, 51.2Gb/s†	16, 32Gb/s*
Bus/Link Bandwidth (Full Duplex)	5, 20, 60Gb/s*	Half Duplex	2Gb, 20Gb	25.6, 51.2, 102Gb/s†	32, 64Gb/s*
Pin Count	4, 16, 48*	90	4(CbE), 8(10GbE-XAU)	55,103,197†	40,76*
Transport Media	PCB, Copper & Fiber	PCB Only	PCB, Copper & Fiber	PCB Only	PCB Only
Max Signal Length PCB/Copper Cable	30in, 17in	inches	20in, 17m	inches	inches
Max Packet Payload	4KB	Not Packet Based	2KB	64 bytes	256 bytes

2. Mellanox사 InfiniBand 네트워크 어댑터

InfiniBand 네트워크를 구성하기 위한 요소로 앞서 언급한바와 같이 스위치, 노드, 망관리 S/W 등의 존재하지만, 본 절에서는 노드와 노드를 연결하기위한 네트워크 어댑터 중 관련기술을 선도하고 있는 Mellanox사의 InfiniBand 네트워크 어댑터에 대해 설명한다. Mellanox사의 InfiniHost Architecture는 Port당 10Gbps (InfiniBand 4x)의 대역을 가지는 구조를 가지며, Modular Server I/O 연결 및 Physical Layer의 Serializer/Deserializer(Ser/Des) Interface를 포함한다. InfiniHost는 Bus Interface 방식에 따라 PCI-X 및 PCI-Express의 형태로 적용되어 있으며, System Bus와 External Serial 데이터 전송을 위해 DDR RAM이 적용되어 있다 [7-8].

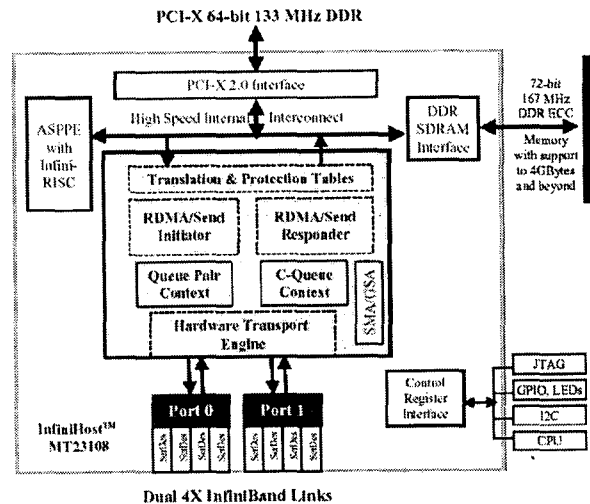


그림 2. InfiniHost Block Diagram.

Mellanox사는 기존 PCI-X의 내부 Interface를 가지는 Dual Ports 어댑터에서부터, 효율적인 고속의 데이터 처리를 위한 PCI-Express 8x 제품을 발표하였으며, 그림 3. 에서와 같이 기존 Ethernet 망에서는 두 System Bus가 비슷한 성능을 내었으나, InfiniBand 네트워크에서는 확연한 차이를 나타내고 있다 [7].

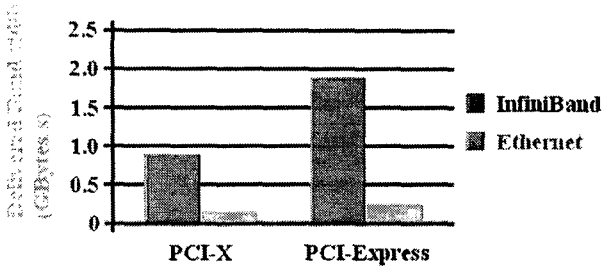


그림 3. Bandwidth on PCI-X and PCI-Express.

III. High Speed Network Adapter for DCP

1. 12x 네트워크 어댑터 구조

앞에서와 같이 High Speed Serial Data 전송 및 고속 I/O 네트워크 구성을 위해서는 외부적인 전송속도나 대역폭도 중요하지만, Local Interface의 적용에 따라 현격한 성능차이를 확인할 수 있었다. 채널기반의 12x 네트워크 어댑터는 외부적으로 2.5Gbps, 12채널인 60Gbps(Full Duplex)의 데이터 전송을 위한 구조와 Memory free 기능구현을 위한 PCI-Express 16x의 Local Interface 적용을 위한 설계이다. 이러한 구조는 기존 InfiniBand 4x 네트워크 어댑터의 성능보다 3배가량 향상된 전송환경이며, 현재 FPGA Level에서 채널기반 입출력 인터페이스 칩셋 개발의 설계 및 검증과 Device Driver 설계를 통한 네트워크 구성, 망관리 응용 S/W의 적용 및 테스트가 진행중이며, 최종적으로 DCP에 적용가능한 제품형태를 목표로 하고있다. 또한, InfiniBand 네트워크 전용 Switch의 4x to 12x Auto Sensing 기능을 통해, 4x의 타기종 네트워크 어댑터가 적용된 노드들과의 네트워크를 구성하여, 호환성검증 및 신뢰성 있는 데이터 전송을 검증할 수 있는 환경이 구축되어있다. 그림 4.는 현재 개발완료된 12x 네트워크 어댑터이다.

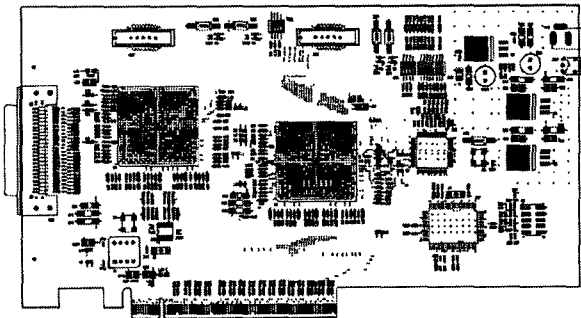


그림 4. 12Channel High Speed Serial Network Adapter.

2. 네트워크 어댑터 세부구조

어댑터에 적용된 두개의 FPGA는 별도의 Processor가 존재하지 않으며, 각각 내장된 20채널의 Serializer/Deserializer (Ser/Des)를 InfiniBand 12x Port와 PCI-Express 16x Interface를 위해 사용된다. 그림 5.는 12x 네트워크 어댑터의 구조이며, 어댑터의 내부 구성요소 및 세부구조는 다음과 같다.

- ALTERA Stratix GX EP1SGX40G : PCI-Express IP 적용 및 채널기반형 입출력 칩셋설계 적용
- Local Interface : PCI-Express 16x
- External Port : 12x MicroGigaCN

- Mictor32 Connector : Logic Analyzing 을 위한 검증 Port
- JTAG : CPLD Configuration
- Clock : 50Mhz, 156.25Mhz
- Miscellaneous Features : Local Reset, Mechanical Switches, User LEDs
- Power : 12V, 3.3V, 1.5V DC

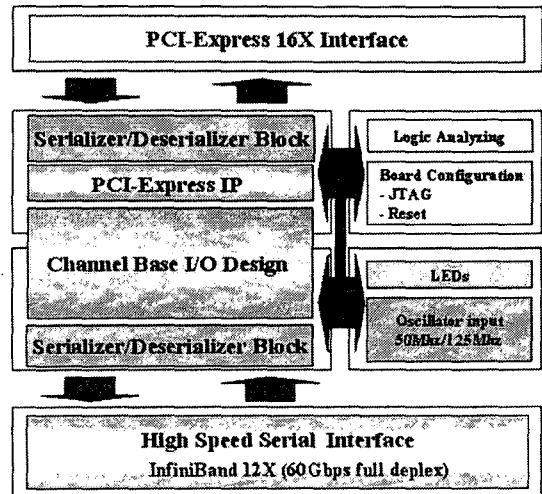


그림 5. 12Channel Network Adapter Block Diagram.

PCI-Express Interface는 각 채널당 “PETp, PETn” 과 “PERp, PERn” 이라는 Differential pair의 Lane Signal 을 가지며, 한 쌍의 Tx, Rx Differential Signal 은 각각 고속 Serial Data를 의미한다. 외부 Port 역시 Differential pair로 존재하는 각각의 Tx, Rx Lane이 존재하며, “MicroGigaCN” 이라는 전용 Board Connector와 12x Cable 을 통해 Switch 및 타 노드와의 네트워크를 구성할 수 있으며, 최대 17미터 까지 Data의 신뢰성을 보장할 수 있다. PCI-Express 와 InfiniBand 는 모두 고속 Serial Data 전송 규격을 갖는다는 공통점 외에 구조적으로 많은 유사점을 가지고 있다. 그러나, 어느하나가 다른하나를 대체하기 보다 서로 내부 System Bus 와 외부 I/O 의 성능향상을 통한 상호 보완적인 관계라 할 수 있다. 두개의 ALTERA FPGA는 각 PCI-Express 와 InfiniBand Interface를 담당하며, Ser/Des Block, PCI-Express IP, InfiniBand Design Block, 내,외부 Interface Block 등이 주요한 칩내부 구성이다. 그 외에, 설계된 Logic을 물리적으로 확인해 보기위한 Analyzing Port 와 Board Configuration Part, 고속 Serial Block의 동작을위한 Gigalink quads clock 등이 적용되어 있다. 본 어댑터의 핵심 설계부분인 고속 Serial 네트워크 인터페이스 칩셋의 구조는 그림 6과 같이 Transport Layer, Network Layer, Link Layer, Physical Layer로 구성되어 있는 프로토콜 스택을 가진다. 각각의 독립적인 송,수신부를 통해 데이터가 전송되는 구조를 가지며, 고속 Serial 데이터를 처리하기 위한 Ser/Des 부를 통해, Physical, Link Layer 에서 Packet 상태가 확인된다. Network Layer에서는 Subnet Manager 를 통해 전체 네트워크의 구성을 확인하며, Transport Layer와 System I/O 를 통해, 응용 Software 를 통한 네트워크 구성 및 관리가 이루어진다.

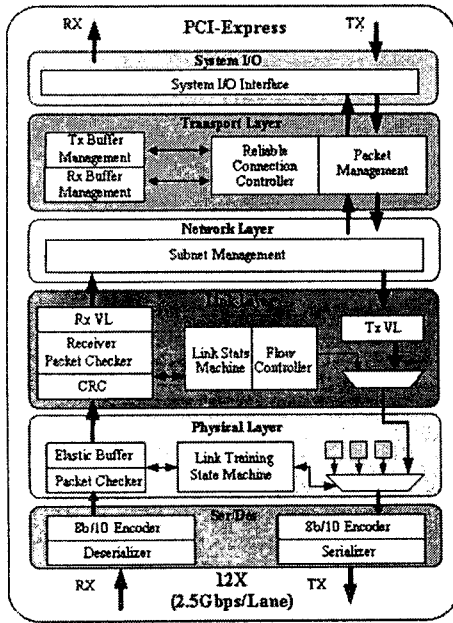


그림 6. High Speed Network Interface Chipset.

IV. 결론

본 논문은 현재 개발진행중인 DCP의 고속 네트워크 환경구현을 위해 FPGA Level에서 설계된 InfiniBand 네트워크 어댑터의 구조를 설명하고, 기존 I/O 네트워크 및 네트워크 어댑터와 어떠한 구조적 차이를 가지는지를 알아보는 것이었다. 현재 어댑터의 성능분석 보다 기능구현에 개발의 초점을 두고 있으며, InfiniBand 규격에 맞는 설계의 적용 및 검증과 Driver 개발 등에 집중할 예정이다. 최종 개발완료될 SoC 형태의 칩셋 적용을 위해 많은 테스트 및 결과분석을 통해, 안정적인 기능구현 및 네트워크 성능향상을 목표로 하고 있다. 또한, 상용화 관점에서의 DCP 개발을 위해, on Board 형태의 Platform을 개발할 예정이다.



전 기 만
 2000년 한양대학교 전기공학과 졸업. 2000년~2001년 삼보컴퓨터 기술연구소. 2001년~현재 전자부품연구원 지능형정보시스템연구센터 전임연구원.
 관심분야 : System Hardware, Storage System.



정 하 중
 1984년 한양대학교 전자공학과 졸업. 1987년 한양대학교 전자공학과 석사. 1989년~1993년 금성사 정보기기연구소. 1993년~현재 전자부품연구원 지능형정보시스템연구센터 수석연구원.
 관심분야 : Smart Home, U-Health Care.

참고문헌

- [1] 박경, 모상만, "InfiniBand: 차세대 시스템연결망", 정보과학회지, vol. 19, no.3, pp. 43-51, 2001. 7
- [2] J. Wu, P. Wyckoff, and D.K. Panda. "PVFS over InfiniBand: Design and Performance Evaluation", *In ICPP*, 2003.
- [3] WP010800170, <http://www.mellanox.com>
- [4] A. Cohen, "A Performance analysis of the socket direct protocol (SDP) with asynchronous I/O over 4X InfiniBand", 2004 IEEE International Conference, pp. 241-246, April 2004.
- [5] P. Balaji, P.Shivam, P.Wyckoff, and D.K. Panda. "High Performance User Level Socket over Gigabit Ethernet", In Cluster Computing, September 2002.
- [6] "Get on the Fabric : InfiniBand Fabric Prototype Demonstration White paper from Fall 2000 IDF", Technical white paper from Intel, http://download.intel.com/design/server/future_server_io/documents/get_on_fabric.pdf, Aug. 2000.
- [7] 2229WP, <http://www.mellanox.com>
- [8] S.N. Damianakis, C. Dubnicji, and E.W. Felten. "Stream Sockets on SHRIMP", In Lecture Notes in Computer Science 1199, pp.16-30, 1997.
- [9] T. von Eiken, A. Basu, V.Buch, and W. Vogels. "U-Net: A User-Level Network Interface for parallel and Distributed Computing". In Proceedings of the ACM Symposium on Operating Systems Principles, PP. 40-53, December 1995.
- [10] T. von Eiken, A. Basu, V.Buch, and W. Vogels. "U-Net: A User-Level Network Interface for parallel and Distributed Computing". In Proceedings of the ACM Symposium on Operating Systems Principles, PP. 40-53, December 1995.
- [11] P. Balaji, S. Narravula, K. Vaidyanathan, S. Krishnamoorthy, J. Wu, and D. K. Panda. "Socket Direct Protocol over infiniband: Is it Beneficial?" Technical Report OSU-CISRC-10/03-TR54, The Ohio State University, 2003.



박 창 원
 1986년 중앙대학교 전자공학과 졸업. 2002년 광운대학교 전자공학과 석사. 2006년 성균관대학교 정보통신공학 박사수료.
 1985~1988년 동양정밀 주임연구원. 1988년~1993년 효성컴퓨터 선임연구원. 1993년~현재 전자부품연구원 지능형정보시스템연구센터 수석연구원.
 관심분야 : Digital Convergence System, U-Health Care.