PCI Express 물리계층의 IP 설계

권영민, 성광수
영남대학교 전자공학과
Email : youngmin@yumain.ac.kr, kssung@yu.ac.kr

Design of PCI Express Physical Layer IP

*Young-min Kwon, Kwang-soo sung
Yeungnam University Electrical engineering & Computer science

Abstract

In this paper, we propose design of PCI Express Physical Layer for IP. The proposed design is compatible with PCI Express Base specification Revision 1.0a. and supports only single Lane. The best feature of this design is that Physical Layer includes Power Management block. Therefor, the entire design of PCI Express component is simplified

In the near future, as optimizing this design and extending Lane, we will redesign Physical Layer.

I. 서론

지난 91년 처음 등장한 PCI 버스는 10년 이상 PC의 표준 버스로 사용되었다. 그러나 인터넷 시대에 접어들면서 폭발적으로 증가하는 데이터 전송량을 처리하기 위해서 보다 넓은 대역폭의 버스 아키텍처가 필요하게 되었다. 대역폭의 문제점을 국소적으로 해결하기 위해 AGP, PCI-X와 같은 것이 발표되었으나 이들 역시 여러 I/O 디바이스가 I/O 버스를 공유(multi-drop)하는 병렬버스(parallel bus technology)구조로 되어 있어 성능향상의 한계를 가지고 있다.

이러한 PCI 버스의 한계를 극복하기 위해 PCI SIG에서 기존의 PCI를 계승하는 새로운 I/O 표준으로 PCI Express를 발표하였다. 기존의 병렬 전송을 하는 PCI가 133MB/s의 전송 속도를 가진 것에 반해, PCI Express는 point-to-point 방식을 이용한 직렬 전송 방식으로 2.5GB/s의 전송 속도를 가진다. PCI Express의 기본 아키텍처는 최하위 계층인 Physical Layer에서 교체되기 때문에 기존의 PCI 장치에 사용된 운영체제와 디바이스 드라이버 소프트웨어를 그대로 사용할 수 있다. 이와 같은 장점으로 인해 PCI Express가 차세대 컴퓨터 I/O 시스템의 표준으로 자리잡을 것으로 보인다.

본 논문에서는 PCI Express의 최하위 계층인 Physical Layer를 소개하고, 효과적인 PCI Express Physical Layer의 IP 설계를 제안 하고자 한다.

II. PCI Express의 특징

II-1. PCI Express 물리계층의 특징

PCI Express는 그림1과 같이 전송계층, 데이터 링크 계층 그리고 물리계층으로 구성되어 있다.

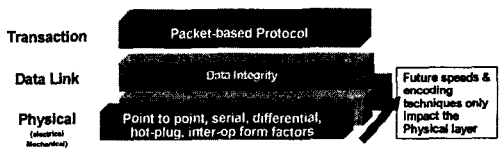

그림 1 PCI Express의 계층구조

PCI Express의 최하위 계층인 물리계층은 링크를 통한 직렬 전송을 담당한다. PCI Express의 링크는 한 쌍의 Differential signal pair로 이루어진 레인으로 구성되며 최대 x32의 레인을 지원함으로써 대역폭을 선형적으로 증가시킬 수 있다. 링크의 초기화 과정에서 컴포넌트 상호간의 레인 수, 전송속도 등을 하드웨어 자율적으로 결정한다.

물리계층은 구조적으로 논리적 블럭과 전기적 기능 블럭으로 구성된다. 논리적 블럭은 물리계층 기능의 제어와 관리를 담당하고, 전기적 기능 블록은 링크를 통해 데이터를 직렬전송하기 위한 전기적 특성들을 담당한다. 본 논문에서는 PCI Express 물리계층의 IP 설계를 목적으로 논리적 블록에 중점을 두고 있다.

논리적 블럭은 심볼 인코딩, 프레이밍, 데이터 스크램블링, 그리고 LTSSM(Link Training and Statue Machine)과 같은 기능을 제공한다.
PCI Express는 심볼 인코딩을 위해 8b/10b 전송 코드를 사용한다. 이 전송코드는 IEEE 802.3에 정의된 것과 동일하다.

프레이밍 메카니즘은 DLLP(데이터 링크 Packet)와 TLP(물리계층 Packet)의 시작을 나타내기 위해 각각 SDP(Special 심볼 K28.5)와 STP(Special 심볼 K27.7)을 사용한다. END(special 심볼 K29.7)로서 TLP와 DLLP의 끝을 알린다.

전송하는 데이터가 특정한 패턴을 가지지 않도록 레인당 하나의 LFSR(Linear Feedback Shift Register)를 사용해서 8bit Data를 스크램블링한다. 특정한 의미를 갖는 Special 심볼은 스크램블링되지 않는다.

그림 2는 PCI Express의 LTSSM이다.

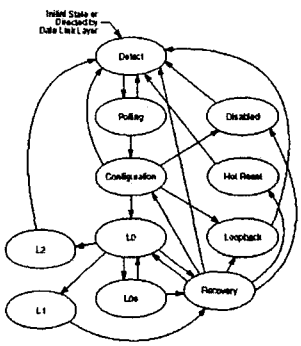

그림 2 Main State Diagram for Link Training and Status State Machine

LTSSM은 링크를 training하는 부분과 Power management, 에러 보정 , 테스트 그리고 링크 power management를 위한 상태로 구성되어 있다. Detect, Polling, Configuration을 거쳐 링크를 training하고 정상적인 동작 상태인 L0 상태로 이동한다. 링크 Power management를 위한 L0s, L1, L2 상태는 L0s에서 L2로 갈수록 더 많은 power를 절약한다. L0 상태에서 링크 에러가 발생할 경우 Recovery 상태로 이동해 링크를 retraining한다. Loopback 상태는 테스트를 위해 Loopback 마스터나 슬레이브로 동작하기 위한 상태이고 Disable과 Hot Rest 상태는 각각 링크를 disable 하거나 링크를 통해 물리계층을 reset하는 상태이다.

II-2 PCI Express Power Management

PCI Express의 Power Management는 기존의 PCI Power Management와 호환 되며, ASPM(Active State Power Management)을 지원한다. PCI Express Power Management가 소프트웨어에 의해 제어되는 것에 반해 ASPM은 하드웨어 자율적으로 power를 절약하는 메카니즘이다.

그림 3은 PCI Express Physical Link Power Management 상태도이다. PCI Express Physical Link Power Management 상태도는 PCI Power Management와 ASPM에 의해 제어된다.

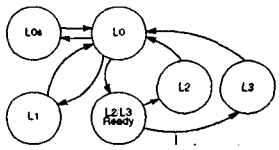

그림 3 PCI Express Physical Link Power Management 상태도

III. 제안된 Physical Layer IP 설계

그림 4는 본 논문에서 제안한 물리계층의 Top 블럭이다.

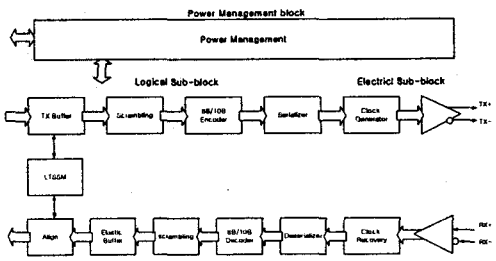

그림 4 물리계층 Top 블럭

제안된 물리계층은 논리적 하위블럭, 전기적 하위블럭 그리고 Power Management 블럭으로 구성되어 있다. 각각의 블럭은 PCI Express Base Specification Revision 1.0a를 기준으로 설계되었다.

물리계층 논리적 하위블럭은 기능적으로 송신단과 수신단으로 구분 지을 수 있다. LTSSM은 링크를 통한 데이터 전송과 power Management 블럭의 제어를 담당한다.

그림 5는 논리적 하위블럭의 수신단이다.

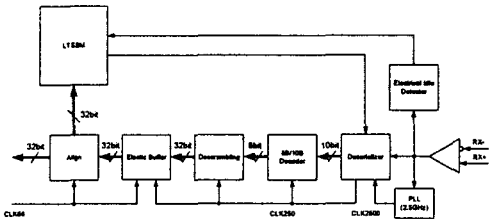

그림 5 수신단 Top 블럭

수신단은 구조적으로 Deserializer, 8b/10b 인코더, 디스크램블링, Elastic 버퍼 그리고 Align 블럭으로 이루어져 있다.

링크로부터 전송 받은 데이터에서 2.5Ghz의 클럭을 추출하고, 2.5Ghz의 clock을 이용하여 250Mhz의 클럭을 생성한다. 2.5Ghz의 클럭과 250Mhz의 클럭을 이용하여 병렬화고, 10bit의 병렬화된 데이터를 8b/10b 인코더를 사용하여 8bit 데이터와 제어 bit으로 변환한다. 8bit 데이터가 하나의 심볼 단위이고 제어 bit은 심볼이 일반 데이터인지 특수한 목적을 가진 special 심볼인지 구분한다.

디스크램블링 블럭에서는 250Mhz의 클럭만을 사용하기 위해 그림 6의 병렬 LFSR을 사용하여 디스크램블링을 처리한다. 그림 5의 병렬 LFSR은 LFSR의 값을 예측하여 1 클럭에 하나의 심볼을 스크램블링 할 수 있도록 만든 로직이다.
또한, 디스램블링 블럭은 이후 단 설계의 유연성을 위해 4개의 심볼을 모아서 Elastic 버퍼에 저장한다.

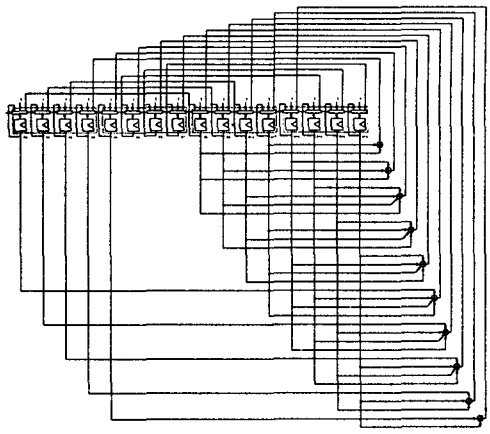

그림 6 병렬 LFSR

Elastic 버퍼는 직렬 데이터에서 추출한 클럭과 내부 클럭 사이의 클럭 오차를 보상해서 항상 일정한 데이터 비율을 유지하는 버퍼이다. 일정한 주기로 전송되는 SKP ordered set(하나의 COM과 3개의 SKP special 심볼로 구성된다.)을 이용하여 Elastic 버퍼를 제어한다. SKP ordered set에서 하나 또는 두 개의 SKP 심볼을 빼거나 더해서 Elastic 버퍼에 채우게 된다. 그림 7은 Elastic 버퍼 관리의 예제이다.

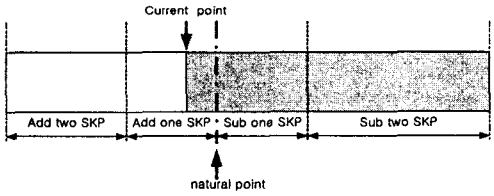

그림 7 Elastic 버퍼 관리

본 논문에서 제시된 Elastic 버퍼는 read 클럭과 write 클럭이 다르고 4개의 심볼씩 Elastic 버퍼에 저장하기 때문에 제어하는 타이밍에 따라 Elastic 버퍼의 상태가 다르게 보인다. 따라서 natural point를 기준으로 일정 영역을 나누어 Elastic 버퍼를 제어한다.

Align 블럭은 수신단으로부터 전송 받은 데이터를 실시간으로 처리해서 데이터 링크 계층이나 LTSSM으로 전송한다. 그림 8은 본 논문에서 제시한 TLP와 DLLP format이다.

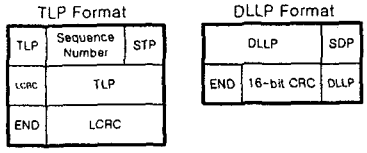

그림 8 TLP & DLLP Align format

그림 8과 같이 프레이밍을 제거하지 않고 데이터 링크 계층로 DLLP와 TLP를 전송함으로서 TLP와 DLLP의 시작과 끝을 따로 알려줄 필요가 없다. 따라서 물리계층의 설계를 좀더 간략히 할 수 있다.
TLP와 DLLP 이외의 ordered set이나 다른 심볼은 align 블럭에서 디코딩해서 LTSSM으로 알려주게 된다.

그림 9은 논리적 하위블럭의 송신단이다.

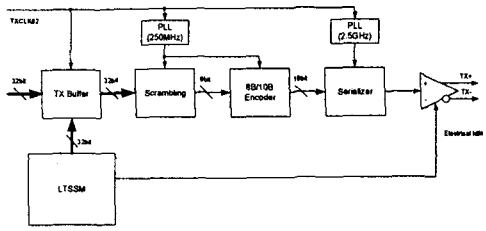

그림 9 송신단 Top 블럭

송신단은 구조적으로 송신 버퍼, 스크램블링, 8b/10b 인코더 그리고 serializer로 이루어져 있다.

송신 버퍼는 상위 계층으로부터 받은 데이터와 LTSSM으로부터 받은 데이터 중 한쪽을 선택해서 스크램블링 블럭으로 전송한다. 스크램블링 블럭 이후단은 수신단의 역순으로 데이터를 처리한다.

본 논문에서 제안된 물리계층은 하나의 레인을 만을 지원하게 설계되었다. 하나의 레인만을 지원하는 물리계층의 경우 복수개의 레인을 지원하는 경우와는 달리 모든 패킷을 하나의 레인에서 직렬 전송한다. 따라서 프레이밍 위치에 대한 규칙을 적용 받지 않는다. 수신단에서와 같이 프레이밍을 물리계층에서 하지 않고 프레이밍된 데이터를 상위 계층에서 전송 받음으로써 물리계층의 설계를 간략히 할 수 있다. 상위 계층의 설계에서 TLP와 DLLP의 시작과 끝을 알려줄 필요가 없으므로 상위 계층의 복잡성이 증가되지는 않는다.

본 논문이 제시한 물리계층의 IP 설계 특징 중 하나는 물리계층에 Power Management 블럭을 포함하여 LTSSM이 power 관리를 수행한다. 기본적으로 LTSSM은 Link Power Management를 위해 L0, L0s, L1, L2와 같은 상태를 가지고 있다. 여기에 추가 적으로 디바이스 power 관리에 대한 제어를 LTSSM에서 담당함으로써 중복된 기능을 가지는 블럭을 줄이고 PCI Express 전체의 설계를 보다 간략히 할 수 있다.

VI. 결론 및 추후 연구

본 논문이 제한한 Physical Layer는 PCI Express Base Specification Revision 1.0a의 물리계층과 Power Management 블럭만을 고려해서 설계되었다.
향후 전송계층과 데이터 링크 계층의 설계가 완료되면 전체 PCI Express 컴포넌트의 최적화를 위해 설계 변경이 필요할 것으로 보인다.
그리고 좀더 높은 대역폭의 지원을 위해 추후 설계에서는 복수개의 레인을 지원할 예정이다.

참고문헌(또는 Reference)

[1] PCI-SIG PCI Express Base Specification Revision 1.0a
[2] PCI-SIG PCI Local Bus Specification Revision 2.3
[3] PCI-SIG PCI Bus Power Management Interface Specification Revision 1.1
[4] PCI-SIG Errata for the PCI Express Base Specification Revision 1.0a
[5] PCI-SIG Endpoint Compliance Checklist for PCI Express Base 1.0a Specification Revision 1.0
[6] 정호성 microsoftware 2003. 3. p.230 - p.231
[7] 백현기 microsoftware 2003. 3. p245 - p.247