Technical report

DMP Processing Architectures based on FPGA and PCIe

Item, NTNU

Version: Draft 1.03

Project: Collaboration Surfaces

Author: Leif Arne Rønningen

Date: 30Aug11

File: PCIeArch2

The DMP Architecture

The Distributed Multimedia Plays architecture (DMP) is a telepresence system of the future, which shall provide near-natural virtual networked auto-stereoscopic continuous (multi-) view video and multichannel sound collaboration between distributed located users (and users and servers [RON11], [RON07]. The end-to-end maximum time delay is guaranteed (< 10-30 ms) by allowing the quality of audio-visual content and the scene composition vary with time and traffic. The sequences of control packets are also guaranteed. The adaptation scheme is called Quality Shaping. The scheme uses traffic classes, controlled dropping of packets according to content, traffic measurements, forecasting of traffic and feedback control. Adaptive parameters are the maximum end-to-end delay, the number of 3D scene sub-objects, object- temporal and spatial resolution, sub-object adaptive and scalable compression, and the number of spatial views. The scheme also includes scene object behaviour analysis. The architecture supports pertinent security and graceful degradation of quality. The architecture introduces independent parallelism by definition. The most basic critical part of DMP, the regeneration of objects from sub-objects by interpolation when sub-objects are dropped in the network, has been extensively tested. Very good interpolation and segmentation methods has been developed and presented in a number of Master's theses and PhD papers at the Dept. of telematics, e.g., [RON09].

Collaboration Surfaces and Spaces

A Collaboration Surface is a technical interface with cameras, displays, microphones, load-speakers and other built-in devices. One or more Collaboration Surfaces combined with people, animals, input devices, natural scenes or scenography, constitute a Collaboration Space, CS. A CS can have a planned, transient or stationary stochastic behavior. The objects of the space can be more or less dynamic or static.

The aim is to develop and test the quality of a modular, quality-enhanced (beyond state-of-the-art) CS. This includes experiments to evaluate the Quality of Experience and establish user requirements and system quality specifications when applying the CS in virtual music, musical theatre, arts and other media. Moreover, the CS will be applied for new gaming concepts, to obtain real feeling by using stereoscopy in serious games, and to enhance the satisfaction of First Person Shooter (FPS) games. The 'Collaboration Surface Project' is funded by the NTNU/NFR AVIT program.

The AppTraNetLFC protocol

A new protocol called the AppTraNetLFC protocol [RON11] is an extension of the AppTraNet protocol described in [RON07], and combines the application layer, transport layer,network layer and partly link layer protocols into one protocol, with only one common header. The protocol handles the setup and management of collaborative distributed scenes and content transfer with adaptive quality. A priority queuing hardware mechanism is used in network nodes to handle controlled dropping and delay of content packets and guaranteed transport

(but not delay) of control packets. The dropping mechanism has been verified by simulation [RON06].

Generic Processing Architecture

This architectural study is based on the CCAB-01 FPGA board developed by Pico Computing and Dept. of telematics, NTNU, 20-slot PCIe Gen 2 Backplane with switching, SSDs with PCIe, the Xilinx m605 FPGA board with mezzanine FMC boards with SFP+ I/O for 10G Ethernet fiber transmission. In another study [RON11a], a functional structure for handling the Camera Cluster Array (CCA) is presented, and in this study we will show how these functions and the DMP Quality Shaper [RON11] can be allocated to physical boards and FPGAs. First, a short review of PCIe is given.

PCI Express

A short introduction to PCIe can be found in [PLX11] and a good textbook is PCI Express System Architecture [BUD10].

PCIe is the latest generation I/O bus for interconnecting peripheral devices in ICT systems. Figure 1 shows a basic architecture of a high-end system.

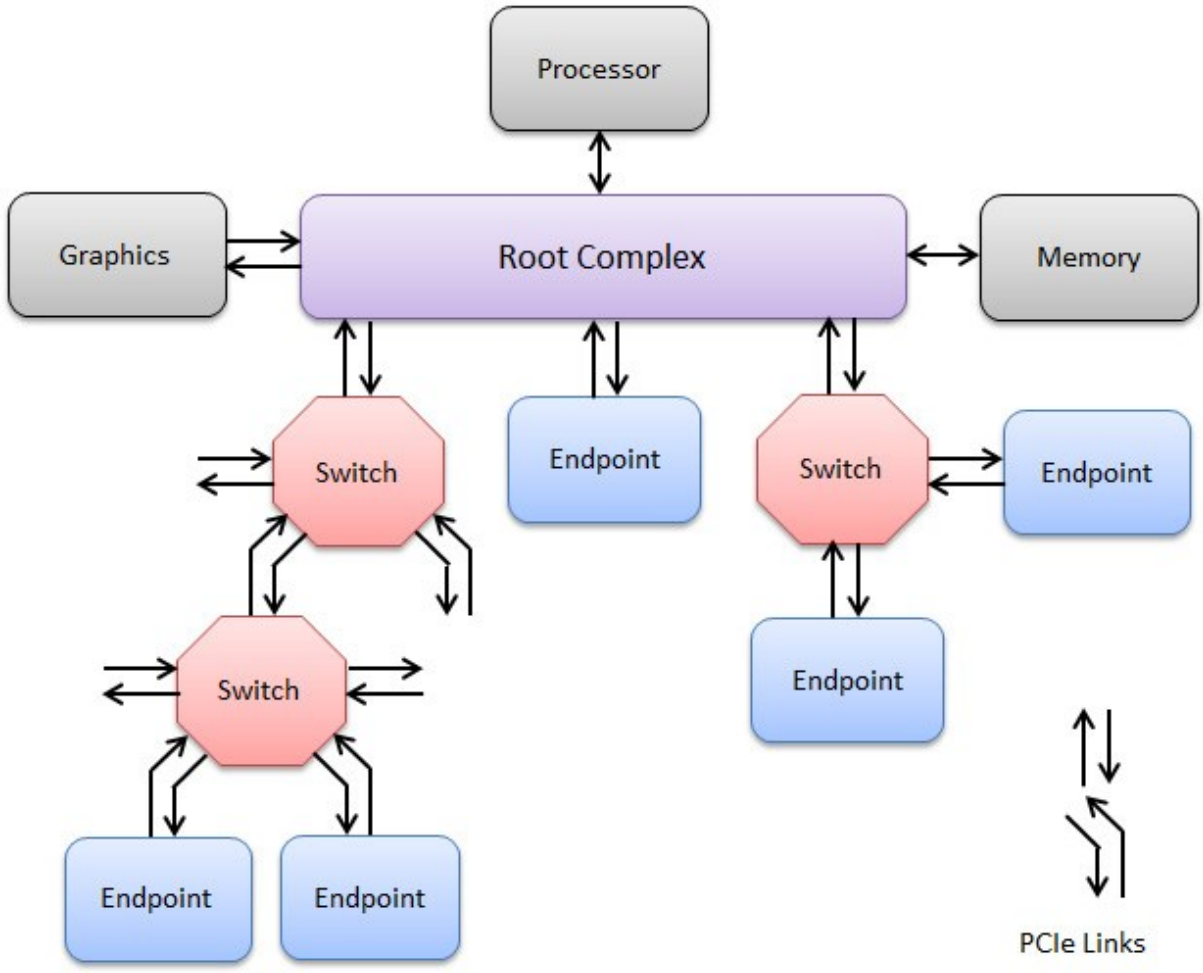

Figure 1. Basic PCIe Architecture

In Figure 1, three basic building blocks are shown. Root Complexes and Endpoints transfer data between each other via Switches, using packets and three protocol layers. Root Complexes and Endpoints can both be Requesters or Completers. The Root Complex is responsible for configuring the physical system. The layers are denoted Transaction layer, Data Link layer and Physical layer. Above the Transaction layer some system- or application software initiates and ends transactions.

Physical layer

This layer provides a point-to-point, dual simplex, differential signaling link between two nodes. The data rates are 5.0Gbits/sec/lane in each direction for PCIe 2.0, and 2.5Gbits/sec/lane for PCIe 1.0. The number of lanes can be x1, x2, x4, x8, x16 and x32. Byte Striping and Scrambling is provided, and the coding is 8b/10b [BUD10].

Data Link layer

The PCIe Link layer supports window flow control with Ack, Nak and other packets, and power management.

Transaction layer

There are eight transactions types (four categories), which can be posted or non-posted:

- Memory Read
- Memory Write
- Memory Read Lock
- IO Read
- IO Write
- Configuration Read
- Configuration Write
- Message

Types Memory Write and Message are posted, meaning that when a Requester sends a packet (TLP) no response is sent from the Completer. The other types are non-posted, that is, the Completer sends a completion notification to inform that a packet has been received successfully.

The Transaction Layer supports Address Routing, ID Routing and Implicit Routing, see [BUD10].

PCIe as applied in DMP

In DMP,TLP packets, Posted Message with Data (MsgD), Address Routing with 32 bits (maybe 64) address included in the packet header, are used. This means that no Ack or Nak is sent by the PCIe Link Layer. The Link Layer just adds/removes and checks the Start, End and LCRC parameters in the DMP application.

Figure 2 shows a Message Request and the TLP framing and Header. TLPs with 32-bit address are denoted 3DW. When the Type field indicates Address Routing, an Endpoint checks the Address field with its Base Address Registers (BAR), and decides to receive or reject the incoming packet. A Switch makes two checks, when the Type field indicates Address Routing, it compares the header Address with its two BARs (configured, stored). If the Address is in the range given by the two BAR addresses, the Switch accepts and forwards the packet. If not, other checks are performed, see [BUD10].

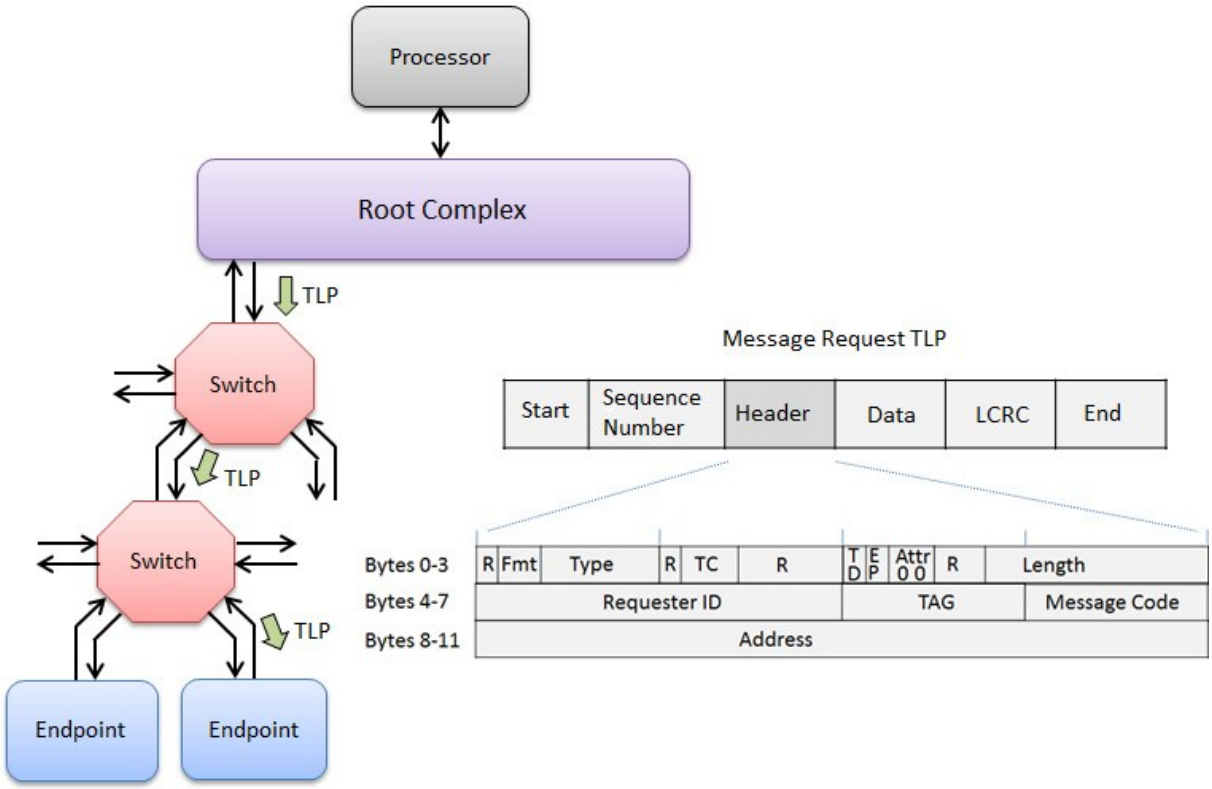

Figure2. Message Request, TLP Header for Address Routing

Processing Architecture, DMP Collaboration Surface, Camera

The Camera Cluster Array (CCA) feasibilities and functionality are described in [RON11] and [RON11a]. Here, the hardware processing architecture as shown in Figure 3 is focused. Each camera outputs image data at 0.8 Gbit/sec, that is, 10 bits parallel at 80 MHz on a 22-lead FFC bus. Three FCC-busses connects three cameras to a processing board, CCAB01, with a Xilinx Vertex 6 FPGA onboard. The output from the CCAB01 is a PCIe x4 generation 1.0 bus, which can transfer date at 2.5 Gbit/sec/lane, overhead included, and 2.0 Gbit/sec/lane of content data, giving 8.0 Gbit/sec in total. Three cameras give 2.4 Gbit/sec, meaning that the PCIe is not a bottleneck in this configuration. In theory, the PCIe x4 gen. 1 bus could handle nine cameras, but it is not known yet how many cameras the FPGA onboard the CCAB01 has capacity to process. The present version has inputs for four cameras. A maximum array of five clusters (45 cameras) can be handled by the chosen backplane, utilizing fifteen PCIe x4 inputs. The Backplane can take 17 PCIe x4 inputs maximum. It is assumed that one Xilinx ml605 board implements IPsec and a small router, and carries an FMC Mezzanine board for a 10G Ethernet long haul fiber. The small router could be implemented by means of the free SFP+ Tranceiver of the FMC board. With a fiber this is connected to a corresponding board of the Display Backplane. The fiber link towards the Access Node is now the bottleneck of the system, since the PCIe x8 generation 1.0 of ml605 can carry 18 Gbit/sec. But note that a full configuration of five clusters will give a data rate of 36 Gbit/sec, which is the double of the maximum input rate for ml605. There are at least two solutions to this. There are slots for two ml605s in the backplane. In this case the Quality Shaping mechanism described in [RON11] can be implemented on the ml605 only, reducing the data rate from 36 Gbit/sec to 20 Gbit/sec, load shared by two ml605s and two fibers. Another solution is to use only one ml605 and implement dropping of content packets in all CCAB01s

and reduce the data rate to 10 Gbit/sec into ml605. The buffering and flow control of control packets have to be implemented in the ml605. If desirable, the dropping of packets can be implemented both in CCAB01 and in ml605.

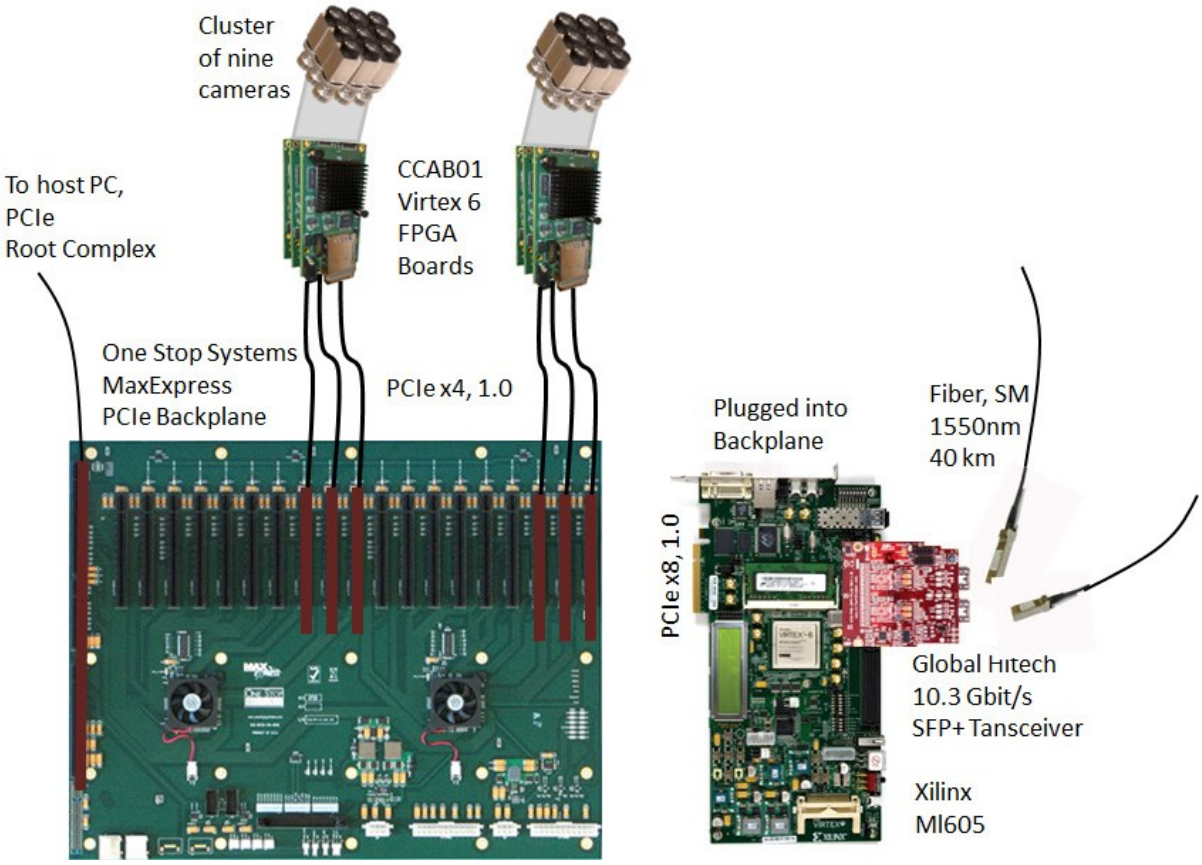

Figure3. Hardware components for DMP Collaboration Surface, Camera Cluster Array Processing

Processing Architecture, DMP Collaboration Surface, Display

The receiver and display feasibilities and functionality are described in [RON11] and [RON11a]. Figure 4 reuses the hardware components from the camera part in Figure 3, only the camera clusters are substituted with display units (TBD). The internal functionality of the FPGAs of course is different. The data rates for busses are the same as in Figure 3.

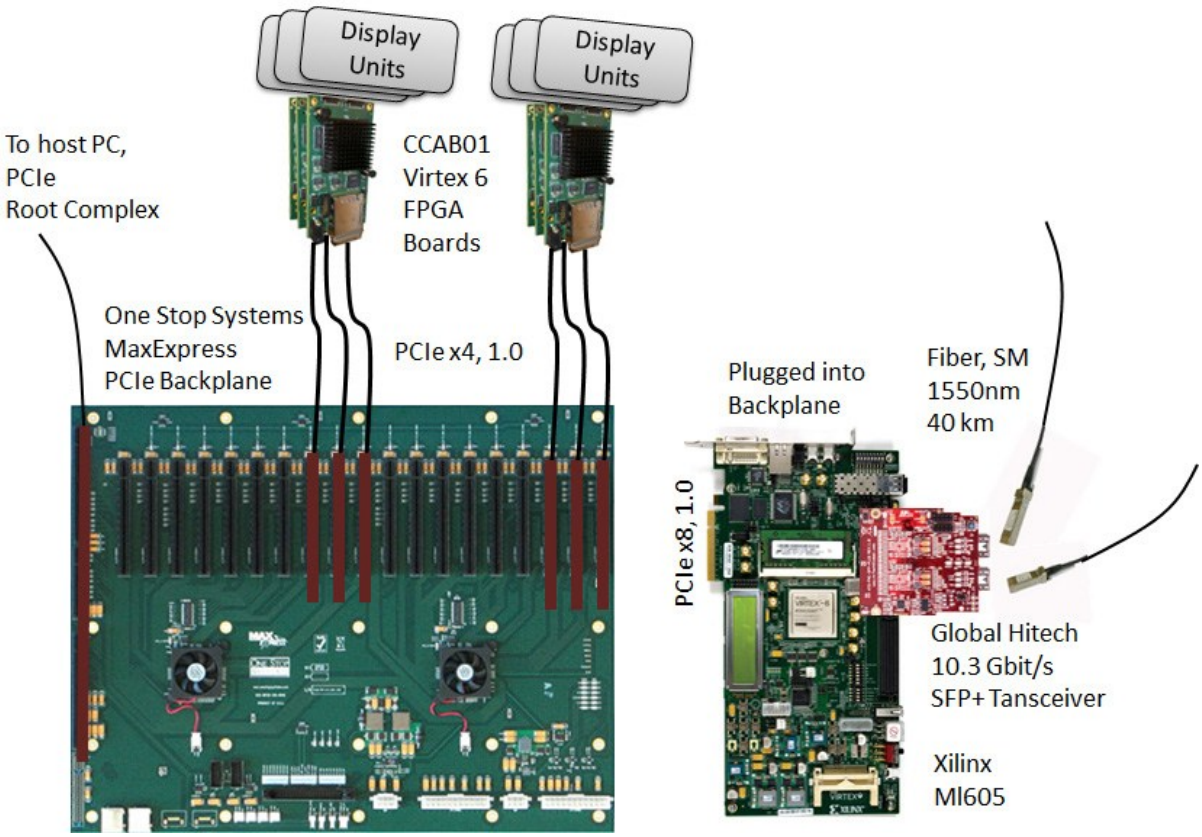

Figure4. Hardware components for DMP Collaboration Surface, Display Processing

Processing Architecture, DMP Network Node

Figure 5 shows a general network node of DMP. Access Node Servers are not shown, but can be implemented in different ways. Tasks without real-time requirements can be accomplished by the host PC, while the ml605 performs real-time processing. In principle another backplane with ml605s, large Solid State Drives and database processors, can be interconnected via fiber.

As an alternative, the Network Node can also be realized by a backplane and processing boards other than ml605. The FPGA board should provide several PCIe x4, 2.0 for Solid State Drives, at least 8 GBytes of fast onboard DRAM, PCIe x4, 2.0 for the backplane, and several SFP+ IOs for 10 and 40 Gbit/sec Ethernet fiber.

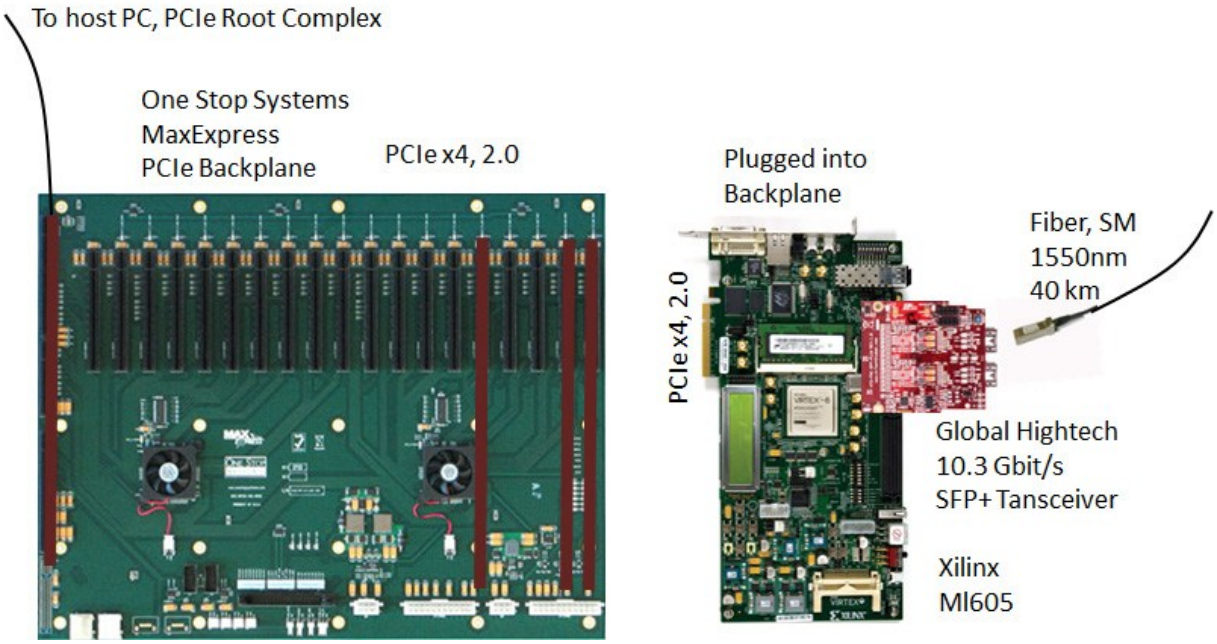

Figure 5. Hardware components for DMP Network Node

References

[BUD10] Budruk, R., Anderson, D., Shanley, T. "PCI Express System Architecture". Addison-Wesley, 11th Printing 2010.

[PLX11] PCI Express Overview, PLX Technology, August 2011.

http://www.plxtech.com/files/pdf/technical/expresslane/TechnologyBrief_PCIExpress_Q4-2003.pdf

[RON06] Rønningen, L A. "Node output queue model. Dropping and prioritizing". Presentation, DEI, Univ. of Padova, 10Nov06.

[RON07] Rønningen, L A. 'The DMP System and Physical Architecure'. Technical Web report, NTNU 2007. http://www.item.ntnu.no/~leifarne/The DMP 14Sep07/The DMP System and Physical Architecture.htm

RON07] Rønningen, L.A. "A protocol for futuristic multimedia". ICSPCS'2007, Gold Coast, Australia 2007

[RON09] Rønningen, L A, Heiberg, E. 'Perception of Time Variable Quality of Scene Objects'. Electronic Imaging 2009

[RON11] Rønningen, L. A. "The Distributed Multimedia Plays Architecture". Technical Report, NTNU (2007, 2009) 2011. http://www.item.ntnu.no/people/personalpages/fac/leifarne/the_dmp_architecture

[RON11a] Rønningen, L.A. "CCA Functional Structure & Architecture". Technical report, NTNU (2009) 2011. http://www.item.ntnu.no/_media/people/personalpages/fac/leifarne/cca_functional_structure_architecture.pdf

,