ADDAPT

Adaptive Data and Power Aware Transceivers for Optical Communications

Deliverable Report D 2.2

Market study, evaluation of applications and product specifications

Small or medium scale focused research project (STREP)

ICT-2013.3.2 Photonics

Project Start Date: 1st November 2013

Duration: 42 months

Project reference: 619197

FP7-ICT-2013-11 April 30, 2014 – Version 1.0

Project co-funded by the European Commission
within the Seventh Framework Programme (2007 - 2013)

Dissemination Level: Public

Document information

Title	D2.2 - Market study, evaluation of applications and product specifications
Work package	WP2 – Market studies, exploitation and dissemination, standardization
Responsible	TE Connectivity (TE)
Due date	Project M06 (April 2014)
Type	Report
Status	Version 1.0
Security	Public
Authors	Pieter van Leeuwen, TE Jeroen Duis, TE Michael Georgiades, PTL Savvas Charalambides, PTL Thomas Toifl, IBM Ronny Henker, TUD
Project URL	www.addapt-fp7.eu

Confirmation

Any work or result described in this report is either genuinely a result of this project or properly referenced.

Table of contents

List of Figures

List of Tables

Version Management

Version	Description	Author	Released
V0.1	First release	P. van Leeuwen	28-apr-14
V0.2	Added ecosystem, SDH and conclusion	P. van Leeuwen	29-apr-14
V0.3	Improved some wording. Added content on IBM Power7 and Infiniband. Improved references and abbreviations. Added Executive summary. Included feedback from [TUD], [TE], [IBM] and [PTL]	P. van Leeuwen	30-apr-14
V0.4	Improved the IBM775 cable calculation, and adapted the report accordingly	P. van Leeuwen	30-apr-14
V1.0	Layout adjustment, final Revision and approval	R. Henker	30-apr-14

Table 1 List of Revisions

Executive Summary

ADDAPT is a technology project, co-funded by the European Commission within the Seventh Framework Programme. ADDAPT aims to adapt the cable speed in the data-center to the offered dataload, by interpreting the datastream and detect the idle data in it. This report is not claiming to be a scientific report. It merely used already existing market surveys from leading suppliers to assess market development and the relevance of the ADDAPT technology in those markets.

In chapter 2, a first inventory is made with respect to the applications. Target applications are high-capacity inter-switching (as in data-centers) such as happens between servers, between racks, between co-locations and to storage devices. In supercomputers, the target application is the more direct communication between processors and also to storage devices. Based on those markets, a few first requirements are given for ADDAPT cables. Based on findings in chapter 3 (data-centers), 4 (supercomputing) and 5 (protocols), a small set of requirements for ADDAPT cables is made, and a first assessment of volume is made. We believe the ADDAPT technology leads to an identified potential market of 25 million cables, with protocols Ethernet, Infiniband and some proprietary protocols. Niche applications might exist for interfacing to SDH networks. (see paragraph 5).
The reachable market can be summarized as (in million cables):

Application	2015	2016	2017	2018	2019	2020
10 GbE	18	15	6			
40 GbE	2	6	16	23	26	23
100 GbE			0.3	1	2	6

Table 2 Potential market for Ethernet cables

# in Million cables	50 cm	100 cm	2 – 10 m	totals
PCIe / SAS		1.3		
SMP	0.8	2.7	1.8	5.3
Proprietary + Ethernet + infiniband	2.4	6.4	4.3	13.1

Table 3 Estimated installed base 'internal cabling' in supercomputers, per 2013

In chapter 3 and 4, the domain of data-centers and supercomputers are better defined, and in paragraph 4, a model is made to find an indication for the potential market of ADDAPT cables for supercomputers. In chapter 5, the development of the markets for the relevant protocols are described, and based on research by Cisco, an estimation of ADDAPT technology cables for the Ethernet domain is made. Finally, in paragraph 6, conclusions are given.

1 Introduction

Future communication networks, which include data centers, computing, and core networks, demand a significant improvement of its performance and flexibility while the power consumption for their operation needs to be drastically reduced. Additionally, more and more data is generated and processed in the chips. The challenge is to transmit this huge amount of data outside the chips. Today, the power consumption of data processing and storage systems is enormous. In a big multistage network with a 50 meter diameter and multiple transceiver hops, the transport consumes around 1000 times more energy compared to the processing of the same data.

Hence, a massive reduction of the power consumption for data transport is mandatory in the high-speed networks, e.g. optical packet switching systems and passive optical networks, already today.

This trend is rapidly leading to a critical need for an extremely significant reduction of energy consumption of optical transmission and server in-/outputs. Especially at high data rates and large link distances, optical links have lower losses than copper links. Most links in today's optical communication systems are statically driven with its maximum performance and provide the maximum data speed required for data traffic peaks. Therefore, almost the same high power is consumed independently of the data load. This makes the links and networks inflexible with respect to user demands and data loads variations over time. [ADDAPT]

ADDAPT aims at introducing flexibility to optical networks by scaling the performance and in turn the power consumption of optical transceivers on *system, optical device, IC and transistor* level.

An important 'buzzword' in the industry is PUE[1] (Power Use Effectiveness). ADDAPT reduces the primary power of the data-center, and as a consequence, cooling can be reduced which will help to improve the PUE factor. Maybe a better indication would be a KPI taking the total power consumption per TF/s into account.

This report investigates the markets in which the ADDAPT technology might play a role, and the requirements those markets impose upon products carrying this technology.

[1] Power Use Effectiveness. It is the relation between power used for actual data-center operations (servers, discs, datacommunication) and "the rest" such as lighting, ventilation, cooling, power supply etc). Present Benchmark is a PUE between 1.8 and 2.9. A stand-alone unit is offered claiming a PUE of 1.14 (ictroom.com). Google claims PUE of 1.12.

2 Market

This chapter touches briefly upon the application markets for ADDAPT technology, addresses briefly the issue of network protocols, and then focusses on volumes for the various markets.

2.1 *Applications*

The envisaged market being addressed by ADDAPT is two-fold:

High-capacity inter-switching (as in data-centers)

- intra-rack communication
- rack-to-rack communication
- co-location communication
- communication to storage devices.

Supercomputers (HPC)

- inter-processor communication
- communication to storage devices

Both markets (data-centers and supercomputers) will be elucidated in a separate chapter. This chapter will focus on volumes and requirements with respect to ADDAPT type of products for the applications and protocols described below.

2.2 *Requirements for ADDAPT technology in data-centers*

2.2.1 Some considerations on protocols

In order for the ADDAPT technology to understand the volume of actual data content ("load") that is being transmitted, there are two ways to go:

- Protocol-sensitive: ADDAPT must be able to identify the transmitting protocol and based on this select if the data speed should be increased or decreased.
- Protocol-agnostic. In this case, there is an algorithm allowing to scale up or down the dataspeed, and the decision how much to scale up or down is made outside the ADDAPT transceiver.

The choice has been made to make the demonstrator version of ADDAPT protocol-sensitive. With speeds of 10s to 100ds of Gbps, idle periods are in µseconds, and in most applications latency is critical. As a consequence, detecting the idle periods must be very close to the mechanism, where the idle datastream is taken out. This is half the core-mechanism of ADDAPT. (The other half is to manage the VCSEL and TIA such, that they actually perform over a wide range of data speeds on an energy-efficient way). For reasons of simplicity, the choice had been made to implement the 'take-idle-data-out' mechanism in the same chip as the 'detect-idle-periods' mechanism.

In WP 3, T3.3 Adaptivity control and T3.4 Data protocol, it is investigated how ADDAPT shall react to data-load changes, and how to communicate data-load changes to ADDAPT transceivers.

For prototype and demonstration purposes, an implementation of a specific protocol shall be made. At this moment, Infiniband is chosen for its relative easy implementation.

In the exploitation, it shall be understood which of the existing and future relevant protocols shall be made available. In the exploitation reports (D2.3 – D 2.5) this will be addressed more deeply.

The envisaged protocols where ADDAPT might be relevant are:

- Ethernet, including various flavours of Ethernet, such as:
  - Fibre Channel over Ethernet
  - Data Center Bridging[2]
  - iWarp over Ethernet
  - RoCE (RDMA over Converged Ethernet),

  These versions are targeting the data-center space.

  On top of that, various interfaces are developed to enable Ethernet over SDH:

  - Packet over SDH (POS)
  - Ethernet over SDH (EOS)
  - 10 GbE WAN interface

  That are targeting the telecommunication between data-centers.
- Fibre Channel
- Infiniband
- SMP (a proprietary IBM protocol)
- PCIexpress
- SCSI probably less relevant for ADDAPT, and has evaluated in the protocols below:
  - SATA, (Serial ATA, bus that connects host bus adapters to mass storage devices, SATA goes presently until SATA revision 3.2 - 16 Gbit/s)
  - SAS, (**Serial Attached SCSI** is a point-to-point serial protocol that moves data to and from computer storage devices such as hard drives and tape drives. SAS replaces the older Parallel SCSI)

Not believed relevant for ADDAPT:

- USB (smaller devices to host)
- RapidIO. (obsolete)
- FireWire (1394) (a sort of USB, but small marketshare)

In the chapter “Protocols”, we will focus deeper on the considered relevant protocols.

[2] **Data center bridging** (DCB) refers to a set of enhancements to Ethernet local area networks for use in data center environments. Specifically, DCB goals are, for selected traffic, to eliminate loss due to queue overflow (sometimes called **lossless Ethernet**) and to be able to allocate bandwidth on links. The higher level goal is to use a single set of Ethernet physical devices or adapters for computers to talk to a Storage Area Network, Local Area network and InfiniBand fabric. [Gai]

2.2.2 Ethernet

With present understanding, it seems as if the more 'moderate' speed of 40 GbE will be the winner for the coming years, over 100 GbE. The latter is seen as complex to implement, expensive in its technology, and potentially reducing redundancy in the network (as you need only a few, or even only one connection). Also, the PCIe infrastructure to easily support those speeds (of 100 GbE) in the server is not defined yet. [PCI-SIG]
Cisco however sees 100 GbE as the near future enabler for the fabric network for the increasing East-West traffic. Products are available in the routers. [CISCO]
For the Server to Router communication, ADDAPT technology must be available for 40 Gbps, later (2017), 100 Gbps.

2.2.3 Infiniband

Infiniband has a strong foothold (42%) in the supercomputing domain, and is penetrating into specialized data-centers. After Ethernet, it will be the biggest market for ADDAPT technology.
The prototype implementation of ADDAPT will be Infiniband.

2.2.4 SMP

This is the proprietary protocol used inside (members of) the IBM Power 7 family. It is designed with low latency in mind, thus increasing the performance for HPC

2.2.5 PCIe, SAS and SATA for SAN[3]

PCIe is a high-speed serial computer expansion bus standard in the servers, to which the network cards interface. There is a debate going on if PCIe 4.0 will be strong enough to support 100 GbE speeds. Standardisation is due for end of 2015[PCI-SIG] (see also par 5.2.1 PCIe 4.0). The future standard of PCIe4.0 will take some time to implement, and will run parallel to the development of ADDAPT based product.
PCIe standards are also used to directly communicate to storage and other I/O devices. PCIe 3.0 speeds up to max 1064 MB/sec (appr 8 Gbps). It appears that PCIe as storage communication protocol (so, outside the server) is increasingly important.
More investigation is needed to compare this with market penetration of more or less competing standards such as SATA and SAS.

2.2.6 Redundancy

An important issue is redundancy (as a means to increase reliability). At network level, the Ethernet switches often provide for redundancy. This is at the cost of expensive Ethernet ports. However, within the racks, on cable level, redundancy can also be provided by the cable itself. For e.g. the

[3] SAN: Storage Area Network, a network dedicated to manage the traffic between storage devices and servers.

Power775, all trunks have a 1 over 6 redundancy, allowing 1 fibre over 6 to fail without any consequence for throughput. It must be considered to also build in the ADDAPT technology (the TxRx chip) not only the adaptability feature, but also to combine this with a redundancy feature. The ADDAPT cable does have by definition more fibres per cable for the higher speeds (for 100 Gbps plus). Until 56 Gbps one fiber will suffice, taking out redundancy options. A good solution might be to have fibers running at 25 Gpbs and lower (adaptable by ADDAPT technology), depending on the application.

2.2.7 Ethernet Standardisation Alignment

Fibre assignment

As can be seen in chapter ”5.3.5 Ethernet cable standardization“, the fibre layout for Ethernet cables 40GbE and 100GbE is different than foreseen for ADDAPT. It must be investigated if, and how, an ADDAPT implementation can be acceptable in the context of this Ethernet Standardisation.

Energy Efficient Ethernet Standardisation

Moreover, a number of standardisation activities such as in IEEE802. Energy Efficient Ethernet Committee[4] shall be pursued to enable the application of this (ADDAPT) technology on a standardized way. [IEEE802]

2.3 Market Volume ADDAPT cables for data-centers

2.3.1 Ethernet inside the data-center

Please refer to Figure 5 Projected Timeline Showing Mainstream Adoption of 40 Gigabit Ethernet-Capable Switching Equipment [WP40GbE] showing the expected volume in 40GbEe ports. This is the most important application for ADDAPT cables, because more uniform than HTC cables (see above).

10 GbE	It can also be noted that the market for 10 Gbps Ethernet Server Ports is really huge until 2016 (like 15 mill ports/year for server only) afterwards fast declining. It shall be investigated if 10 Gbps is still an option for ADDAPT technology. If we can make it cheap enough, due to the volume, the potential power savings are huge.
40 GbE	This diagram suggests a potential market of 5 mill ports in 2016. The market will start to pick up per 2015.
100 GbE	The market for 100 Gbps will definitely exist. Volume is expected to pick up per 2018 with a volume of several 100ds of thousands.

Table 4 Potential market for Ethernet ADDAPT cables

[4] This subcommittee has developed baseline proposals accepted by the IEEE 802.3az Task Force mechanisms to reduce energy consumption by networking equipment, and communicate state and control information through the network to enable/disable energy efficient modes of operation

In other words, there is a potential market of approximately 6 million ADDAPT cables in the Ethernet world.

2.3.2 Communication between data-centers (co-locations)

For reasons of Point-of-Presence, security and backup, data-centers may want to communicate their data to co-locations. There are two basic ways the market is implementing this:

1. Using a dark fibre, either self-installed and owned, or rented from a dark-fibre company. This fibre is connected to own network equipment, running a network protocol of own choice, such as Ethernet or FibreChannel. Expected are high speed protocols such as 100 GbE or 400 GbE. Distances are between 200 m – 2 km. It has to be investigated how ADDAPT technology is applicable in this market.
2. Connecting to a telecom standard interface, provided by a telecom provider, such as ATM or SDH (protocol agnostic) via the POS technology and some other Ethernet technologies (See 5.5 SDH.) We expect there is a niche market for ADDAPT technology in the cables interfacing between the Ethernet router, and the SDH public interface.

2.4 *Market Volume ADDAPT cables for supercomputers*

If we assume that

1. the architecture of IBM's Power 775 is representative for all supercomputer, and if we assume
2. the TOP500 list (see 4.3 Supercomputer Top 500) is representing 80% of all supercomputers,

then the amount of potential ADDAPT type of cable can be estimated with reference to this IBM Power775:

- 670 cables @ 200 Gbps of 50 cm (within the CEC drawer),
- and another 4100 cables @ 60 – 120 Gbps between drawers (100 cm – 10 meter), outside the CEC drawer.
- for a peak performance of 100 TF/s

For all Top 500 supercomputer systems, including the 25 % not in this list (100 – 80) / (80), see the top500 list, we see an installed computing power of 625 * 729 TF/s average = 455 PF/s. Assuming the above mentioned relation, and assuming all cables replaced by ADDAPT, this leads to a potential market of 20 million ADDAPT cables.

Manufacturer	Avg Of Rpeak (1000)	Count Of Rank	Best Of Rank	cables 50 cm (x1000) inter process	cables >50cm (x1000) inter process	PCI express (*1000)
Hewlett-Packard	347	196	33	457	2609	196
IBM	643	164	3	708	4046	303
Cray Inc.	1205	48	2	389	2221	167

Manufacturer	Avg Of Rpeak (1000)	Count Of Rank	Best Of Rank	cables 50 cm (x1000) inter process	cables >50cm (x1000) inter process	PCI express (*1000)
SGI	612	17	14	70	399	30
Bull SA	551	14	20	52	296	22
Fujitsu	1873	8	4	101	575	43
Dell	1424	7	7	67	383	29
NUDT	15320	4	1	412	2353	176
Hitachi	264	3	152	5	30	2
Megware	226	3	181	5	26	2
Supermicro	199	3	398	4	23	2
NEC	182	3	210	4	21	2
National Research Center	619	2	40	8	48	4
RSC Group	499	2	84	7	38	3
Self-made	474	2	64	6	36	3
Dawning	1609	2	21	22	124	9
Sun Microsystems	402	2	71	5	31	2
Itautec	460	2	156	6	35	3
Others (within Top 500)	792	13	13	69	395	30
rest (20%, outside top500)	729	125	<500	612	3500	262
Total	455 PF/s	625		3000	17000	1300

Table 5 Estimation of HTC ADDAPT cables

IBM Power775 uses SMP for its inter-processor communication. According to the table above, IBM has a 27% market-share, capacity-wise.

Other manufacturers will also apply proprietary protocols for reasons of low latency, other may adhere to standards such as Infiniband and Ethernet. Summarizing (based on 2013 numbers):

# in Million cables	50 cm	100 cm	2 – 10 m	totals
PCIe / SAS		1.3		

SMP	0.8	2.7	1.8	5.3
Proprietary + Ethernet + infiniband	2.4	6.4	4.3	13.1

Table 4 cables for internal HPC

Based on the Top500 information there is also a networking 'outside' the supercomputers. It is not clear at this point, to what extend this information overlaps or is exclusive to the 'internal' cabling.
Interesting enough, according to this market research, the penetration of 'Custom Interconnect' an IBM protocol, has a capacity based market penetration of even 42%, Infiniband 29%, Cray 12%, Ethernet 9 % and other proprietary protocols 8%. See paragraph 3.2.3 Data-traffic in and around the Datacenter.

2.4.1 Potential Power Saving in HPC

Assuming an average power consumption of 25 pJ/bit, we saw in par. 4.2.1 IBM Power775, that 100 TF/s HPC capacity consumes approximately 25 kW in those cables.
For the complete supercomputing world according to the above table, this means a power consumption of 110 MW consumed in HTC cabling worldwide. (Only the inter-HTC cabling, this is exclusive of backups, communication to internet etc.)
With ADDAPT technology and state-of-the-art electro-optics technology, this could reduce with 80%.

2.5 *Data-center Ecosystem*

Like all industries, also the data-center industry has its own ecosystem. Take care, a particular organisation can be vertically integrated, such as AT&T. It can be depicted as follows:

supplier	Description	Examples
1-tier internet provider (ISP)	Global internet transit provider, connection 2-tier ISPs via Transit: Tier-2 ISPs pay to be routed over a tier-1 ISP network	AT&T, Verizon, Sprint, Deutsche Telecom
Internet Exchange (IEX)	Network hub connecting equal-tier ISPs via Peering: rerouting other equal-tier ISP traffic for free. The member ISPs pay for the IEX	Amsterdam Internet Exchange
2-tier internet provider (ISP)	Internet provider routing to the end-users	Kpn, Vodafone, XS4ALL, etc

supplier	Description	Examples
Hosting	Providing computing and storage power to customers. A number of flavours exist in the marketplace.	
	Providing servers, storage capacity. Customer uses dedicated devices as his own	
	Infrastructure as a Service (IaaS), virtual servers, pay per use, managed by datacenter operator	
	Cloud Giants. Cloud Services	Google, Amazon, Yahoo, Microsoft.
	Enterprise Hosters. Big enterprises, providing added services such as integration	HP, BT, Origin
	Mass Market hoster. Targeting SMEs	Leaseweb
ICT System Integrators	Connecting it all together. Also provider of management tools, virtualisation tools etc.	
ICT Installer	Physical installation of racks, servers, cables etc	
Housing	Providing physical infrastructure (power, air-conditioning, cooling, security). A number of flavours exist in the market	
	Wholesale colocation	Providing completely equipped datacenters, leased per m2	Digital Realty, Global Switch
	Retail Co-location	Completely equipped, but including network connections of ISPs to connect to. Far more flexibility then wholesale co-location	Equinix, Telecity, Interxion
	Carrier-biased co-location	One ISP offering (his) connectivity and completely equipped datacenter space.	KPN, Level3, Colt
	Infrastructure as a service (IaaS)		
Physical System Integrators	Connecting the physical infrastructure together.	

supplier	Description	Examples
Physical Installer	Physical installation of pipes, cooling, lighting, cables etc	
ICT Suppliers	All suppliers of hardware, such as servers, discs, routers, racks, cables etc	Too much to mention
Builders		

Table 6 Ecosystem Data centers

The ADDAPT customers can be found in the ICT layer of the ecosystem, but their decision will be influenced by their customers in the upper layers.

3 Data-centers

From a functional point of view, you can discriminate within a data center two types of servers:

- front-end servers: commodity servers running common operating systems such as Linux, Windows, and FreeBSD. Those servers have direct interaction to users. Physically they are in a more accessible environment.
- back-end servers: will typically host databases and other persistent storage servers such as file servers. These rely on more scalable and robust multi-processor operating systems such as HP-UX, Sun Solaris, IBM AIX and MVS, and Windows 2000 Data Center servers. Those servers normally don't have interaction with users. Physically they are in a less accessible environment.

This issue needs further investigation: It has to be understood if, and what, this means for data communication needs. We expect the front-end servers will be closer to 'the spine', the back-end servers will be linked more close to each other and have strong interaction with the front-end servers. Physical accessibility will be different, although virtualisation might counter this development. This will impact the (need for) manual handling of the cables, and might have influence on the specification of the cable.

3.1 *Applications*

This section lists a number of applications that are highly dependent on a fast delivery data center support. Some of those will be more sensitive on high data throughputs, other more on fast processing of small transactions. We discriminated between Operation / Networking Applications, and Software Applications.

3.1.1 Operation / Networking Applications

Those applications running in the data-centers are mainly technology focused. They need technical infrastructure. Application software is provided by the customer. Middle Ware is provided by the data-center. Management such as capacity planning, virtualization etc. is provided by the datacentre.

- application hosting
- IT management services
- Supporting Telecommunication Operations Architecture
- Interconnection Services
- High Capacity Storage Provisioning
- Virtualisation Support
- Cloud Services
- High Capacity Inter-Switching
- Operations and Monitoring Tools

3.1.2 Software Applications

Those applications not only need technical infrastructure, but also heavy application software to perform well.

- Financial Transactions

- Big Data
- Mobile Apps
- mission-critical applications
- Internet of Things
- Internet of Everything
- Internet Service Provider Service
- Streaming Services (like Netflix and Spotify)
- etc.

3.1.3 Web hosting

Generally speaking, today's web pages are more complex than the source to server and back again requests, leading to data accesses for a single web page from multiple locations within the data center. This issue needs further investigation: It has to be understood if, and what, this means for data communications needs. We expect the front-end servers will be closer to 'the spine', the back-end servers will be linked more close to each other and have strong interaction with the front-end servers.

3.1.4 Financial Transactions

Applications such as high-speed financial trading, analytics and credit card transaction sorting are creating lots of small communications that stay within the rack and need high-speed connectivity. [ComWeekly]
This will lead to different idle / load patterns than typical patterns show. ADDAPT will have to cope with this.

3.1.5 Cloud computing

Growth of workloads in cloud data centers is expected to be five times the growth in traditional workloads between 2012 and 2017. Traditionally, one server carried one workload. However, with increasing server computing capacity and virtualization, multiple workloads per physical server are common in cloud architectures. Cloud economics, including server cost, resiliency, scalability, and product lifespan, are promoting migration of workloads across servers, both inside the data center and across data centers (even data centers in different geographic areas). Often an end-user application can be supported by several workloads distributed across servers. This approach can generate multiple streams of traffic within and between data centers, in addition to traffic to and from the end user. Table 6 provides details on the shift of workloads from traditional data centers to cloud data centers. [CISCO]

	2012	2013	2014	2015	2016	2017	CAGR 2012-2017
Traditional data center workloads	51.2	53.5	58.4	62.3	66.3	69.7	6%
Cloud data center workloads	32.2	45.7	61.1	78.1	96.8	118.5	30%
Total data center workloads	83.4	99.3	119.5	140.4	163.2	188.2	18%
Cloud workloads as a percentage of total data center workloads	39%	46%	51%	56%	59%	63%	NA
Traditional workloads as a percentage of total data center workloads	61%	54%	49%	44%	41%	37%	NA

Table 7 Data center workload in millions [CISCO]

3.2 *Changing network configuration*

3.2.1 Traditional configuration

Large Datacenters contain typically 10,000-50,000 servers with high-speed network connections to other data centers and the Internet. The large scale and rapid evolution of these data centers offers significant challenges for the data center operators.

Traditionally, traffic consisted of external sources making requests from a server inside the datacenter, followed by the server's response. This model led to three-tier networks, with access, aggregation and core switch layers becoming the default datacenter network topology, and traffic density being lowest at the access layer, becoming greater as you moved closer to the core. [ComWeek]

3.2.2 Virtualization

One of the main factors affecting this migration of workloads from traditional data centers to cloud data centers is the greater degree of virtualization in the cloud space, which allows dynamic deployment of workloads in the cloud to meet the dynamic demand of cloud services. Nowadays, Virtual application environments (VAE) are implemented or being implemented. VAEs provide a model for encapsulating system resources in a way that supports the management of these very large systems. VAE presents an environment to an application that is consistent with the application's configuration requirements. For example, every application has certain requirements regarding network connectivity, file systems, middleware, and storage services. It can have explicit layers of servers and many local area networks.

It is believed, that this trend towards cloud computing on Virtual Machines will continue. This leads from an architecture designed to take traffic from servers to the edge of the network and vice versa – known as north-south traffic – to a fabric-like network, with all nodes interconnected, enabling high-

speed bandwidth for east-west connectivity over the traditional traffic to/from the datacenter (north / south). [ComWeekly]. See figure 1 below.

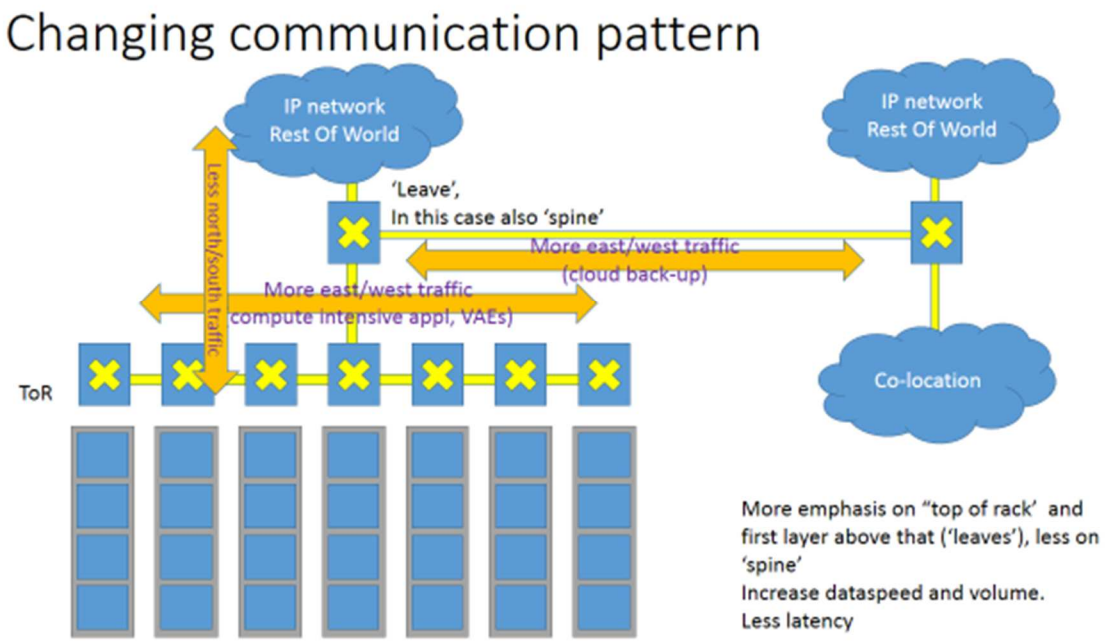

Figure 1 Changing Network Flows

Cisco's global cloud index report showed 76% of traffic now stays within the data-center. Much of the traffic now relates to the separation of functions allocated to different servers inside the facility, such as applications, storage and databases. These then generate traffic for backup and replication and read/write traffic across the datacenter. [CISCO]

3.2.3 Data-traffic in and around the Datacenter

The main qualitative drivers for cloud adoption include faster delivery of services and data, increased application performance, as well as improved operational efficiencies. While security and integration with existing IT environments continue to represent concerns for some potential cloud-based applications, a growing range of consumer and business cloud services are currently available. Today's cloud services address varying customer requirements (for example, privacy, mobility, and multiple device access) and support near-term opportunities as well as long-term strategic priorities for network operators, both public and private.

The following sections summarize not only the volume and growth of traffic entering and exiting the data center, but also the traffic carried between different functional units within the data center, cloud versus traditional data center segments, and business versus consumer cloud segments. [CISCO]

	2012	2013	2014	2015	2016	2017	CAGR 2012-2017
By Type (EB per Year) (ExaByte = 10^18 Byte)							
Data center to user	427	560	711	883	1,086	1,317	25%
Data center to data center	167	221	281	352	435	530	26%
Within data center	1,971	2,560	3,223	3,978	4,867	5,879	24%
By Segment (EB per Year)							
Consumer	1,952	2,585	3,301	4,123	5,097	6,211	26%
Business	613	756	914	1,091	1,291	1,515	20%
By Type (EB per Year)							
Cloud data center	1,177	1,755	2,419	3,224	4,178	5,313	35%
Traditional data center	1,389	1,586	1,795	1,989	2,210	2,413	12%
Total (EB per Year)							
Total data center traffic	2,565	3,341	4,215	5,214	6,387	7,726	31%

Table 8 IP traffic in data centers [CISCO]

4 Supercomputers (HPC)

4.1 *HPC in the cloud*

The use of High Performance Computing (HPC) in commercial and consumer IT applications is becoming popular. HPC users need the ability to gain rapid and scalable access to high-end computing capabilities. Cloud computing promises to deliver such a computing infrastructure using data centers such that HPC users can access applications and data from a Cloud anywhere in the world on demand and pay based on what they use. This might mean there is 'corner' in the datacentre with a different (HTC) infrastructure (infiniband, proprietary protocols), but connected to the internet like the rest of the center. For less demanding HTC customers, there will just be strong back-end server infrastructure based on regular server technology and regular infrastructure (Ethernet dominant, some infiniband). We do not believe that this trend will be fundamentally different from the developments described above, (data-centers), with the exception that for the actual server performing the supercomputing function, considerations as in the next chapter (Dedicated HPC supercomputers) will apply.

4.2 *Dedicated HPC supercomputers*

These are clusters of dedicated supercomputers: servers where the focus is on computing, less on I/O, webhandling, storage, file serving etc. Supercomputer architectures are designed for high-speed number crunching, and thus have a parallel multi-processor processing capability, efficient and multi-level cache handling, low latency. They are used for applications such as weather prediction, geological modelling, behaviour of complex biological systems, quantum simulations of Nano Systems and Biomolecules, simulation of the evolution of the cosmos and many other computing intensive tasks. Also visualization of processes (like the mergers of black-hole binaries with very small ratios and very high spins) is an important application.

4.2.1 IBM Power775

Based on the IBM Power775 supercomputer infrastructure, one can say that:
One full rack consists of 384 Power775 processors (each consisting of 8 P7 cores), each with a capacity of 0.25 TF/s[5]

- 4 Power7 processors are assembled in a QCM module
- Each QCM module is copper connected to 7 other QCM modules in the same CEC blade via 24 GBps[6] (per direction) each.

[5] Tera Flop per second. 10^12 arithmetic 8 Byte calculations per second
[6] GBps = Giga Byte per second, 10^12 8 bit words per seond. In this case (IBM Power7, intra-CEC) we mean the gross communication speed. In effective load it is * 8/10 (per 8 bits payload, 10 bits are used for error correction and load balancing)

- Each QCM module is fibre connected to 24 other QCMs in the same 'supernode' via 7.5 GBps[7] (per direction) each. A supernode consists of 4 CECs.
- Each QCM module is fibre connected to 16 remote QCMs via 15 GBps[7] (per direction) each.
- Each QCM module is copper connected to 3 PCIe links: 2 links of 10 GBps[8] and 1 of 5 GBps (per direction) each.

If this system is a reference, then a supercomputer with a power of 100 TF/s has a potential need of:

Protocol	Number cables	length	speed	Application
SPM	672	50 cm	200 Gbps	Intra-CEC
SPM	2304	100 cm	60 Gbps	Intra-supernode
SPM	1536	2 - 10 meter	120 Gbps	between supernodes
SPM	Total 4512	50 cm – 10 m	60 – 200 Gbps	Inter-hub
PCIe	288	100 cm	80 Gbps	storage
	Total 4800	50 cm – 10 m	60 – 200 Gbps	

Table 9 Potential ADDAPT cables in a Power775

This is one of the target markets for ADDAPT cables.

Within the IBM Power775, those links are built upon a 3, 5 and 10 Mbps connection hierarchy and compliant to industry standards. For other supercomputer brands this might be slightly different. A way must be found to integrate the ADDAPT cables in this connection hierarchy.

Assuming an average power efficiency of 25 pJ/bit, than this supercomputing rack consumes 25 kW in these cables.

PS: this is exclusive of networked applications to the world 'outside' the supercomputer. See paragraph 4.3.1 Supercomputing intra-networking

4.3 *Supercomputer Top 500*

A group of people within the supercomputing community, produces twice a year a top500 of supercomputer centers. The TOP500 table shows the 500 most powerful commercially available computer systems known to this community. In addition to cross checking different sources of information, they select randomly a statistical representative sample of the first 500 systems of our database. For these systems they ask the supplier of the information to establish direct contact between the installation sites and to verify the given information. As the TOP500 should provide a

[7] This data runs over fibre @ 10Gbps. There is a 8/10 ECC mechanism, and a 5/6 fibre redundancy. Effective data payload is 6 * 10 * 5/6 * 8/10 = 40 Gbps = 5 GBps. The same mechanism is used for the remote connection: effective data payload = 10 GBps.
[8] 16 lines @ 2.5 GHz @ DDR = 10 GBps.

basis for statistics on the market of high-performance computers, they limit the number of systems installed at vendor sites. [Top500]

Below you will find an aggregation of this list, with this those suppliers that appear more than once on this list. The list is ordered by [count of rank]

- [Rmax] is the maximal LINPACK performance achieved, and
- [Rpeak] the theoretical peak performance
- [Count of Rank] is the number of time this supplier has a position in the top500 list.

Manufacturer	Avg Of Total Cores (1000)	Avg Of Rmax (TF/S)	Avg Of Rpeak (TF/s)	Max Of Rpeak (TF/s)	Avg Of Rank	Count Of Rank	Best Of Rank
Hewlett-Packard	19	198	347	1341	282	196	33
IBM	44	482	643	20133	279	164	3
Cray Inc.	51	871	1205	27113	156	48	2
SGI	33	524	612	2296	152	17	14
Bull SA	32	453	551	1667	178	14	20
Fujitsu	114	1715	1873	11280	119	8	4
Dell	79	897	1424	8520	232	7	7
NUDT	843	9353	15320	54902	63	4	1
Hitachi	9	206	264	316	263	3	152
Megware	22	186	226	285	241	3	181
Supermicro	11	128	199	238	426	3	398
NEC	5	155	182	218	331	3	210
National Research Center	76	471	619	1070	185	2	40
RSC Group	28	332	499	524	106	2	84
Self-made	22	362	474	594	115	2	64
Dawning	76	726	1609	2984	129	2	21
Sun Microsystems	34	354	402	497	104	2	71
Itautec	14	206	460	563	218	2	156
Other suppliers (1 appearance in list)	39	502	792	3388	199	13	13

Table 10 List of Top500 supercomputer centers.

4.3.1 Supercomputing intra-networking

The Top500 statistics show also the share of various network protocols in supercomputing, ranked against performance (not against number of systems).

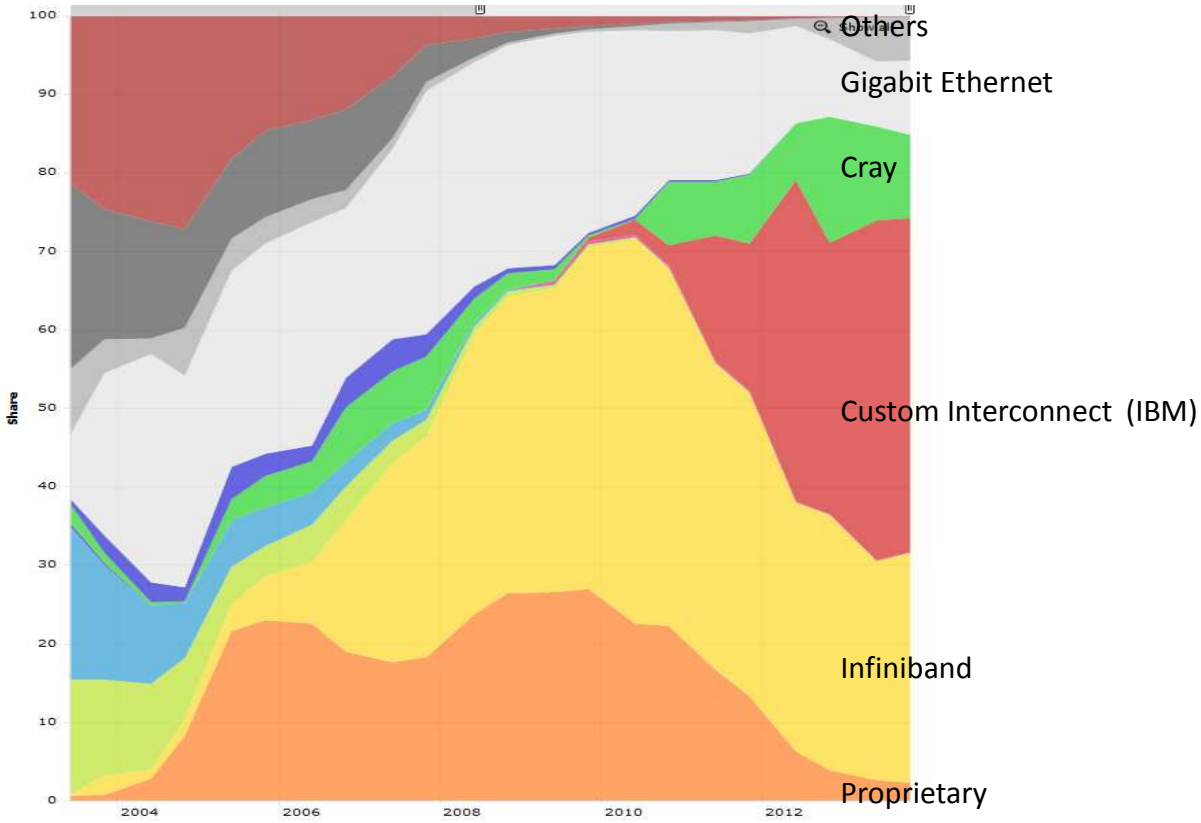

Figure 2 Share of network protocols in HTC [Top500]

Protocol penetration	against performance	Against # systems
Custom Interconnect (IBM)	42 %	9 %
Infiniband	29%	41 %
Cray	12%	5 %
Gigabit Ethernet	9 %	43 %
Other Proprietary	8%	2 %
	100 %	100 %

Table 11 Share of protocols in HTC (based on [Top500])

Take care: those are networks ‘outside’ the supercomputer, and not fully comparable to the internal networking as discussed in 4.2.1 IBM Power775. Those ‘internal’ networks are less visible as such, they are supplied with the supercomputer.

5 Protocols

5.1 *Infiniband*

Like Fibre Channel, PCI Express, Serial ATA, and many other modern interconnects, InfiniBand offers point-to-point bidirectional serial links intended for the connection of processors with high-speed peripherals such as disks. On top of the point to point capabilities, InfiniBand also offers multicast operations. It supports several signalling rates and, as with PCI Express, links can be bonded together for additional throughput. The technology is promoted by the InfiniBand Trade Association.

Effective unidirectional theoretical throughput in Gb/s; actual data rate, not signaling rate						
	SDR	DDR	QDR	FDR-10	FDR	EDR
1X	2	4	8	9.67	13.64	25
4X	8	16	32	38.79	54.54	100
12X	24	48	96	116.36	163.64	300

Table 12 Infiniband speeds

Larger systems with 12X links are typically used for cluster and supercomputer interconnects and for inter-switch connections. [Wikipedia]
The strong points of Infiniband are its ability to directly interface to the memory (RMA), and the low latency in the switching fabric.
The number of InfiniBand-based supercomputers increased to 42 percent of the TOP500 (but 29 % based on performance). It is starting to penetrate into the data-center environment where communication intensive processes run (technical computing, cloud computing, Web 2.0, and enterprise data centers.) [Top500]

5.2 *PCI Express*

PCI-E (Peripheral Component Interconnect Express) is a high speed serial computer expansion bus. PCI-E eliminates a lot of the short comings of standard PCI including and provides more bandwidth as well as compatibility with existing systems. The PCI-E has a number of beneficial features including: PCI Transparency, high bandwidth provisioning, good flow control, reliable and robust link layer, error reporting, fault isolation support, power management, spread spectrum clock offering high speed serial connectivity on a single physical layer channel. Through this PCI-E offers a large number of benefits including buffer size flexibility over provisioning for QoS, end-to-end reliable transport for service availability and with no interoperability issues and operations and management support.

5.2.1 PCIe 4.0

PCIe 4.0 is the next evolution of the PCI Express I/O specification. At 16 GT/s[9] bit rate, the interconnect performance bandwidth will be doubled over the PCIe 3.0 specification, while preserving compatibility with software and mechanical interfaces. The key requirement for evolving the PCIe architecture is to continue to provide performance scaling consistent with bandwidth demand from a variety of applications with low cost, low power and minimal perturbations at the platform level. The final PCIe 4.0 specifications, including form factor specification updates, are expected to be available in late 2015. [PCI-SIG]. It was suggested that PCIe 11 4.0 would enable dual 100Gbps Ethernet server ports starting in 2015, but there is some doubt in the market if PCIe 4.0 is able to support 100 GbE at all [ComWeekly]. Intel believes a new replacement for PCIe 4.0 that won't arrive before 2017-18 which will better handle the requirements of higher-bandwidth network technologies. [ComWeekly]

5.3 *Ethernet*

Over time, 100GbE will become the core to replace 10GbE, with 10GbE becoming the replacement for 1GbE. Today 10GbE switch (Layer 2/3) revenue increased 10.7% year over year while 10GbE port shipments grew a remarkable 61.4% year over year to just under 3.5 million ports in 3Q12, and continue to be the main driver of the overall Ethernet switch market. [ComWeekly]
Currently, there is a debate going on if 40 or 100 GbpE is the speed to go for this fabric. Cisco believes "The need to deal with increasing east-west traffic will mean that 100GbE is the best way to deal with latency in the datacenter, provided the right tools are in place to ensure that it all works effectively," Cisco says, but this is not agreed with by all parties. [CISCO]
New traffic patterns and network fabrics are driving demand for bandwidth – but are network technologies ready to meet this challenge? When it comes to servers, the answer is no [ComWeekly].
See also PCIe ($ 5.2.1 PCIe 4.0)

5.3.1 Considerations of NCSA (Blue Waters) on Ethernet

Aggregating 40GbE technology is fraught with "gotchas", according to Tim Boerner, senior network engineer at the University of Illinois's National Center for Supercomputing Applications (NCSA).
The NCSA's Blue Waters project houses some 300TB and needs to move 300Gbps out of the server clusters across an Ethernet fabric to an archiving system or to an external destination. It uses Extreme Networks' 40GbE technology and the link aggregation protocol (LACP) to move its large volumes of data. Boerner says aggregation can deliver "good [performance] numbers with sufficient transfer", but "what you run into isn't a limit of the network but the difficulty of juggling how I/O is assigned to different CPUs and PCI slots". Despite the difficulties associated with aggregation, the NCSA

[9] This is typical PCI speak: a Transfer/second is speed per second, including the overhead. PCI uses 8b/10b coding, hence 1 GT/second equals 800 Mbps payload (payload might be a full Ethernet stack).

made a conscious decision not to go for 100GbE. "It would be a huge expense to go for 100GbE and to have all the peripherals on that system," says Boerner. [ComWeekly]

5.3.2 40 and 100Gbps Ethernet in data-centers

Major network suppliers, such as Cisco, Juniper and Brocade, already sell 100GbE-capable switches and routers, mostly aimed at datacenter fabrics for users such as telecoms operators and cloud providers. Large enterprises now have the tools to create fabric networks, and as the technology exists, it is possible to build such a high-speed network. Enterprises are only now starting to buy 10Gbps ports in significant numbers, so prices of faster ports are likely to remain high in the near future. Additionally, the cost of peripherals and technologies such as packet inspectors and management tools will also stay up and will need to be made 100GbE-capable.
So, 40GbE is a more cost-effective route in the medium term. Also, datacenter network managers may be expecting the enterprise to outsource significant volumes of traffic to a variety of cloud providers, relieving them of the need to think about and pay for expensive network upgrades.
As illustrated by Figure 3 Leading Edge Deployment Trends; General Market Deployment Is Typically Delayed by Several Years, I/O data transfer rates within the access layer are doubling every 24 months, while transfer rates at the core layer double approximately once every 18 months. A primary driver behind the push to 40Gigabit Ethernet is a new generation of high-speed, high-demand, computing applications and technologies. These include the spreading deployment of virtual servers and cloud computing. [WP40GbE]

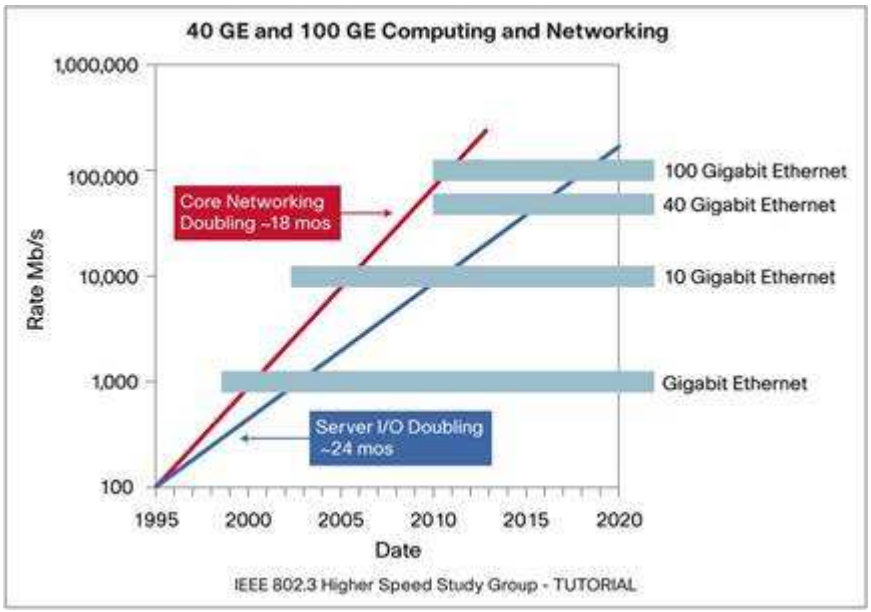

Figure 3 Leading Edge Deployment Trends; General Market Deployment Is Typically Delayed by Several Years [WP40GbE]

There has been some debate as to whether IT managers should hold off on deploying the 40 Gigabit Ethernet technology and bide their time waiting for 100 Gigabit Ethernet to become commercially

available. But that question is fast becoming moot because 40 Gigabit Ethernet provides design flexibility and cost advantage over 100 Gigabit Ethernet. 40 Gigabit Ethernet can be effectively deployed today in aggregation links in data center networks. By 2016, 40 Gigabit Ethernet will also be commonly applied to access links to connect servers, as Figure 4 Gigabit Ethernet for Multimode and Single-Mode indicates. [WP40GbE]

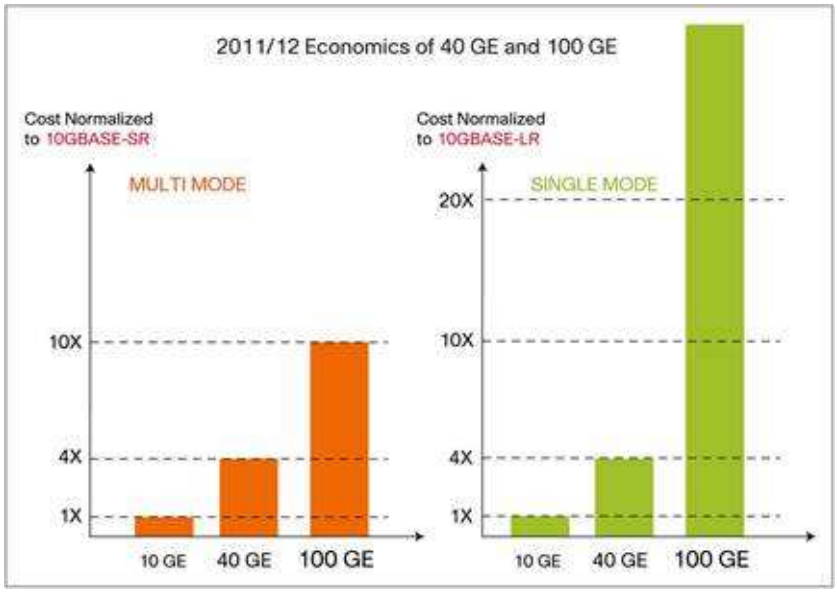

Figure 4 Gigabit Ethernet for Multimode and Single-Mode [WP40GbE]

As "Figure 5 Projected Timeline Showing Mainstream Adoption of 40 Gigabit Ethernet-Capable Switching Equipment" illustrates, 40 Gigabit Ethernet is still a couple of years from wide-scale adoption, which gives IT managers and CIOs time to start their migration planning. [WP40GbE]

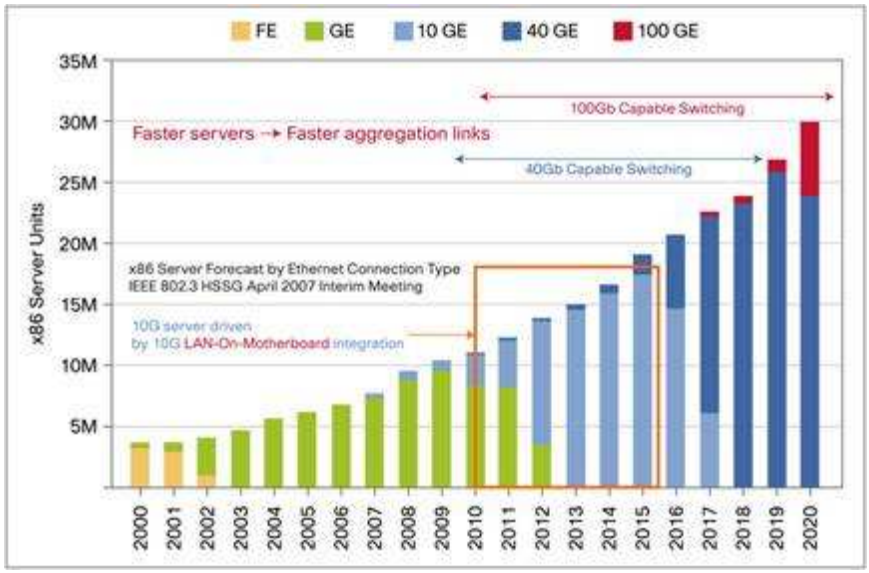

Figure 5 Projected Timeline Showing Mainstream Adoption of 40 Gigabit Ethernet-Capable Switching Equipment [WP40GbE]

5.3.3 Form Factor

Media	Reach	Speed	CFP	QSFP	CXP
Single-mode	10Km	100G			
		40G		Future?	
	40Km	100G			
Multimode (OM3)	100m	100G		Future?	
		40G			
Multimode (OM4)	150m	100G		Future?	
		40G			
Copper	3-7m	100G		Future?	
		40G			

Planned for 1st Generation Not Planned for 1st Generation

Figure 6 Transceiver Form Factors Planned for 1st Generation Implementation [WP40GbE]

5.3.4 Transceivers

40 Gigabit Ethernet transceivers (see Figure 6) are being developed along several standard form factors. TheCForm-Factor Pluggable (CFP) transceiver features 12 transmit and 12 receive 10-Gbps lanes to support one 100 Gigabit Ethernet port, or up to three 40 Gigabit Ethernet ports. Its larger size is suitable for the needs of single-mode optics and can easily serve multimode optics or copper as well. The CXP transceiver form factor also provides 12 lanes in each direction, but is much smaller than the CFP and serves the needs of multimode optics and copper. The Quad Small-Form-Factor Pluggable (QSFP) is similar in size to the CXP and provides four transmit and four receive lanes to support 40 Gigabit Ethernet applications for multimode fiber and copper today and may serve single-mode in the future. Another future role for the QSFP may be to serve 100 Gigabit Ethernet when lane rates increase to 25 Gbps. [WP40GbE]

5.3.5 Ethernet cable standardization

Traditionally, the Ethernet standard has relied upon duplex fiber cabling with each channel using one fiber to transmit and the other to receive. However, the 802.3ab standard requires multiple lanes of traffic per channel. To get multiple lanes, the 40 and 100 Gigabit Ethernet standard uses parallel optics, as indicated in Figure 7.

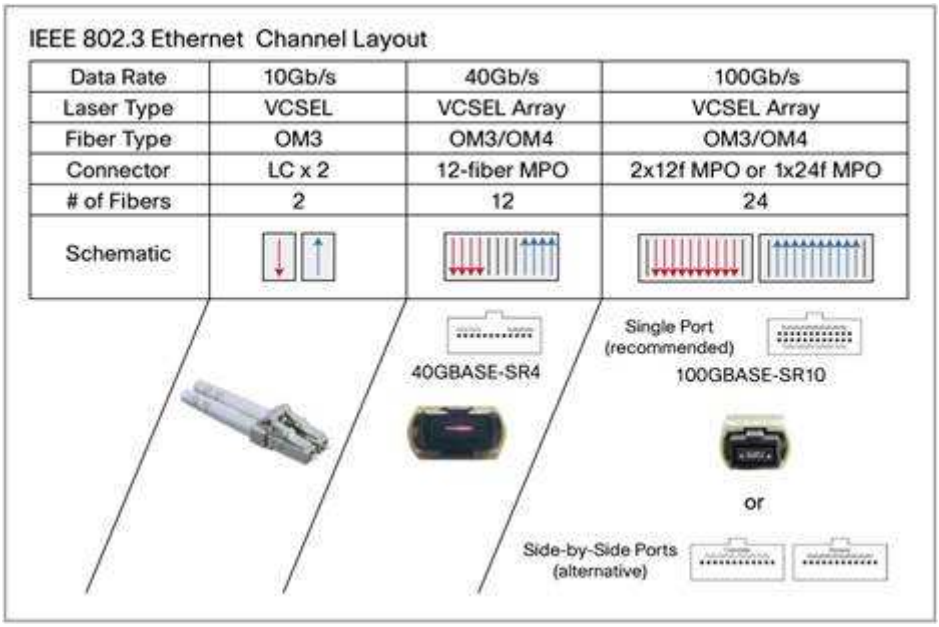

Figure 7 Layout Showing Ethernet Channel Distribution for 10/40/100 Gigabit Ethernet Using Multimode Fiber [IEEE802]

For 100 Gigabit Ethernet, several interface variants have been described with the preferred option being a single 24-fiber MPO connector. Alternatively, two 12-fiber connectors can be positioned either vertically or side by side to make up the channel. [WP40GbE]. 25 Gbps lanes are being discussed [IEEE802]. In fact, the standardization of fiber usage is not in the IEEE, but is done within an MSA.

5.3.6 Tunneling protocols

Also, there is a trend to apply more tunnel protocols, tunneling IP and Ethernet inside the different virtual machines running in the datacenter. (IP packets are encapsulated inside IP Packets). It has to be investigated if this needs to be taken into account for the ADDAPT technology. Please refer to "2.2.1 Some considerations on protocols" (about protocol dependency). The present assumption is: not relevant because tunneling is embedded within the protocol.

5.4 *Converged Network Adapters (CNAs)*

The more virtual the data center is, the more difficult it becomes to pre-provision appropriate levels of I/O. This makes Converged Network Adapters (CNAs) important to support convergence all the way to the edge of the network. Fiber Channel, Fiber Channel over Ethernet (FCoE), or even iSCSI network access server (NAS), need a network infrastructure capable of handling the characteristics of storage requirements as well. The ADDAPT project needs more understanding if this impacts a potential application of ADDAPT in this domain.

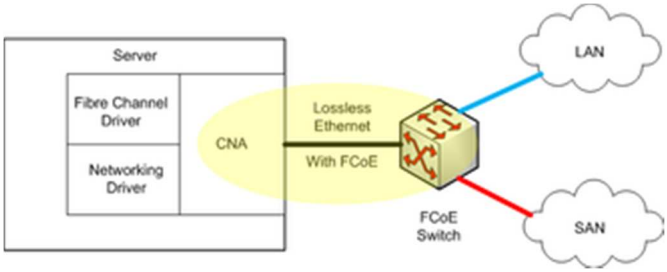

Figure 8 CNA Converged Network Adapter [Wikipedia]

5.5 *SDH*

SDH is a telecom standard [10] used by the Telecom industry.
In the context of data-centers, the normal application of SDH for Internet service provider (ISP) is to provide inter-Point-of- Presence (inter-POP) and backbone connectivity, carrier backbone, networks, high-speed cable operator backbones, and private enterprise networks.

5.5.1 Packet over SDH (POS)

The basic application is the transfer of IP packets, for which technology is developed: "Packet over SONET/SDH" (POS) technology is ideally suited for Internet and/or IP networks, because it provides superior bandwidth utilization efficiency over other transport methods.
The datacentre normally interface directly to this interface via a SONET/SDH Port Adapter (POS-PA). It is specifically suited for building high-speed and fault-tolerant IP networks that runs over Gigabit Router platforms. The POS PA leverages the router IP class-of-service (CoS) capabilities for service providers to provide differentiated services while also providing advanced SDH/SONET features such as automatic protection switching (APS).
The Gigabit router takes care of packaging the IP packets from Ethernet to SDH/SONET and vice versa.
A niche application of ADDAPT technology could be to incorporate this standard into the cables carrying this POS traffic, between the telecom's interface, and the router POS-PA interface. Relevant speeds for this application range from 0.6 Gbps to 40 Gpbs. (the standard goes further), so ADDAPT may only be relevant for the latter case.

[10] **Synchronous Optical Networking** (**SONET**) and **Synchronous Digital Hierarchy** (**SDH**) are standardized protocols that transfer multiple digital bit streams over optical fiber using lasers or highly coherent light from light-emitting diodes (LEDs). At low transmission rates data can also be transferred via an electrical interface. The method was developed to replace the Plesiochronous Digital Hierarchy (PDH) system for transporting large amounts of telephone calls and data traffic over the same fiber without synchronization problems. SONET generic criteria are detailed in Telcordia Technologies Generic Requirements document GR-253-CORE.[1] Generic criteria applicable to SONET and other transmission systems (e.g., asynchronous fiber optic systems or digital radio systems) are found in Telcordia GR-499-CORE.[2]

SONET Optical Carrier level	SONET frame format	SDH level and frame format	Payload bandwidth (Gbit/s) (appr)
OC-12	STS-12	STM-4	0.6
OC-24	STS-24	–	1.2
OC-48	STS-48	STM-16	2.4
OC-192	STS-192	STM-64	10
OC-768	STS-768	STM-256	40

Table 13 Standards within SDH / SONET

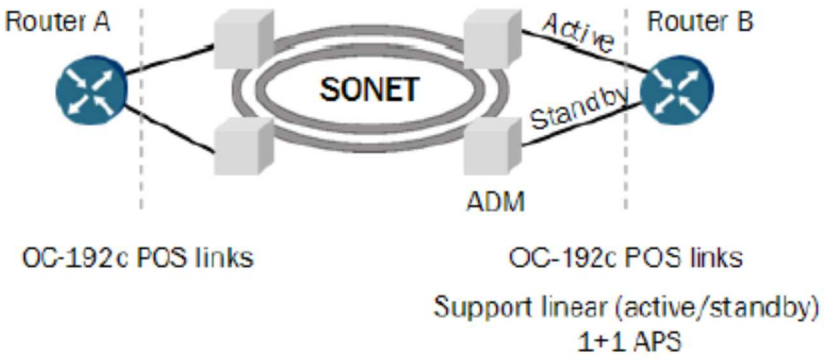

Figure 9 Application of POS in SDH [Brocades]

5.5.2 10 GbE WAN interface

In some cases, the data-center operator can use dark fibre to interconnect relative close data-centers (200 m – 2 km). Applications look like this:

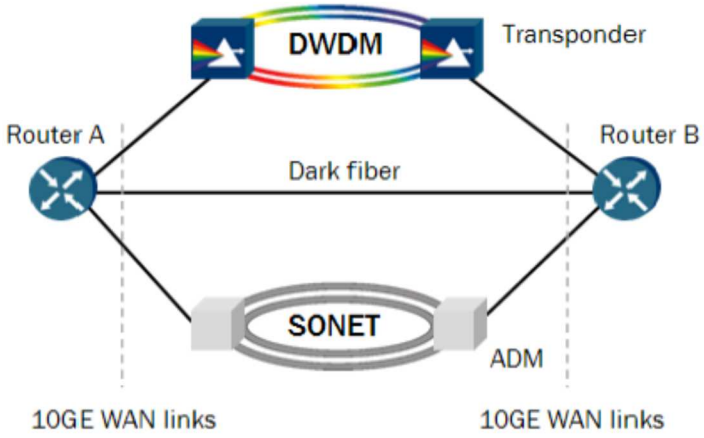

Figure 10 Application of 10GE WAN [Brocades]

In this case, a different standard is used: 10 GE WAN. It provides a WAN interface sublayer (WIS) to provide a simplified SONET/SDH framer function. The purpose of the WIS is to allow the 10 GbE WAN PHI to generate Ethernet data streams that are mapped directly to STS-192c or VC-4-64c streams at the PHY level. See Figure 10 Application of 10GE WAN. [Brocades][Wikipedia]

5.5.3 Ethernet over SDH/SONET

This is a third way to send Ethernet frames over SDH link. They are sent through an "encapsulation" block to create a synchronous stream of data from the asynchronous Ethernet packets. The synchronous stream of encapsulated data is then passed through a mapping block which typically uses virtual concatenation (VCAT) to route the stream of bits over one or more SDH paths. After traversing SDH paths, the traffic is processed in the reverse fashion: virtual concatenation path processing to recreate the original synchronous byte stream, followed by decapsulation to converting the synchronous data stream to an asynchronous stream of Ethernet frames. The standard allows for 10 MbE to 1 GpE. EoS also drops the "idle" packets of the Ethernet frame before encapsulating the Ethernet frame to GFP, which is recreated at the other end during decapsulation process. Hence this provide a better throughput compared to native Ethernet transport.
This is exactly the same approach as ADDAPT is taking, with the difference that ADDAPT is switching down the dataspeed. EoS is targeted at running the SDH link more efficiently.
Theoretically, the ADDAPT technology can somehow be integrated with this EoS technology.

6 Conclusion

It is clear that 'sustainability' also entered the data-center world. A lot of focus is on making the data-center less power-hungry. KPI's as PUE have emerged from this awareness. ADDAPT clearly can contribute to this movement, and the capability of a cable to be 'data-adaptive' will also be a key selling argument.
First market investigations show, that the potential market volume for these kinds of cables (containing ADDAPT technology) is in the range of 20 million.

6.1 *Supercomputers*

For the internal CPU-CPU communication, vendors use proprietary protocols such as SMP and Custom Interconnect by IBM, also Cray and some others have proprietary standards, and/or use public standards such as Infiniband and Ethernet, or a mix thereof. If we consider the IBM implementation as the benchmark, then the total supercomputing community has an installed base of of 20 million (internal) cables, based on the top500 inventarisation of 2013. The market per year is a derivative of this, taking into account depreciation and market growth.

# in Million cables	50 cm	100 cm	2 – 10 m	totals
PCIe / SAS		1.3		
SMP / Custom Interconnect	0.8	2.7	1.8	5.3
Proprietary + Ethernet + infiniband	2.4	6.4	4.3	13.1

Table 14 estimated installed base 'internal cabling' in supercomputers, per 2013

6.2 *Data-centers:*

Below are market estimates for Ethernet Ports. Those ports appear in Routers and Servers. Every connection between a server and a router or between 2 routers is potentially suitable for ADDAPT technology.

Application	2015	2016	2017	2018	2019	2020
10 GbE	18	15	6			
40 GbE	2	6	16	23	26	23
100 GbE			0.3	1	2	6

Table 15 potential market for Ethernet cables

6.2.1 Co-location communication

This domain is using SDH technology, and/or Ethernet WAN standards. It is not clear yet, if and how ADDAPT technology is applicable in this domain. It is a potential market. See also paragraph 5.5.

6.3 *Specification*

At least the following issues needs to be further addressed:

6.3.1 Protocols

Definitely Ethernet and Infiniband Protocols are very relevant. Proprietary protocols such as SMP, Custom Interconnect (IBM) and Cray in the supercomputer domain seem relevant also.
PCIe is believed to be 'nice to have'.
The added value for applying ADDAPT technology in cables connecting from Ethernet or Infiniband domain to SDH as carrier needs to be investigated further.

6.3.2 Redundancy

This is believed to be an important feature. Some cable (in fact all higher Ethernet and Infiniband cables) implementations have redundancy built in because a lower level of communication speed is used. The cable aggregates this. If ADDAPT provides this higher speed interface on a single fiber, than redundancy (at the cable level) is lost. A feature in the ADDAPT technology to maintain the inherent redundancy given by the separate fibres, seems easy to implement and a no extra costs for higher speeds (100 Gpbs plus). For speeds at 50 Gbps or lower, multiple fibers should be considered for reasons of redundancy.

6.3.3 Integration

The easy way is interfacing to existing electrical interfaces. This however means serious losses in the electrical path between CPU and interface. Also, the integration in racks, serverblades etc. needs to be better understood.

6.3.4 Standardisation

Ethernet MSA standardisation describes how the fibres are used in Ethernet cables. It needs to be discussed in MSA context if and how the ADDAPT technology can be standardized in this context. The application of ADDAPT technology inside the Ethernet standardisation [IEEE802] must be investigated.

References

[ADDAPT]	Grant Agreement no 619197 for Collaborative Project, Annex 1 - "Description of Work"
[Brocades]	10 Gigabit ethernet WAN PHY Capabilities
[CISCO]	Cisco Cloud Computing Index 2013
[ComWeekly]	http://www.computerweekly.com/feature/100Gbps-Ethernet-is-it-time-to-move
[Gai]	Silvano Gai, *Data Center Networks and Fibre Channel over Ethernet (FCoE)* (Nuova Systems, 2008)
[HP01]	Adaptive Internet Data Centers, Jerome Rolia, Sharad Singhal, Richard Friedrich, Hewlett Packard Labs, Palo Alto, CA, USA
[IEEE802]	IEEE P802.3az Energy Efficient Ethernet Task Force (http://www.ieee802.org/3/az/index.html).
[InfoWeek]	InformationWeek, January 29th, 2009, reported by W. David Gardner
[Mellanox]	Mellanox Technologies White Paper. TOP500 Results and Analysis http://www.mellanox.com/page/top_500
[PCI-SIG]	the steering body for PCIe, on the availability of PCIe 4.0
[Top500]	http://www.top500.org/lists/2013/11/
[Wikipedia]	The well-known online open web-based, free-content encyclopaedia project supported by the Wikimedia Foundation
[WP40GbE]	The Market Need for40 Gigabit Ethernet. White Paper 2012, Cisco, Gautam Chanda

Acronyms

Acronym	Definition
100GbE	100 Gigabit per second Ethernet
40GbE	40 Gigabit per second Ethernet
DBC	Data Center Bridging. An Ethernet specification for the data-center environment
EoS or EoSDH	Ethernet over SDH. Refers to a set of protocols which allow Ethernet traffic to be carried over synchronous digital hierarchy networks in an efficient and flexible way. The same functions are available using SONET (a predominantly North American standard).
HPC	High Performance Computing (supercomputing)
IEEE	Institute of Electrical and Electronics Engineers. It amongst others standardized the Ethernet protocols
iWARP	The Internet Wide Area RDMA Protocol (iWARP) is a computer networking protocol for transferring data efficiently. It is sometimes referred to simply as "RDMA", though RDMA is not a feature exclusive to iWARP.
KPI	Key Performance Indicator. An easy metric that shows a performance on a particular subject.
MSA	Multi-source agreements (MSAs) are not official standards organizations. Rather, they are agreements that equipment vendors assume when developing form factors for communications interfaces.
NCSA	National Centre for Supercomputing Applications, University of Illinois
PoCE	RDMA over Converged Ethernet (RoCE) is a network protocol that allows remote direct memory access over an Ethernet network. RoCE is a link layer protocol and hence allows communication between any two hosts in the same Ethernet broadcast domain. Although the RoCE protocol benefits from the characteristics of a converged Ethernet network, the protocol can also be used on a traditional or non-converged Ethernet network. [Wikipedia]
PoS	Packet over SDH, a way to transport (IP) packets over SDH / SONET
PUE	Power Use Effectiveness. A KPI developed by The Green Grid Association , a nonprofit, open industry consortium
SMP	Internal IBM protocol between processor of the Power7 family