

Cees de Laat

EU
COMMIT
UvA

NWO

PID/EFRO

SURFnet

TNO

NCF



Science Faculty @ UvA

Informatics Institute



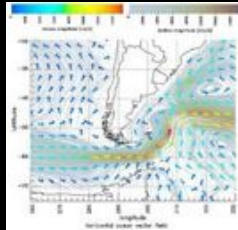
- CSA: Computer Systems Architecture (dr. A.D. Pimentel)
- FCN: Federated Collaborative Networks (Prof. dr. H. Afsarmanesh)
- IAS: Intelligent Autonomous Systems (Prof. dr. ir. F.C.A. Groen)
- ILPS: Information and Language Processing Systems (Prof. dr. M. de Rijke)
- ISIS: Intelligent Sensory Information Systems (Prof. dr. ir. A.W.M. Smeulders)
- SCS: Section Computational Science (Prof. dr. P.M.A. Sloot)
- SNE: System and Network Engineering (Prof. dr. ir. C.T.A.M. de Laat)
- TCS: Theory of Computer Science (Prof. dr. J.A. Bergstra)



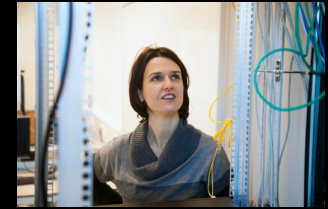
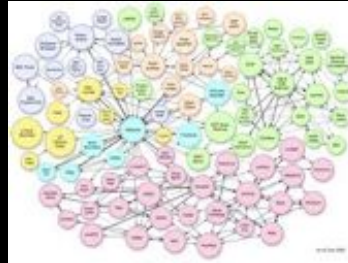
... more data!

Internet developments

Google



DATA



... more realtime!



twitter



myspace
a place for freedom



Linked in



SchoolBANK

Hyves

flickr
from YAHOO!



... more users!

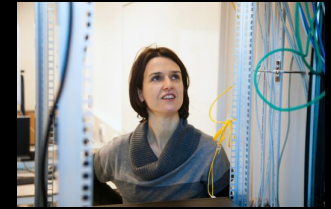
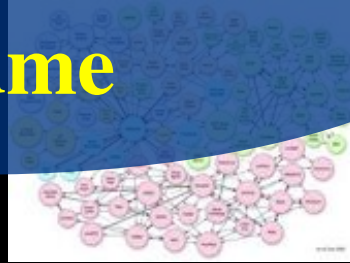
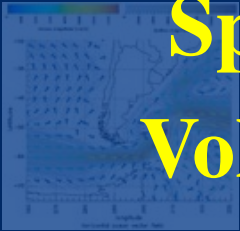
... more data!

Internet developments

Google

Speed
Volume

DATA



Deterministic

Real-time



twitter



Scalable

Secure

Linked in

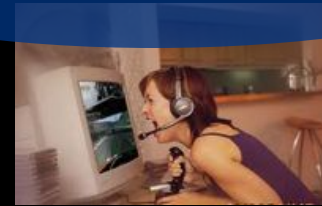


myspace

SchoolBANK

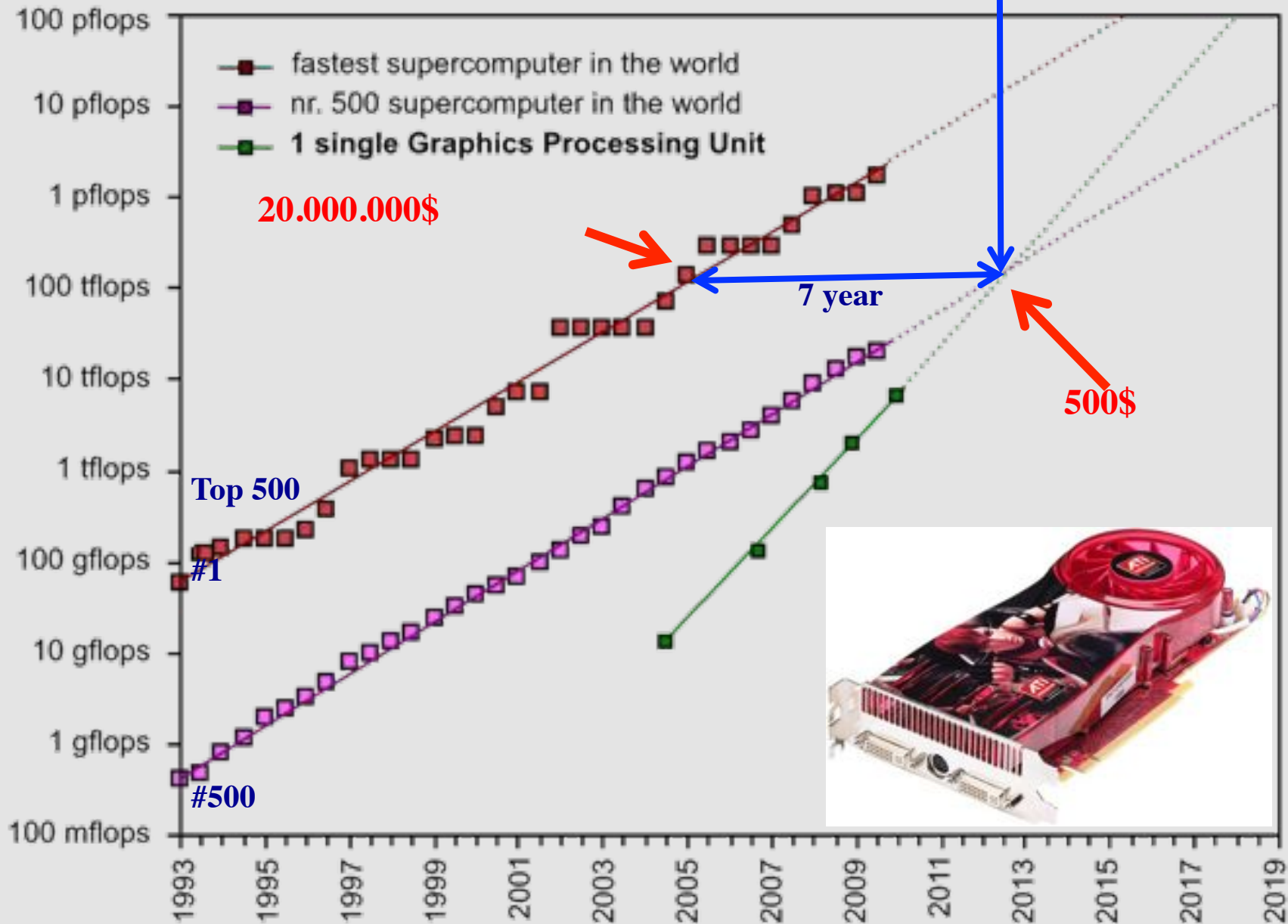
Hyves

flickr

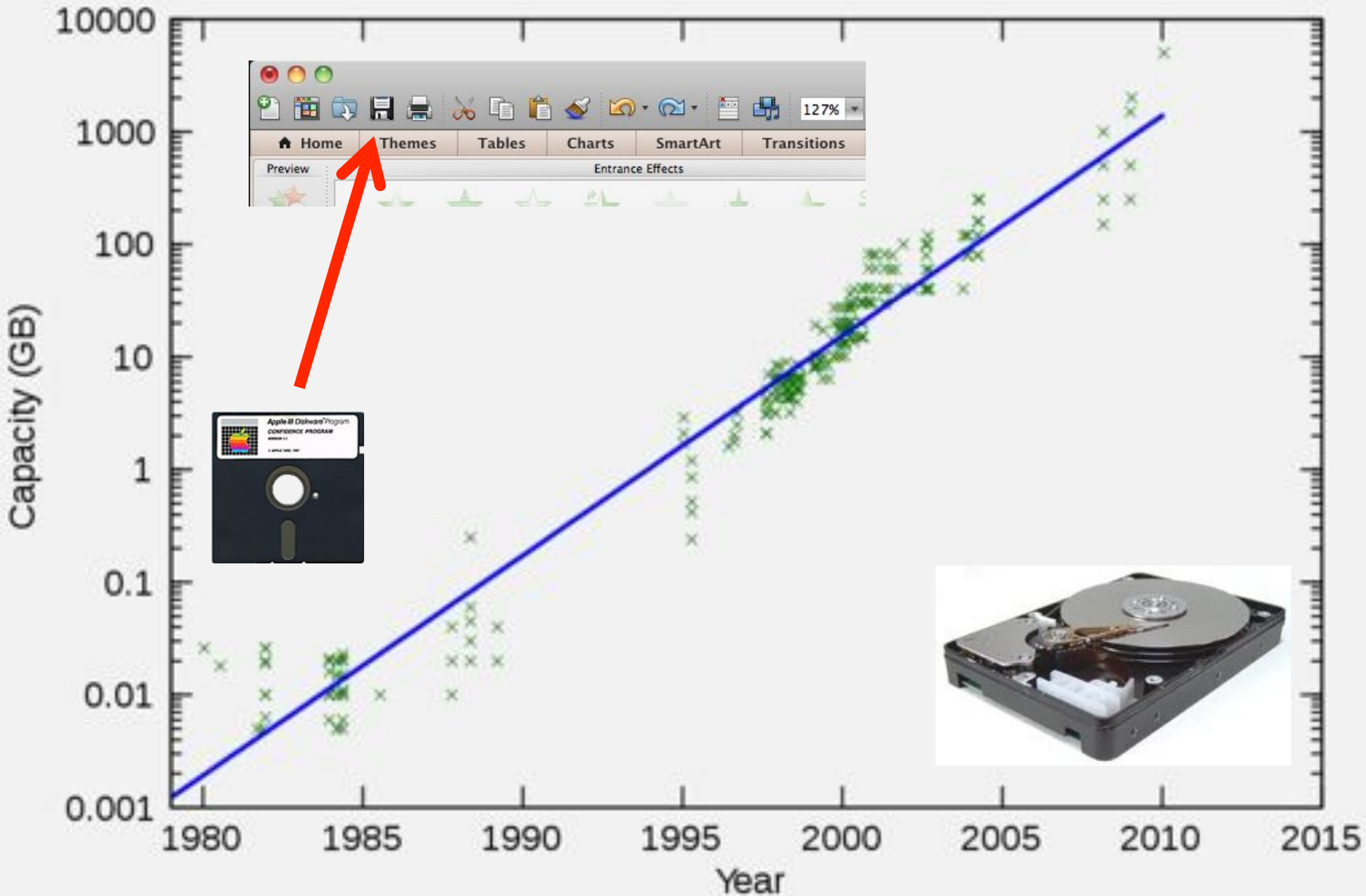


... more users!

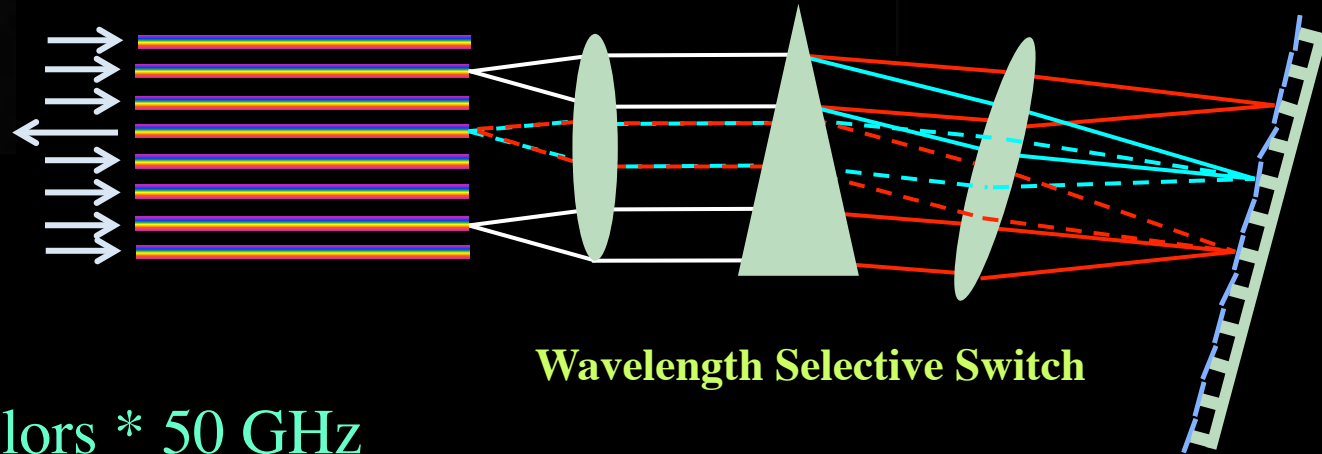
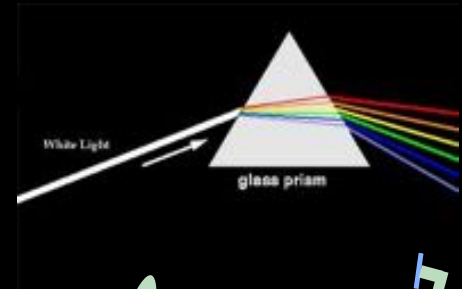
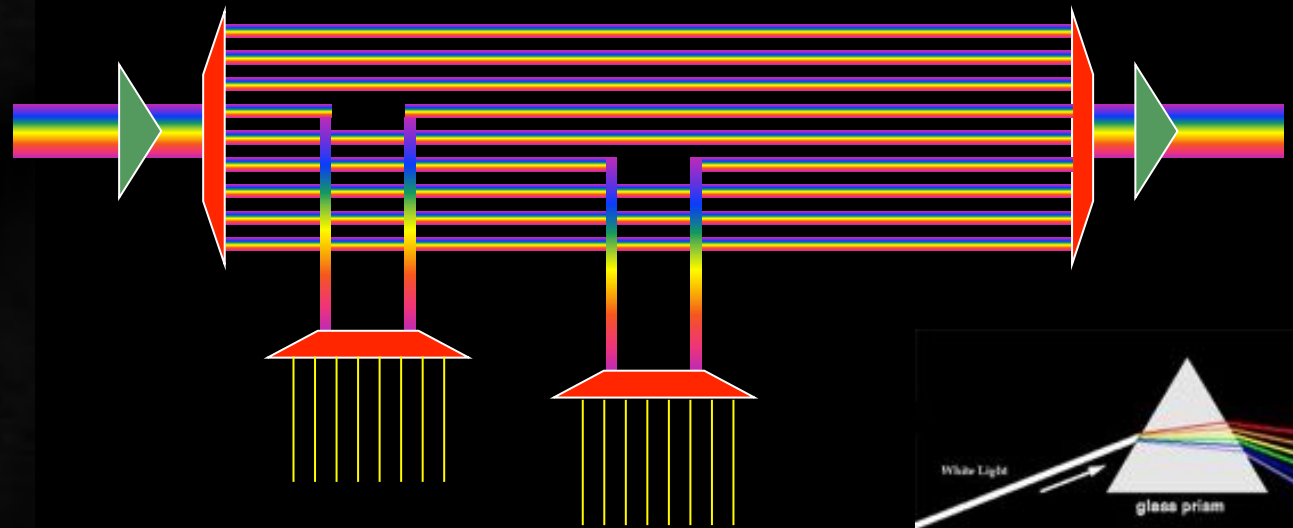
GPU cards are disruptive!



Data storage: doubling every 1.5 year!



Multiple colors / Fiber



Wavelength Selective Switch

Per fiber: $\sim 80-100$ colors * 50 GHz

Per color: 10 – 40 – 100 Gbit/s

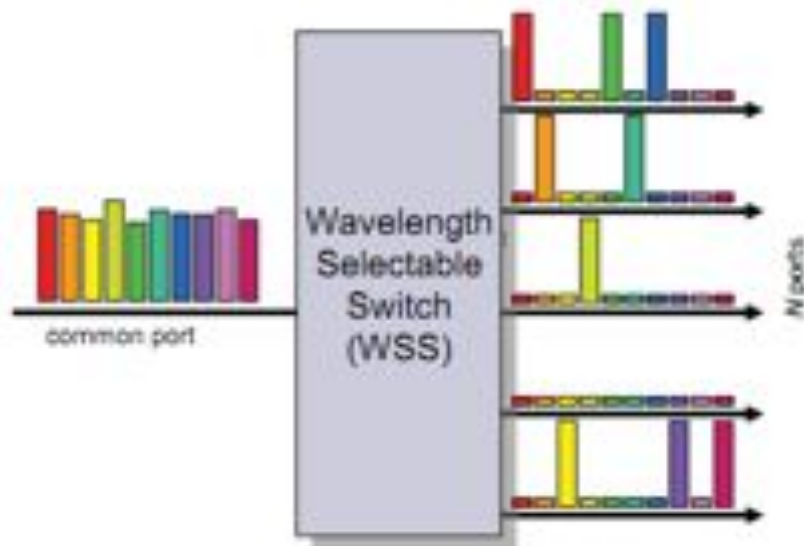
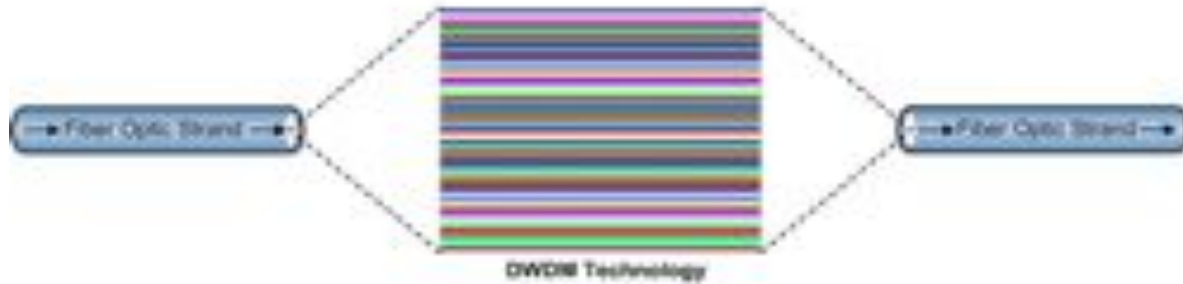
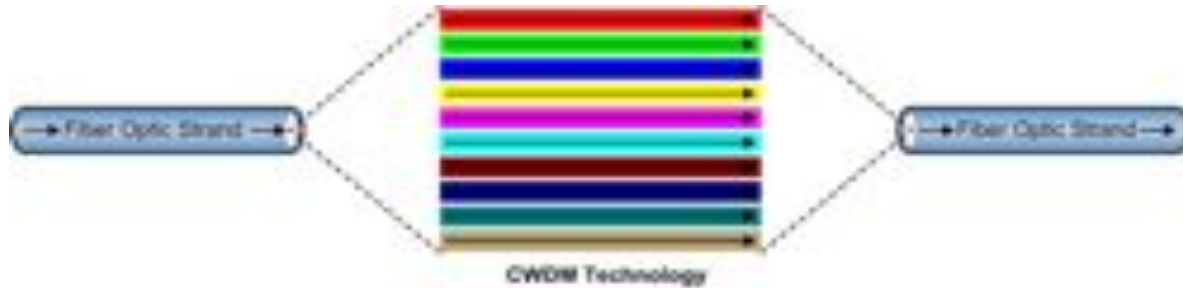
BW * Distance $\sim 2 * 10^{17}$ bm/s

New: Hollow Fiber!

➔ less RTT!

Optical transmission

... more possibilities



Virtualization



u
s
e
r
s

A. Lightweight users, browsing, mailing, home use

Need full Internet routing, one to all

B. Business/grid applications, multicast, streaming, VO's, mostly LAN

Need VPN services and full Internet routing, several to several + uplink to all

C. E-Science applications, distributed data processing, all sorts of grids

Need very fat pipes, limited multiple Virtual Organizations, P2P, few to few

For the Netherlands 2011

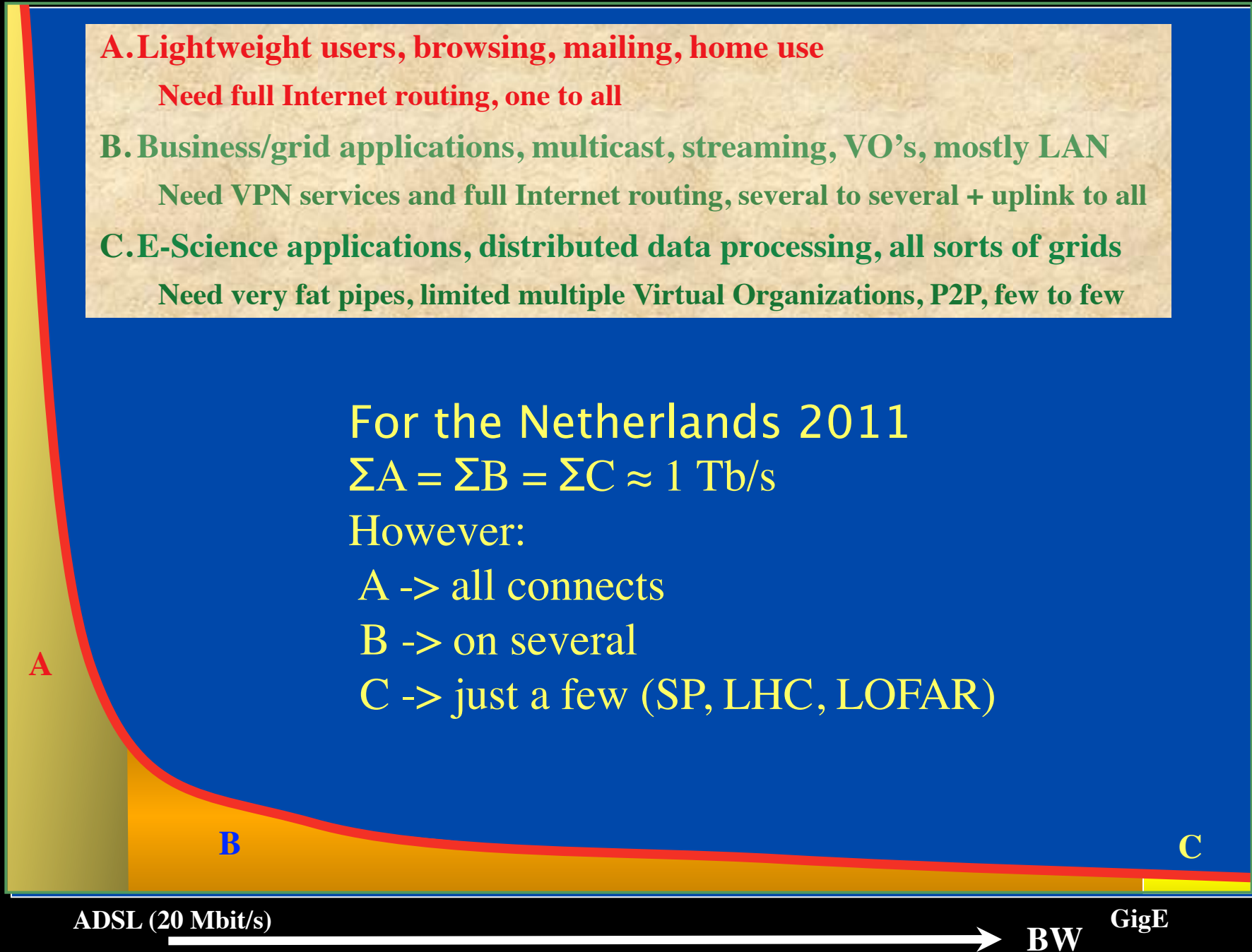
$\Sigma A = \Sigma B = \Sigma C \approx 1 \text{ Tb/s}$

However:

A -> all connects

B -> on several

C -> just a few (SP, LHC, LOFAR)



ADSL (20 Mbit/s)

BW

GigE

Towards Hybrid Networking!

- Costs of photonic equipment 10% of switching 10 % of full routing
 - for same throughput!
 - Photonic vs Optical (optical used for SONET, etc, 10-50 k\$/port)
 - DWDM lasers for long reach expensive, 10-50 k\$
- Bottom line: look for a hybrid architecture which serves all classes in a cost effective way
 - map A -> L3 , B -> L2 , C -> L1 and L2
- Give each packet in the network the service it needs, but no more !

L1 \approx 2-3 k\$/port



L2 \approx 5-8 k\$/port



L3 \approx 75+ k\$/port



SNE @ UvA

Speed
Volume

Deterministic
Real-time

Scalable
Secure

Ijkdijk/Urban Flood

Medical

LifeWatch/ENVRI

CosmoGrid/eVLBI

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud

Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

Topology/Architecture

Optical Photonic

X X

X

X

X X

X X

X

X

X

X

X

X

X

X

X

X

X

X

X

SNE @ UvA

Speed
Volume

Deterministic
Real-time

Scalable
Secure

Ijkdijk/Urban Flood

Medical

LifeWatch/ENVRI

CosmoGrid/eVLBI

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud

Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

Topology/Architecture

Optical Photonic

X X

X

X X

X

X

X

X X

X

X

X X X

X

X

X

X

Where when will it happen?

SNE @ UvA



Ijkdijk/Urban Flood

Medical

LifeWatch/ENVRI

CosmoGrid/eVLBI

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud

Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

Topology/Architecture

Optical Photonic

X

X

X

X

X

X X

X

X

X

X

X X

X X

X

X

X

X

X

X

X

X

X

X

X

X

X



IJKDIJK

Sensors: 15000km* 800 bps/m ->12 Gbit/s to cover all Dutch dikes

Sensor grid: instrument the dikes

First controlled breach occurred on sept 27th '08:



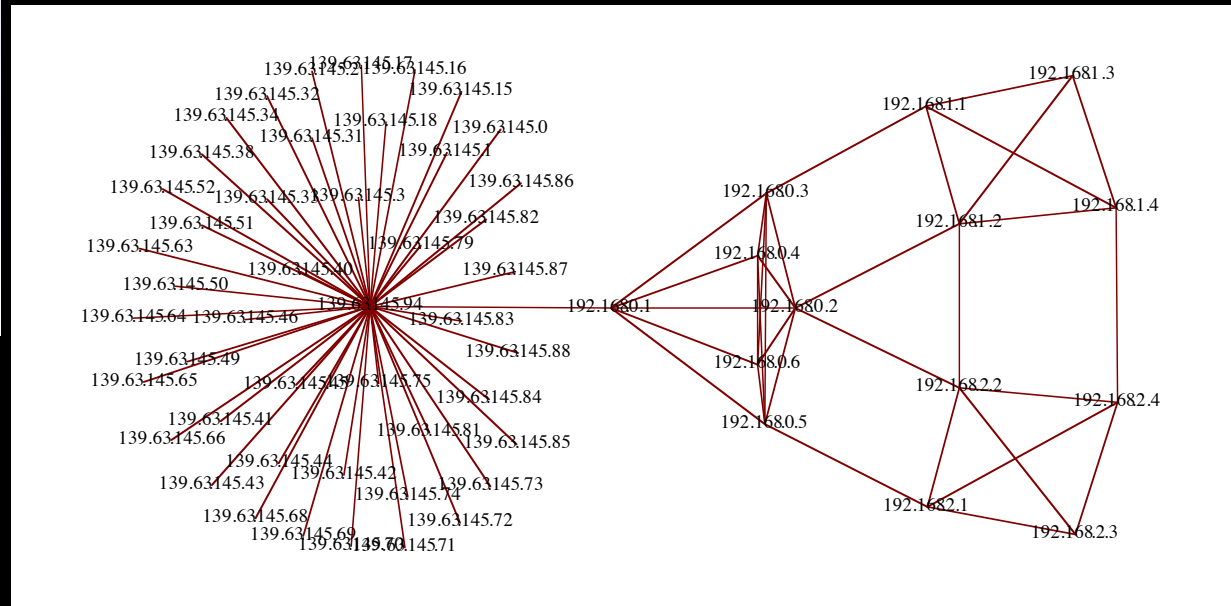
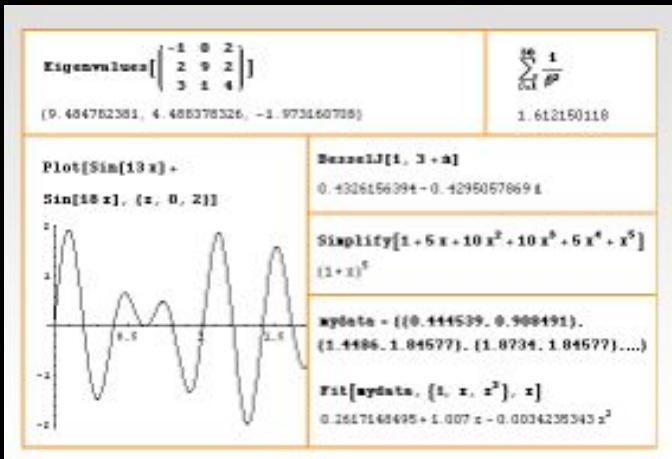
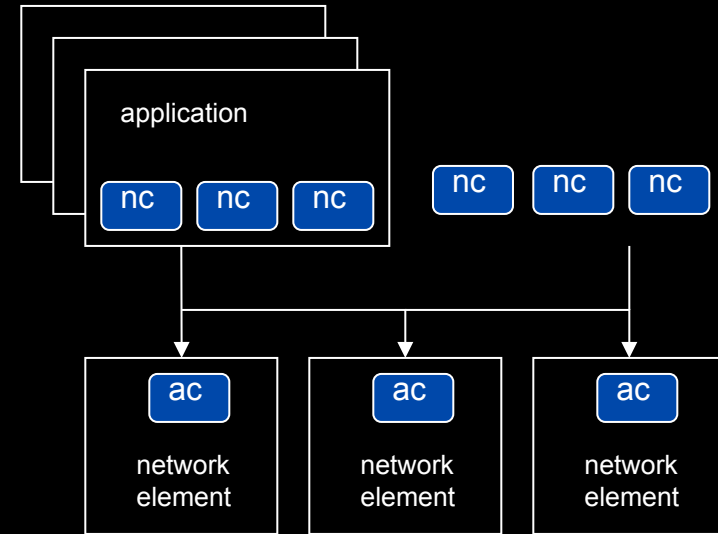
Many Pflops/s

Many small flows -> 12 Gb/s

User Programmable Virtualized Networks.

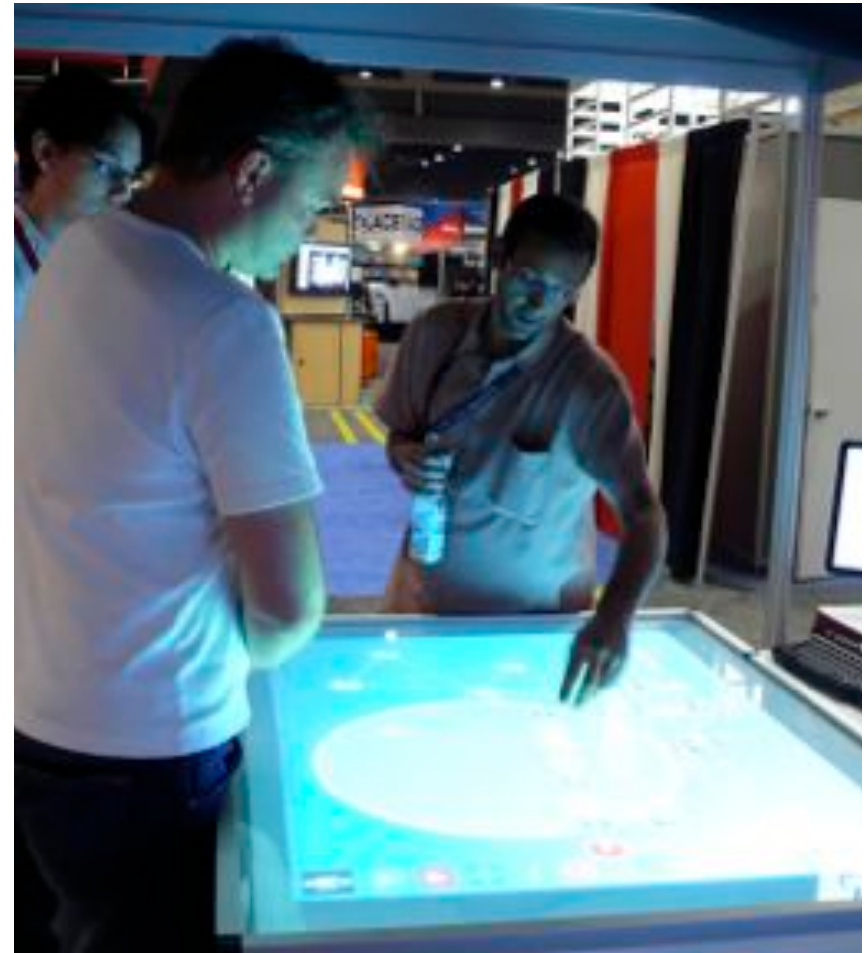
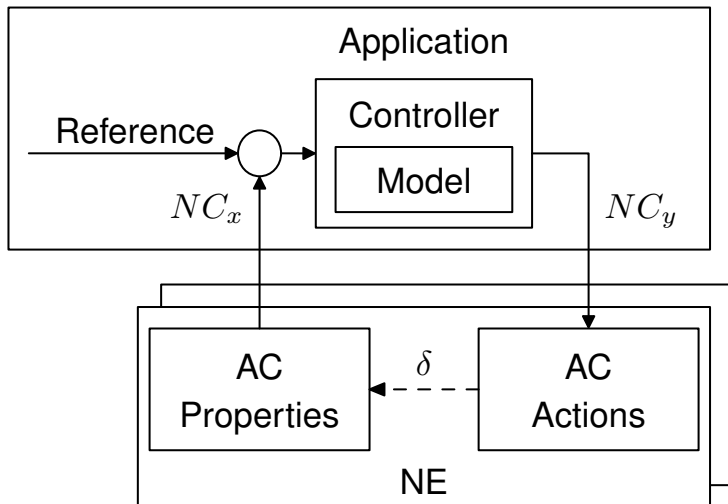
The network is virtualized as a collection of resources
 UPVNs enable network resources to be programmed
 as part of the application

Mathematica interacts with virtualized networks using
 UPVNs and optimize network + computation



In the Intercloud virtual servers and networks become software

- Virtual Internets adapt to the environment, grow to demand, iterate to specific designs
- Network support for application specific interconnections are merely optimizations: Openflow, active networks, cisco distributed switch
- But how to control the control loop?



Interactive Networks

- SuperComputing 2008
- SuperComputing 2009
(in programmable Grid networks demo)
- SuperComputing 2011
- Next Generation
Telecom networks
workshop 2011
- LHCONE architecture
workshop 2011



Interactive Networks

Rudolf Strijkers^{1,2}

Marc X. Makkes^{1,2}

Mihai Christea¹

Laurence Maller¹

Robert Belleman¹

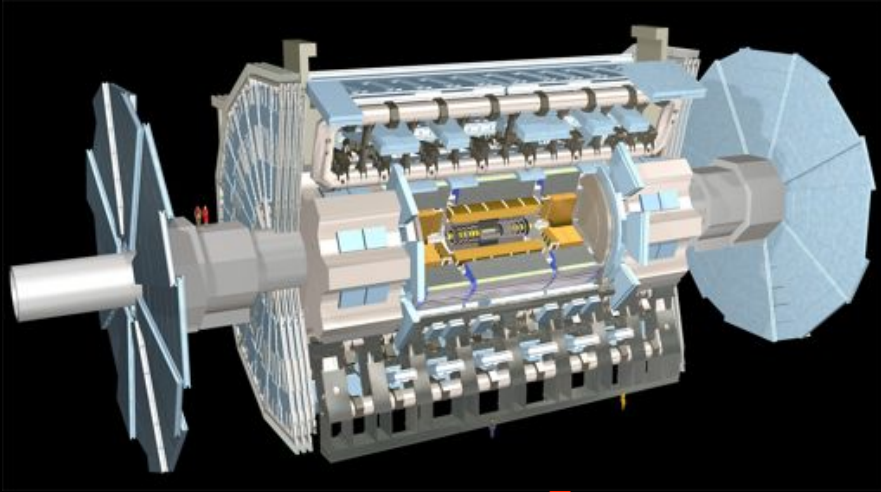
Cees de Laat¹

Robert Meijer^{1,2}

¹ University of Amsterdam, Amsterdam The Netherlands

² TNO Information and Communication Technology, Groningen, The Netherlands

SNE @ UvA



Ijkdijk/Urban Flood

Medical

LifeWatch/ENVRI

CosmoGrid/eVLBI

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud

Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

Topology/Architecture

Optical Photonic

X X

X

X

X X

X X

X

X

X

X

X

X

X

X

X

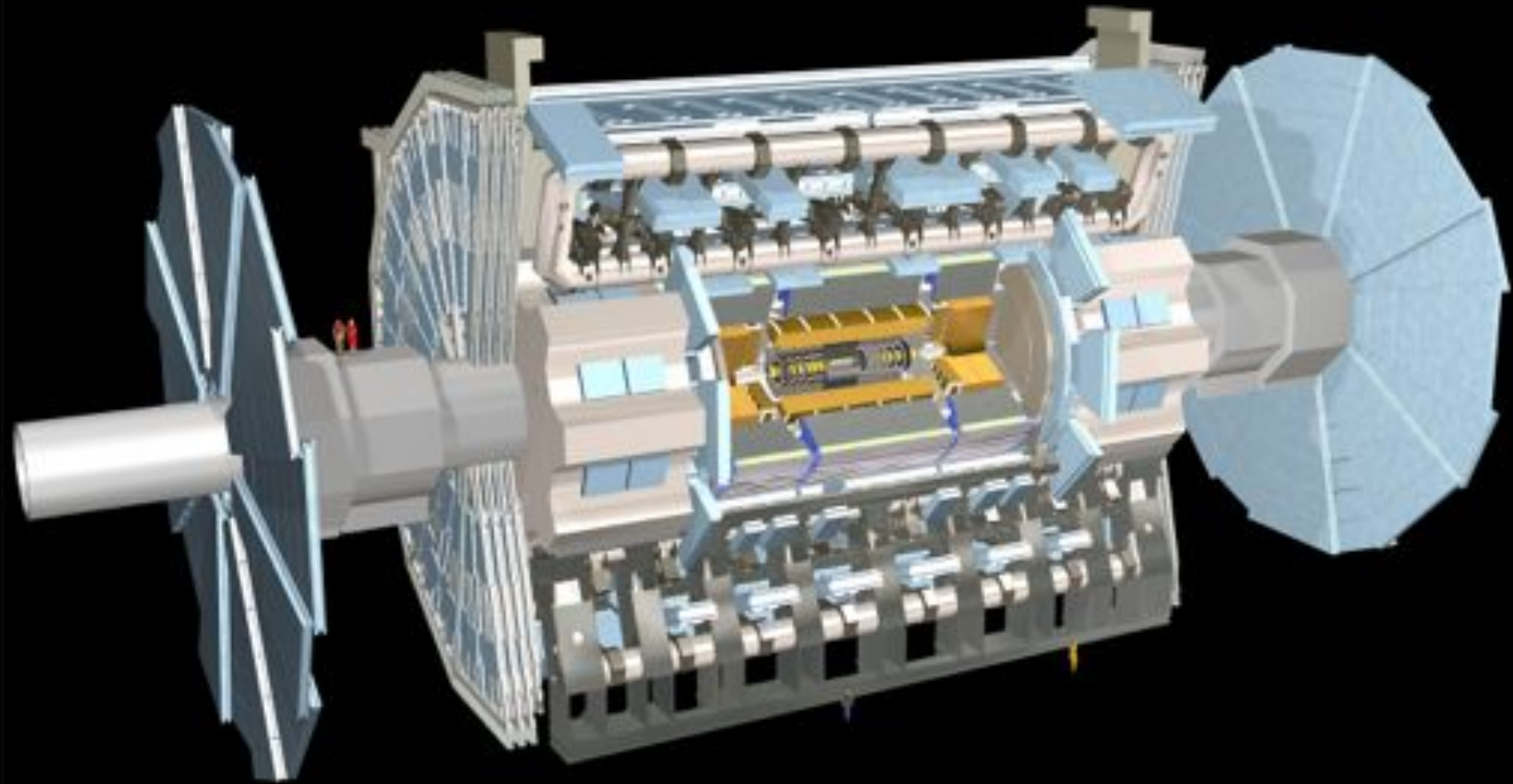
X

X

X

X

ATLAS detector @ CERN Geneve



ATLAS detector @ CERN Geneve





LHC Data Grid Hierarchy

CMS as example, Atlas is similar

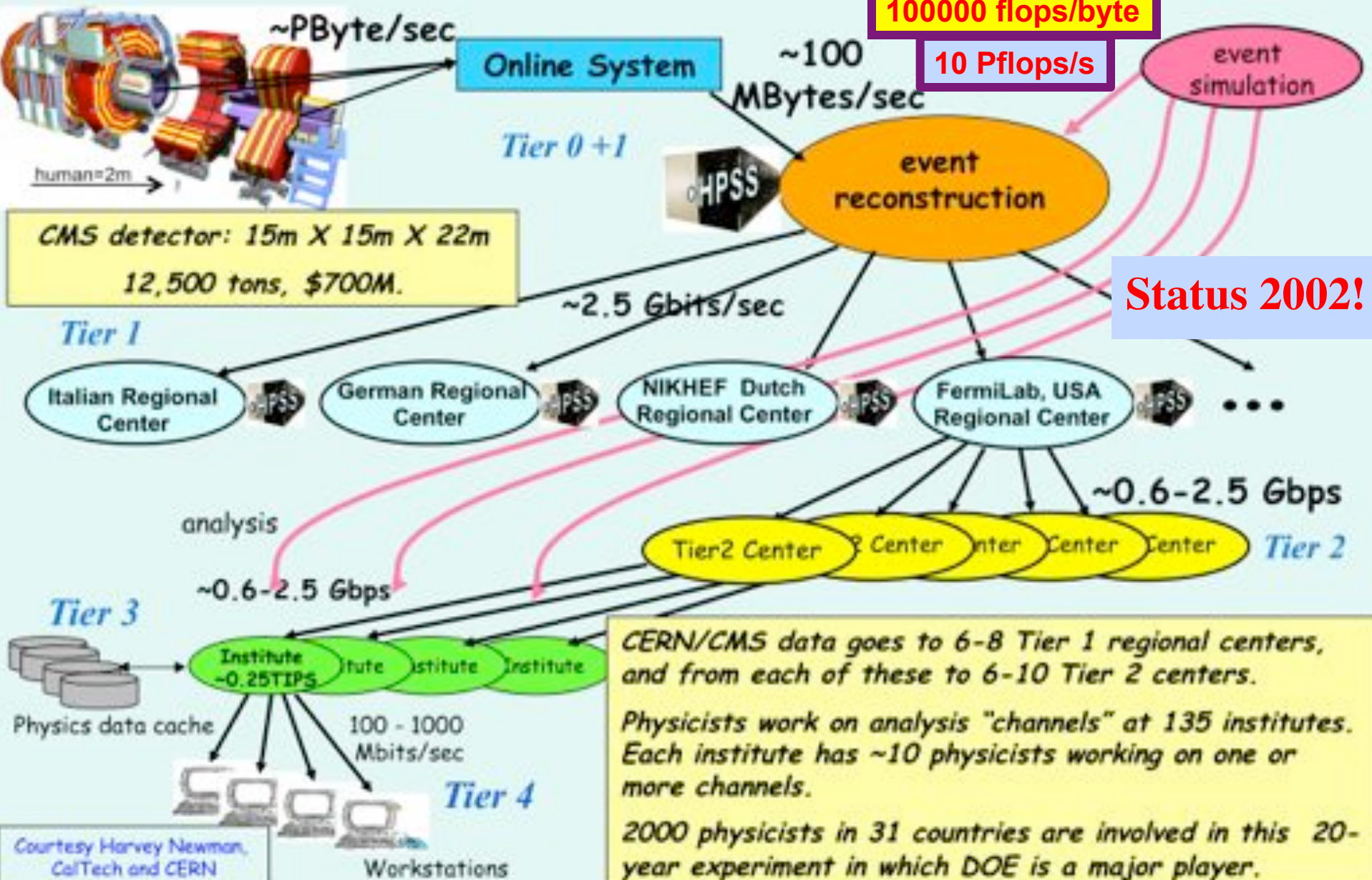
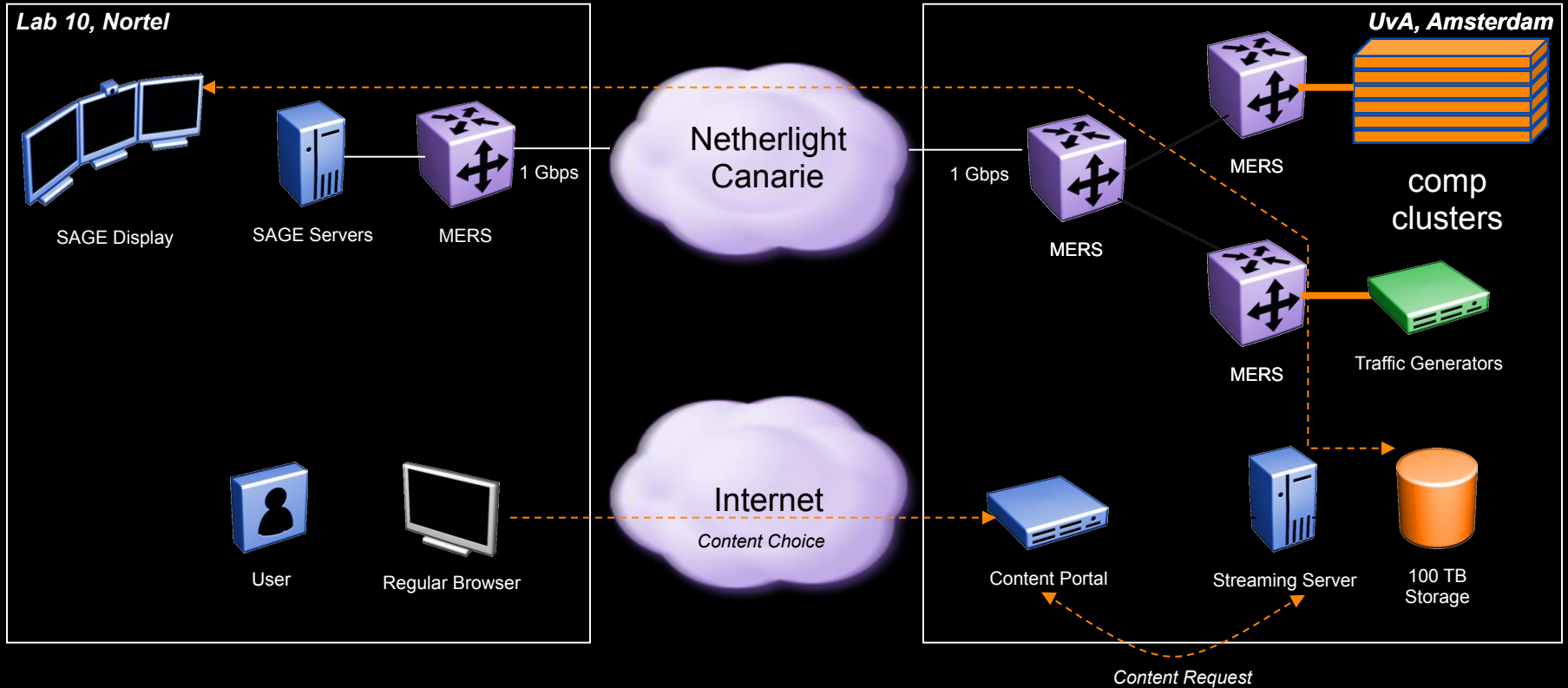




Diagram for SAGE video streaming to ATS

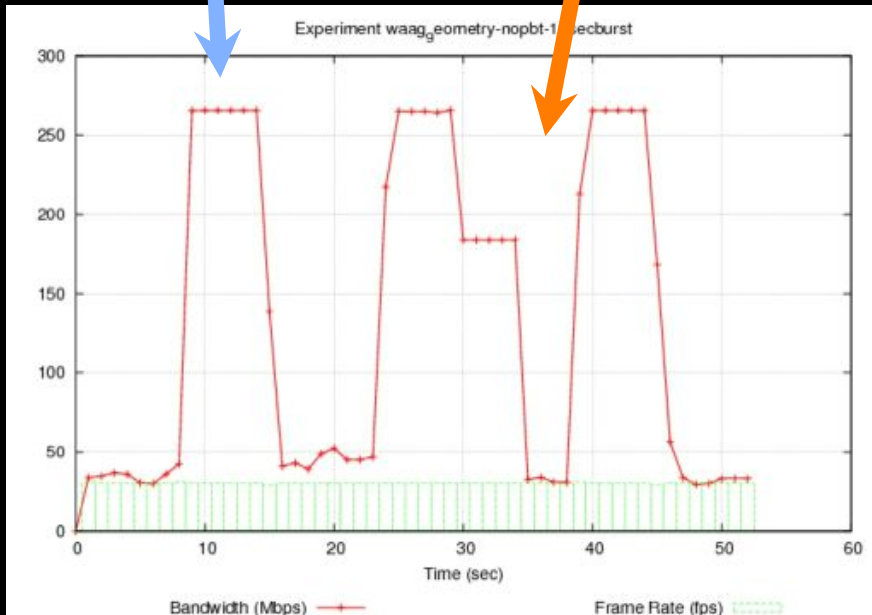


Experimental Data

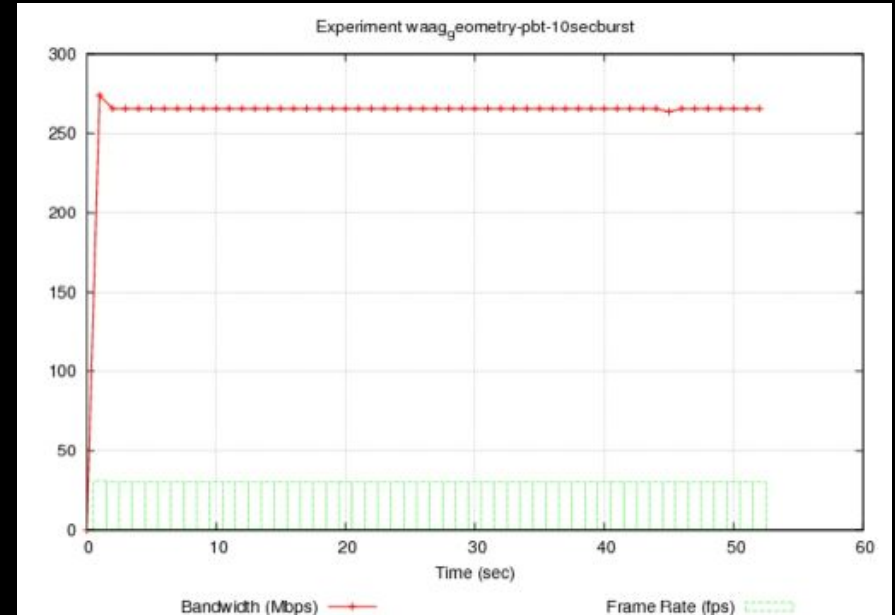


Sage without
background
traffic

Sage with
background
traffic



10 Second Traffic
bursts with No PBT



10 Second Traffic
bursts with PBT

PBT is SIMPLE and EFFECTIVE
technology to build a shared Media-Ready Network



Alien light From idea to realisation!

40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure



Alien wavelength advantages

- Direct connection of customer equipment^[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service^[2] → time savings
- Support of different modulation formats^[3] → extend network lifetime

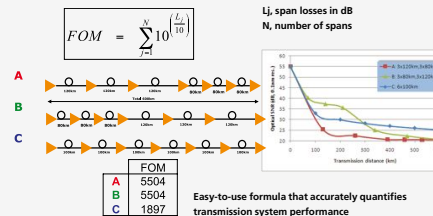
Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

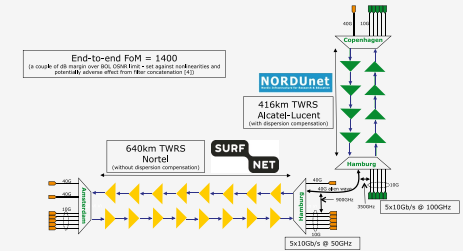
New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

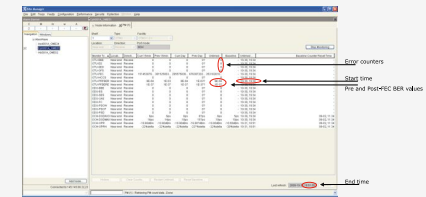


Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3,0 · 10⁻¹⁶

Conclusions

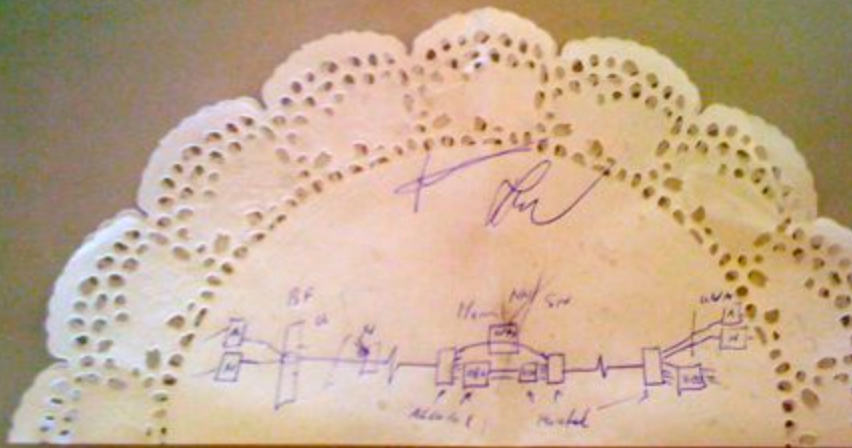
- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10⁻¹⁵) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.



REFERENCES
ACKNOWLEDGEMENTS

[1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL. OFC2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09
[3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECCO2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION
WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

Alien light From idea to realisation!



40Gb/s alien wavelength transmission via a multi-vendor 10Gb/s DWDM infrastructure



Alien wavelength advantages

- Direct connection of customer equipment^[1] → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service^[2] → time savings
- Support of different modulation formats^[3] → extend network lifetime

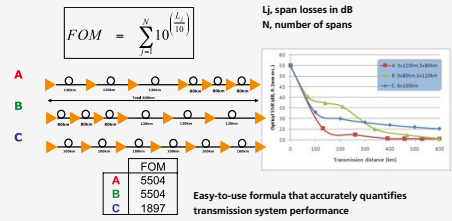
Alien wavelength challenges

- Complex end-to-end optical path engineering in terms of linear (i.e. OSNR, dispersion) and non-linear (FWM, SPM, XPM, Raman) transmission effects for different modulation formats.
- Complex interoperability testing.
- End-to-end monitoring, fault isolation and resolution.
- End-to-end service activation.

In this demonstration we will investigate the performance of a 40Gb/s PM-QPSK alien wavelength installed on a 10Gb/s DWDM infrastructure.

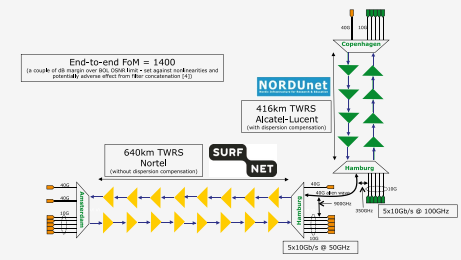
New method to present fiber link quality, FoM (Figure of Merit)

In order to quantify optical link grade, we propose a new method of representing system quality: the FOM (Figure of Merit) for concatenated fiber spans.

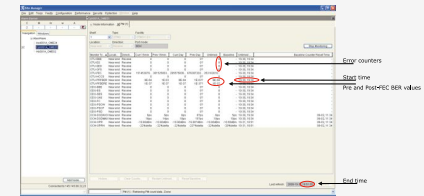


Transmission system setup

JOINT SURFnet/NORDUnet 40Gb/s PM-QPSK alien wavelength DEMONSTRATION.



Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3,0 · 10⁻¹⁶

Conclusions

- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10⁻¹⁵) during a 23 hour period.
- More detailed system performance analysis will be presented in an upcoming paper.



REFERENCES [1] "OPERATIONAL SOLUTIONS FOR AN OPEN DWDM LAYER", O. GERSTEL ET AL. OFC2009 | [2] "AT&T OPTICAL TRANSPORT SERVICES", BARBARA E. SMITH, OFC'09 [3] "OPEX SAVINGS OF ALL-OPTICAL CORE NETWORKS", ANDREW LORD AND CARL ENGINEER, ECCO2009 | [4] NORTEL/SURFNET INTERNAL COMMUNICATION
ACKNOWLEDGEMENTS WE ARE GRATEFUL TO NORDUNET FOR PROVIDING US WITH BANDWIDTH ON THEIR DWDM LINK FOR THIS EXPERIMENT AND ALSO FOR THEIR SUPPORT AND ASSISTANCE DURING THE EXPERIMENTS. WE ALSO ACKNOWLEDGE TELINDUS AND NORTEL FOR THEIR INTEGRATION WORK AND SIMULATION SUPPORT

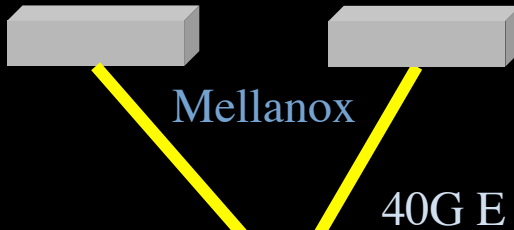
ClearStream @ TNC2011

Setup codename:
FlightCees



UvA

iPerf I7 3.2 GHz Q-core iPerf Amd Ph II 3.6 GHz HexC



Copenhagen

iPerf 2* dual 2.8 GHz Q-core iPerf



CERN

CIENA DWDM

Hamburg

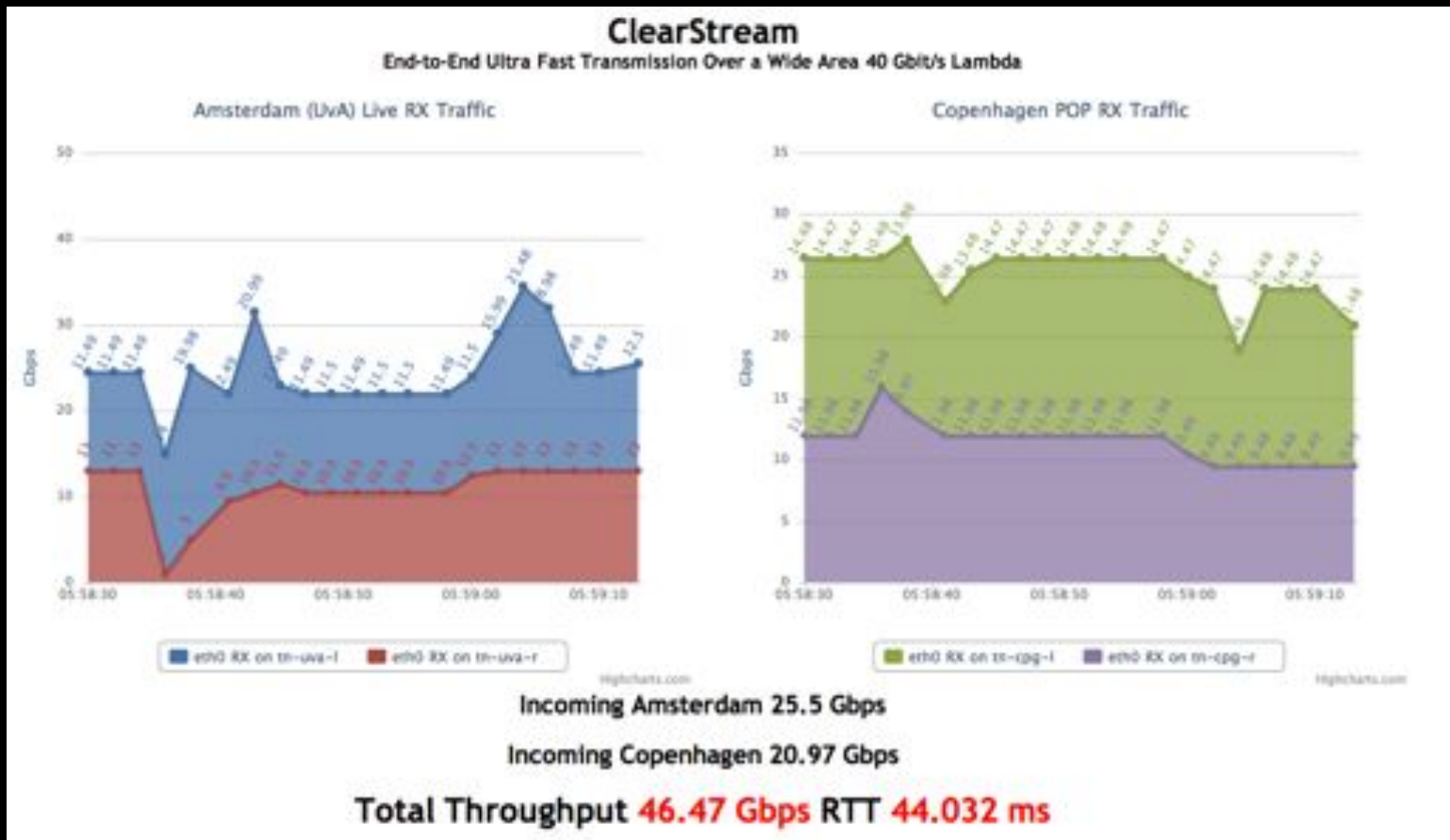
Alcatel DWDM



Amsterdam – Geneva (CERN) – Copenhagen – 4400 km (2700 km alien light)

Visit CIENA Booth

surf to <http://tnc11.delaat.net>



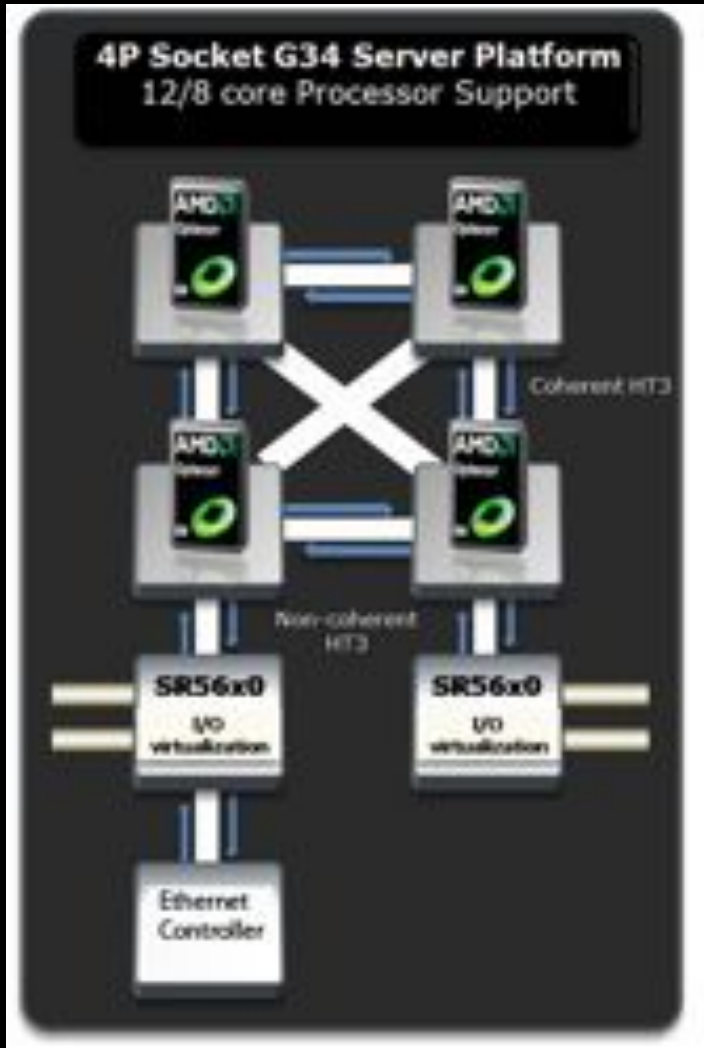
Results (rtt = 17 ms)

- Single flow iPerf 1 core -> 21 Gbps
- Single flow iPerf 1 core <> -> 15+15 Gbps
- Multi flow iPerf 2 cores -> 25 Gbps
- Multi flow iPerf 2 cores <> -> 23+23 Gbps
- DiViNe <> -> 11 Gbps
- Multi flow iPerf + DiVine -> 35 Gbps
- Multi flow iPerf + DiVine <> -> 35 + 35 Gbps

Performance Explained

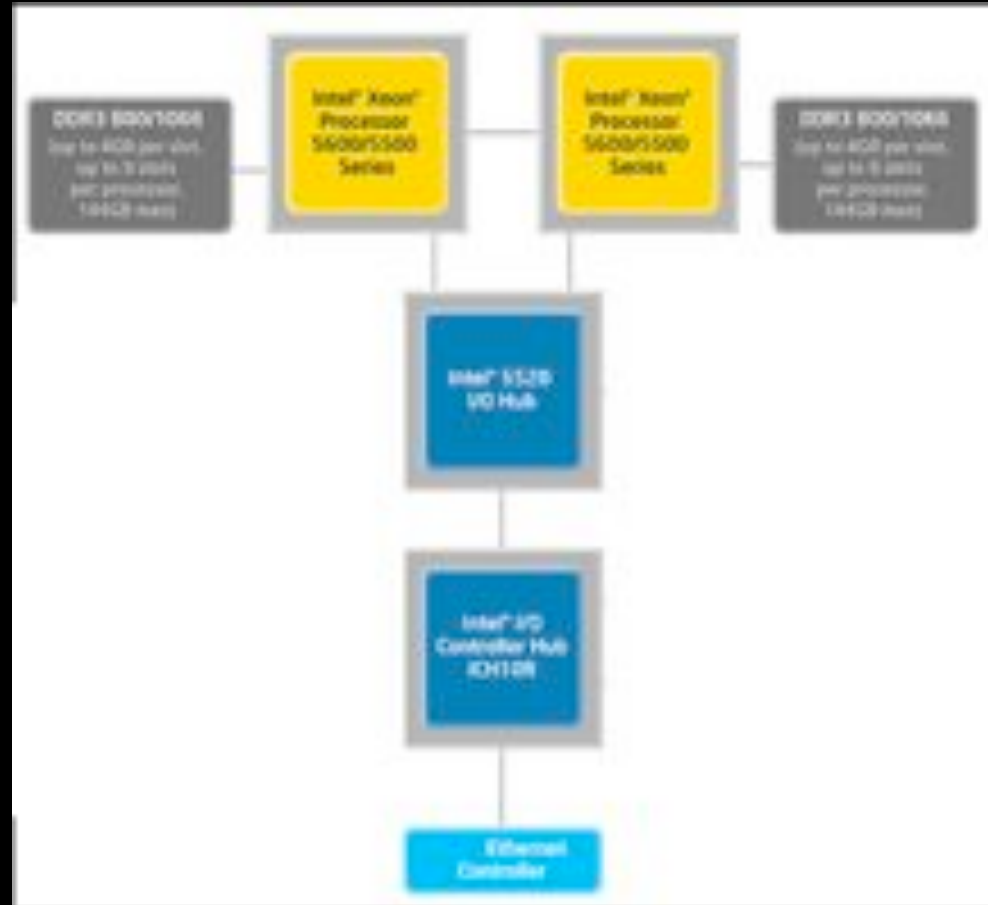
- Mellanox 40GE card is PCI-E 2.0 8x (5GT/s)
- 40Gbit/s raw throughput but
- PCI-E is a network-like protocol
 - 8/10 bit encoding -> 25% overhead -> 32Gbit/s maximum data throughput
 - Routing information
- Extra overhead from IP/Ethernet framing
- Server architecture matters!
 - 4P system performed worse in multithreaded iperf

Server Architecture



DELL R815

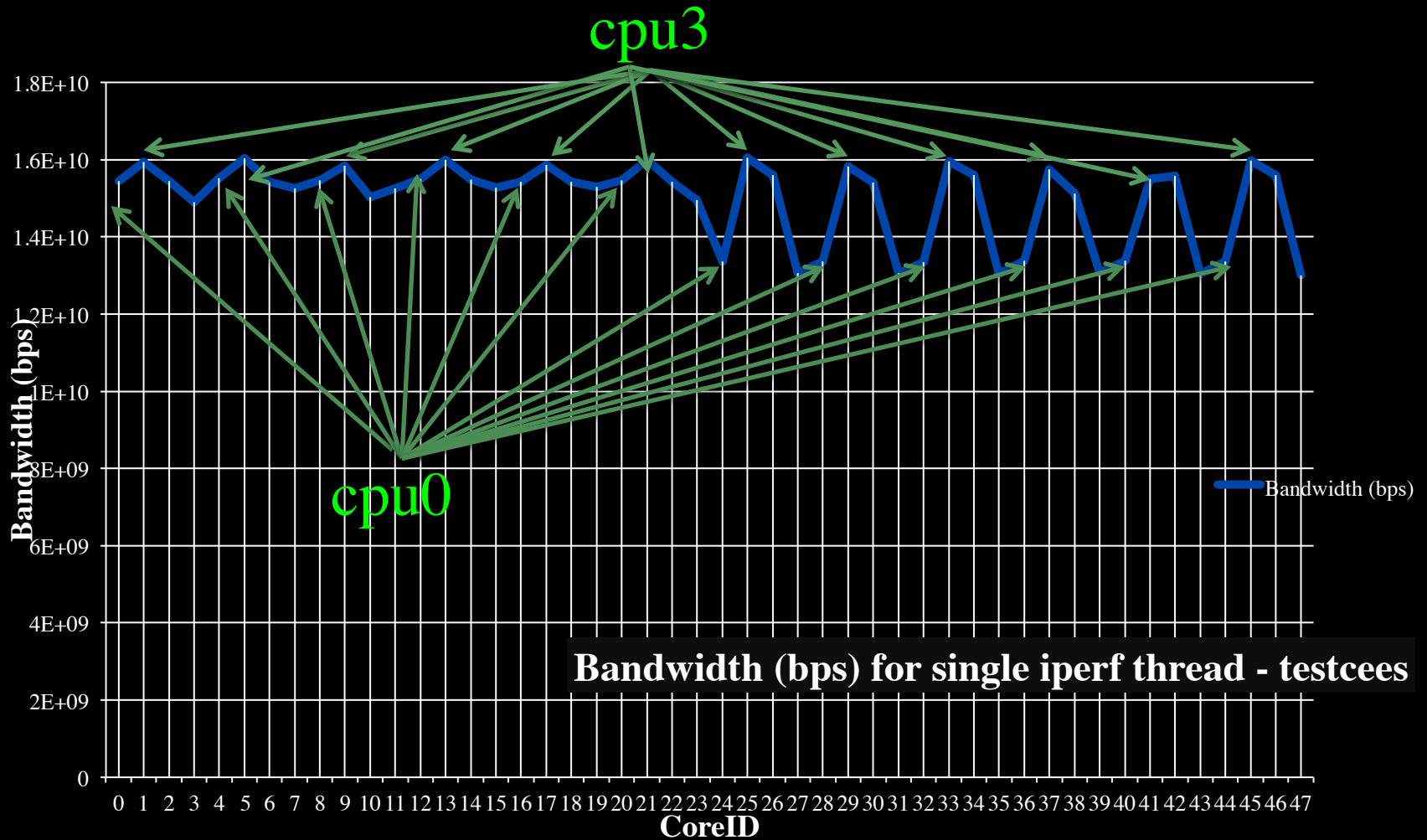
4 x AMD Opteron 6100



Supermicro X8DTT-HIBQF

2 x Intel Xeon

CPU Topology benchmark



We used numactl to bind iperf to cores



SLP Map 2011 - Global Lambda-optimized Paths Visualization by Todd Patterson, NCSA, University of Illinois at Urbana-Champaign. Path Computation by Bertil O. Steen, University of Illinois at Chicago. Network Protocol by JPO Corporation, NCSA. Both Parties: redwood.llnwd.net - www.glif.org



GLIF 2011

**Visualization courtesy of Bob Patterson, NCSA
Data collection by Maxine Brown.**



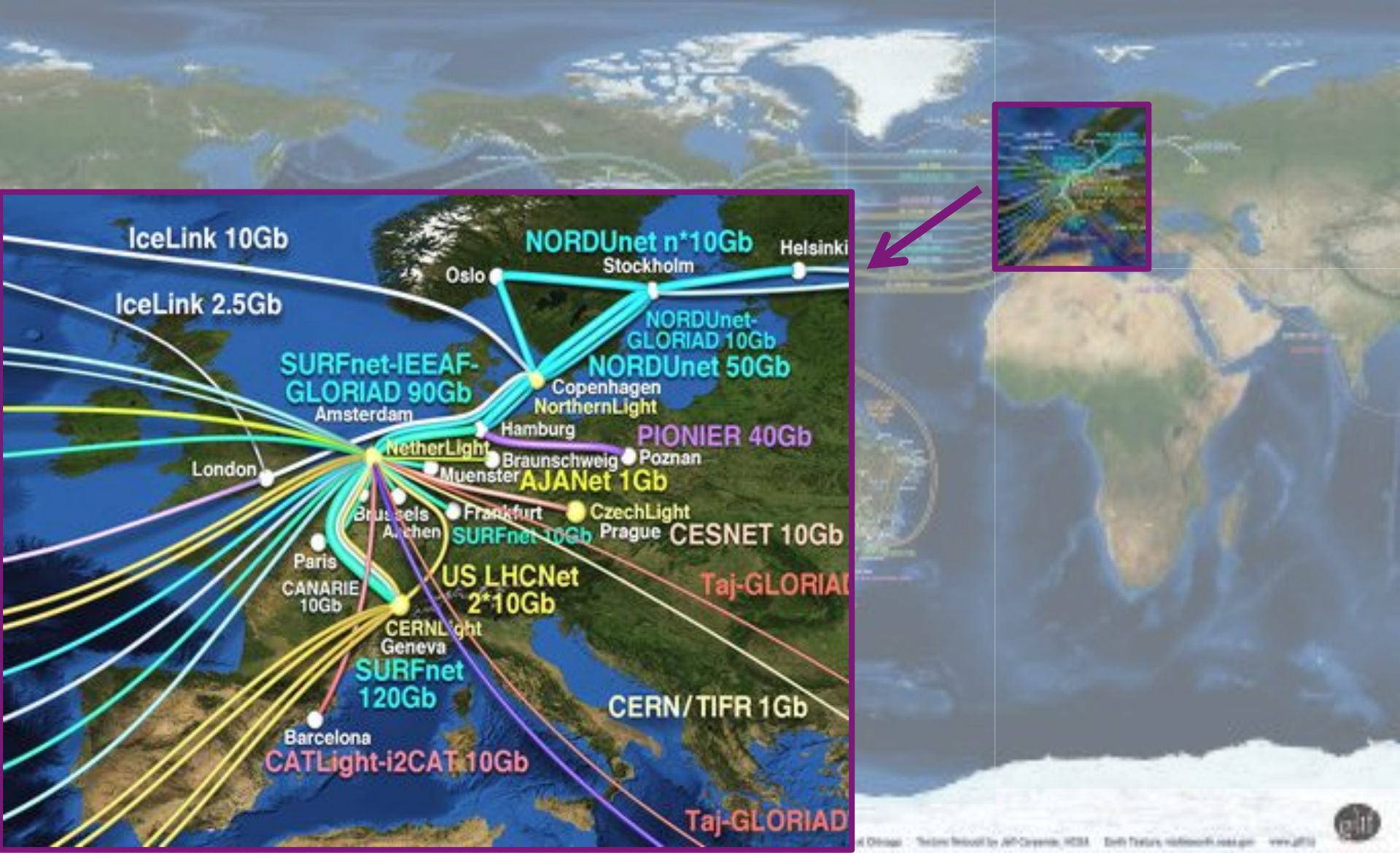
We investigate:
complex networks!



for



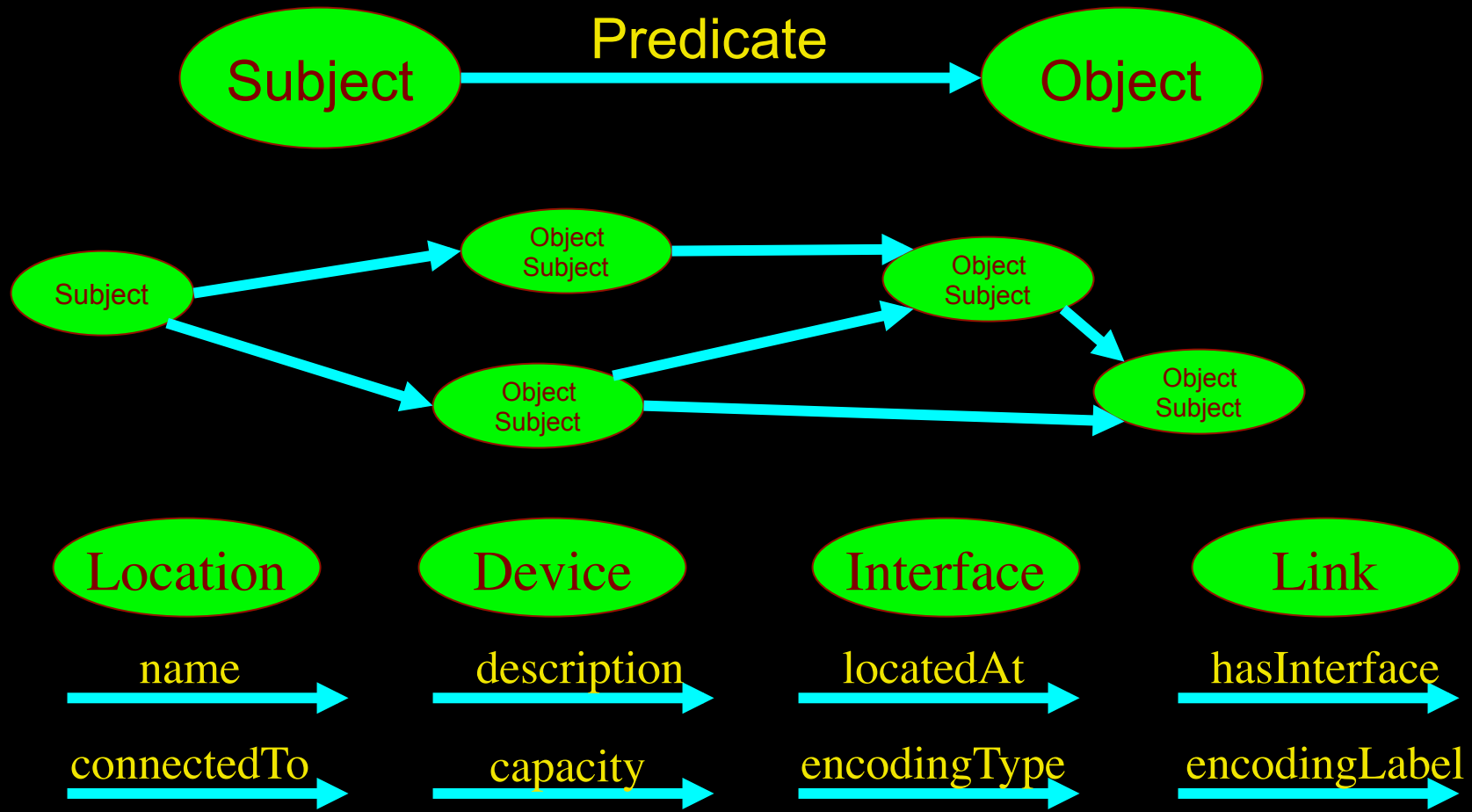
The GLIF – lightpaths around the world



LinkedIn for Infrastructure



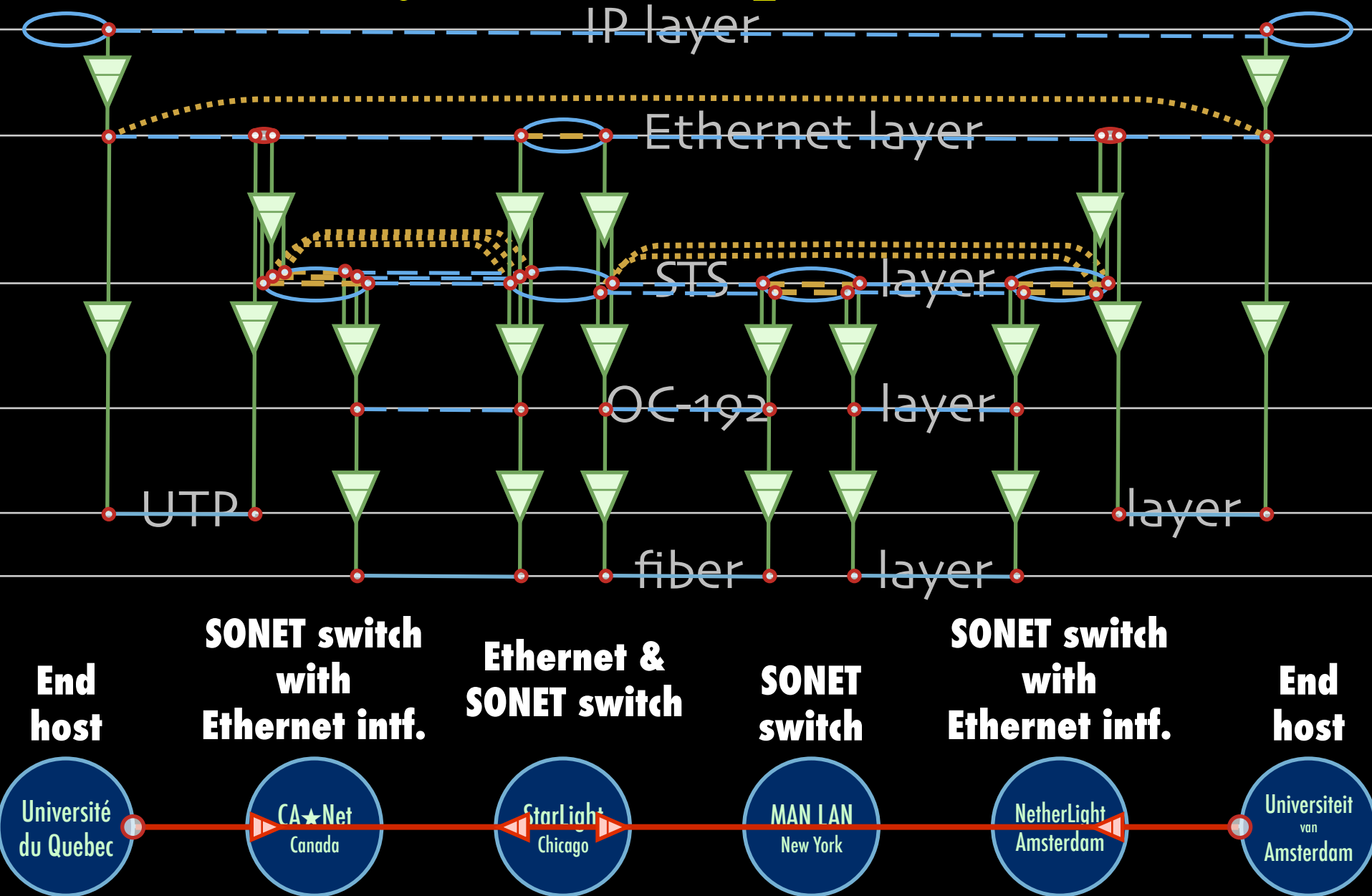
- From semantic Web / Resource Description Framework.
- The RDF uses XML as an interchange syntax.
- Data is described by triplets (Friend of a Friend):



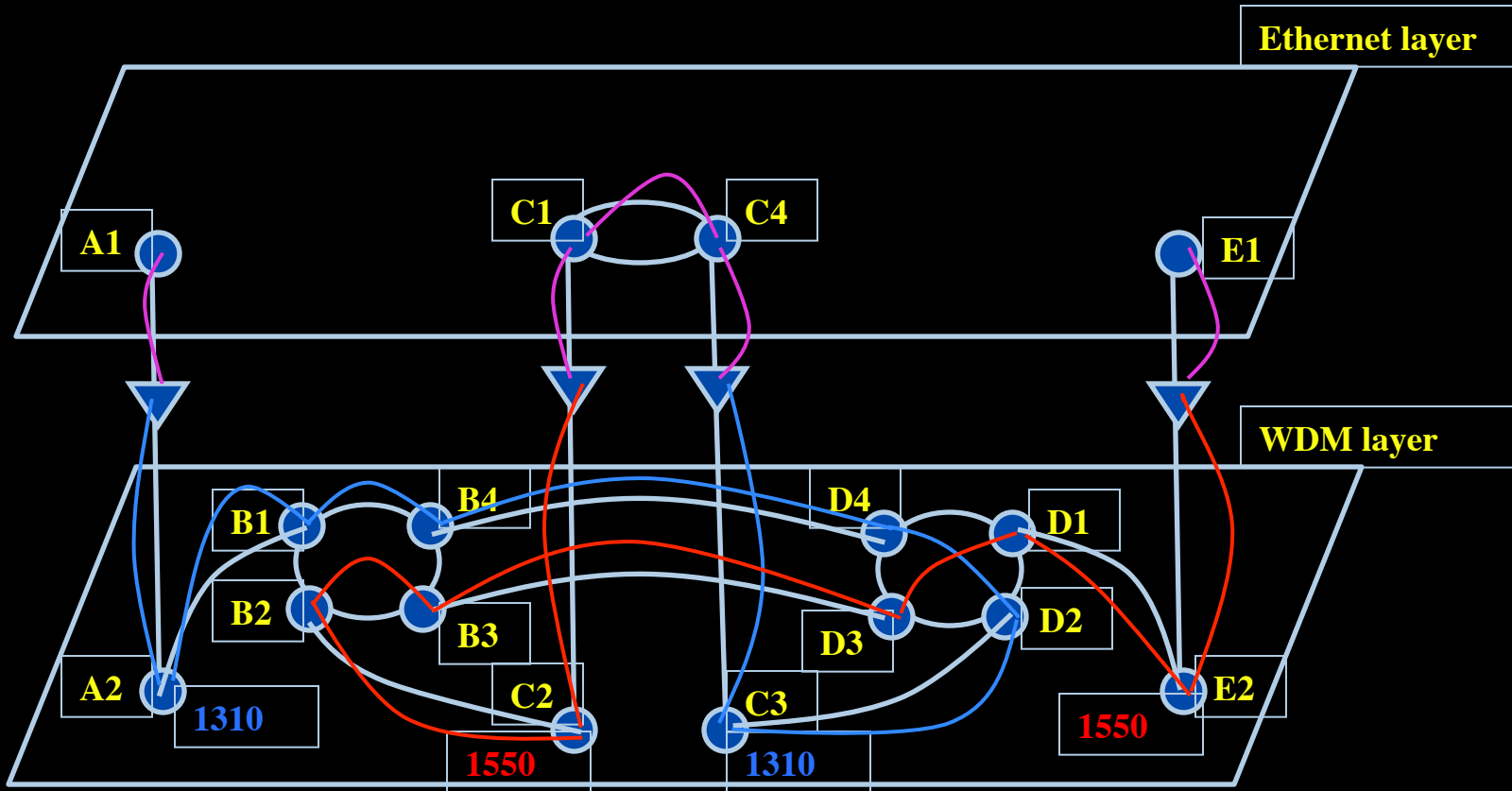
NetherLight in RDF

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:ndl="http://www.science.uva.nl/research/air/ndl#">
  <!-- Description of Netherlight -->
  <ndl:Location rdf:about="#Netherlight">
    <ndl:name>Netherlight Optical Exchange</ndl:name>
  </ndl:Location>
  <!-- TDM3.amsterdam1.netherlight.net -->
  <ndl:Device rdf:about="#tdm3.amsterdam1.netherlight.net">
    <ndl:name>tdm3.amsterdam1.netherlight.net</ndl:name>
    <ndl:locatedAt rdf:resource="#amsterdam1.netherlight.net"/>
    <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/1"/>
    <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/3"/>
    <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:501/4"/>
    <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:503/1"/>
    <ndl:hasInterface rdf:resource="#tdm3.amsterdam1.netherlight.net:503/2"/>
    <!-- all the interfaces of TDM3.amsterdam1.netherlight.net -->
    <ndl:Interface rdf:about="#tdm3.amsterdam1.netherlight.net:501/1">
      <ndl:name>tdm3.amsterdam1.netherlight.net:POS501/1</ndl:name>
      <ndl:connectedTo rdf:resource="#tdm4.amsterdam1.netherlight.net:5/1"/>
    </ndl:Interface>
    <ndl:Interface rdf:about="#tdm3.amsterdam1.netherlight.net:501/2">
      <ndl:name>tdm3.amsterdam1.netherlight.net:POS501/2</ndl:name>
      <ndl:connectedTo rdf:resource="#tdm1.amsterdam1.netherlight.net:12/1"/>
    </ndl:Interface>
```

Multi-layer descriptions in NDL



Multi-layer Network PathFinding

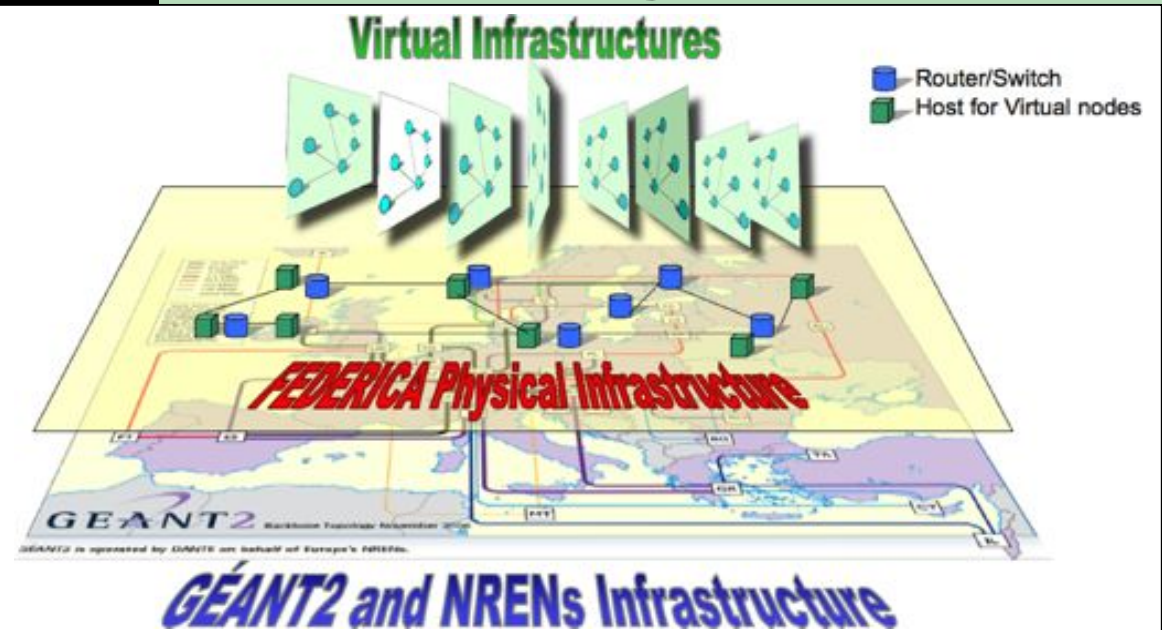
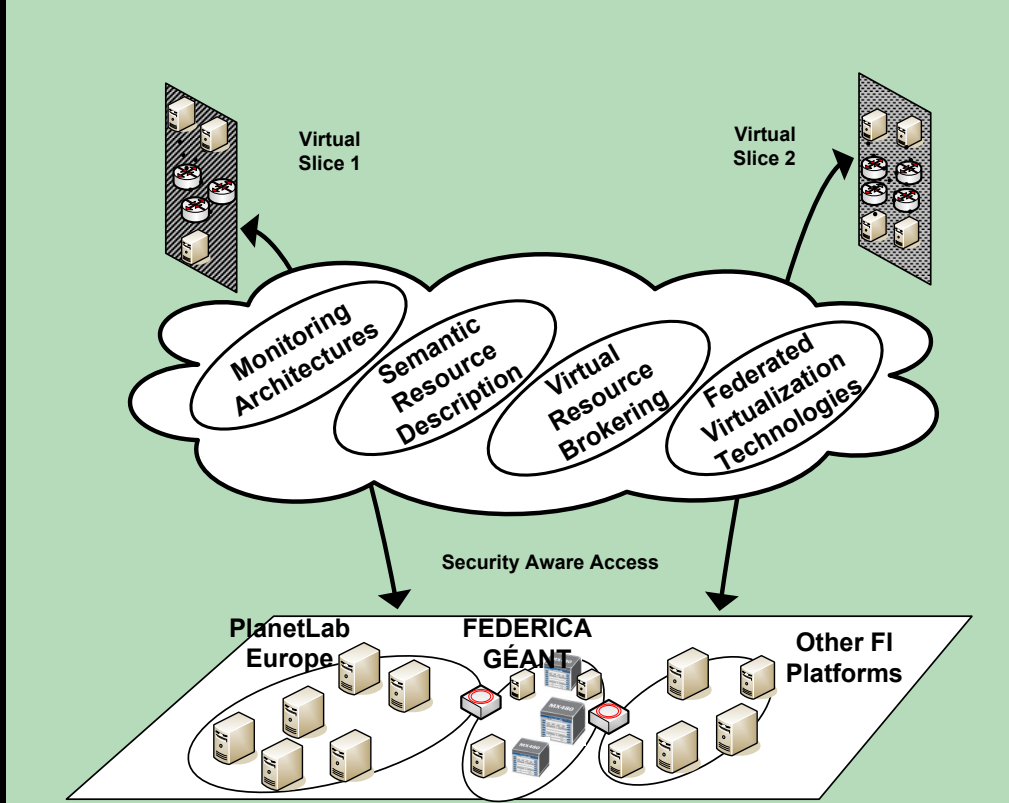
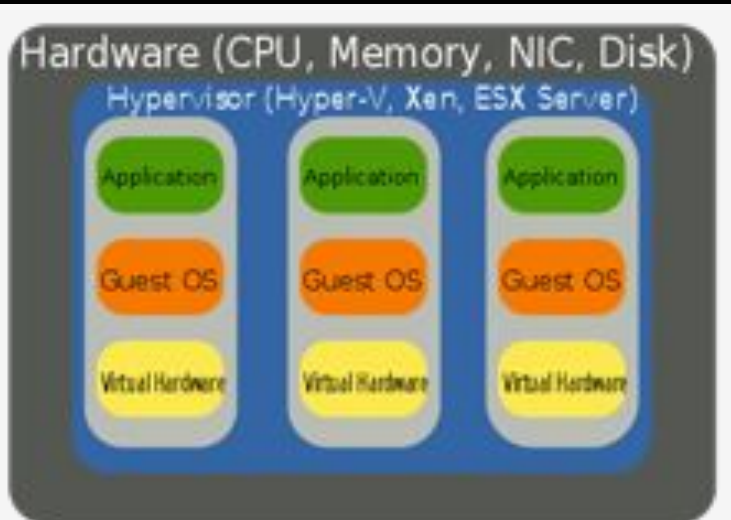


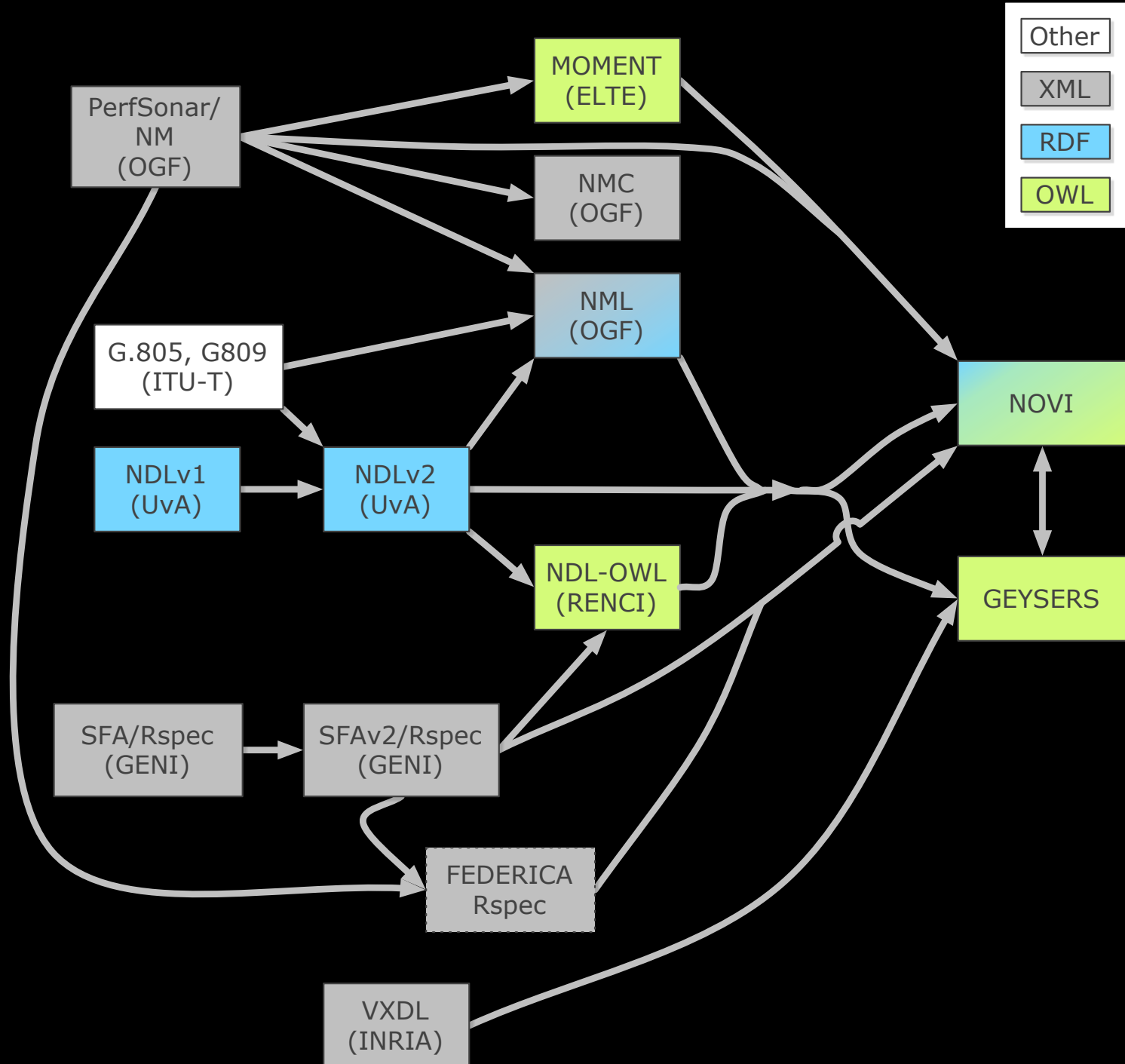
Path between interfaces A1 and E1:

A1-A2-B1-B4-D4-D2-C3-C4-C1-C2-B2-B3-D3-D1-E2-E1

Scaling: Combinatorial problem

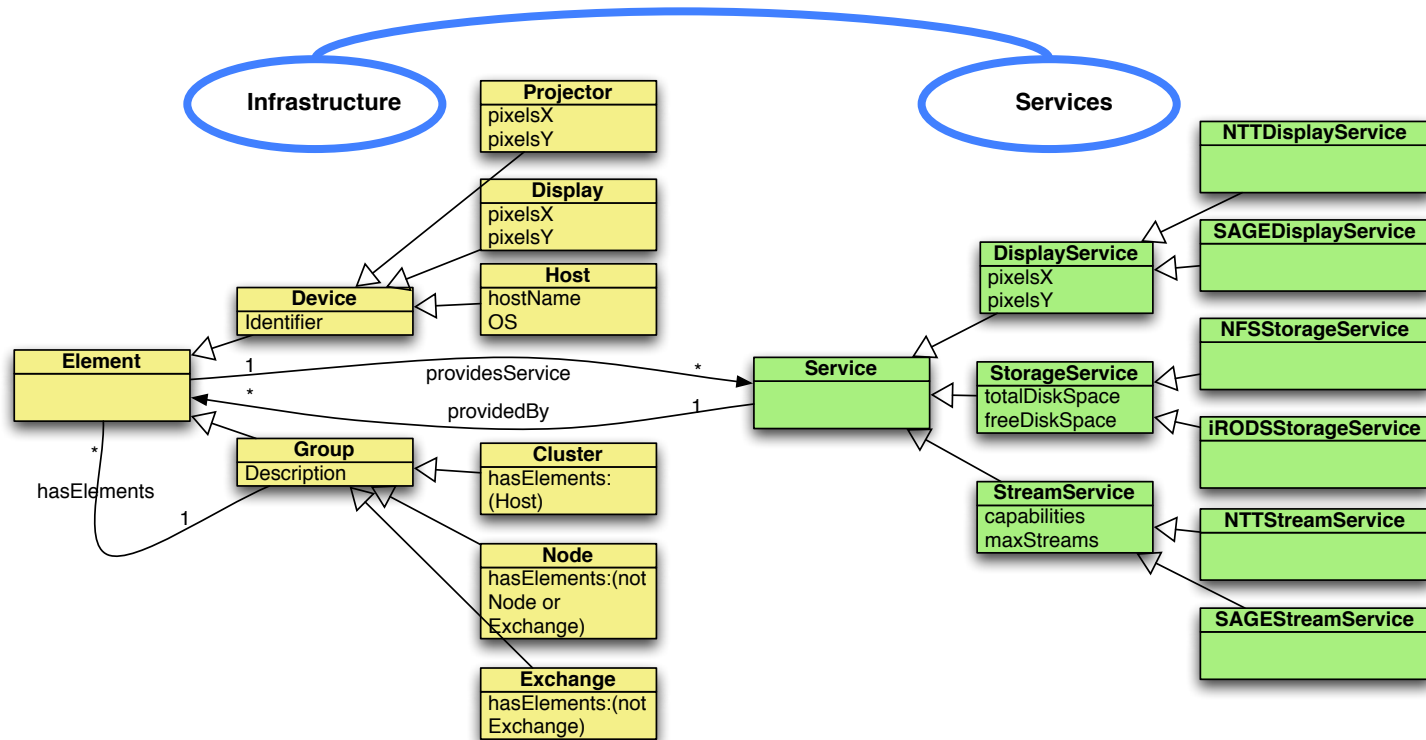
Virtualisatie van infrastructuur & QoS





Information Modeling

Define a common information model for **infrastructures** and **services**.
Base it on Semantic Web.



SNE @ UvA



Ijkdijk/Urban Flood
Medical
LifeWatch/ENVRI
CosmoGrid/eVLBI
CineGrid
EU-GN3/NOVI/Geysers
SURFnet/GLIF/Cloud

Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

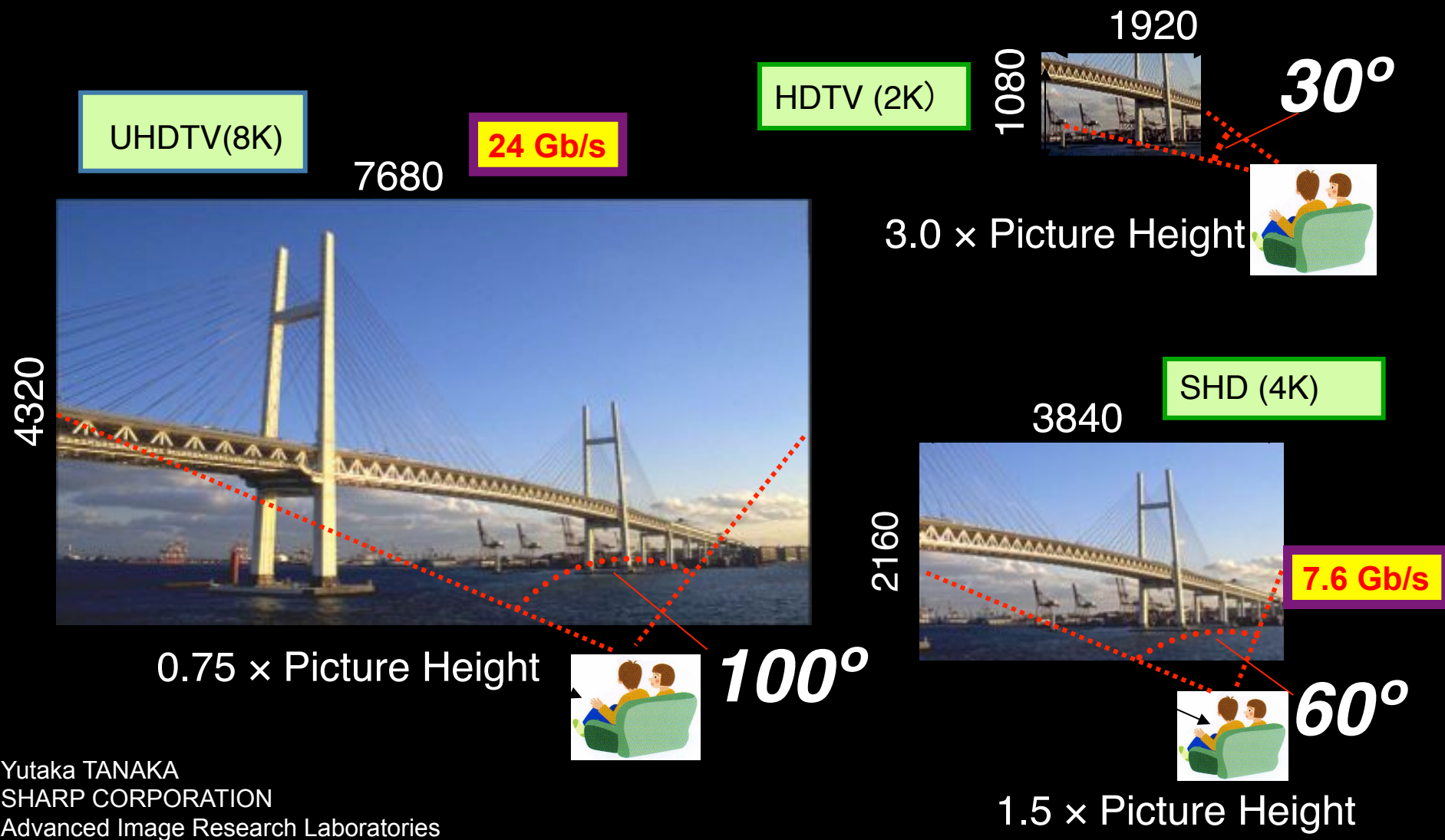
Topology/Architecture

Optical Photonic

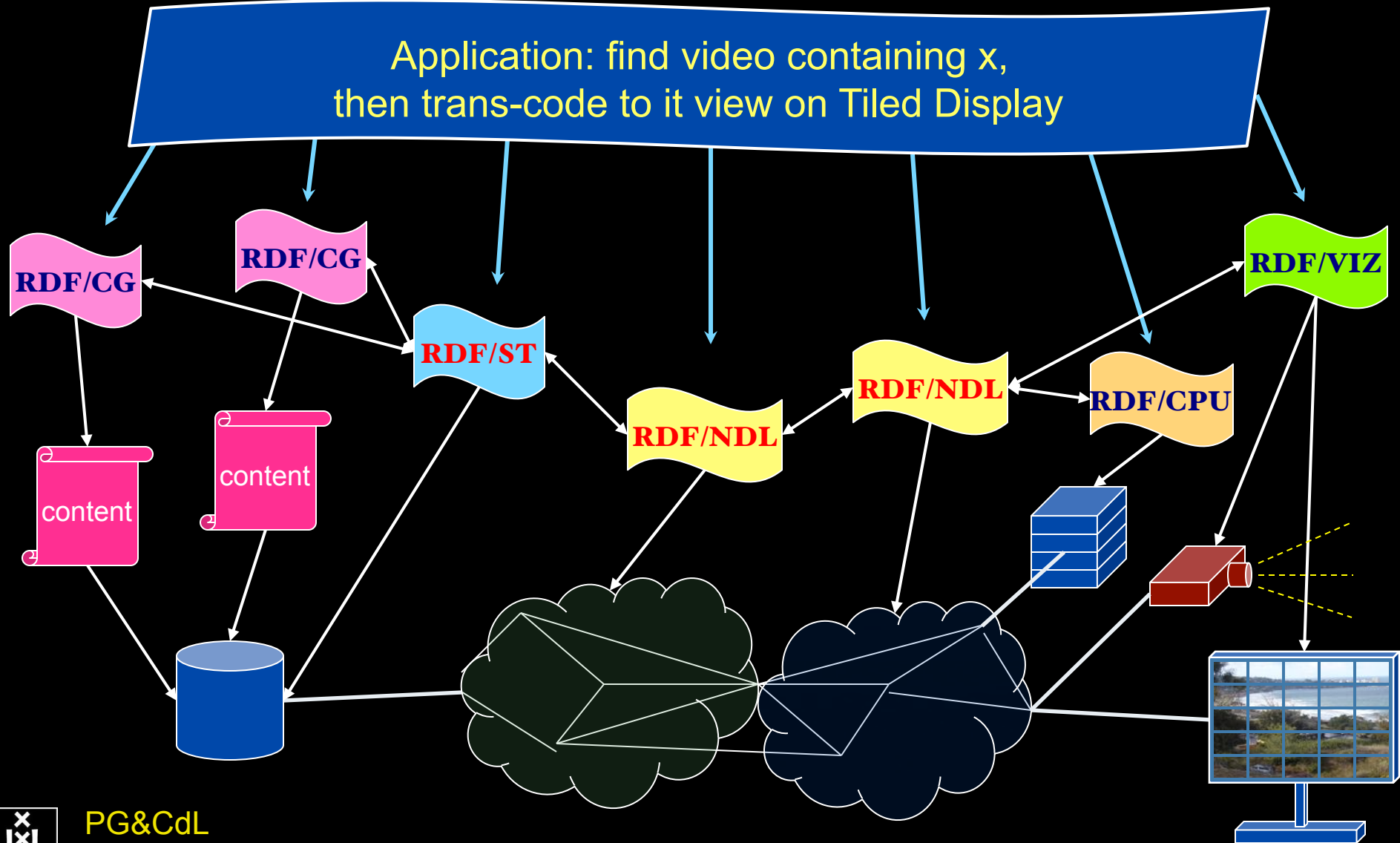
						X	X
		X				X	
		X	X		X	X	
	X		X				
	X		X	X		X	
		X		X	X		
		X	X		X		

Why is more resolution is better?

1. More Resolution Allows Closer Viewing of Larger Image
2. Closer Viewing of Larger Image Increases Viewing Angle
3. Increased Viewing Angle Produces Stronger Emotional Response



RDF describing Infrastructure





Why?



I want to:

“Show Big Bug Bunny in 4K on my Tiled Display using green Infrastructure”

- Big Bugs Bunny can be on multiple servers on the Internet.
- Movie may need processing / recoding to get to 4K for Tiled Display.
- Needs deterministic Green infrastructure for Quality of Experience.
- Consumer / Scientist does not want to know the underlying details.
→ His refrigerator also just works.

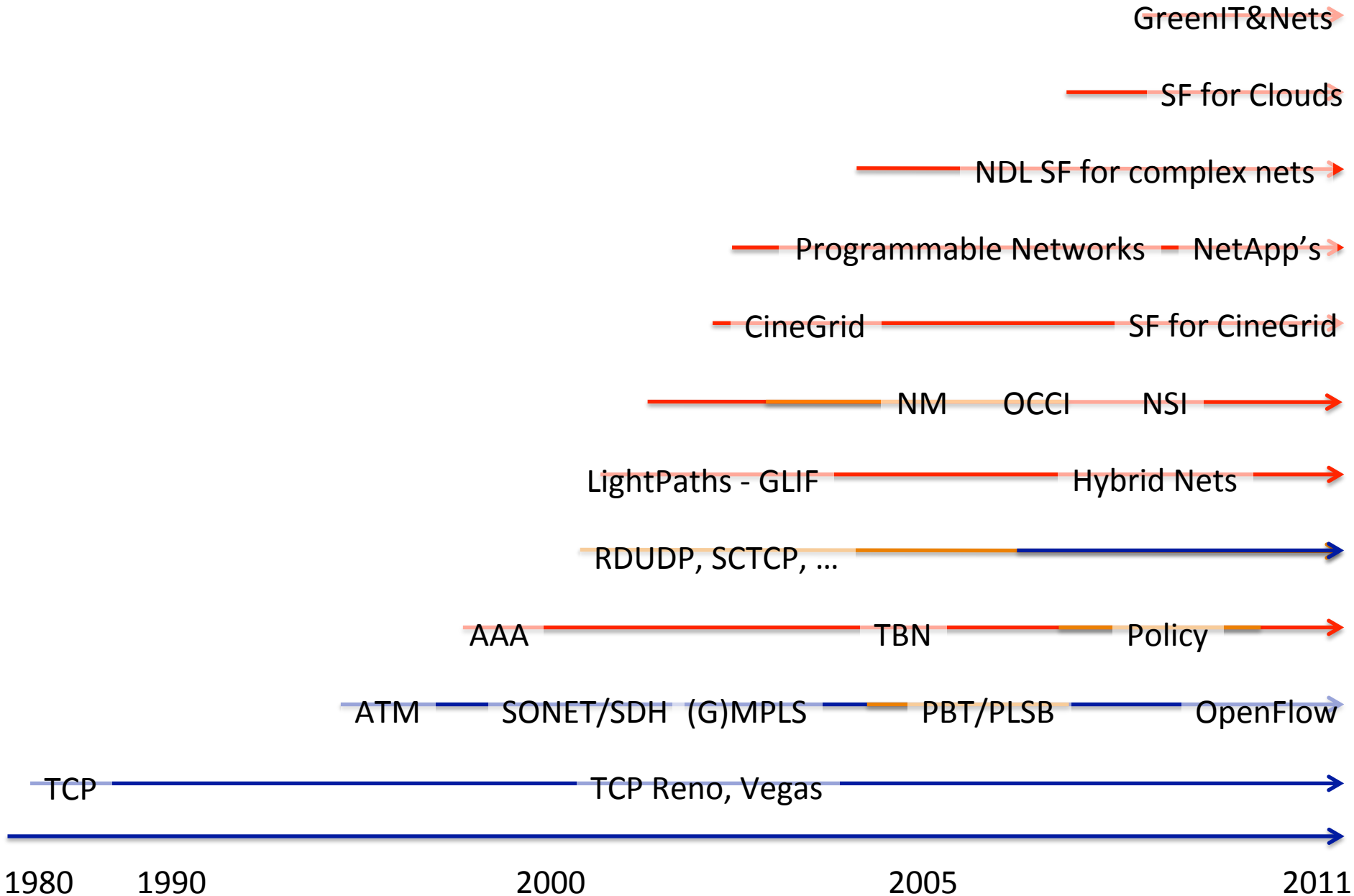
The Ten Problems with the Internet

1. **Energy Efficient Communication**
2. Separation of Identity and Address
3. Location Awareness
4. **Explicit Support for Client-Server Traffic and Distributed Services**
5. Person-to-Person Communication
6. Security
7. **Control, Management, and Data Plane separation**
8. **Isolation**
9. Symmetric/Asymmetric Protocols
10. **Quality of Service**

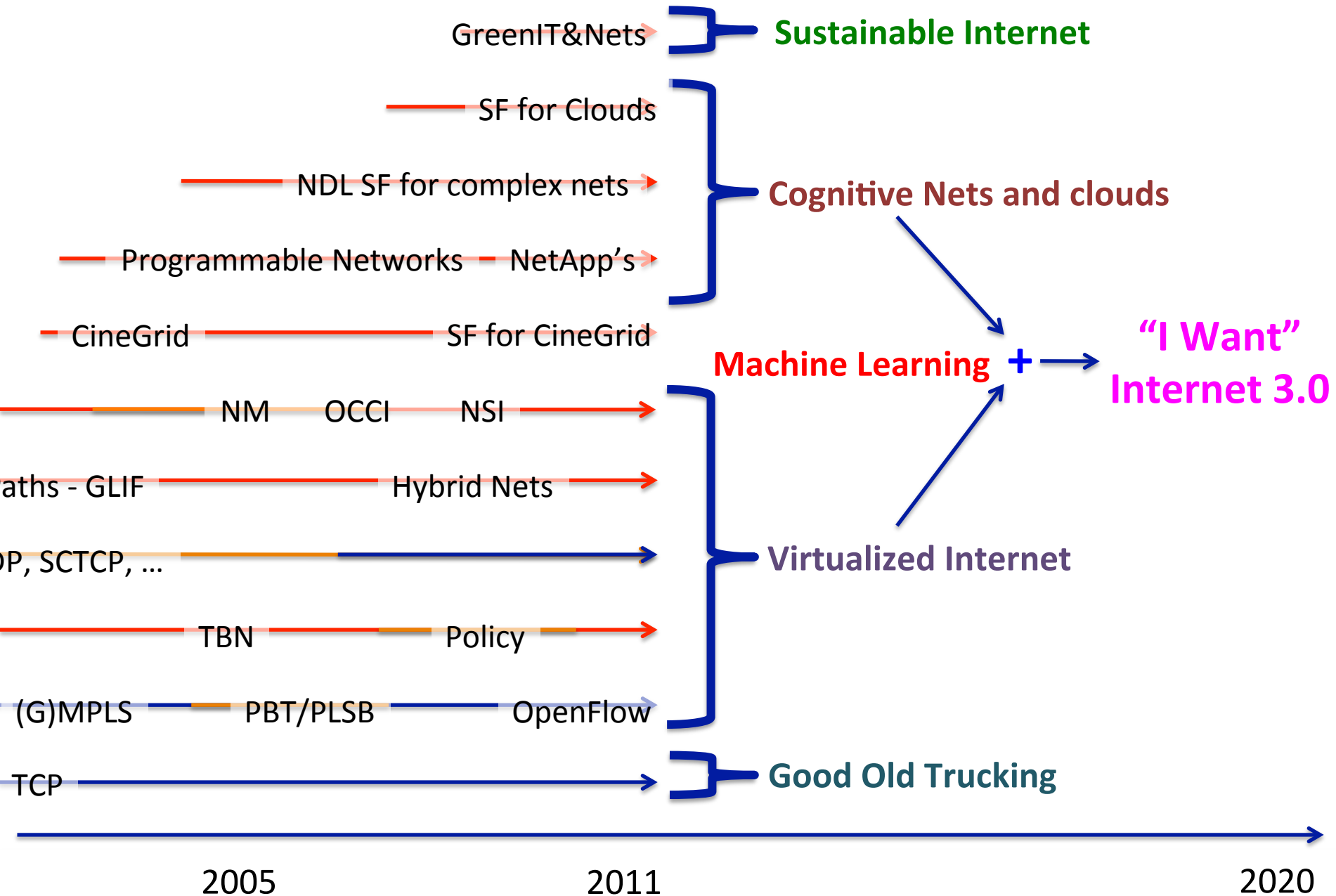
Nice to have:

- Global Routing with Local Control of Naming and Addressing
- **Real Time Services**
- **Cross-Layer Communication**
- Multicast
- Receiver Control
- Support for Data Aggregation and Transformation
- **Support for Streaming Data**
- **Virtualization**

TimeLine



TimeLine



TimeLine

• Sustainable Internet

• Cognitive Nets and clouds

• Machine Learning +

“I Want”
Internet 3.0

• Virtualized Internet

• Good Old Trucking

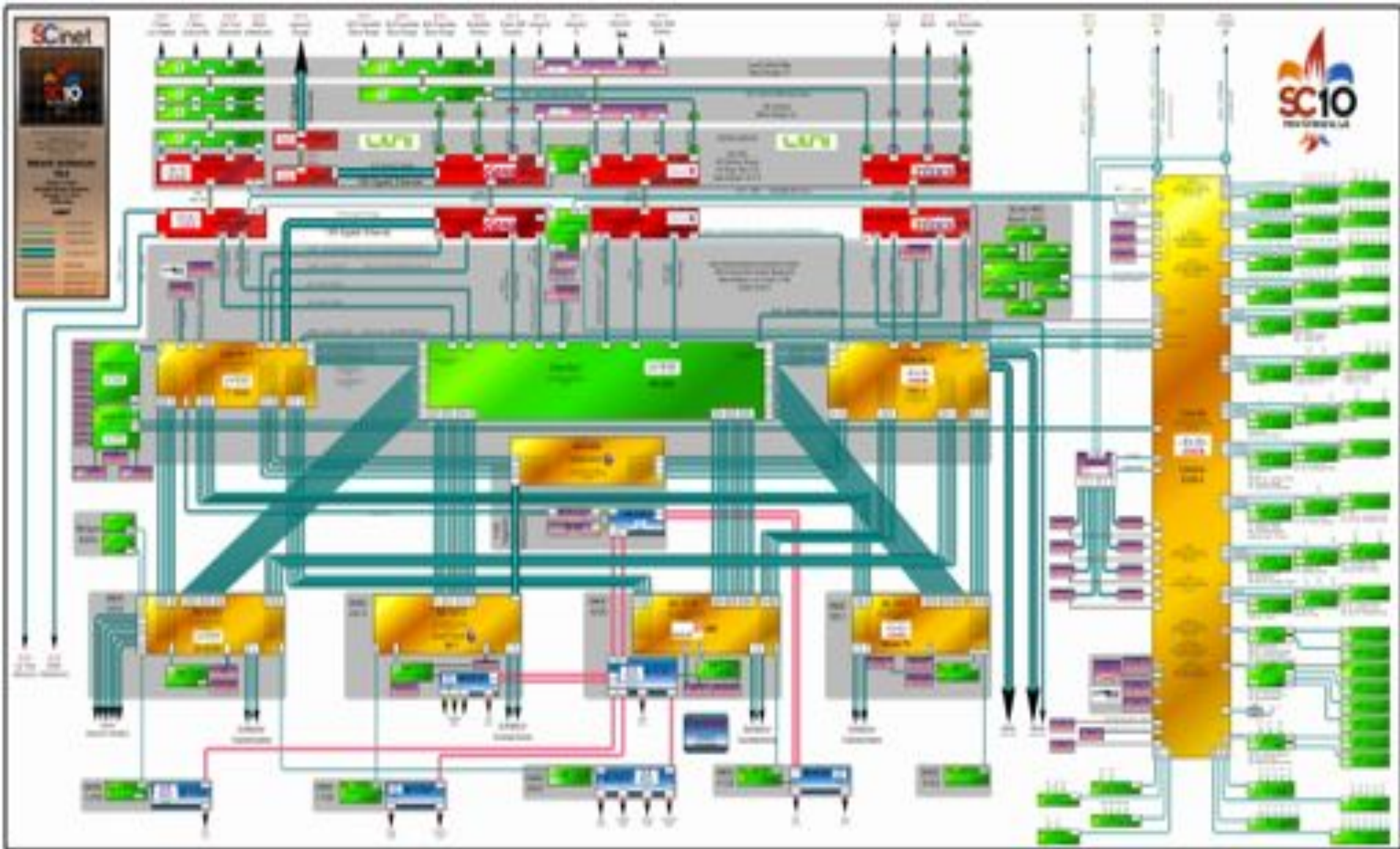


|
retire

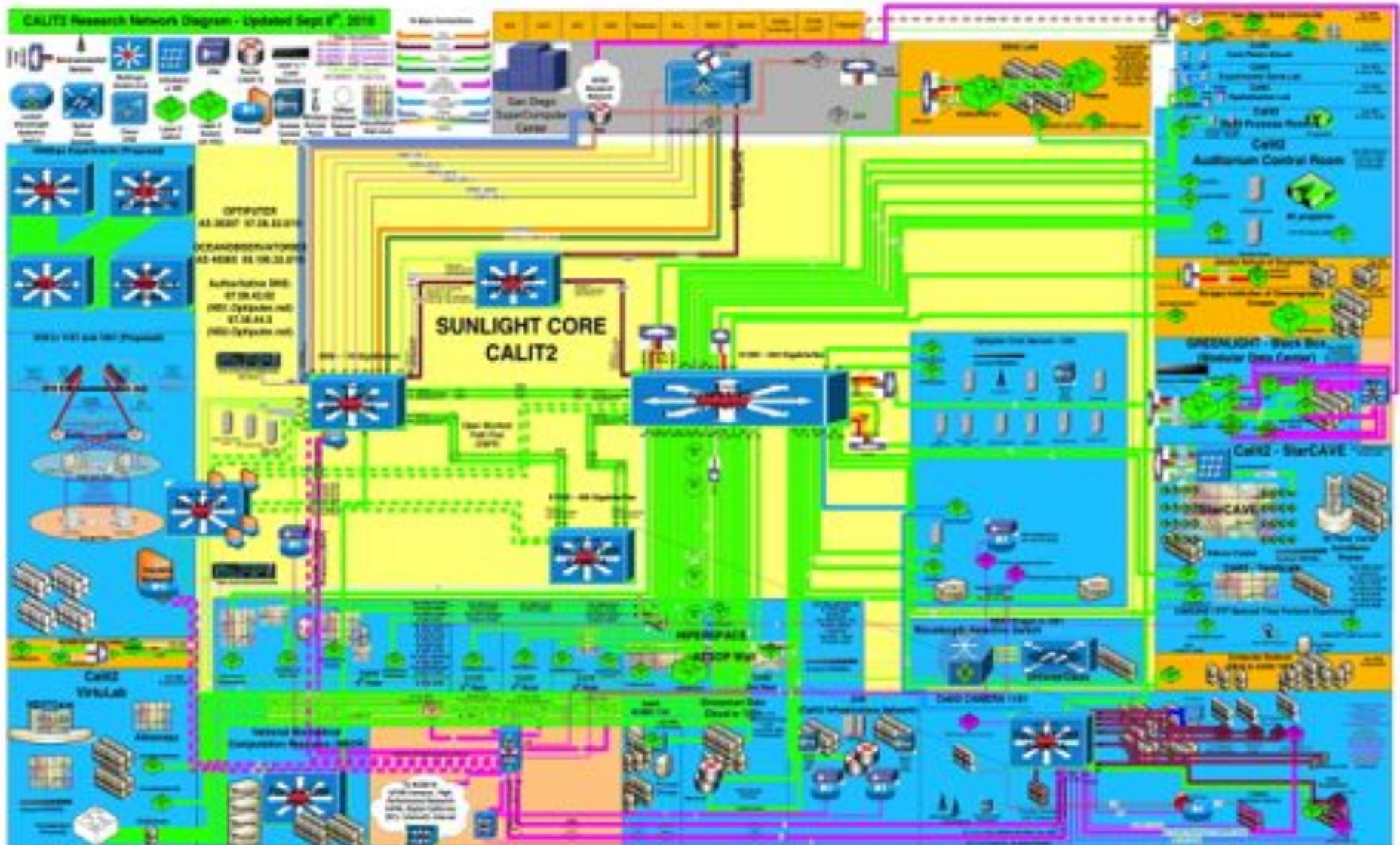
2020

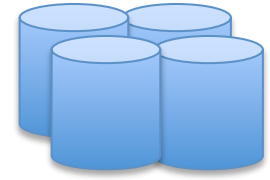
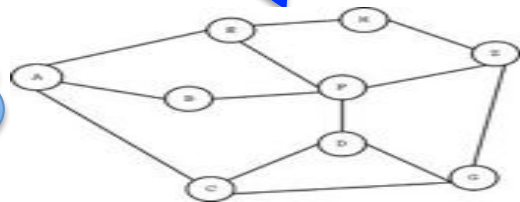
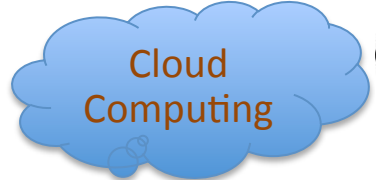
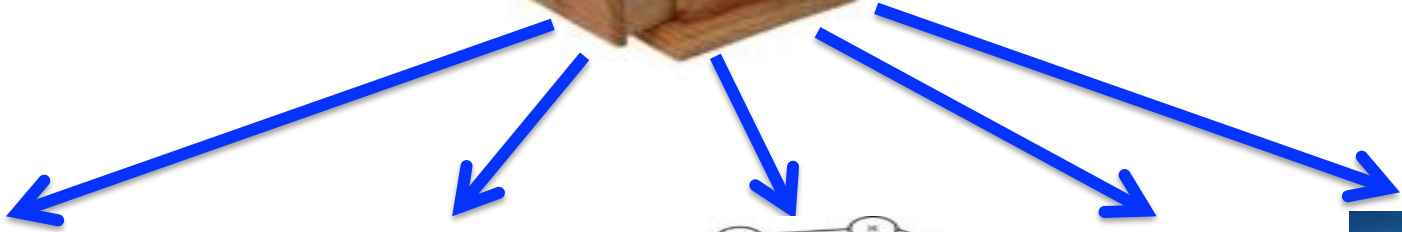
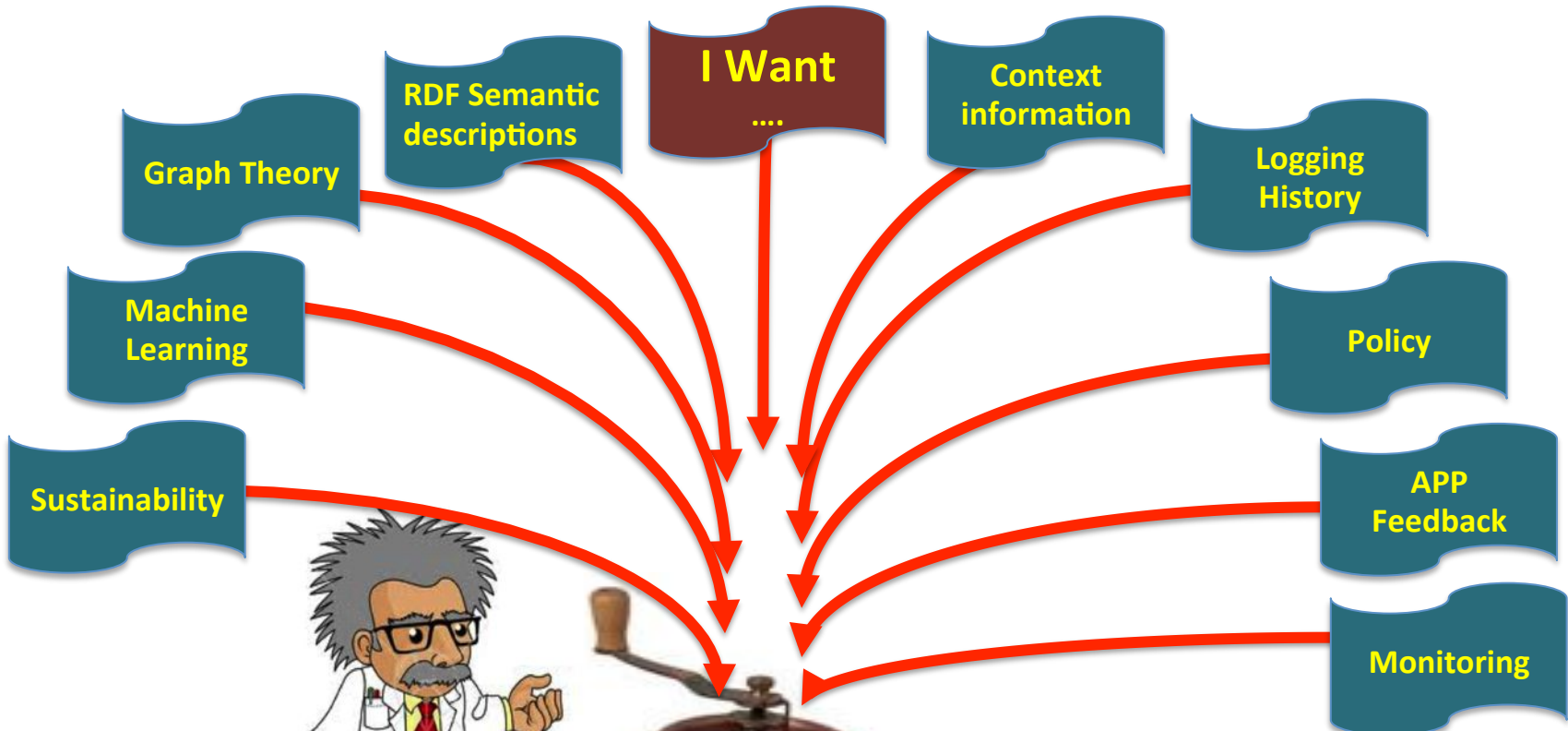
2040

Complex e-Infrastructure!



Complex e-Infrastructure!

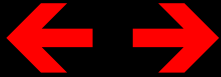




ECO-Scheduling



Hybrid Networking <-> Computing

Routers  Supercomputers

Ethernet switches  Grid & Cloud

Photonic transport  GPU's

What matters:

Energy consumption/multiplication

Energy consumption/bit transported

Challenges

- Data – Data – Data
 - Archiving, publication, searchable, transport, self-describing, DB innovations needed, multi disciplinary use
- Virtualisation
 - Another layer of indeterminism
- Greening the Infrastructure
 - e.g. Department Of Less Energy: http://www.ecrinitiative.org/pdfs/ECR_3_0_1.pdf
- Disruptive developments
 - BufferBloath, Revisiting TCP, influence of SSD's & GPU's
 - Multi layer Glif Open Exchange model
 - Invariants in LightPaths (been there done that ☺)
 - X25, ATM, SONET/SDH, Lambda's, MPLS-TE, VLAN's, PBT, OpenFlow,
 - Authorization & Trust & Security and Privacy



The Way Forward!

- Nowadays scientific computing and data is dwarfed by commercial & cloud, there is also no scientific water, scientific power.
 - Understand how to work with elastic clouds
 - Trust & Policy & Firewalling on VM/Cloud level
- Technology cycles are 3 – 5 year
 - Do not try to unify but prepare for diversity
 - Hybrid computing & networking
 - Compete on implementation & agree on interfaces and protocols
- Limitation on natural resources and disruptive events
 - Energy becomes big issue
 - Follow the sun
 - Avoid single points of failure (aka Amazon, Blackberry, ...)
 - Better very loosely coupled than totally unified integrated...

