

Internet Innovation to support Science & Education.

Cees de Laat

EU  
COMMIT

UvA

NWO

PID/EFRO

SURFnet

NLESC

TNO

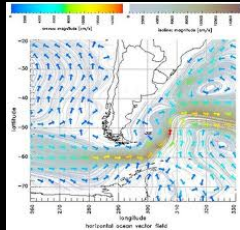
NWO/nf



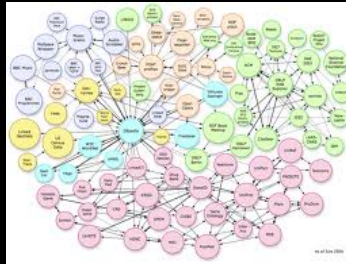
... more data!

# Internet developments

# Google



DATA



... more realtime!



twitter



myspace  
a place for freedom



Linked in



SchoolBANK

Hyves

flickr  
from YAHOO!



... more users!



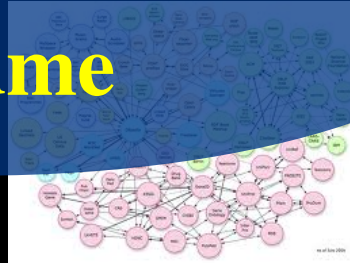
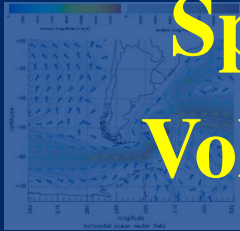
... more data!

Internet developments

Google

Speed  
Volume

DATA



Deterministic

Real-time



twitter



Scalable

Secure

Linked in



myspace  
SchoolBANK

Hyves

flickr  
from YAHOO!



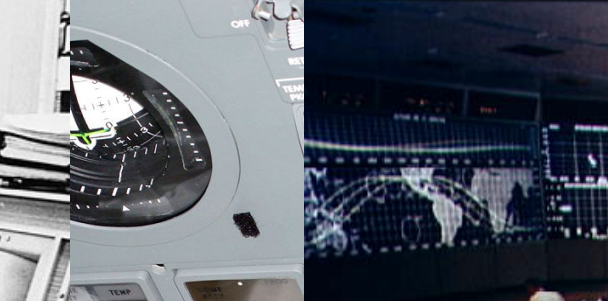
... more users!



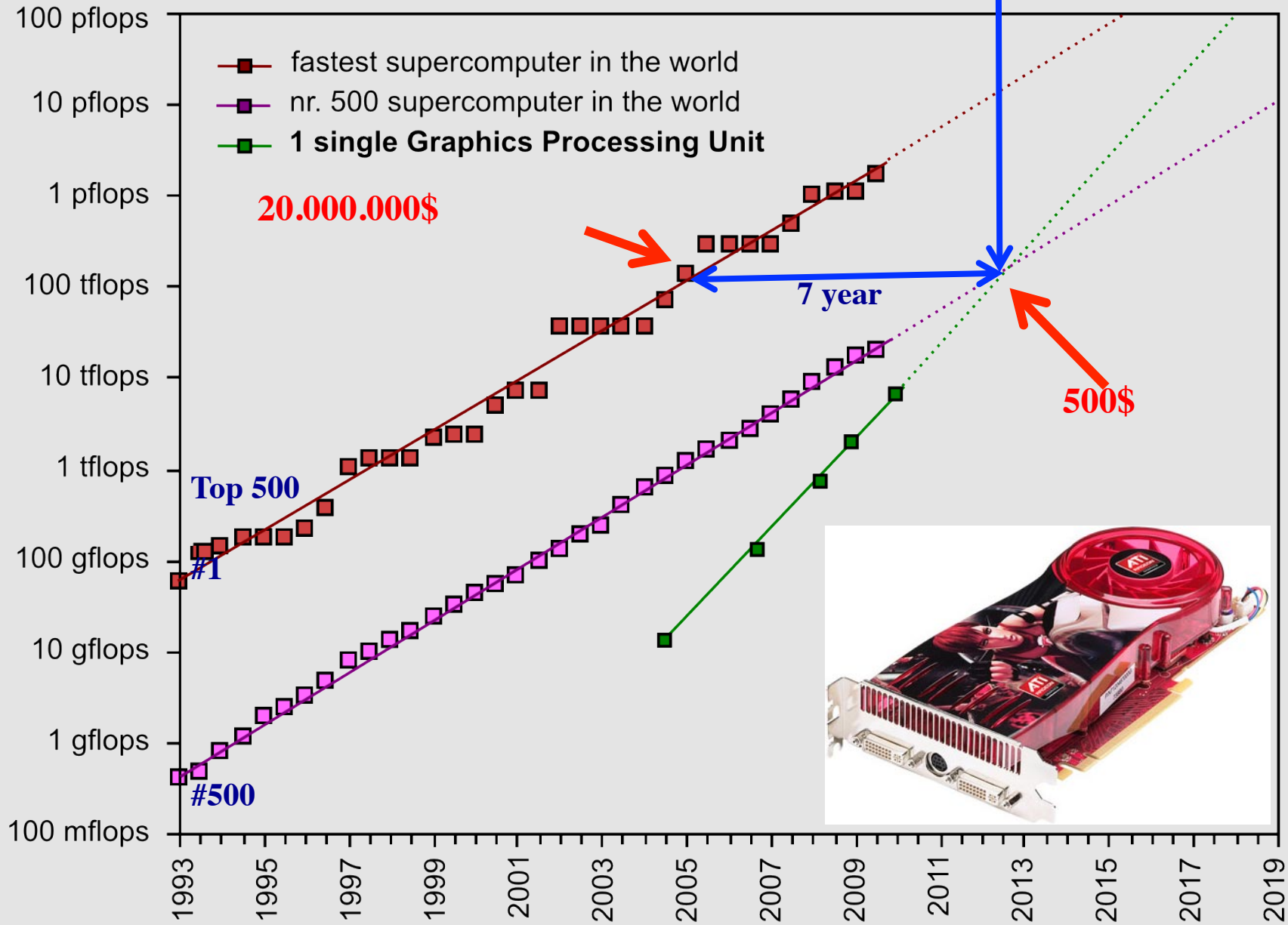




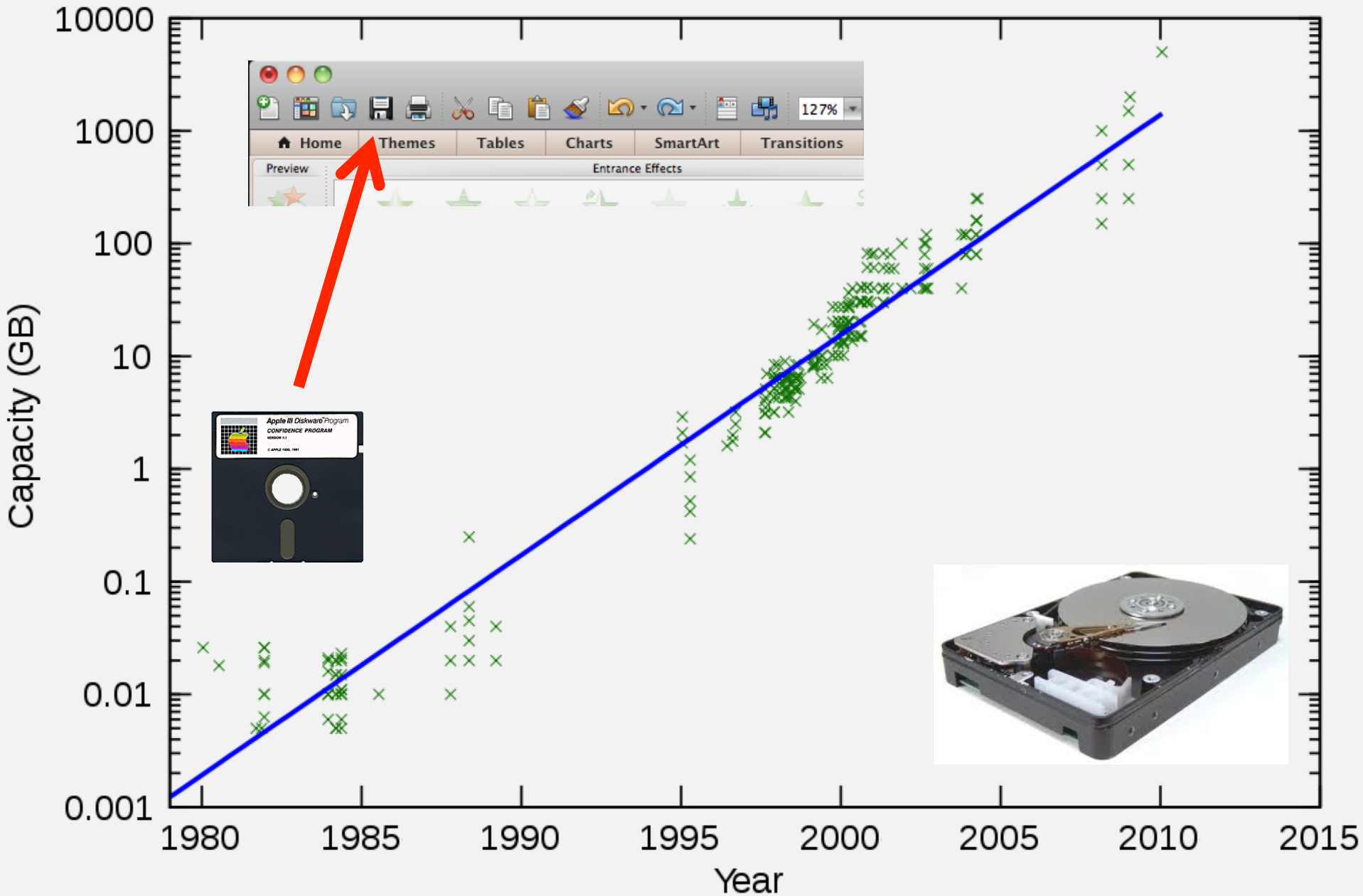




# GPU cards are disruptive!

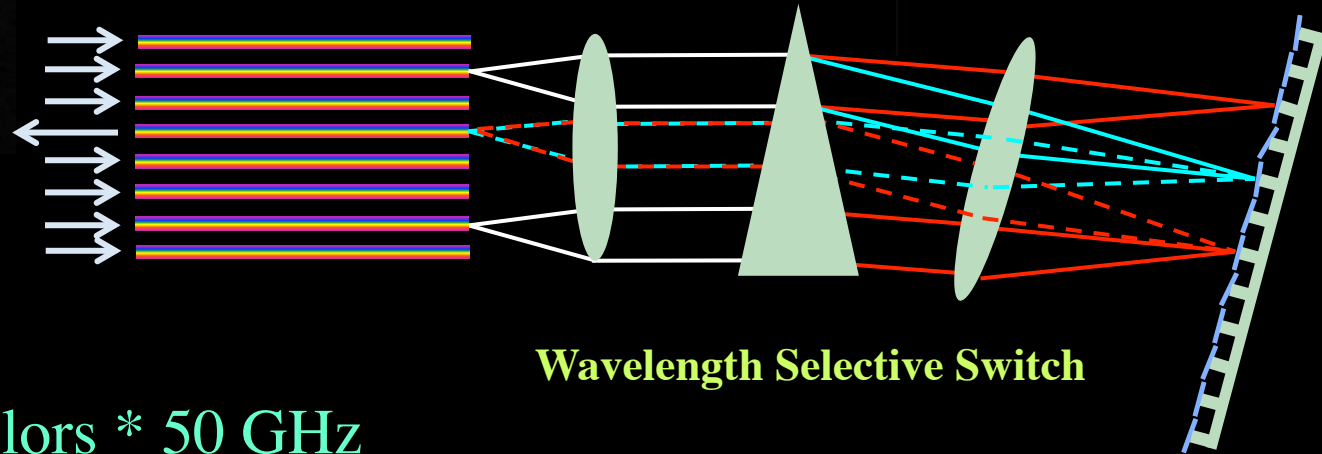
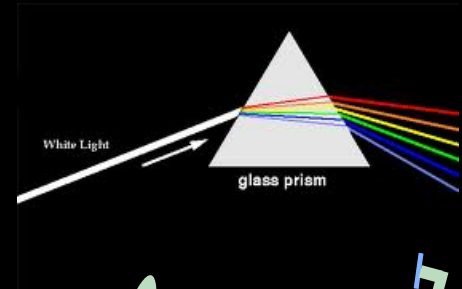
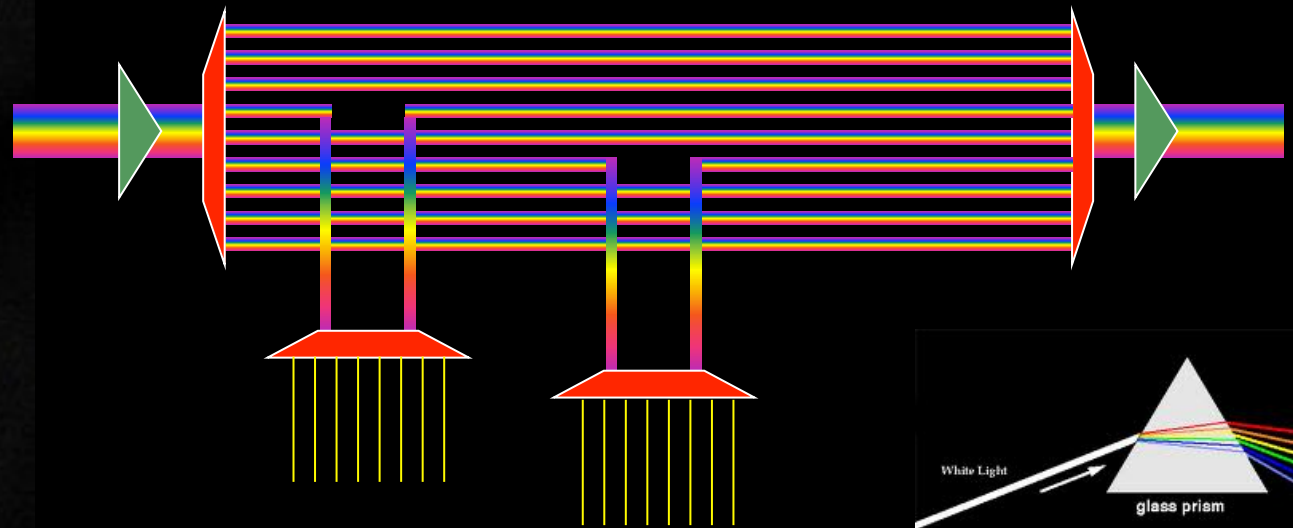


# Data storage: doubling every 1.5 year!





# Multiple colors / Fiber



Per fiber: ~ 80-100 colors \* 50 GHz  
Per color: 10 – 40 – 100 Gbit/s  
BW \* Distance ~  $2 \cdot 10^{17}$  bm/s

New: Hollow Fiber!  
➔ less RTT!

# Wireless Networks



## Digital technology reviews

Tech XO provided latest Digital Technology reviews like digital camera, digital lens reviews, digital camera

[HOME](#)

[CONTACT US](#)

[PRIVACY POLICY](#)

You Are Here : [Digital Technology Reviews](#) » [Network Devices](#) » Next Generation Wireless LAN Technology 802.11ac 1 Gbps throughput with

SEP  
06

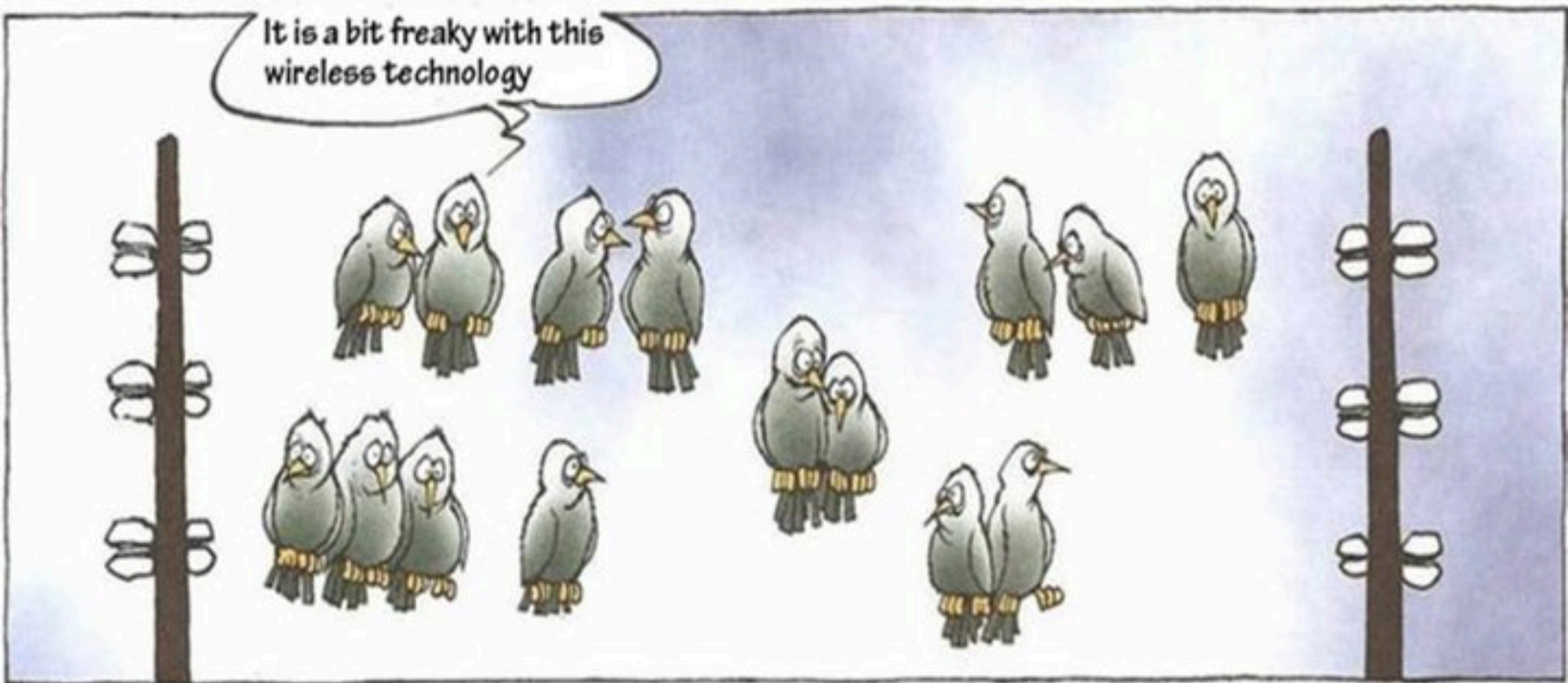
### Next Generation Wireless LAN Technology 802.11ac 1 Gbps throughput with

Published By [admin](#) under [Network Devices](#) Tags: [1gbps throughput](#), [1gbps wireless](#), [1gbps wireless lans](#), [generation](#), [new generation](#), [technologies](#), [technology](#), [throughput](#), [wireless](#), [wireless lan](#)

WiFi is one of the most preferred communication

protocol LAN due to the easy comparison and convenience in the **digital home**. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.

# Wireless Networks



COPYRIGHT : MORTEN INGEMANN

protocol LAN due to the easy comparison and convenience in the **digital home**. While consumer PC products has just started to migrate to a much higher bandwidth of 802.11n wireless LAN now working on next-generation standard definition is already in progress.

# 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

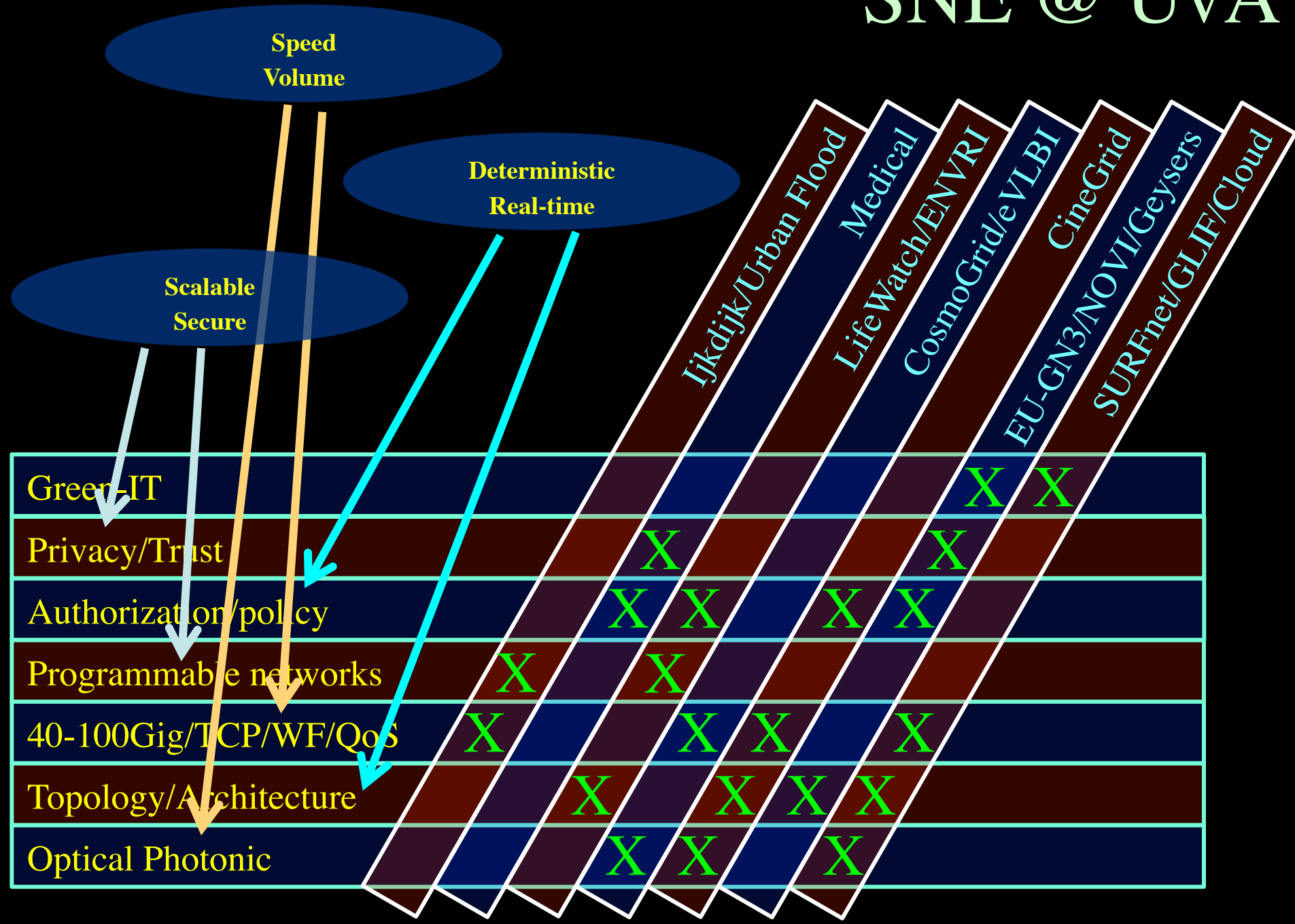
X

X

X

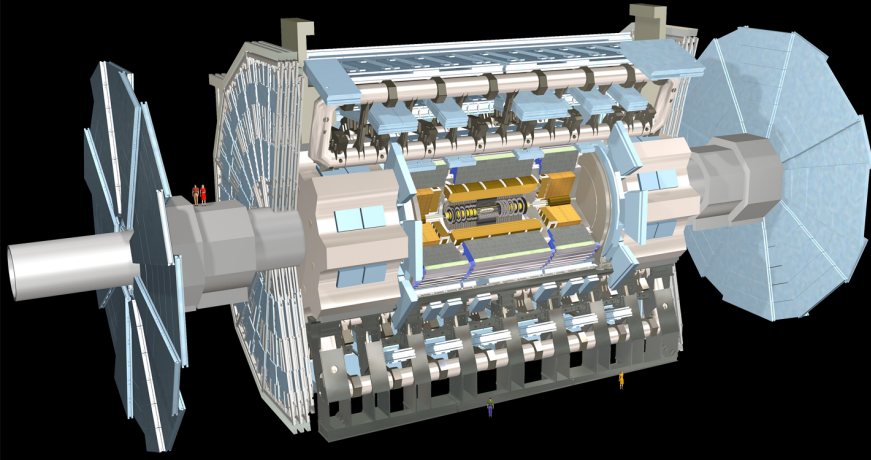
X

# SNE @ UvA





# 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

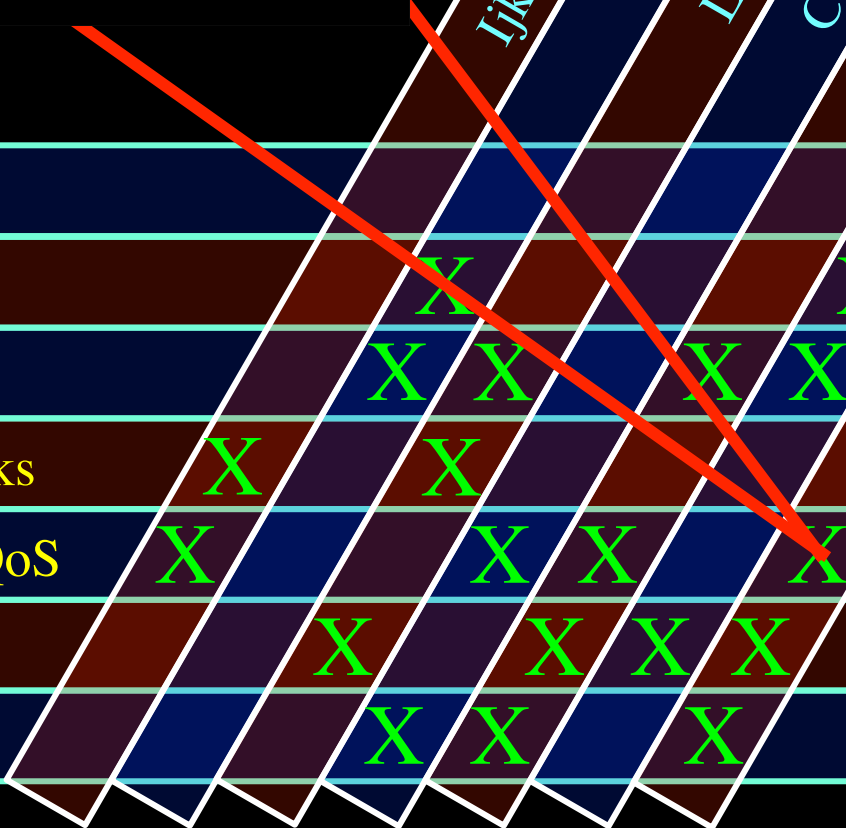
X

X

X

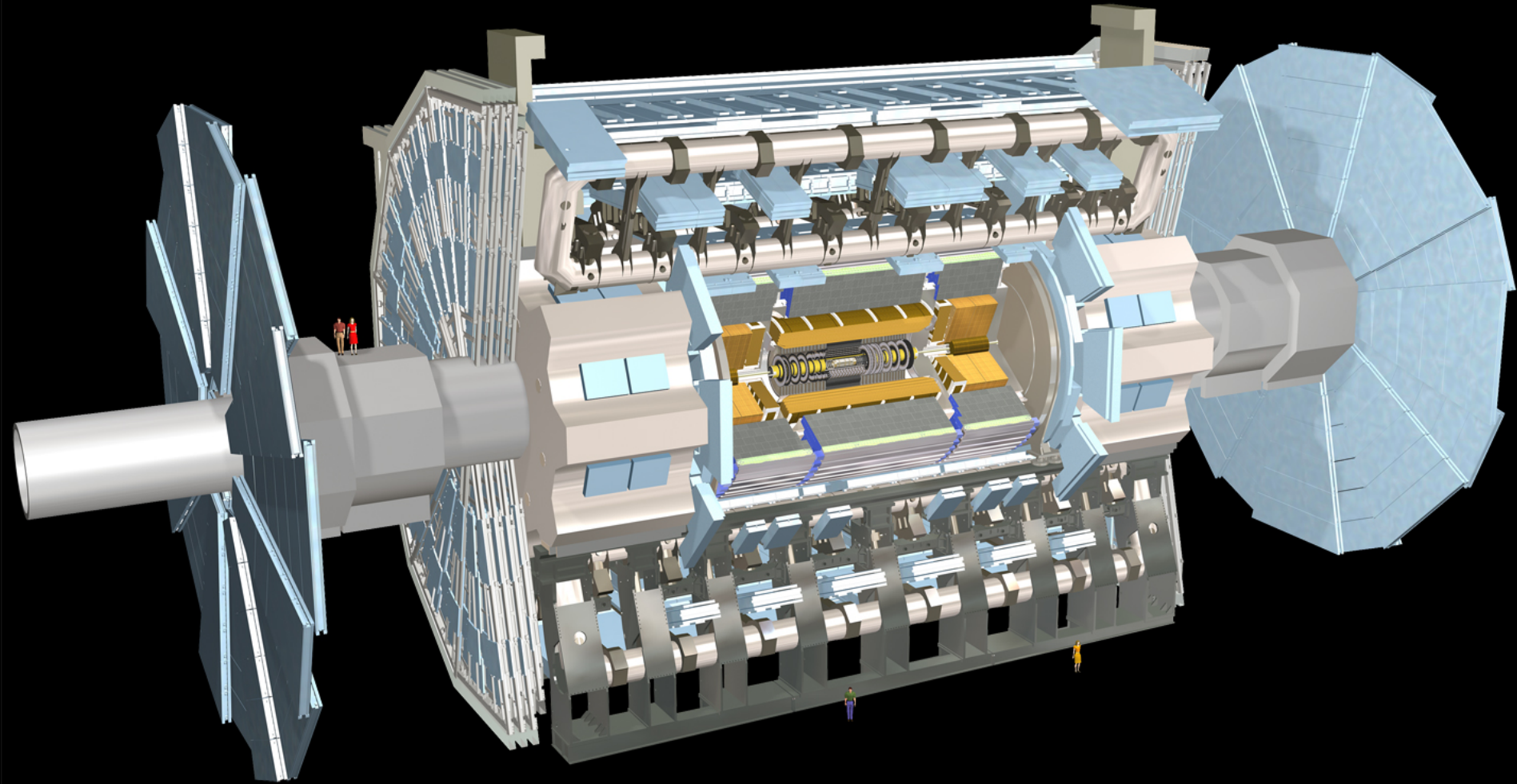
X

X





# ATLAS detector @ CERN Geneve





# ATLAS detector @ CERN Geneve

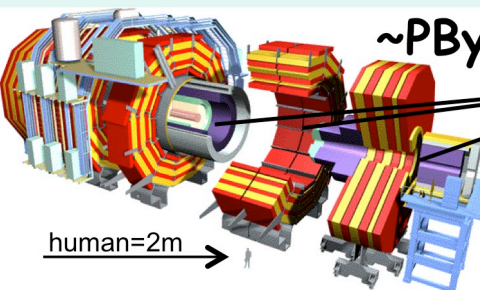
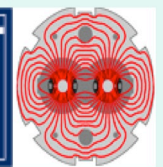






# LHC Data Grid Hierarchy

CMS as example, Atlas is similar



human=2m →

~PByte/sec

Online System

Tier 0 + 1

~100 MBytes/sec

100000 flops/byte

10 Pflops/s

event simulation



event reconstruction

Status 2002!

CMS detector: 15m X 15m X 22m  
12,500 tons, \$700M.

~2.5 Gbits/sec

Tier 1

Italian Regional Center

German Regional Center

NIKHEF Dutch Regional Center

FermiLab, USA Regional Center

...

analysis

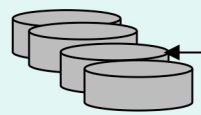
~0.6-2.5 Gbps

Tier 3

~0.6-2.5 Gbps

Tier 2 Center 1, Tier 2 Center 2, Tier 2 Center 3, Tier 2 Center 4, Tier 2 Center 5

Tier 2



Physics data cache

Institute ~0.25TIPS, Institute, Institute, Institute

100 - 1000 Mbits/sec



Workstations

Tier 4

CERN/CMS data goes to 6-8 Tier 1 regional centers, and from each of these to 6-10 Tier 2 centers.

Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

2000 physicists in 31 countries are involved in this 20-year experiment in which DOE is a major player.

#  
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**

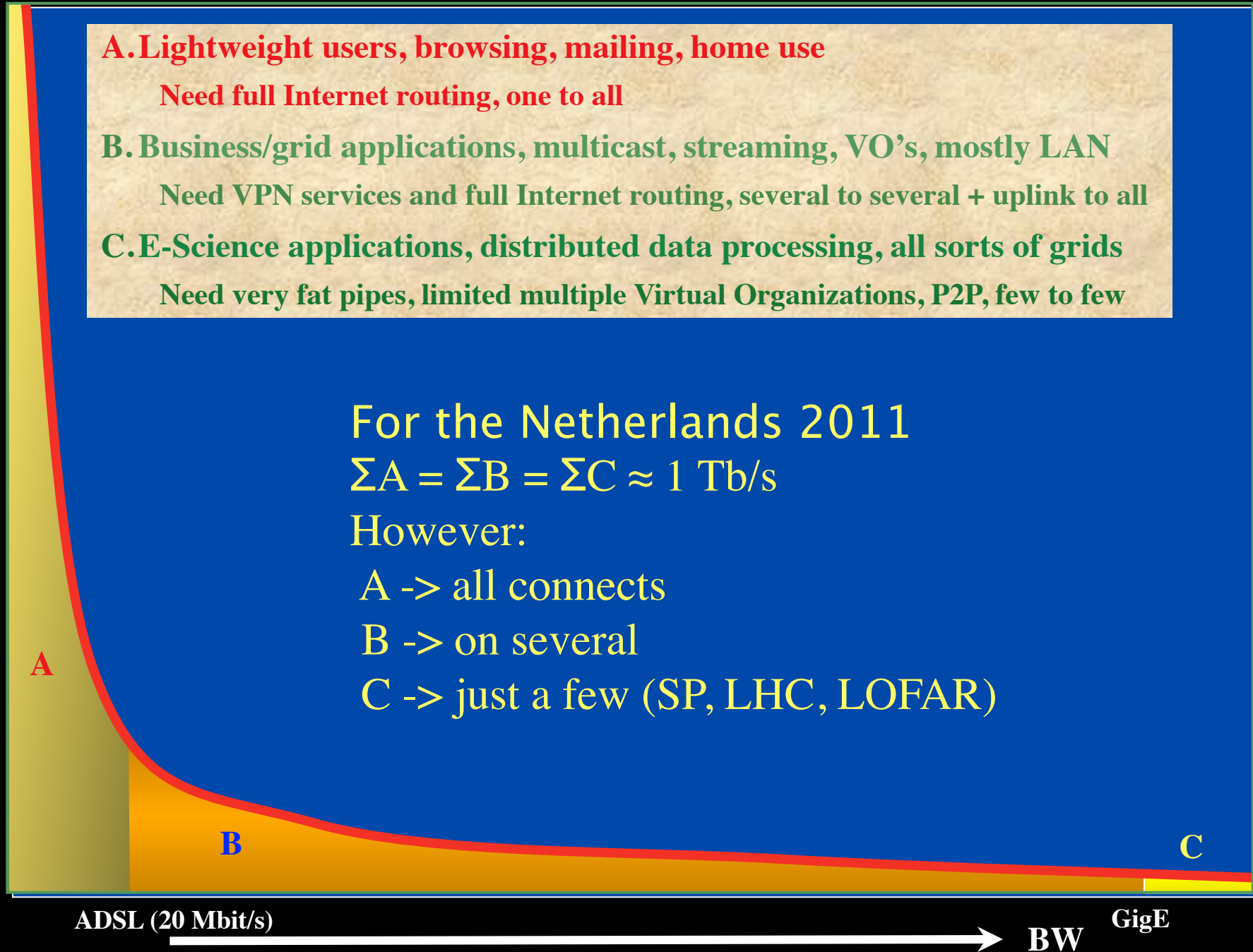
$\Sigma A = \Sigma B = \Sigma C \approx 1 \text{ Tb/s}$

However:

A -> all connects

B -> on several

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



ADSL (20 Mbit/s)

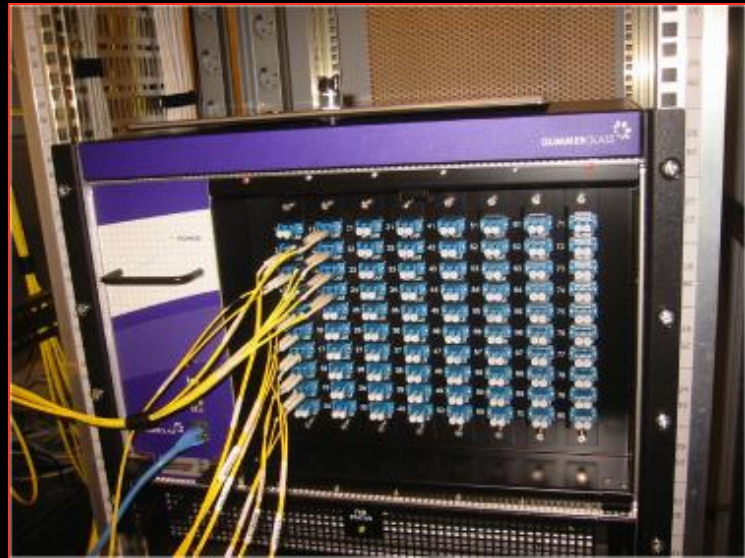
BW

GigE

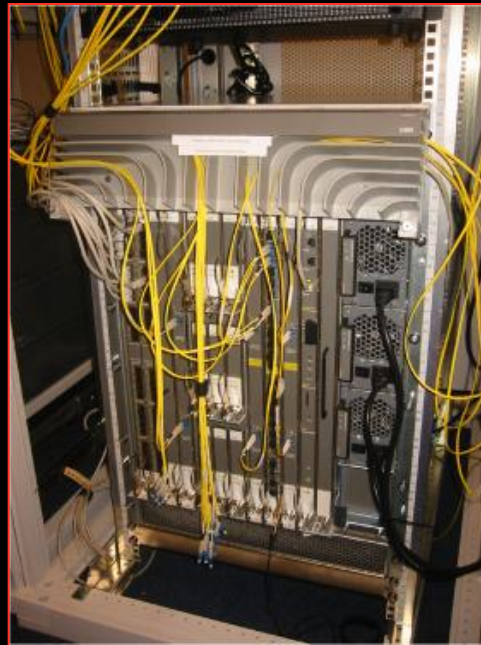
# Towards Hybrid Networking!

- 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

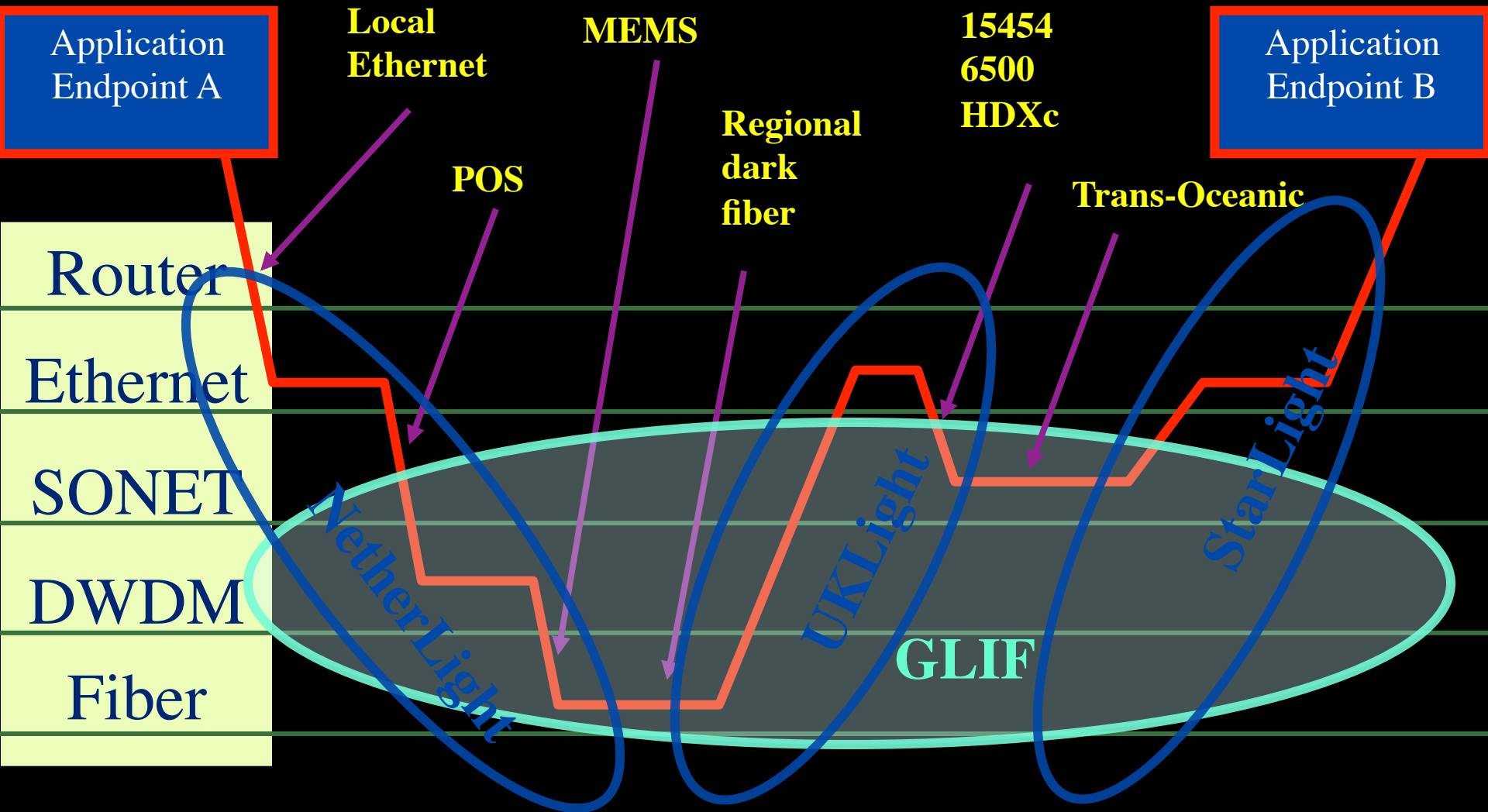


L3  $\approx$  75+ k\$/port



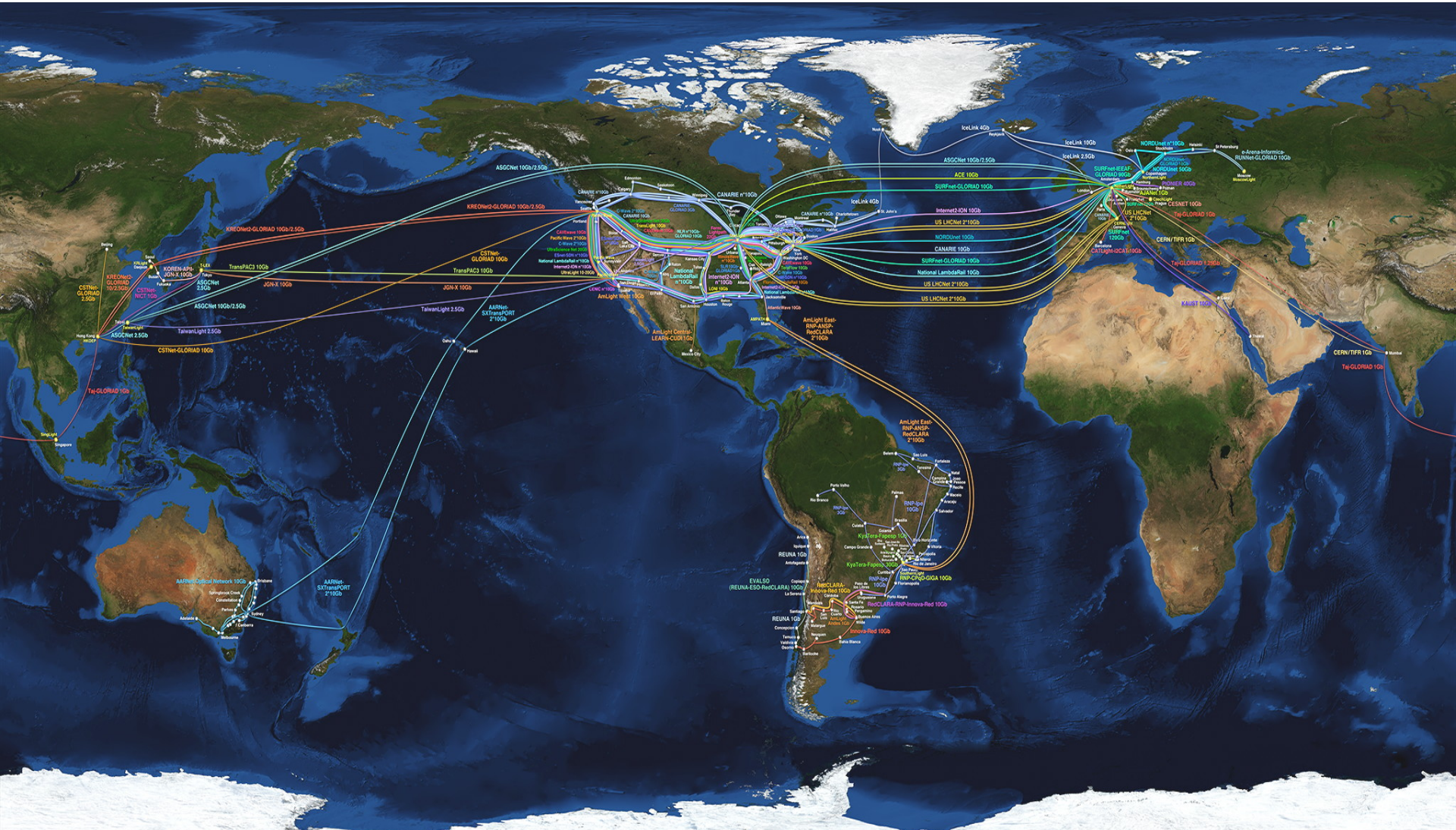


# How low can you go?





# The GLIF – lightpaths around the world









In The Netherlands SURFnet connects between 180:

- universities;
- academic hospitals;
- most polytechnics;
- research centers.

with an indirect ~750K user base

~ 8860 km  
scale  
comparable  
to railway  
system





# 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<sup>[1]</sup> → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service<sup>[2]</sup> → time savings
- Support of different modulation formats<sup>[3]</sup> → extend network lifetime

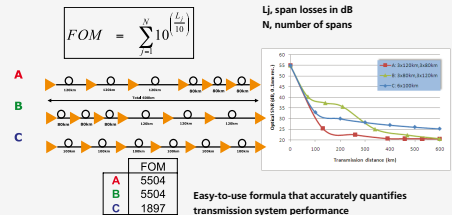
### Alien wavelength challenges

- 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.

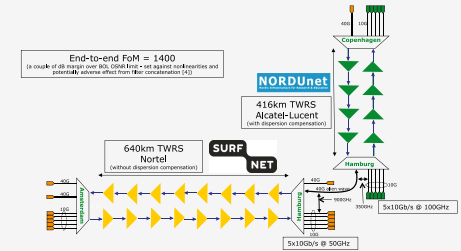
### New method to present fiber link quality, FoM (Figure of Merit)

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.

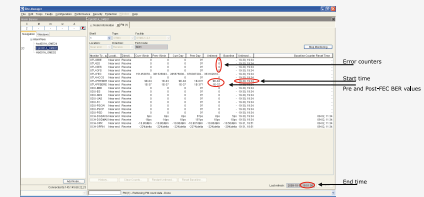


### Transmission system setup

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



### Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3,0 10<sup>-16</sup>

### Conclusions

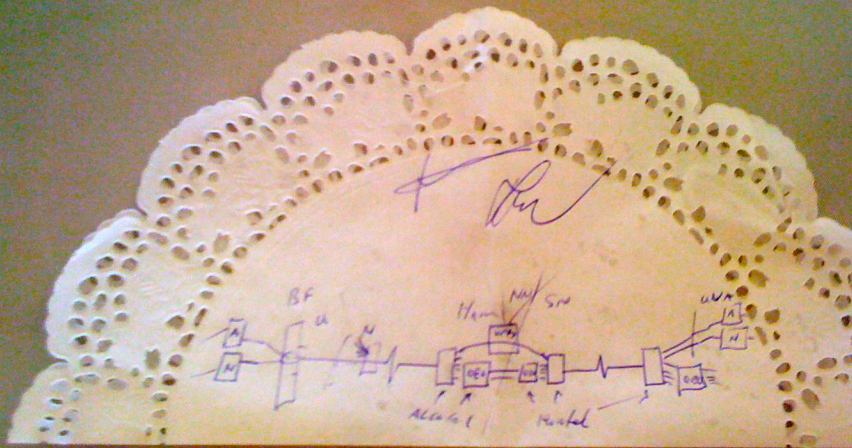
- We have investigated experimentally the all-optical transmission of a 40Gb/s PM-QPSK alien wavelength via a concatenated native and third party DWDM system that both were carrying live 10Gb/s wavelengths.
- The end-to-end transmission system consisted of 1056 km of TWRS (TrueWave Reduced Slope) transmission fiber.
- We demonstrated error-free transmission (i.e. BER below 10<sup>-15</sup>) 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<sup>[1]</sup> → cost savings
- Avoid OEO regeneration → power savings
- Faster time to service<sup>[2]</sup> → time savings
- Support of different modulation formats<sup>[3]</sup> → extend network lifetime

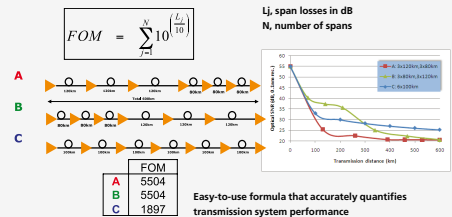
### Alien wavelength challenges

- 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.

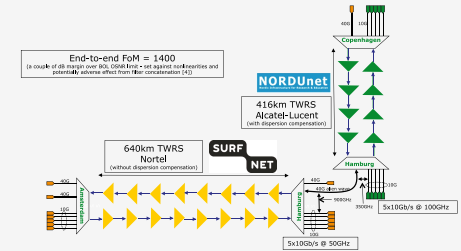
### New method to present fiber link quality, FoM (Figure of Merit)

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.

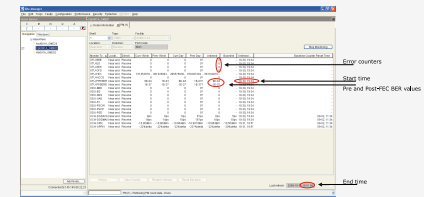


### Transmission system setup

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



### Test results



Error-free transmission for 23 hours, 17 minutes → BER < 3.0 10<sup>-16</sup>

### 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<sup>-15</sup>) 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

# ClearStream @ TNC2011

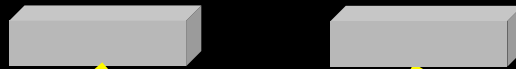
Setup codename:  
FlightCees



## UvA

iPerf  
I7 3.2 GHz Q-core

iPerf  
Amd Ph II 3.6 GHz HexC



Mellanox

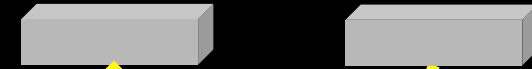
40G E



## Copenhagen

iPerf  
2\* dual 2.8 GHz Q-core

iPerf  
2\* dual 2.8 GHz Q-core

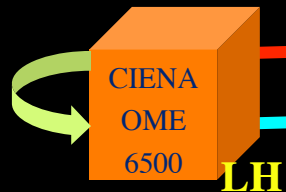


Mellanox



## CERN

CIENA DWDM



LH

17 ms RTT

27 ms RTT

## Hamburg

Alcatel DWDM



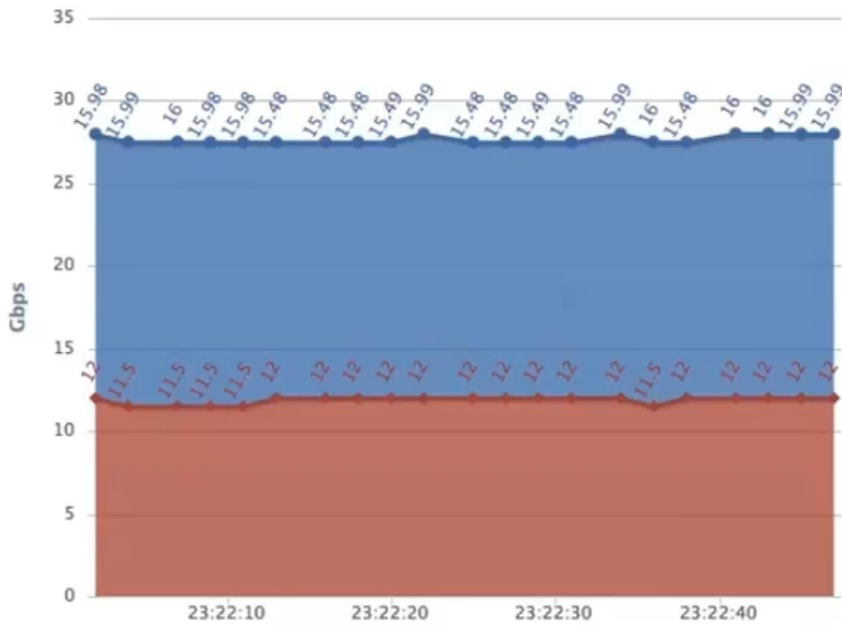
LH

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



# <http://tnc11.delaaat.net>

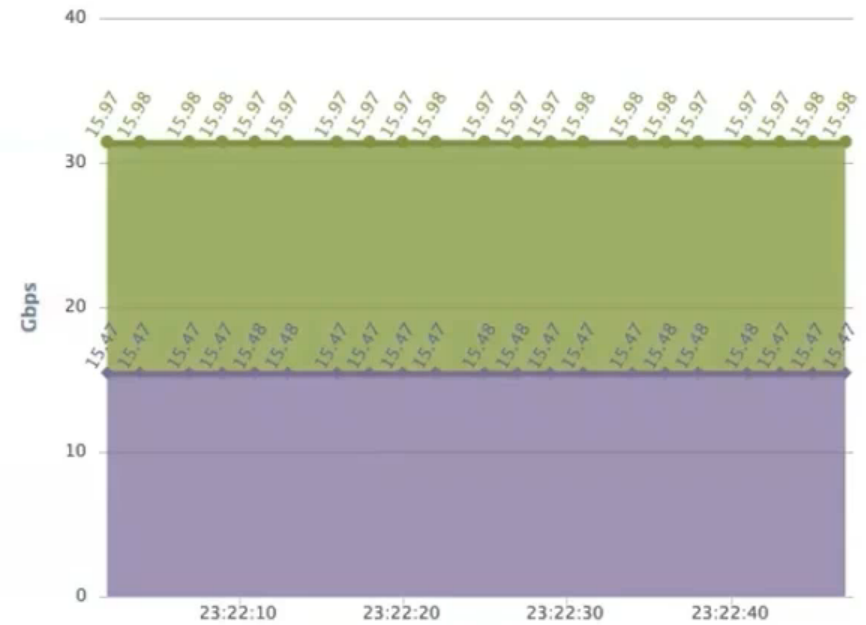
Amsterdam (UvA) Live RX Traffic



■ eth0 RX on tn-uva-l ■ eth0 RX on tn-uva-r

Highcharts.com

Copenhagen POP RX Traffic



■ eth0 RX on tn-cpg-l ■ eth0 RX on tn-cpg-r

Highcharts.com

27.99 Gbps to Amsterdam <-> 31.45 Gbps to Copenhagen

Total Throughput 59.44 Gbps RTT 44.010 ms

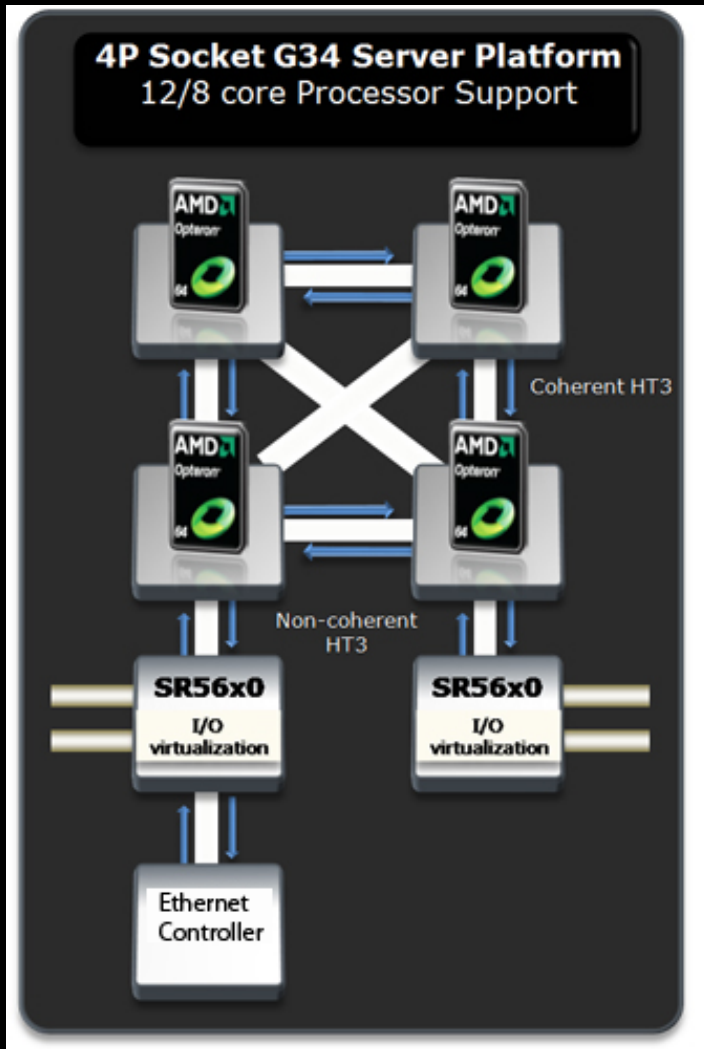
# Results (rtt = 17 ms)

- 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

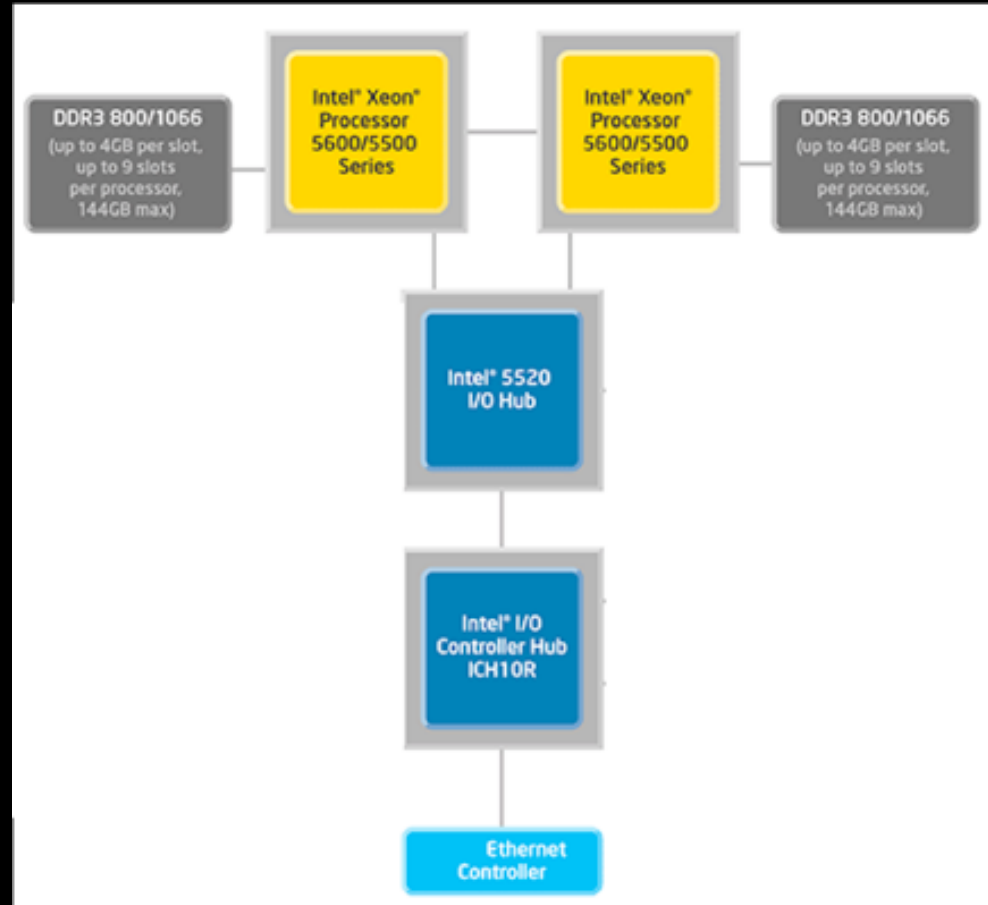
- Mellanox 40GE card is PCI-E 2.0 8x (5GT/s)
- 40Gbit/s raw throughput but ....
- PCI-E is a network-like protocol
  - 8/10 bit encoding -> 25% overhead -> 32Gbit/s maximum data throughput
  - Routing information
- Extra overhead from IP/Ethernet framing
- Server architecture matters!
  - 4P system performed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

# Server Architecture



DELL R815

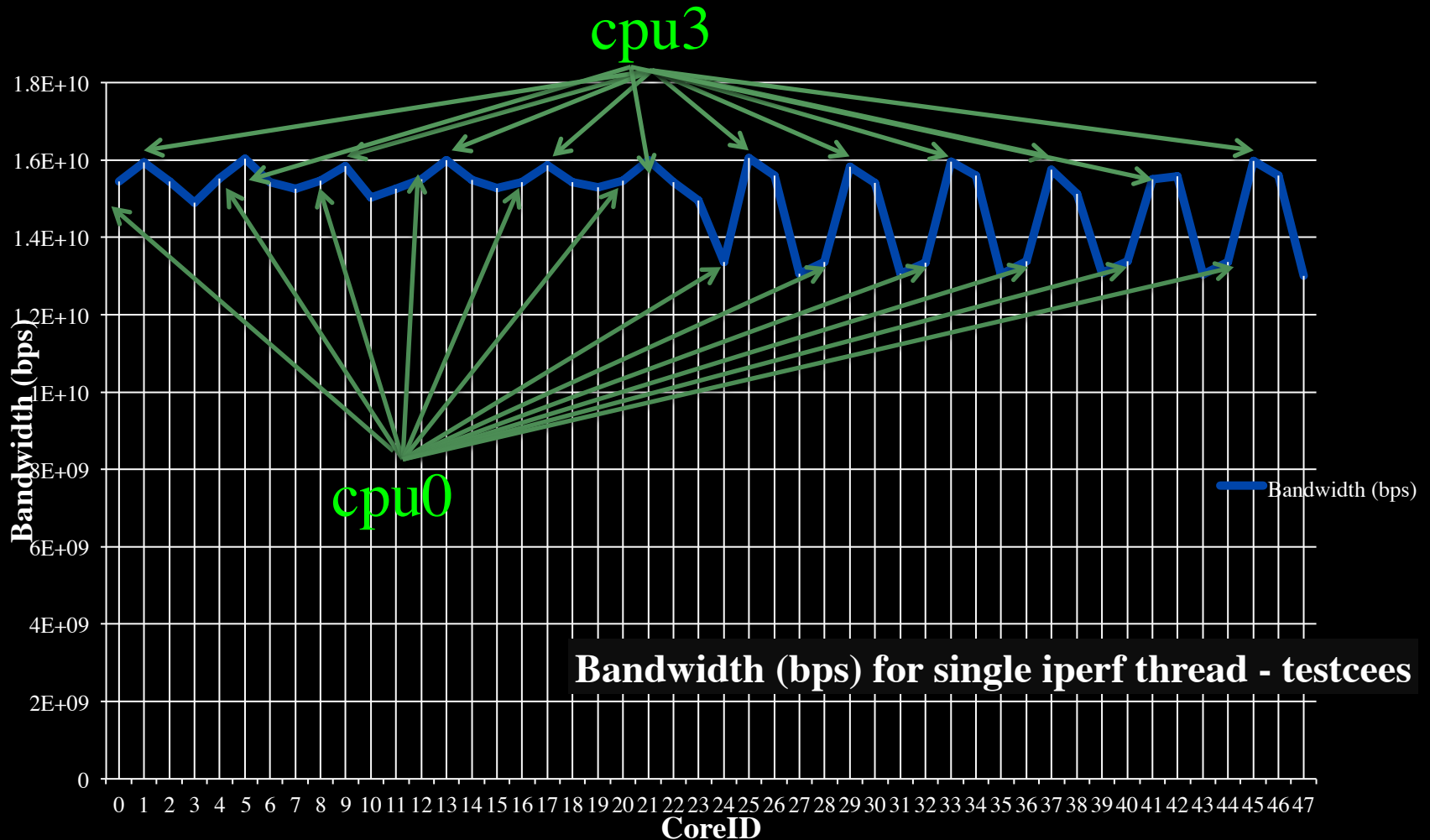
4 x AMD Opteron 6100



Supermicro X8DTT-HIBQF

2 x Intel Xeon

# CPU Topology benchmark

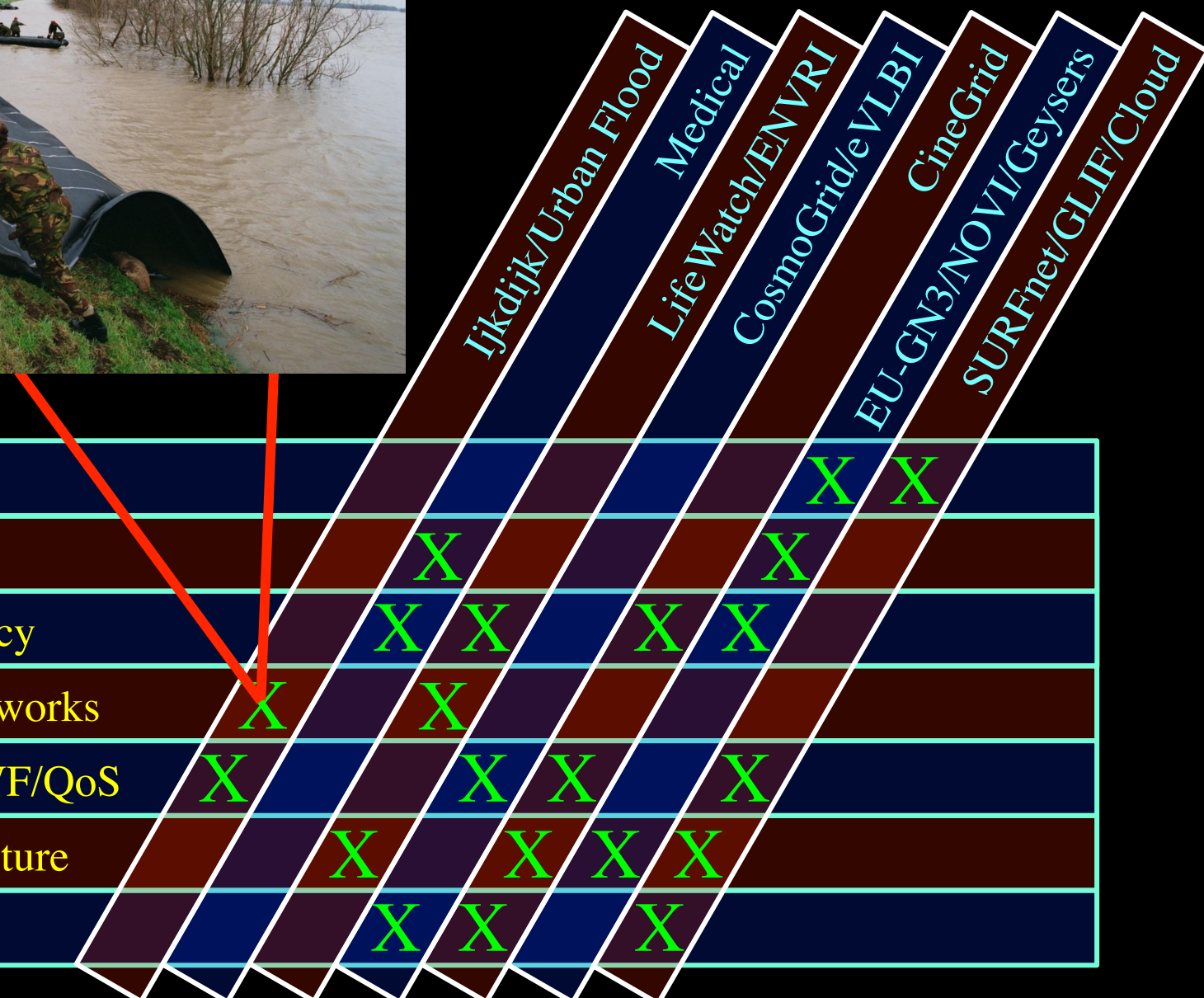


We used numactl to bind iperf to cores



Where when will it happen?

SNE @ UvA



Green-IT

Privacy/Trust

Authorization/policy

Programmable networks

40-100Gig/TCP/WF/QoS

Topology/Architecture

Optical Photonic

Ijkdijk/Urban Flood

Medical

LifeWatch/ENVRI

CosmoGrid/eVLBI

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud

X

X

X X

X

X

X X

X

X X X

X X

X

X X

X

X X

X

CineGrid

EU-GN3/NOVI/Geysers

SURFnet/GLIF/Cloud



## 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

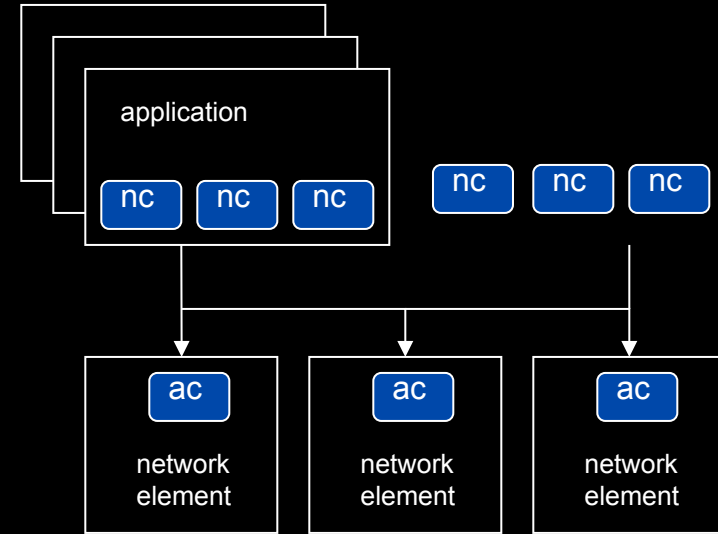
Many small flows -> 12 Gb/s



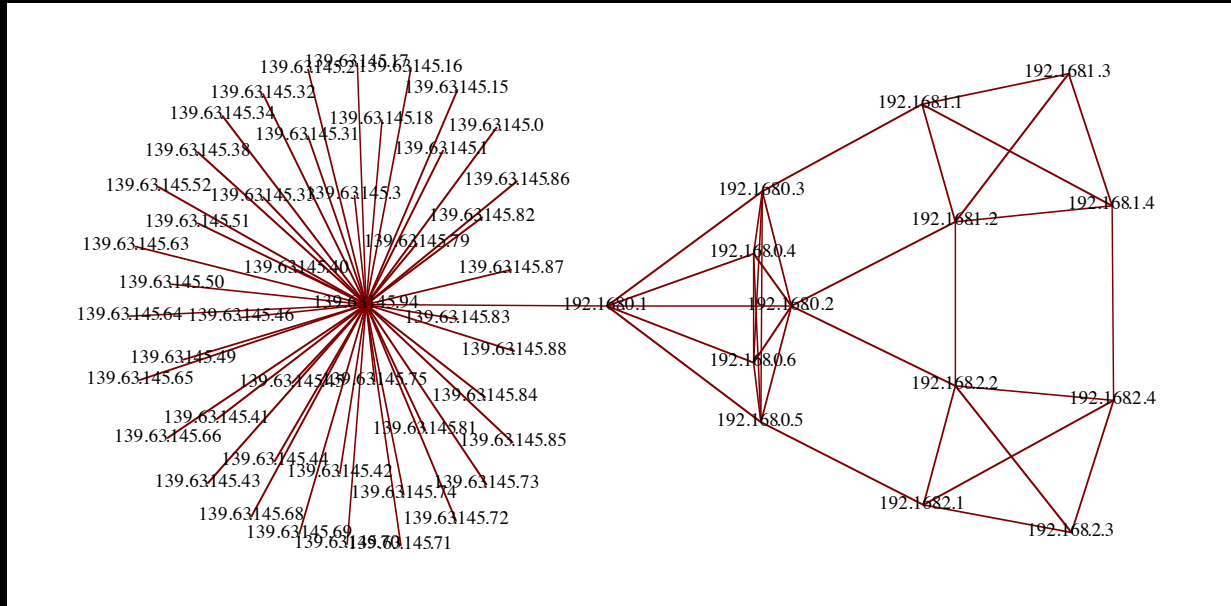
# User Programmable Virtualized Networks.

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



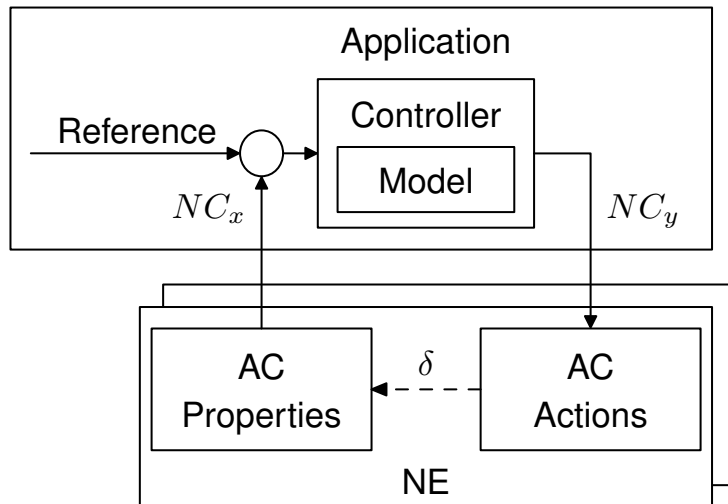
<b>Eigenvalues</b> $\left[\begin{pmatrix} -1 & 0 & 2 \\ 2 & 9 & 2 \\ 3 & 1 & 4 \end{pmatrix}\right]$ {9.484782381, 4.488378326, -1.973160708}	$\sum_{p=1}^{30} \frac{1}{p^2}$ 1.612150118
<b>Plot</b> [Sin[13 x] + Sin[18 x], {x, 0, 2}]	<b>BesselJ</b> [1, 3 + i] 0.4326156394 - 0.4295057869 i
	<b>Simplify</b> [1 + 5 x + 10 x^2 + 10 x^3 + 5 x^4 + x^5] (1 + x)^5
<b>mydata</b> = {{0.444539, 0.908491}, {1.4486, 1.84577}, {1.8734, 1.84577}, ...}	<b>Fit</b> [mydata, {1, x, x^2}, x] 0.2617148495 + 1.007 x - 0.0034235343 x^2



ref: Robert J. Meijer, Rudolf J. Strijkers, Leon Gommans, Cees de Laat, User Programmable Virtualized Networks, accepted for publication to the IEEE e-Science 2006 conference Amsterdam.

# 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

Rudolf Strijkers<sup>1,2</sup>

Marc X. Makkes<sup>1,2</sup>

Mihai Christea<sup>1</sup>

Laurence Muller<sup>1</sup>

Robert Belleman<sup>1</sup>

Cees de Laat<sup>1</sup>

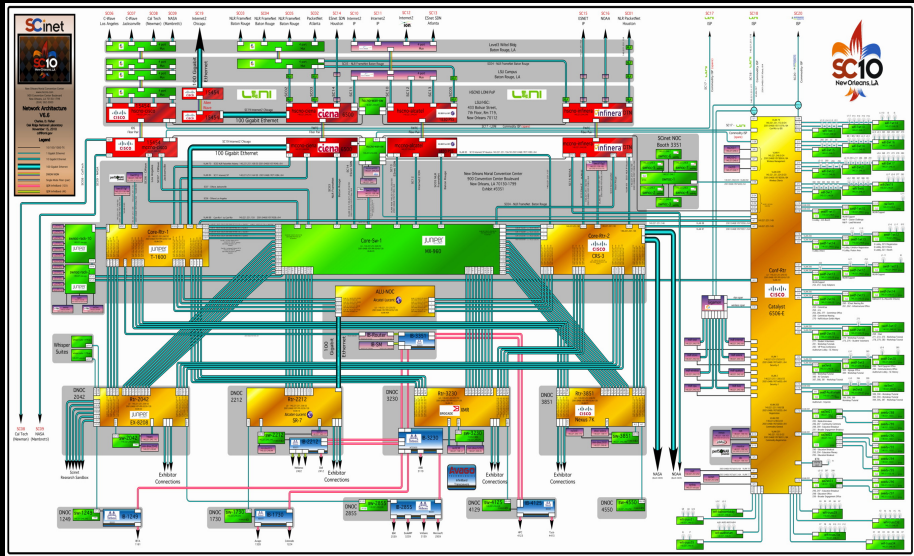
Robert Meijer<sup>1,2</sup>

<sup>1</sup> University of Amsterdam, Amsterdam The Netherlands

<sup>2</sup> TNO Information and Communication Technology, Groningen, The Netherlands

# Mastering Complexity

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

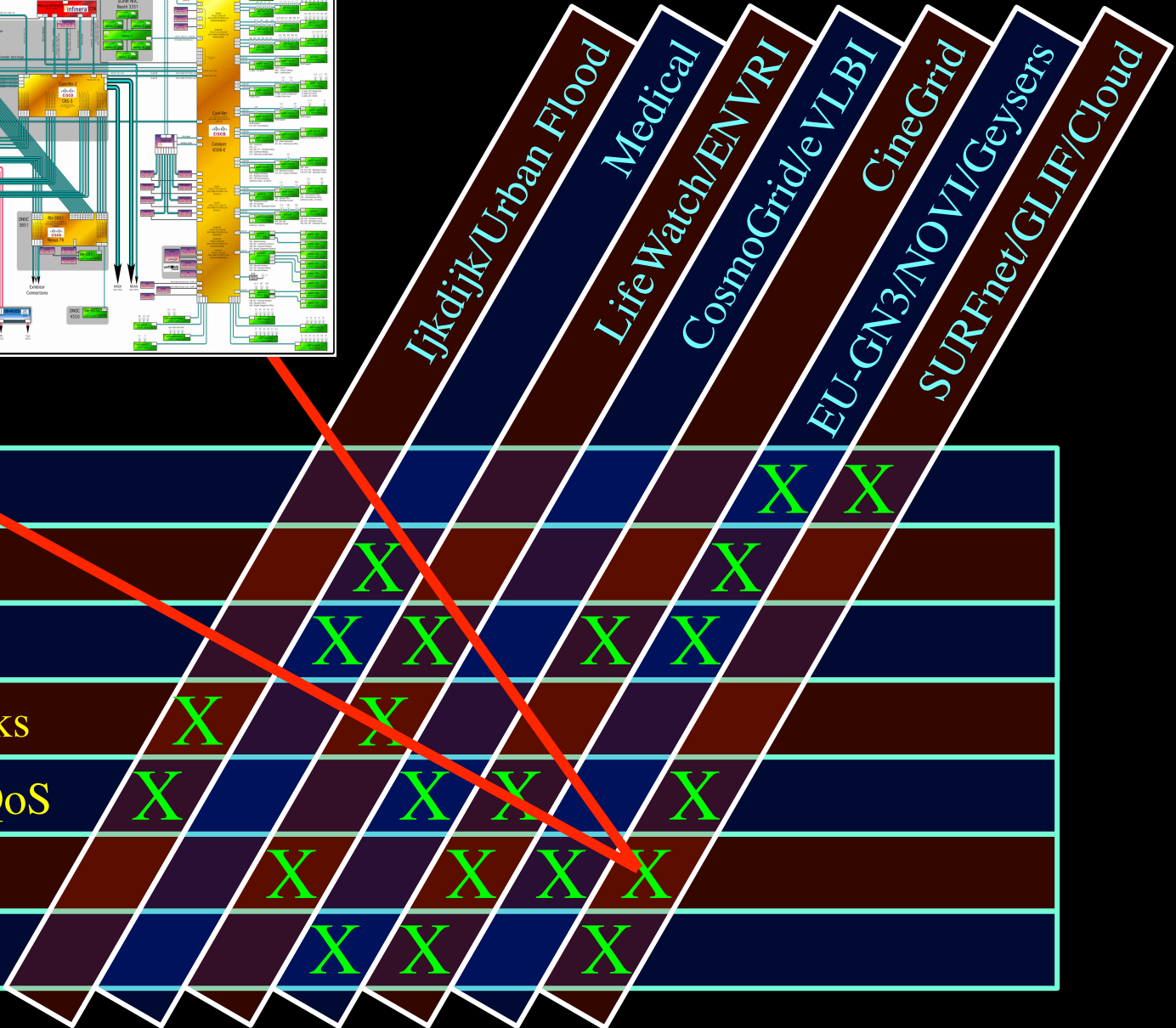
X

X

X

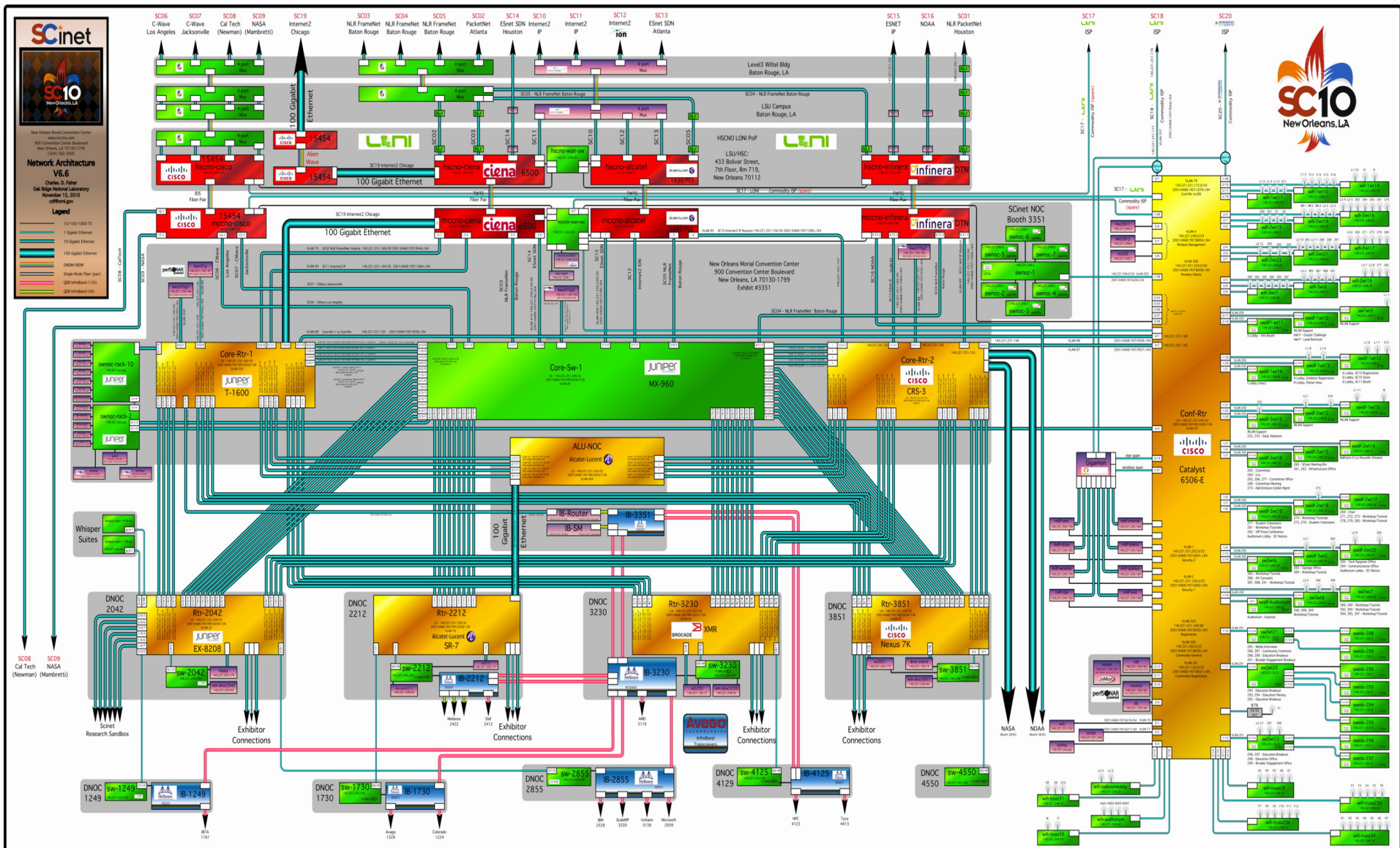
X

X

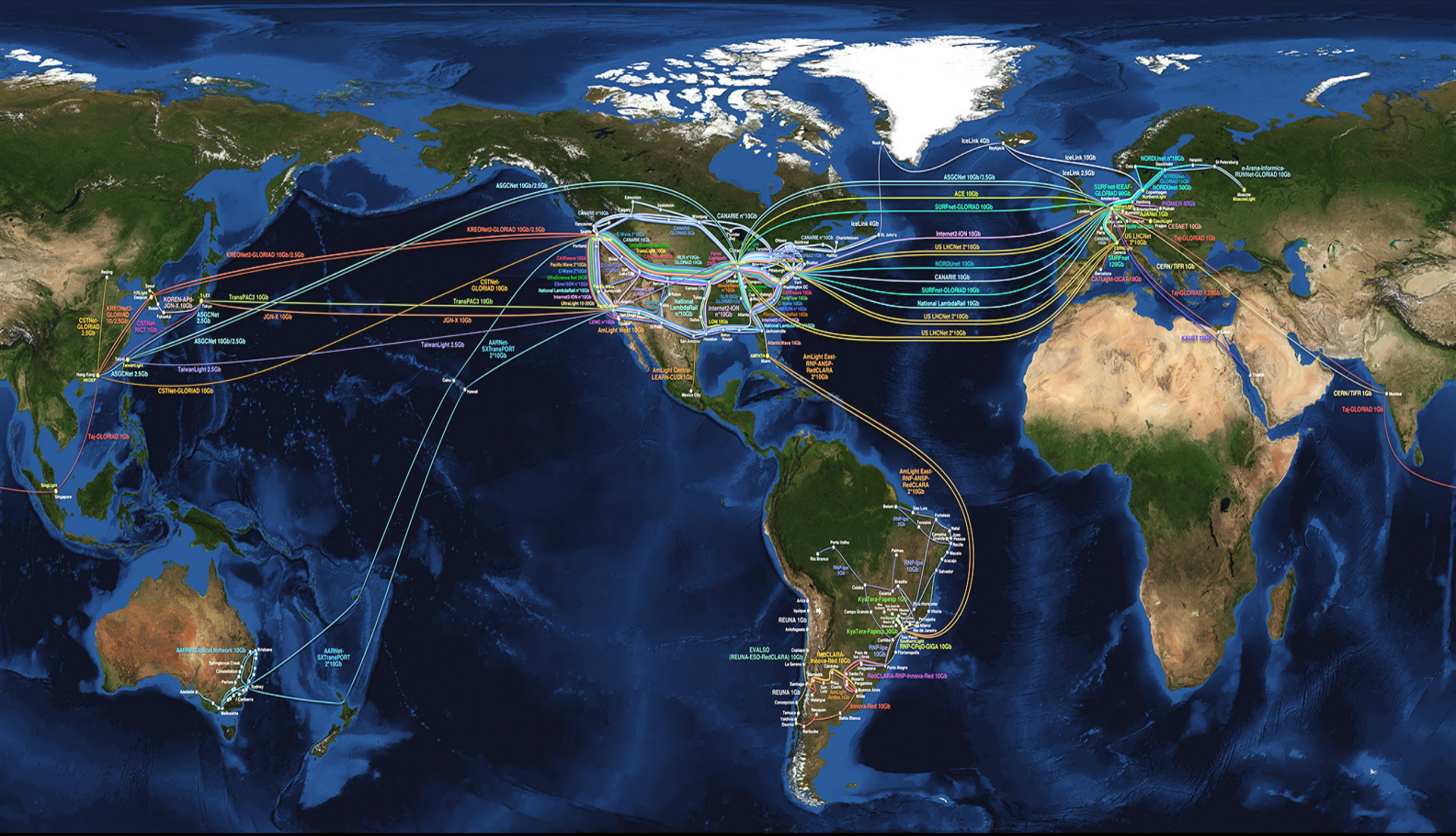




# Complex eInfrastructure @ SC10







We investigate:  
 complex networks!



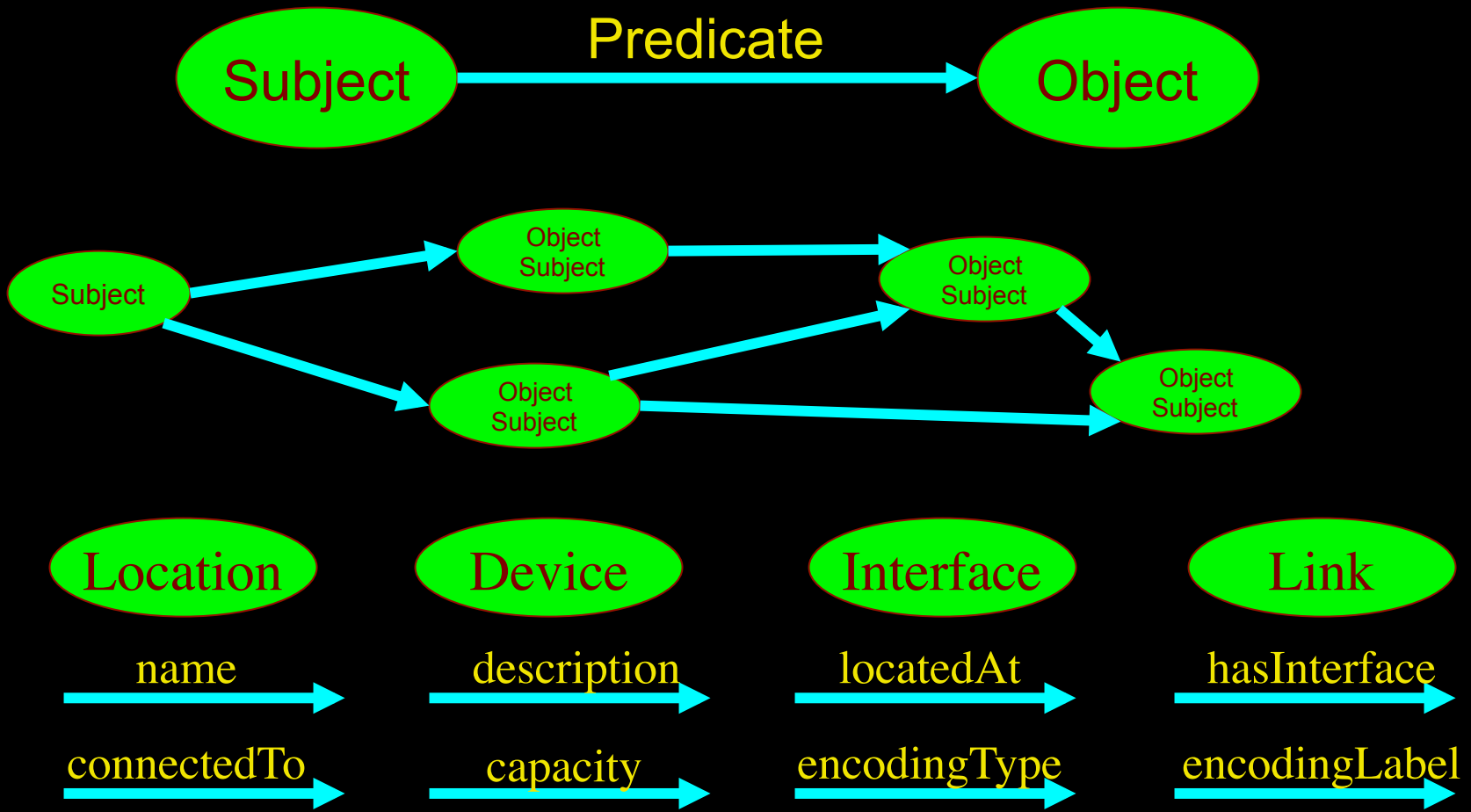
for



# 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):

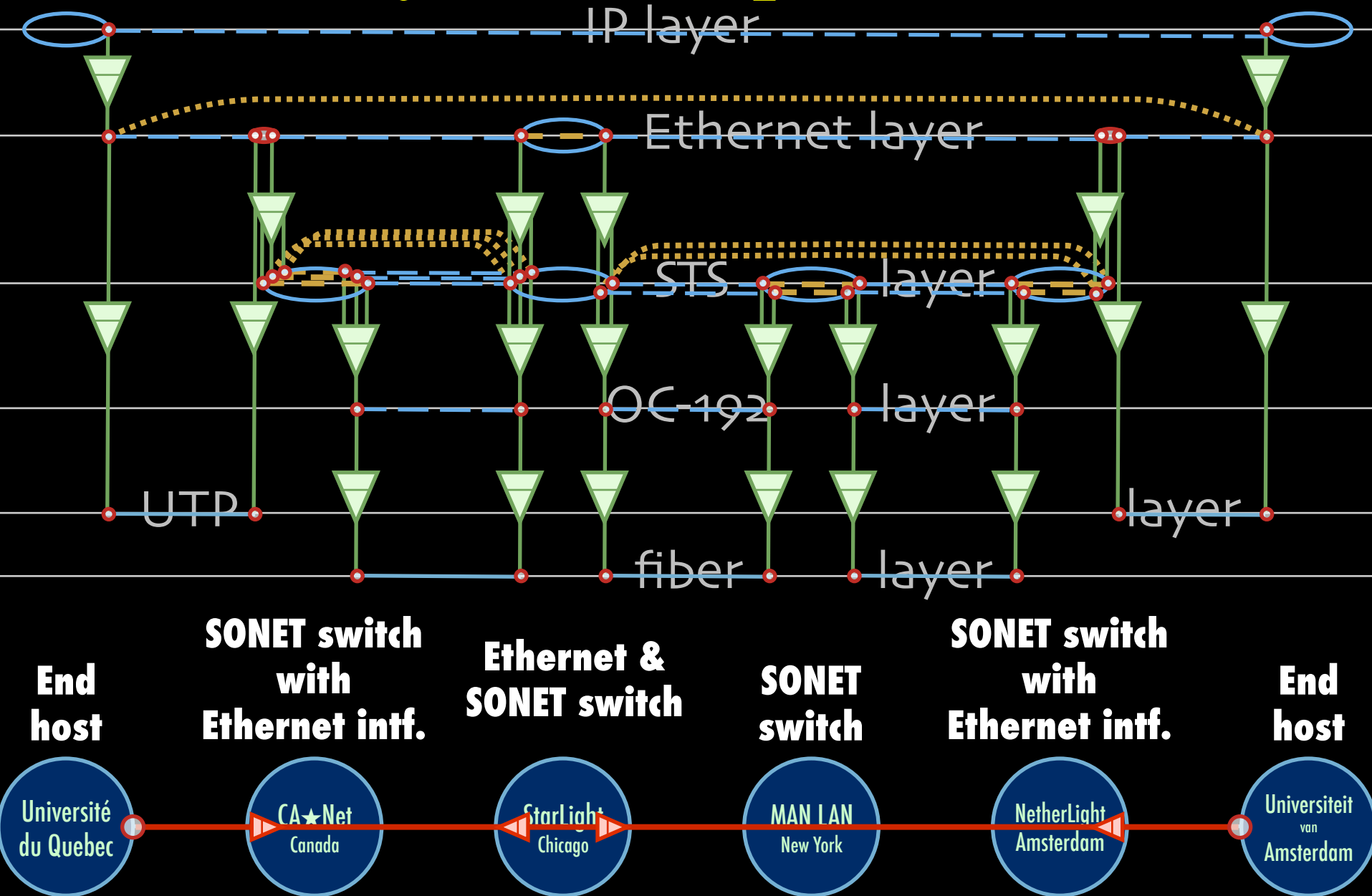




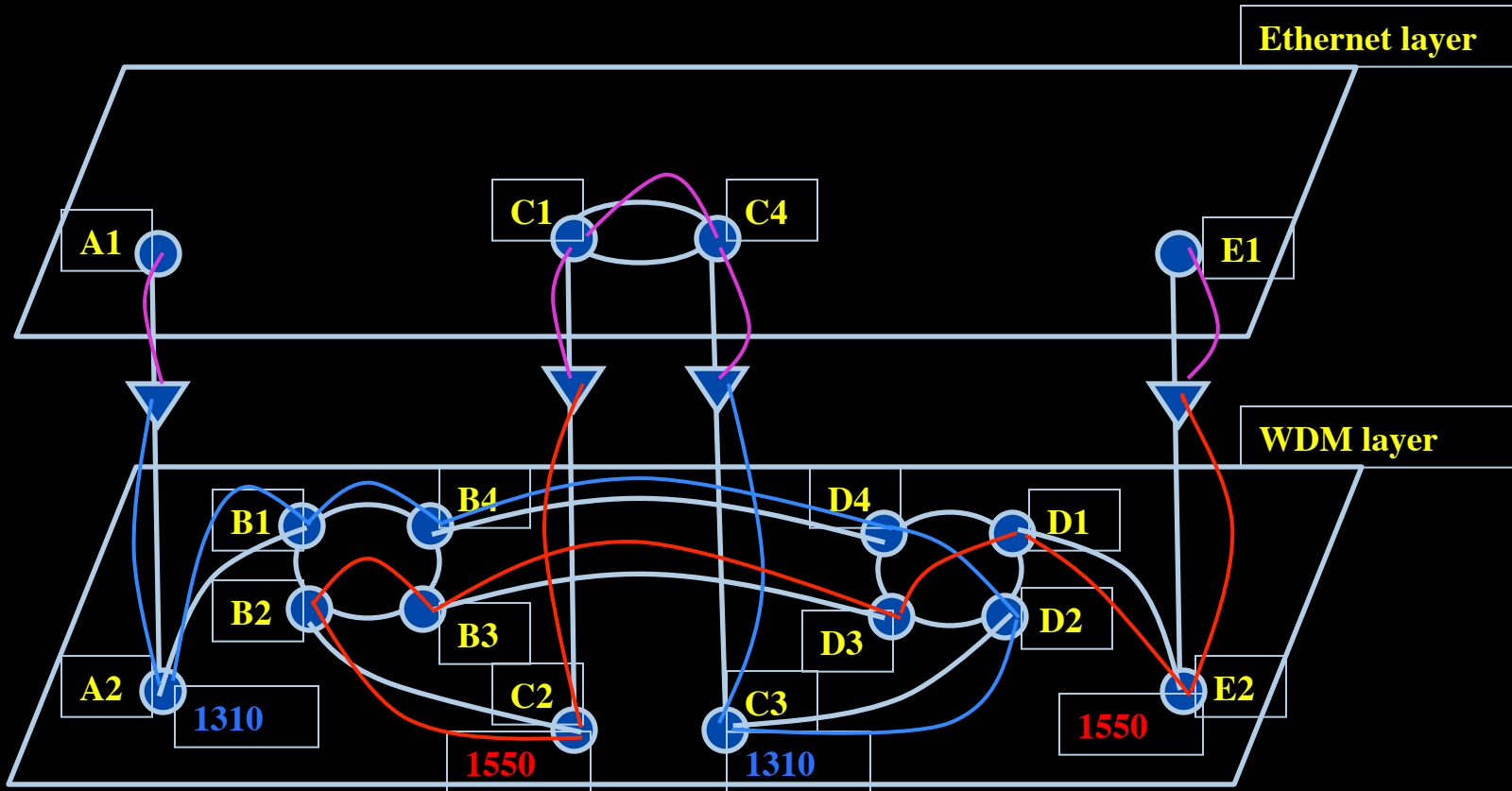
# NetherLight in RDF

```
<?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:

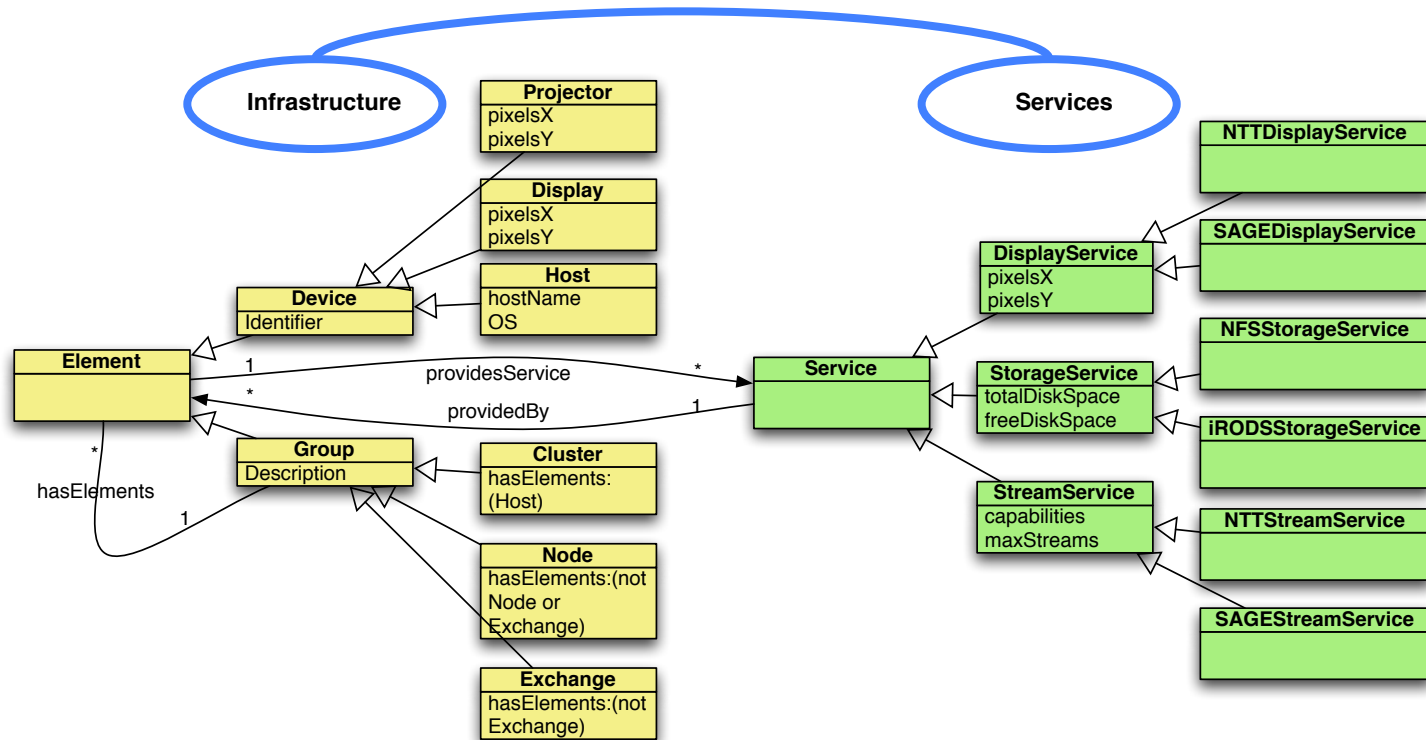
A1-A2-B1-B4-D4-D2-C3-C4-C1-C2-B2-B3-D3-D1-E2-E1

Scaling: Combinatorial problem



# Information Modeling

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





# Why?



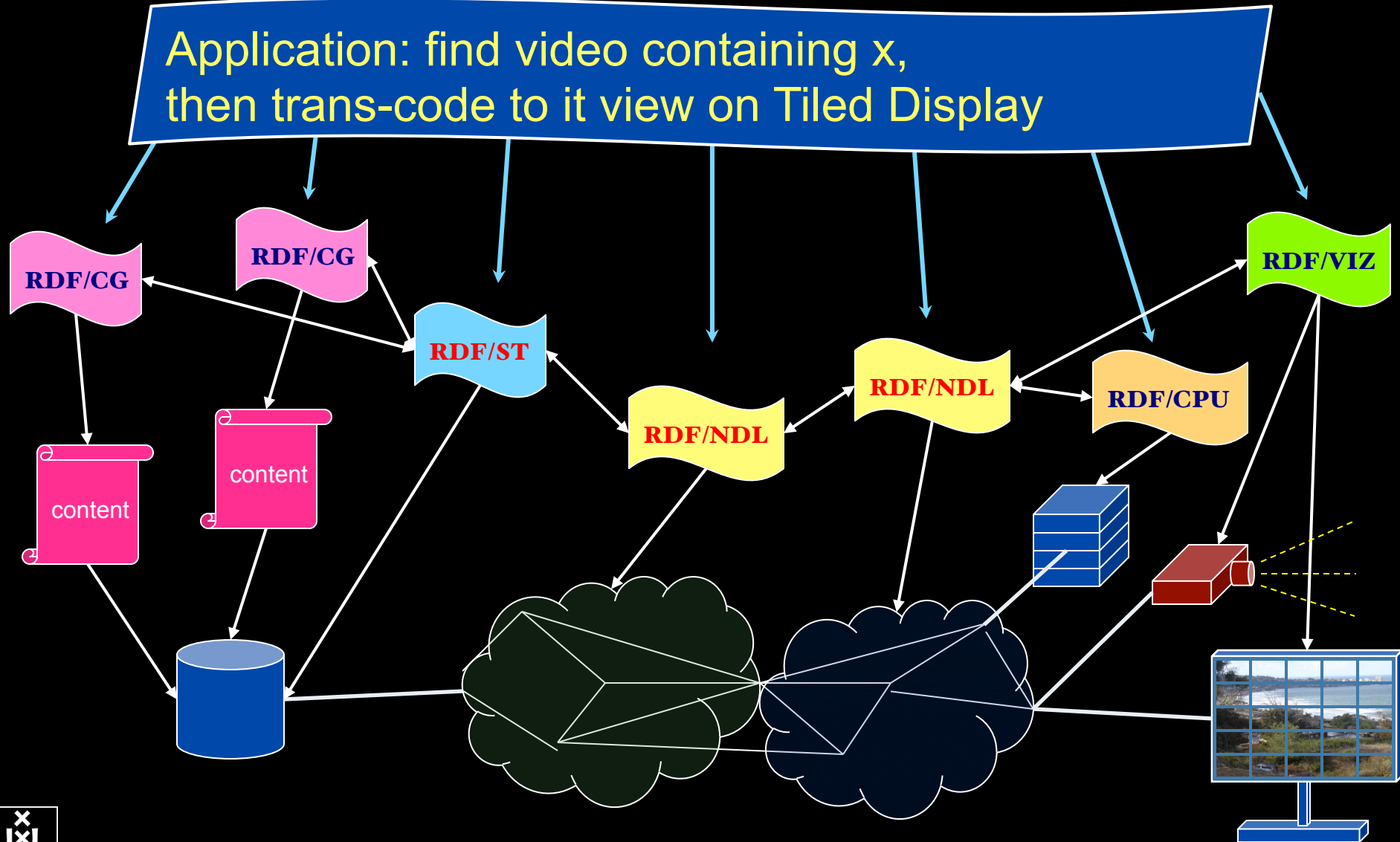
## I want to:

“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.

# RDF describing Infrastructure “I want”

Application: find video containing x,  
then trans-code to it view on Tiled Display





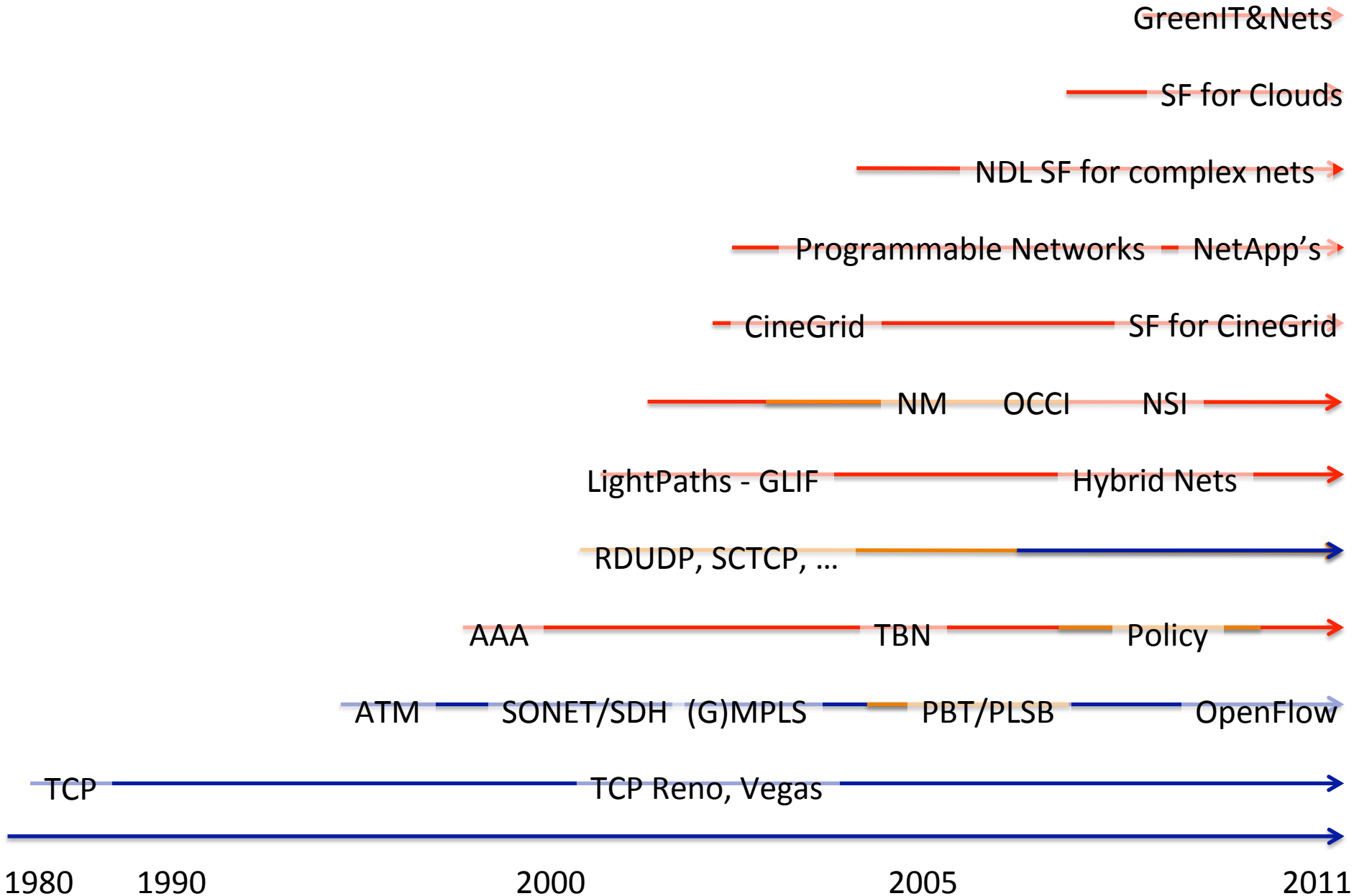
# The Ten Problems with the Internet

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**

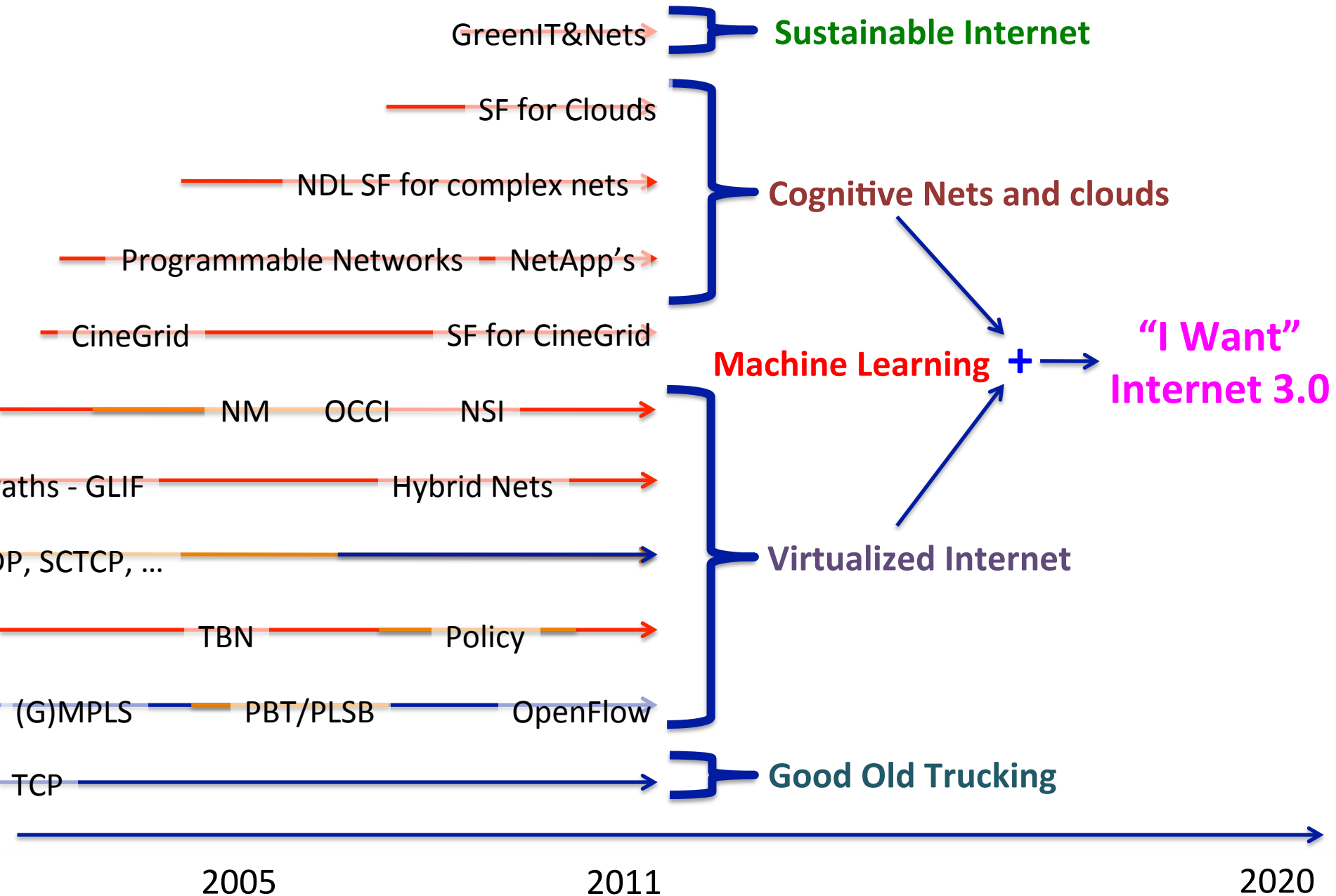
*Nice to have:*

- 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**

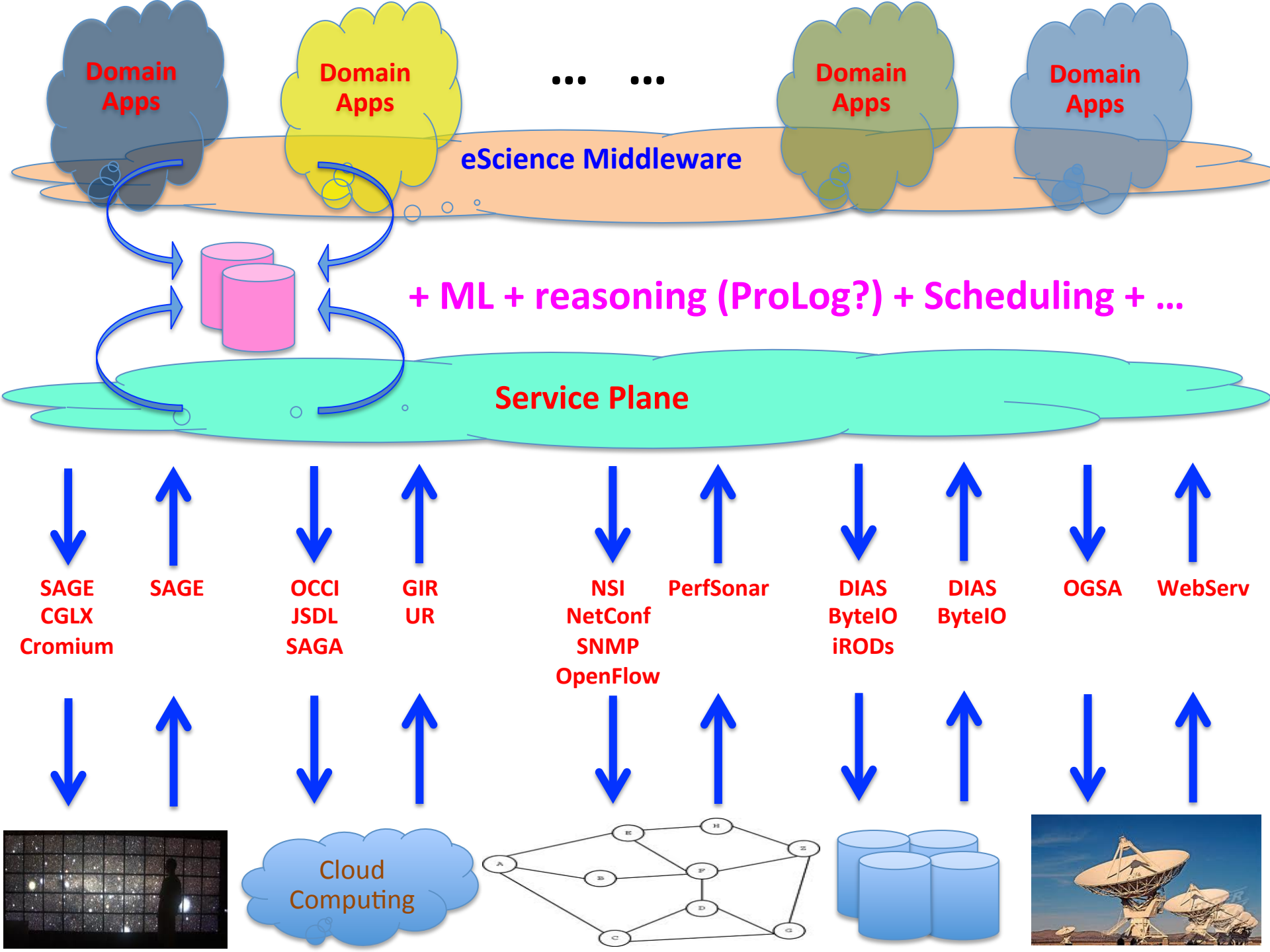
# TimeLine

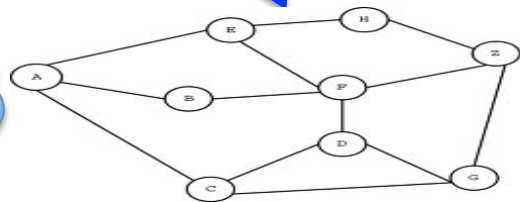
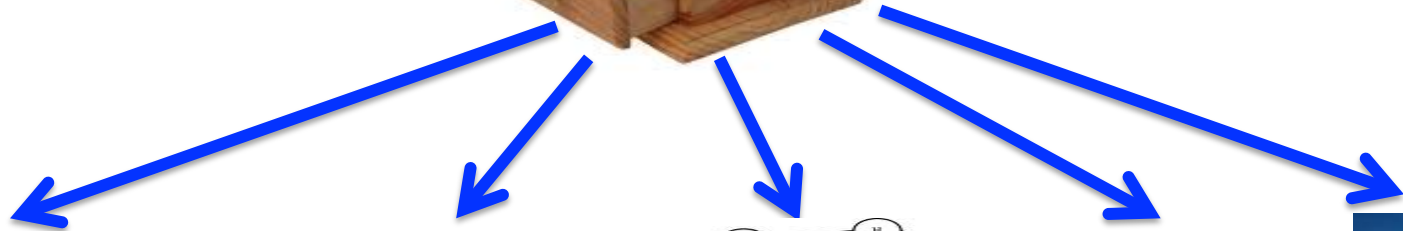
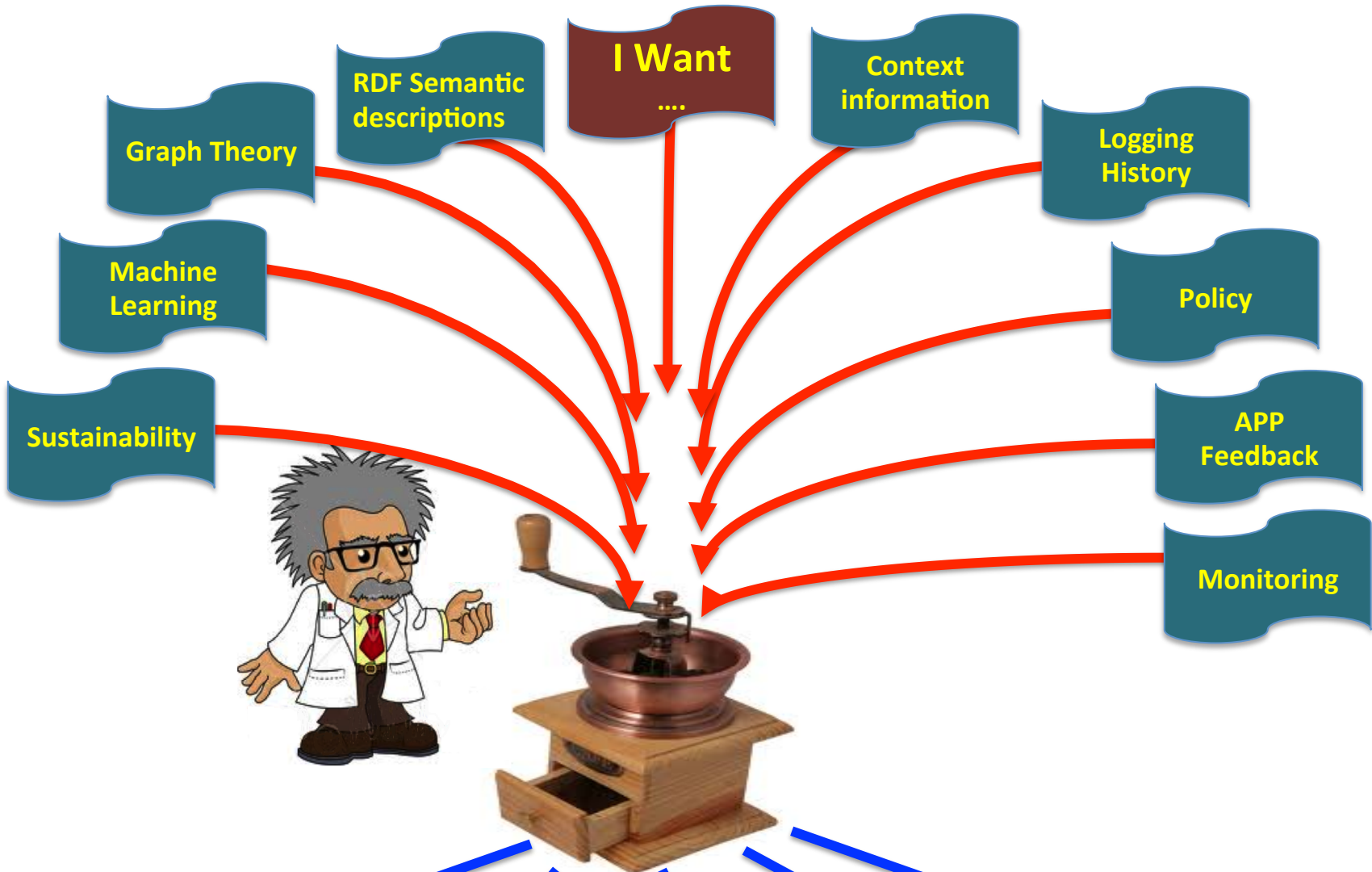


# TimeLine









# 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](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...

# ECO-Scheduling



