

Lambda-Grid developments

Global Lambda Integrated Facility

www.science.uva.nl/~deLaat

Cees de Laat

SURFnet

EU



University of Amsterdam

SARA
NCF



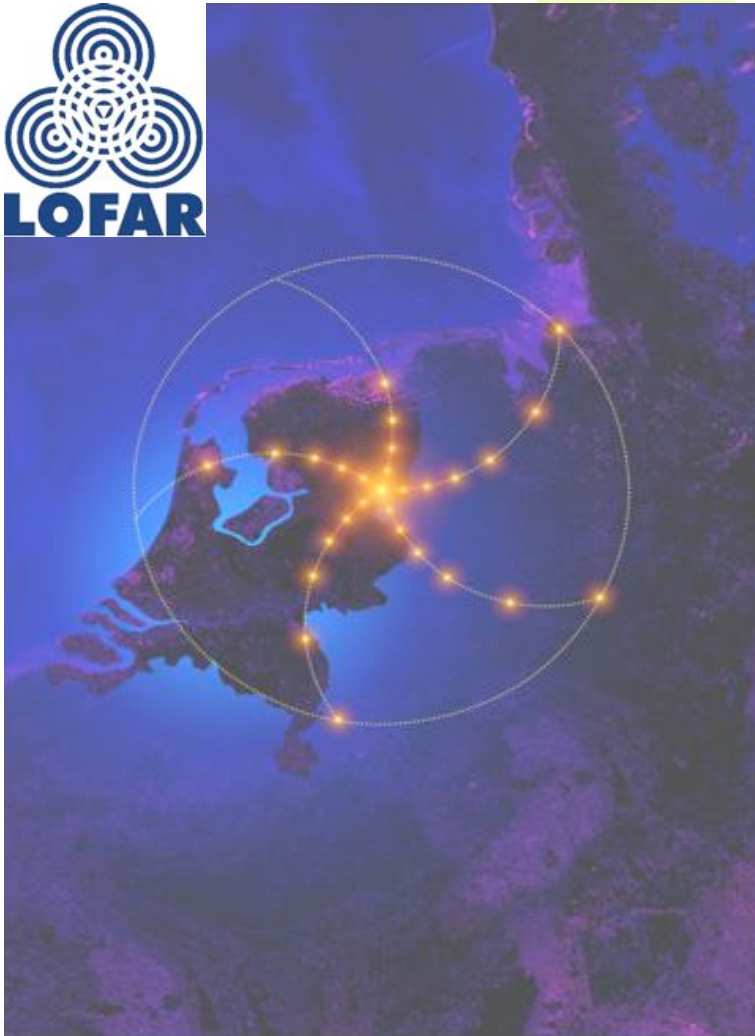
Contents

This page is intentionally left blank

- Ref: www.this-page-intentionally-left-blank.org

Sensor Grids

eVLBI



longer term VLBI is easily capable of generating
The sensitivity of the VLBI array scales with
width (=data-rate) and there is a strong push to mo
dths. Rates of 8Gb/s or more are entirely feasible.
under development. It is expected that parallel
ed correlator will remain the most efficient approach
olves dist
, multi-gig
relator and
g factor.



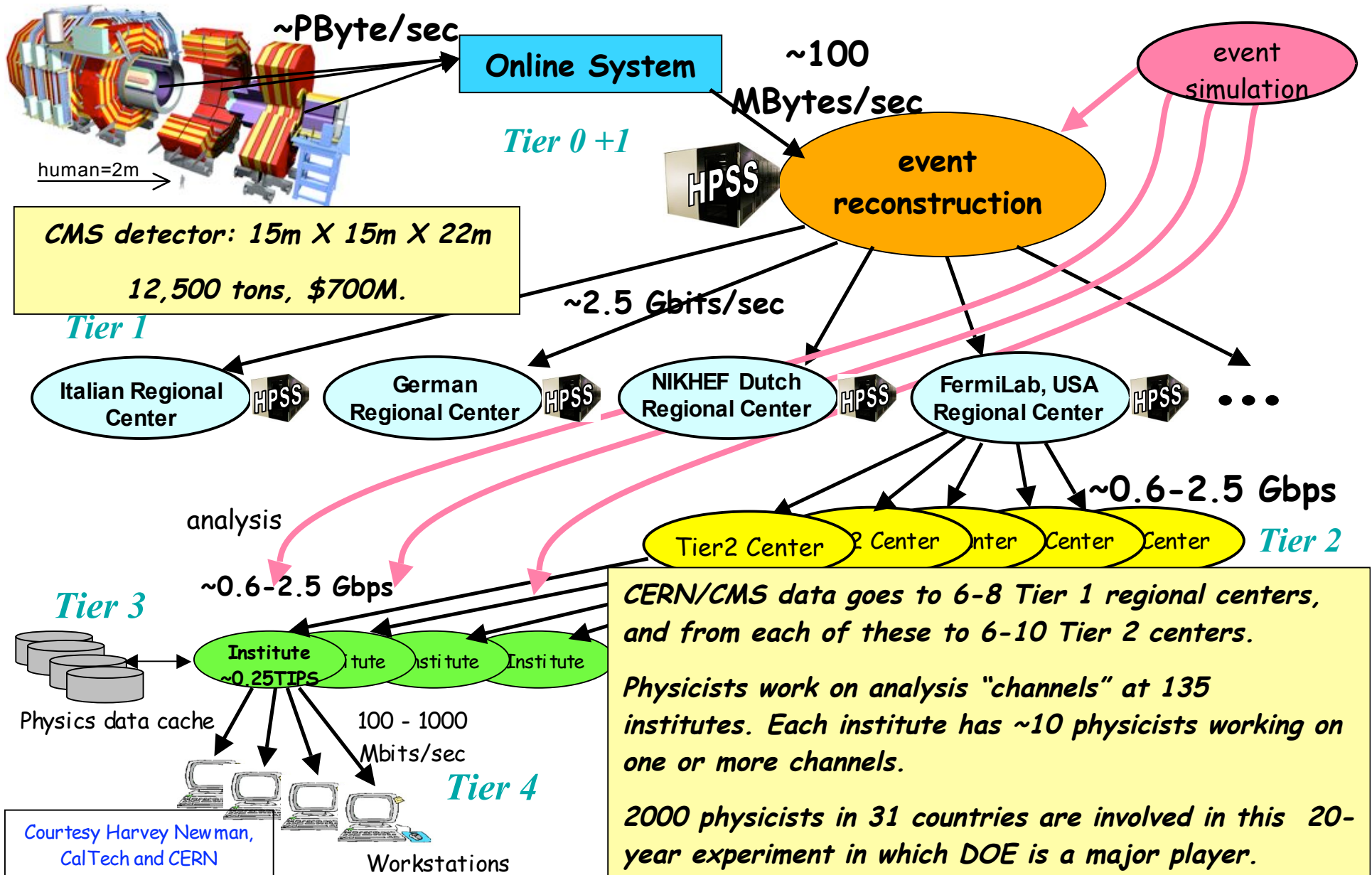
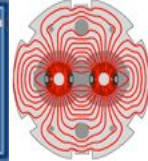
*Westerbork Synthesis Radio Telescope -
Netherlands*

~ 40 Tbit/s
www.lofar.org

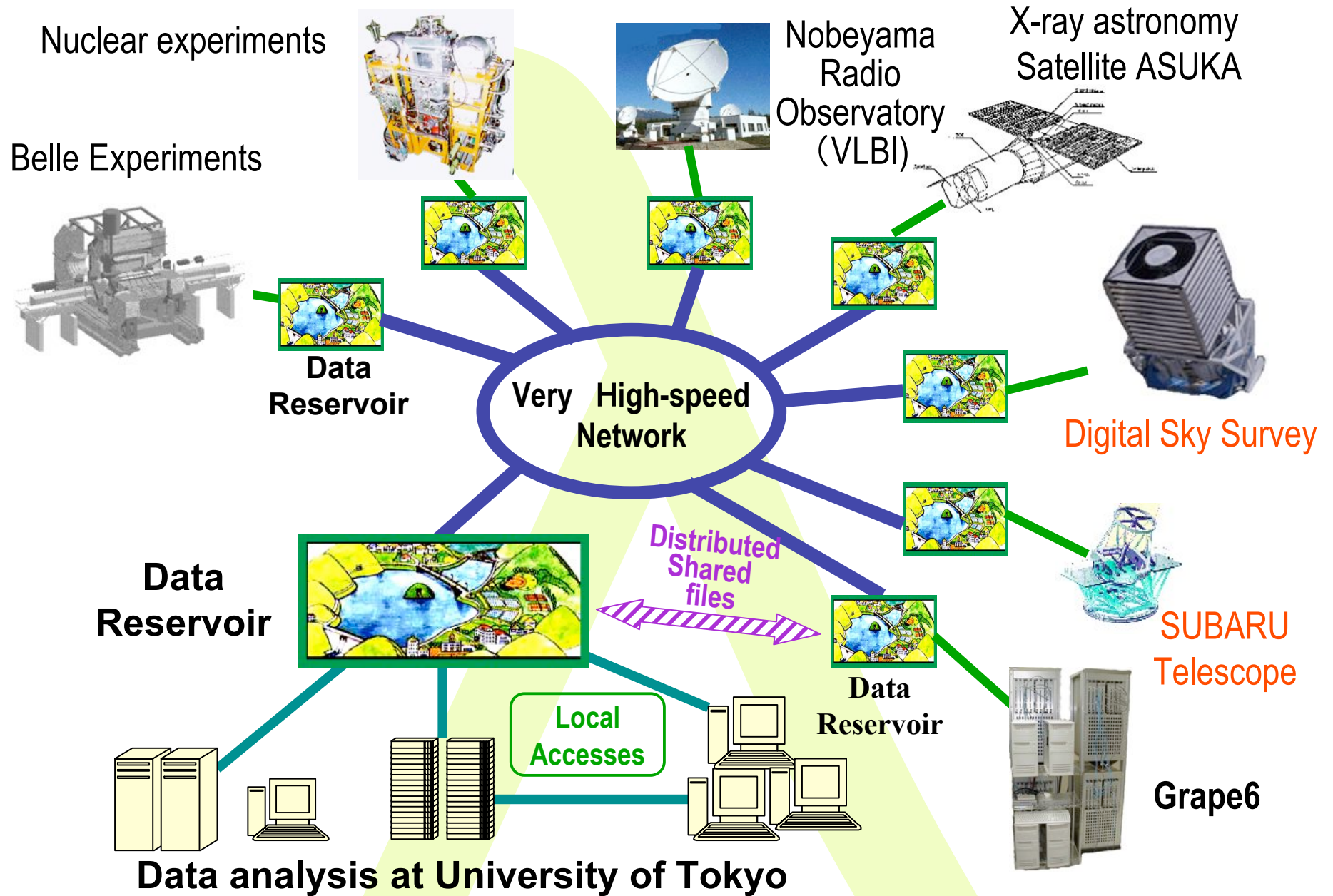


LHC Data Grid Hierarchy

CMS as example, Atlas is similar

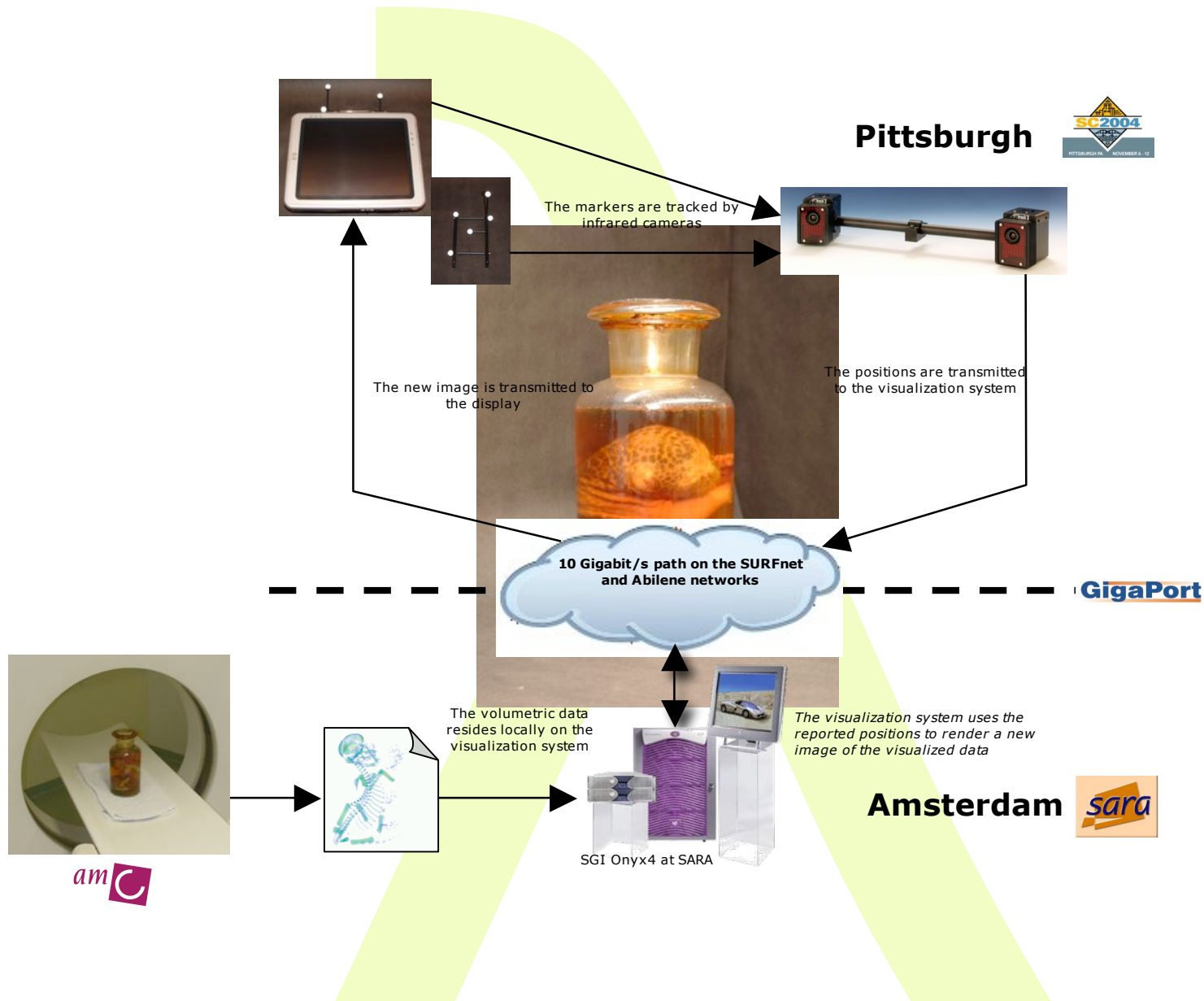


Data intensive scientific computation through global networks





Co-located interactive 3D visualization



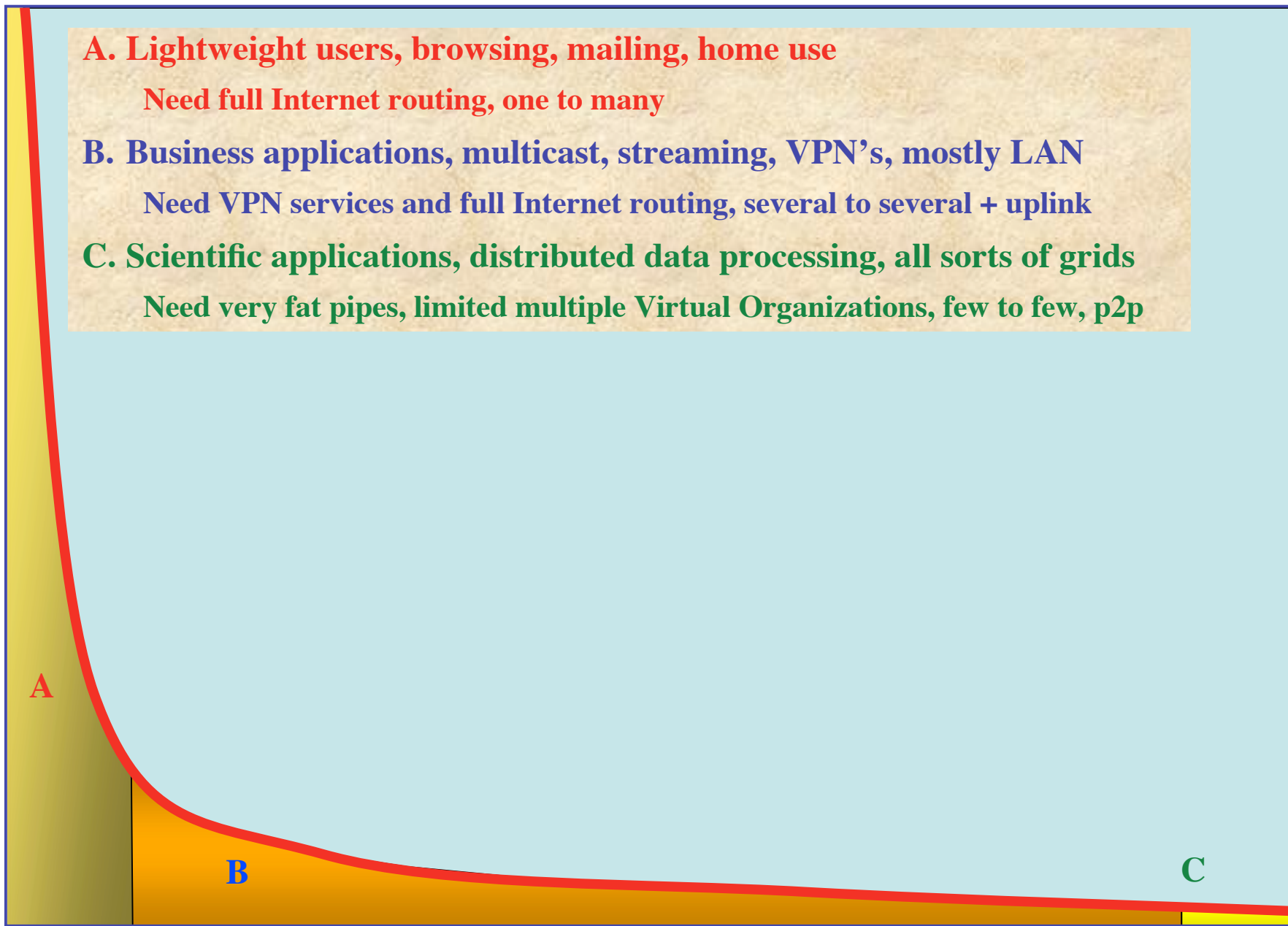


Showed you 5 types of Grids

- Sensor Grids
 - Several massive data sources are coming online
- Computational Grids
 - HEP and LOFAR analysis needs massive CPU capacity
 - Research: dynamic nation wide optical backplane control
- Data (Store) Grids
 - Moving and storing HEP, Bio and Health data sets is major challenge
- Visualization Grids
 - Data object (TByte sized) inspection, anywhere, anytime
- Lambda Grids
 - Hybrid networks

U
S
E
R
S

- A. Lightweight users, browsing, mailing, home use**
Need full Internet routing, one to many
- B. Business applications, multicast, streaming, VPN's, mostly LAN**
Need VPN services and full Internet routing, several to several + uplink
- C. Scientific applications, distributed data processing, all sorts of grids**
Need very fat pipes, limited multiple Virtual Organizations, few to few, p2p



ADSL

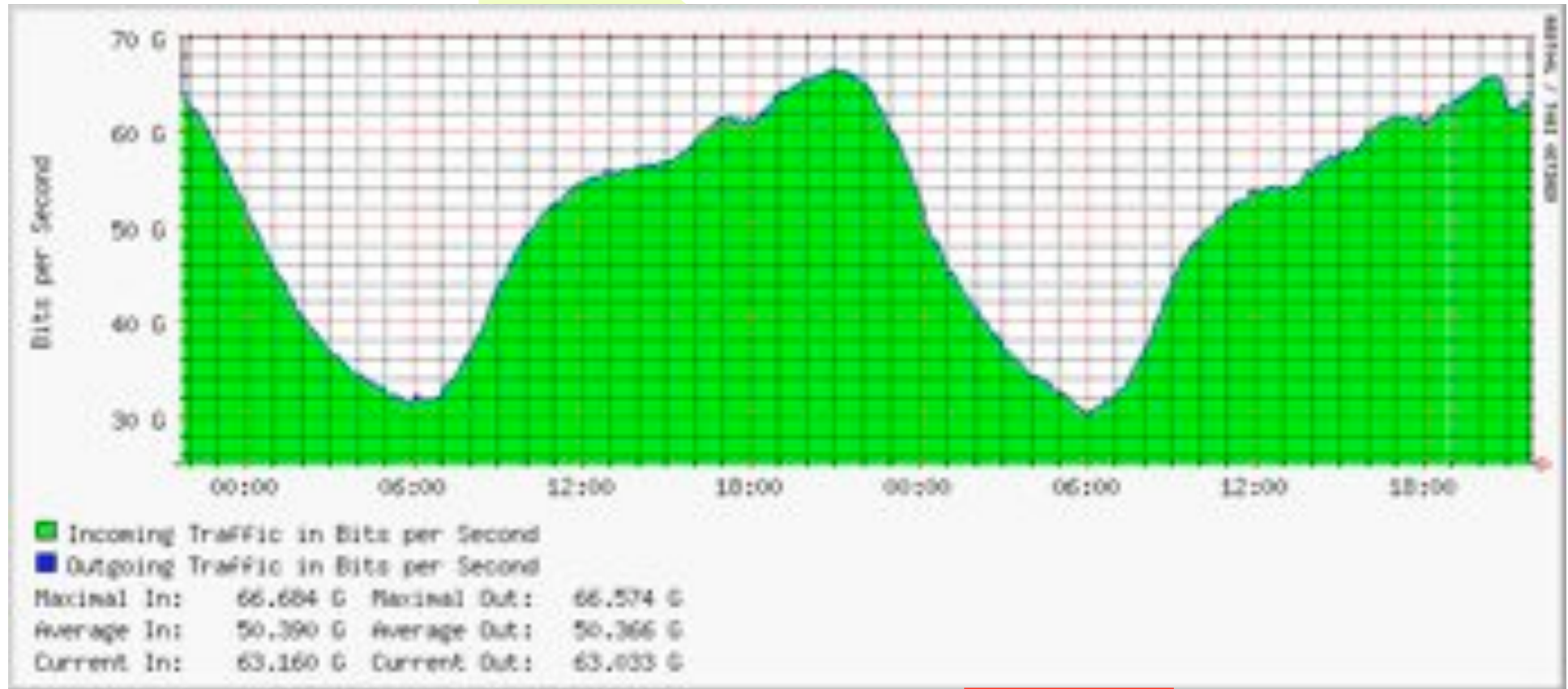
GigE

BW requirements

The Dutch Situation (in 2005)

- **Estimate A**
 - 17 M people, 6.4 M households, 25 % penetration of 0.5 - 8 Mb/s ADSL, 40 times under-provisioning ==> ~ 40 Gb/s

AMS-IX



June 19th 2004

May 2005

Lost :-)

European championship football **Holland -- Czech Republic**

The Dutch Situation (in 2005)

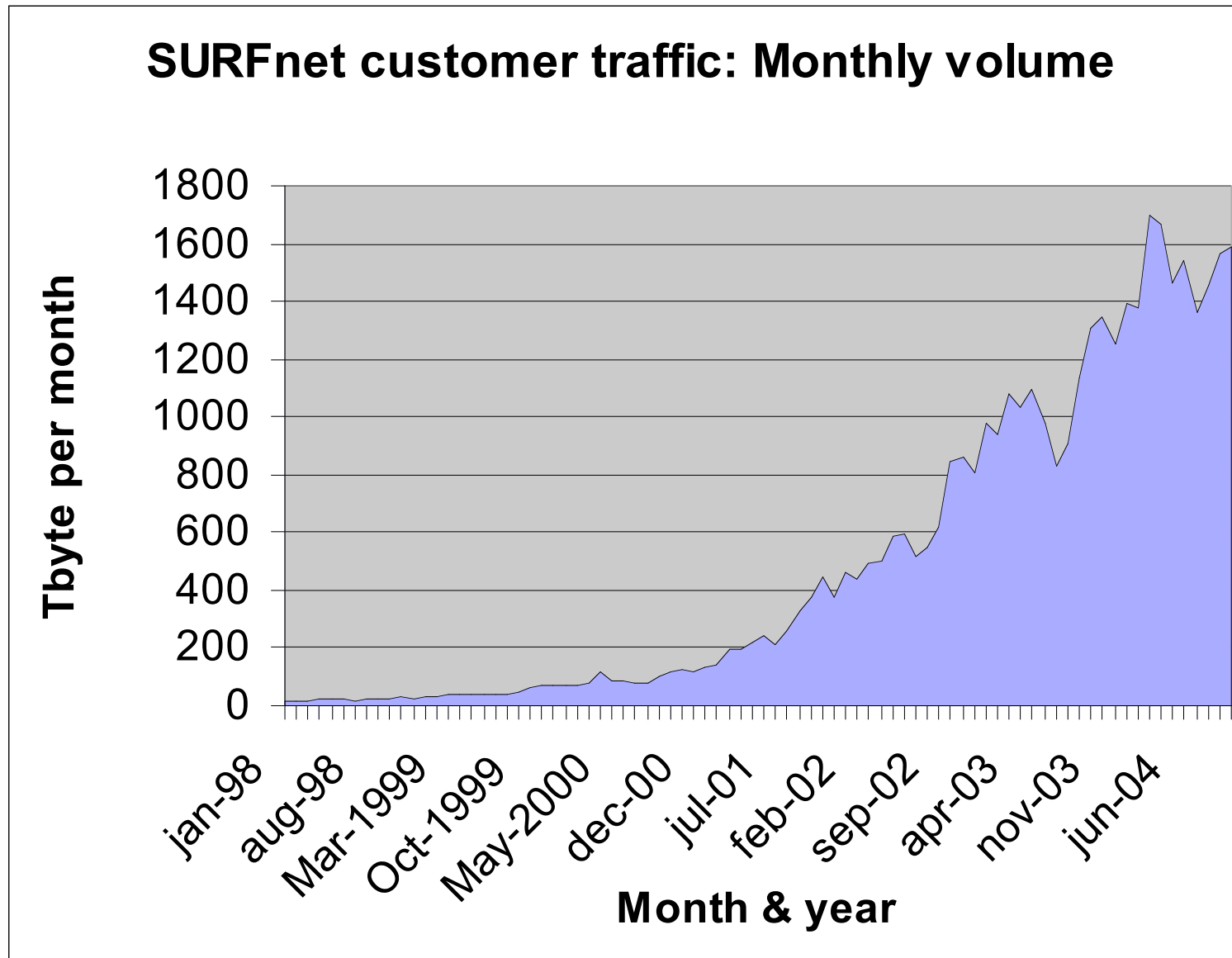
- **Estimate A**

- 17 M people, 6.4 M households, 25 % penetration of 0.5 - 8 Mb/s ADSL, 40 times under-provisioning ==> ~ 40 Gb/s

- **Estimate B**

- SURFnet5 has 2*10 Gb/s to about 15 institutes and 0.1 to 1 Gb/s to 170 customers, estimate same for industry (overestimation) ==> 10-30 Gb/s

Routed L3 traffic growth



1900 TByte/month \approx 6 Gbits/second

Slide courtesy Kees Neggers

The Dutch Situation (in 2005)

- **Estimate A**

- 17 M people, 6.4 M households, 25 % penetration of 0.5 - 8 Mb/s ADSL, 40 times under-provisioning $\implies \sim 40$ Gb/s

- **Estimate B**

- SURFnet5 has 2×10 Gb/s to about 15 institutes and 0.1 to 1 Gb/s to 170 customers, estimate same for industry (overestimation) $\implies 10$ -30 Gb/s

- **Estimate C**

- Leading HEF and ASTRO + rest $\implies 80$ -120 Gb/s
- LOFAR $\implies \approx 37$ Tbit/s $\implies \approx n \times 10$ Gb/s

u
s
e
r
s

A. Lightweight users, browsing, mailing, home use

Need full Internet routing, one to many

B. Business applications, multicast, streaming, VPN's, mostly LAN

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

C. Scientific applications, distributed data processing, all sorts of grids

Need very fat pipes, limited multiple Virtual Organizations, few to few, p2p

$\Sigma C \gg 100 \text{ Gb/s}$ →

$\Sigma B \approx 30 \text{ Gb/s}$

$\Sigma A \approx 40 \text{ Gb/s}$

A

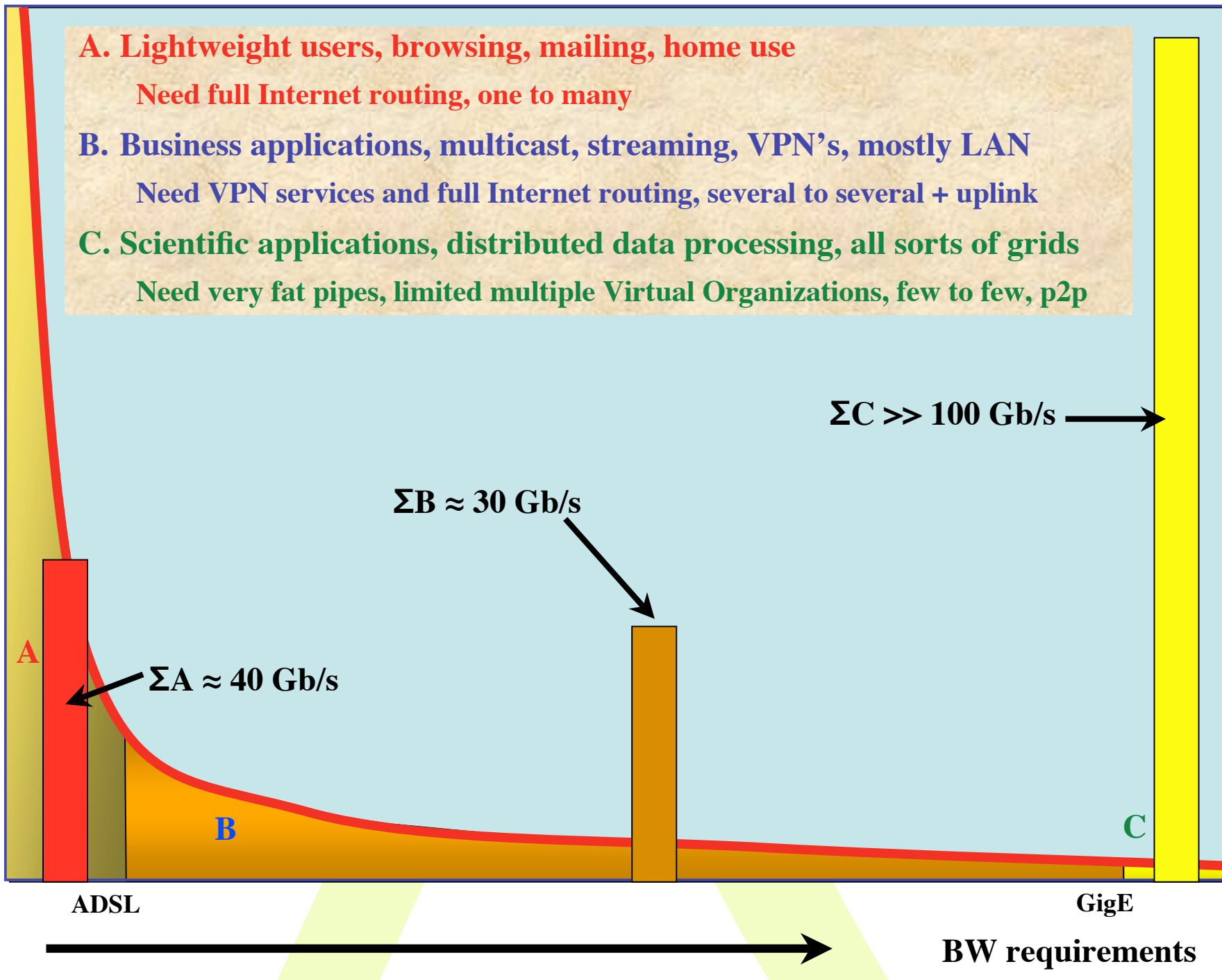
B

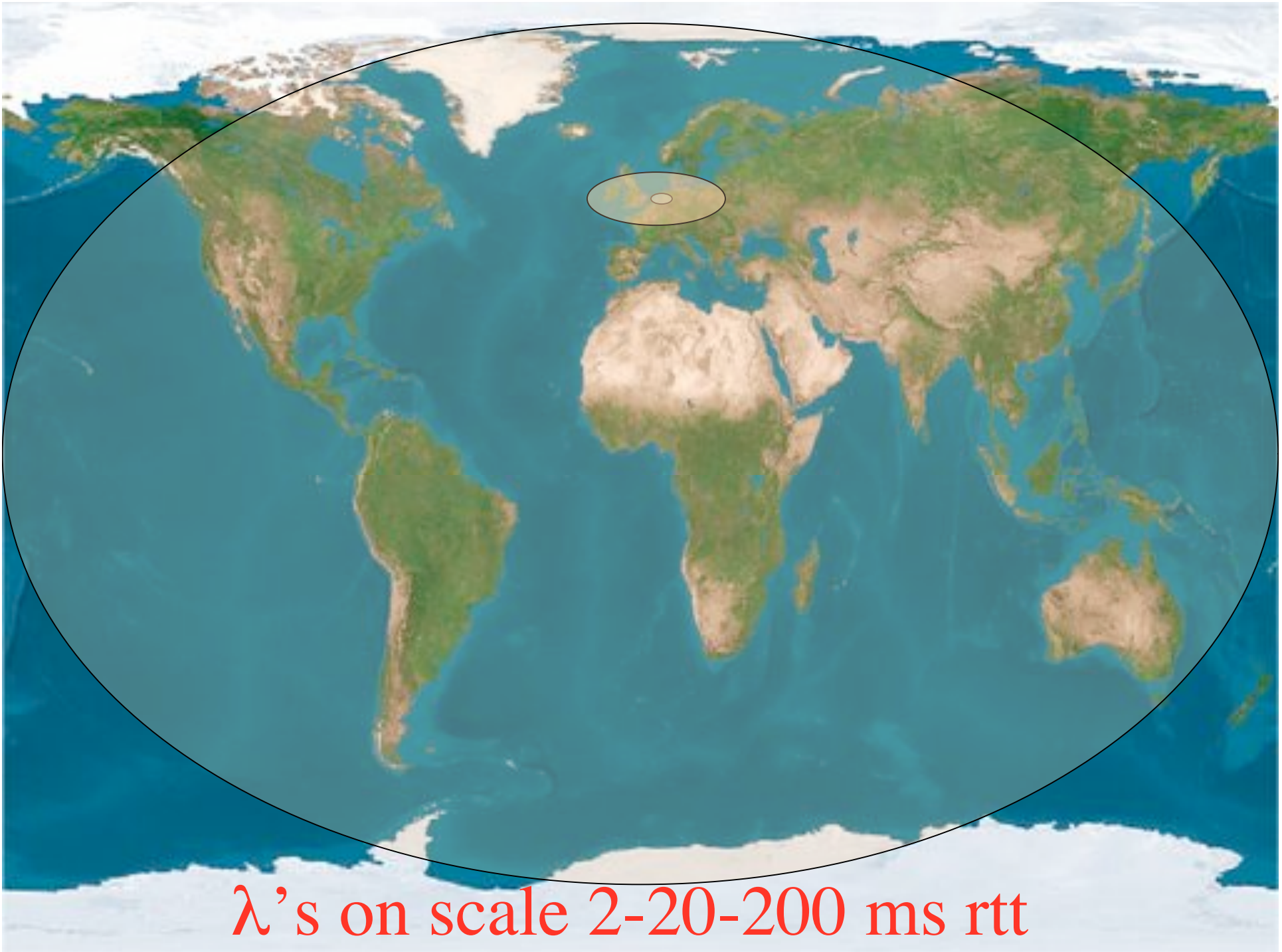
C

ADSL

GigE

→
BW requirements





Towards Hybrid Networking!

- Costs of optical equipment 10% of switching 10 % of full routing equipment for same throughput
 - 10G routerblade -> 100-500 k\$, 10G switch port -> 7-15 k\$, MEMS port -> 1 k\$
 - 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)
- Give each packet in the network the service it needs, but no more !

L1 \approx 1 k\$/port



L2 \approx 7-15 k\$/port

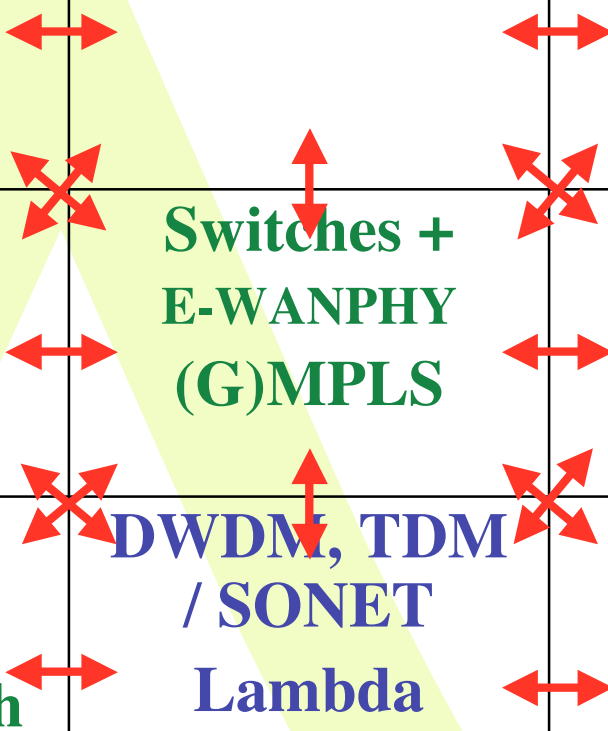
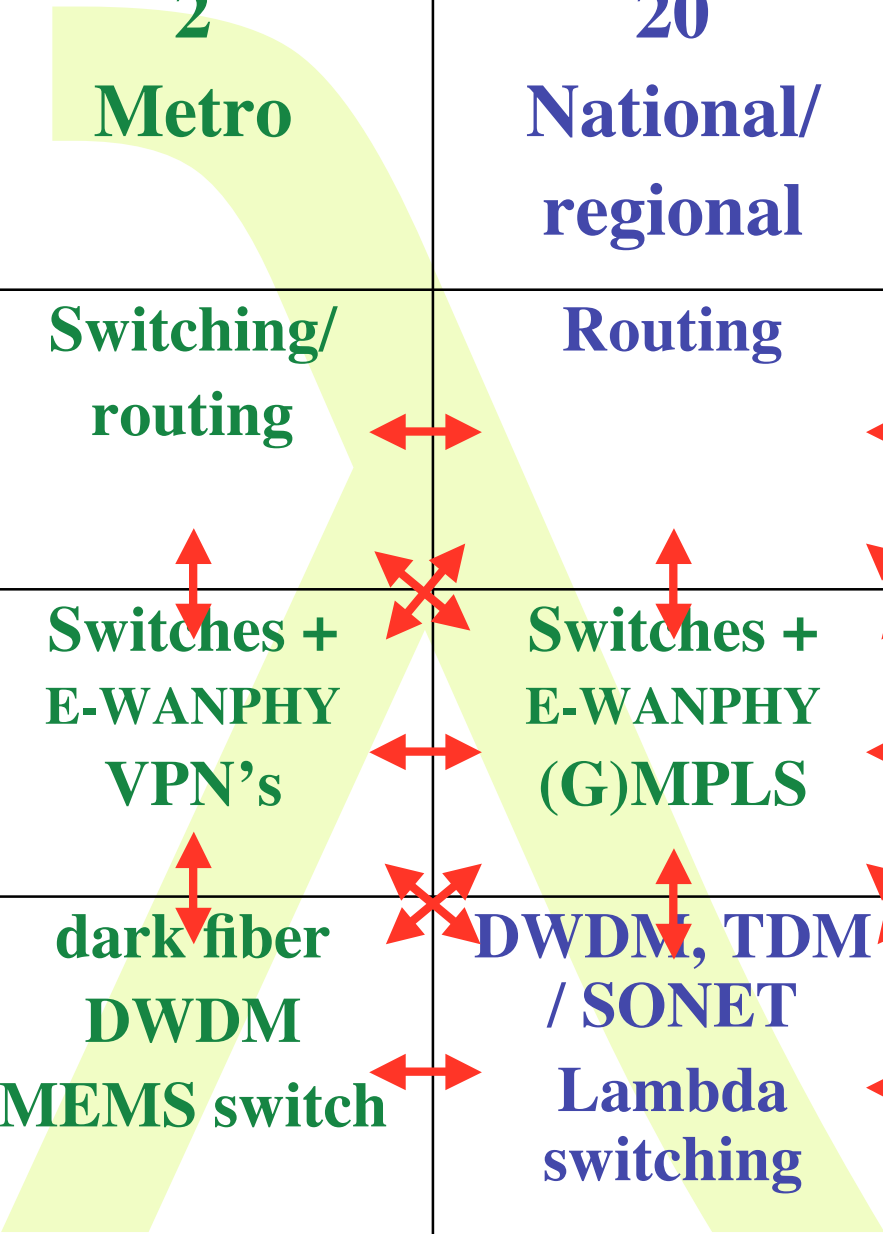


L3 \approx 100+ k\$/port

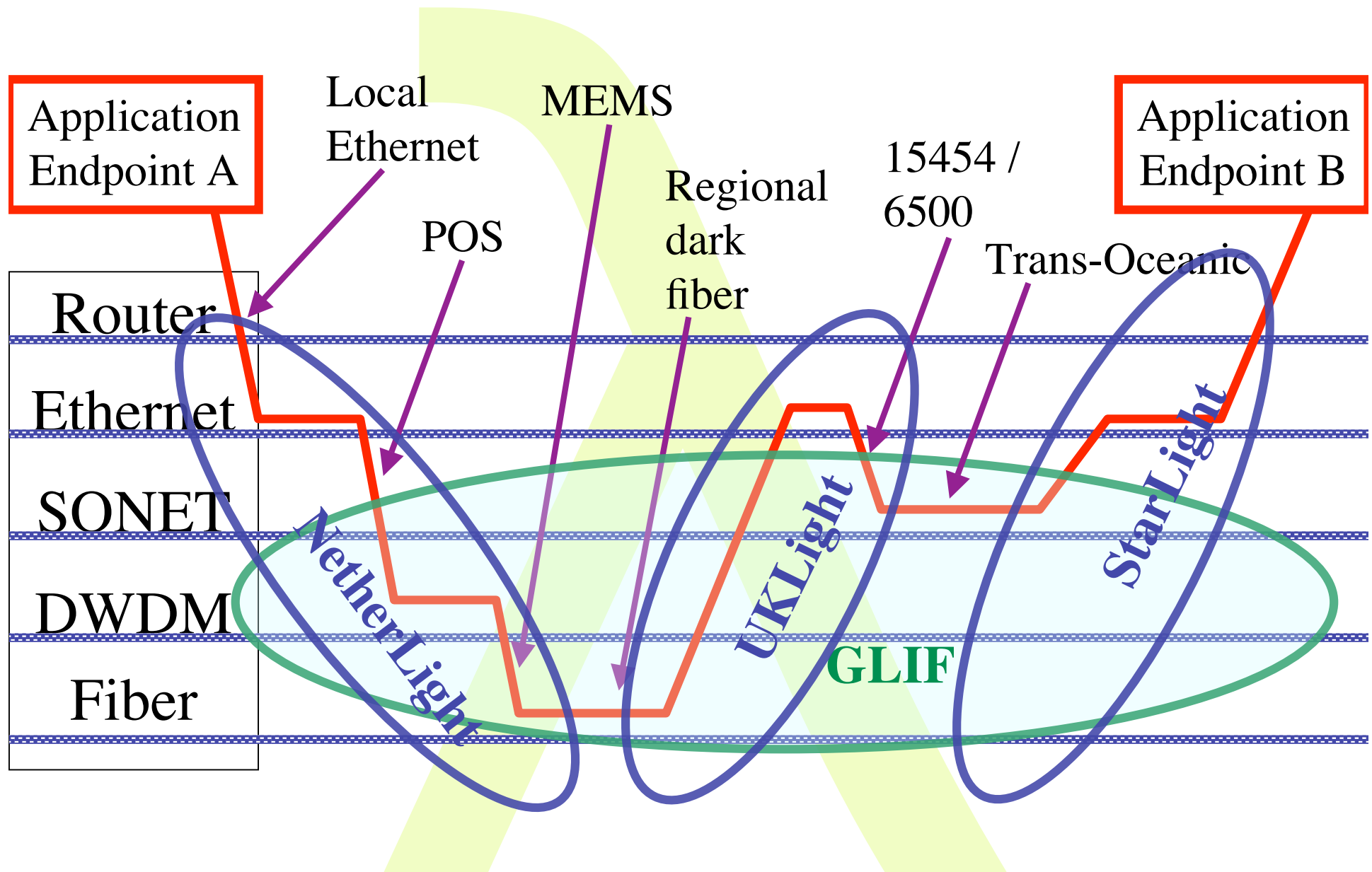


Services

<div style="text-align: right;">SCALE</div> <div style="text-align: left;">CLASS</div>	2 Metro	20 National/ regional	200 World
A	Switching/ routing	Routing	ROUTER\$
B	Switches + E-WANPHY VPN's	Switches + E-WANPHY (G)MPLS	ROUTER\$
C	dark fiber DWDM MEMS switch	DWDM, TDM / SONET Lambda switching	Lambdas, VLAN's SONET Ethernet

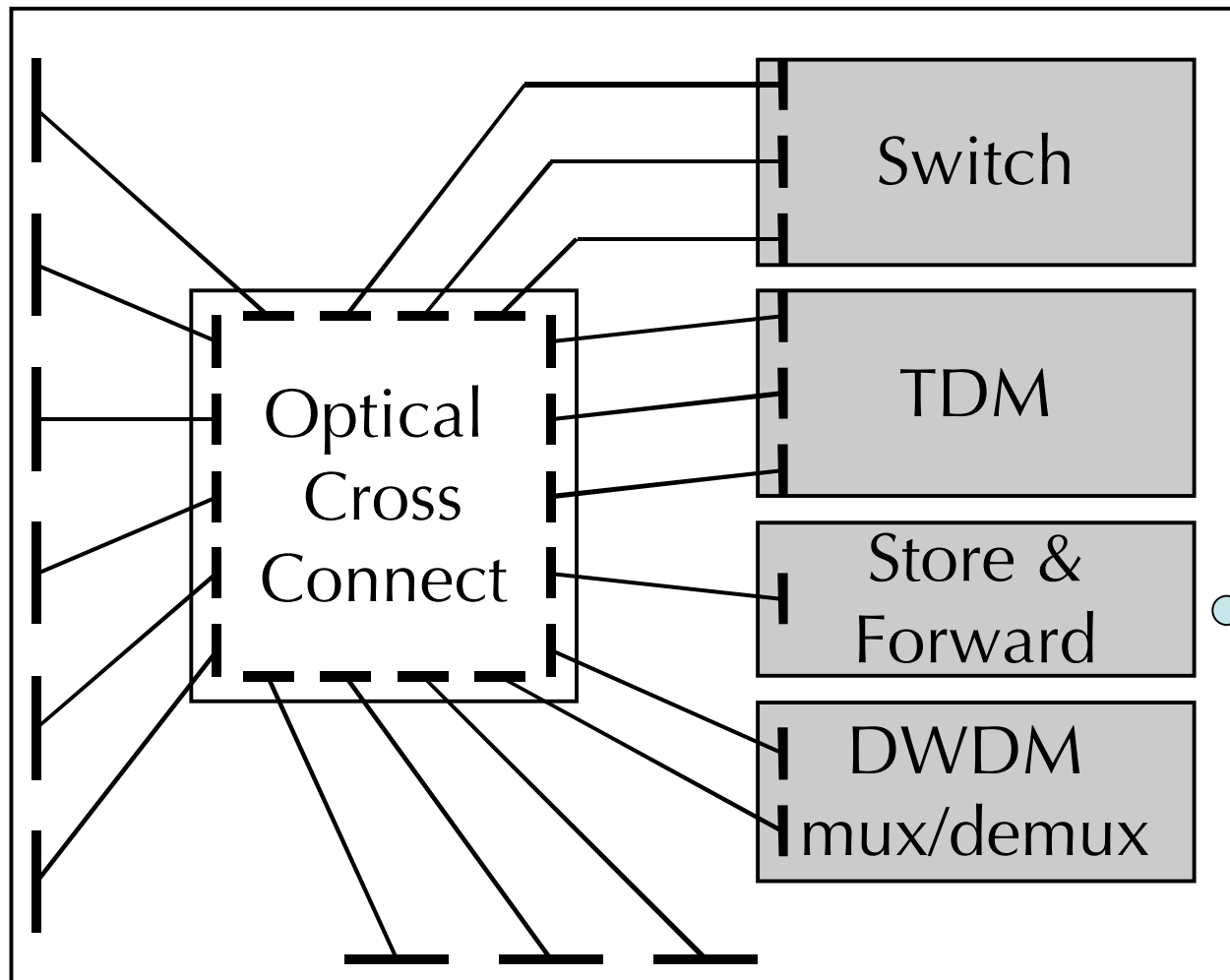


How low can you go?



Optical Exchange as Black Box

Optical Exchange



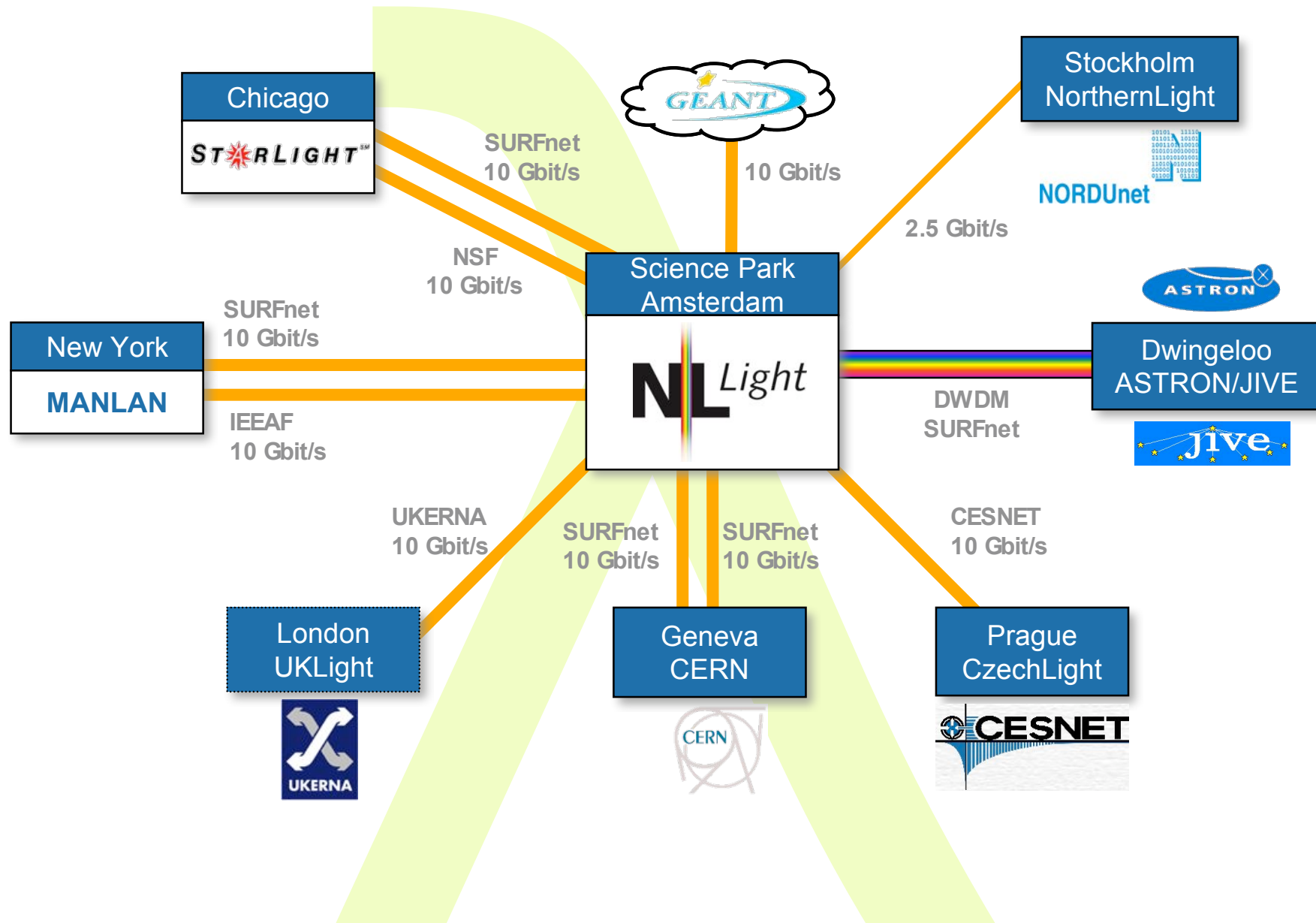
TeraByte
Email
Service



GLIF Q4 2004

Visualization courtesy of Bob Patterson, NCSA.

NetherLight Q1 2005



GLIF History

- **Brainstorming in Antalya at Terena conf. 2001**
- **1th meeting at Terena offices 11-12 sep 2001**
 - **On invitation only (15) + public part**
 - **Thinking, SURFnet test lambda Starlight-Netherlight**
- **2nd meeting appended to iGrid 2002 in Amsterdam**
 - **Public part in track, on invitation only day (22)**
 - **Core testbed brainstorming, idea checks, seeds for Translight**
- **3th meeting Reykjavik, hosted by NORDUnet 2003**
 - **Grid/Lambda track in conference + this meeting (35!)**
 - **Brainstorm applications and showcases**
 - **Technology roadmap**
 - **GLIF established --> glif.is**
- **4th meeting Nottingham (UK), hosted by UKERNA, 2-3 September 2004**
 - **60 participants**
 - **Attendance from China, Japan, Netherlands, Switzerland, US, UK, Taiwan, Australia, Tsjech, Korea, Canada, Ireland, Russia, Belgium, Denmark**
 - **Truly Worldwide!**

GLIF Mission Statement

- **GLIF is a world-scale Lambda-based Laboratory for application and middleware development on emerging LambdaGrids, where applications rely on dynamically configured networks based on optical wavelengths**
- **GLIF is an environment (networking infrastructure, network engineering, system integration, middleware, applications) to accomplish real work**

Working groups

GLIF Governance and policy

Our small-scale Lambda Workshop is now turning into a global activity. TransLight and similar projects contribute to the infrastructure part of GLIF. A good and well understood governance structure is key to the manageability and success of GLIF. Our prime goal is to decide upon and agree to the GLIF governance and infrastructure usage policy.

GLIF Lambda infrastructure and Lambda exchange implementations

A major function for previous Lambda Workshops was to get the network engineers together to discuss and agree on the topology, connectivity and interfaces of the Lambda facility. Technology developments need to be folded into the architecture and the expected outcome of this meeting is an agreed view on the interfaces and services of Lambda exchanges and a connectivity map of Lambdas for the next year, with a focus on iGrid 2005 and the emerging applications.

Persistent Applications and research

Key to the success of the GLIF effort is to connect the major applications to the Facility. We, therefore, need a list of prime applications to focus on and a roadmap to work with those applications to get them up to speed. The demonstrations at SC2004 and iGrid 2005 can be determined in this meeting.

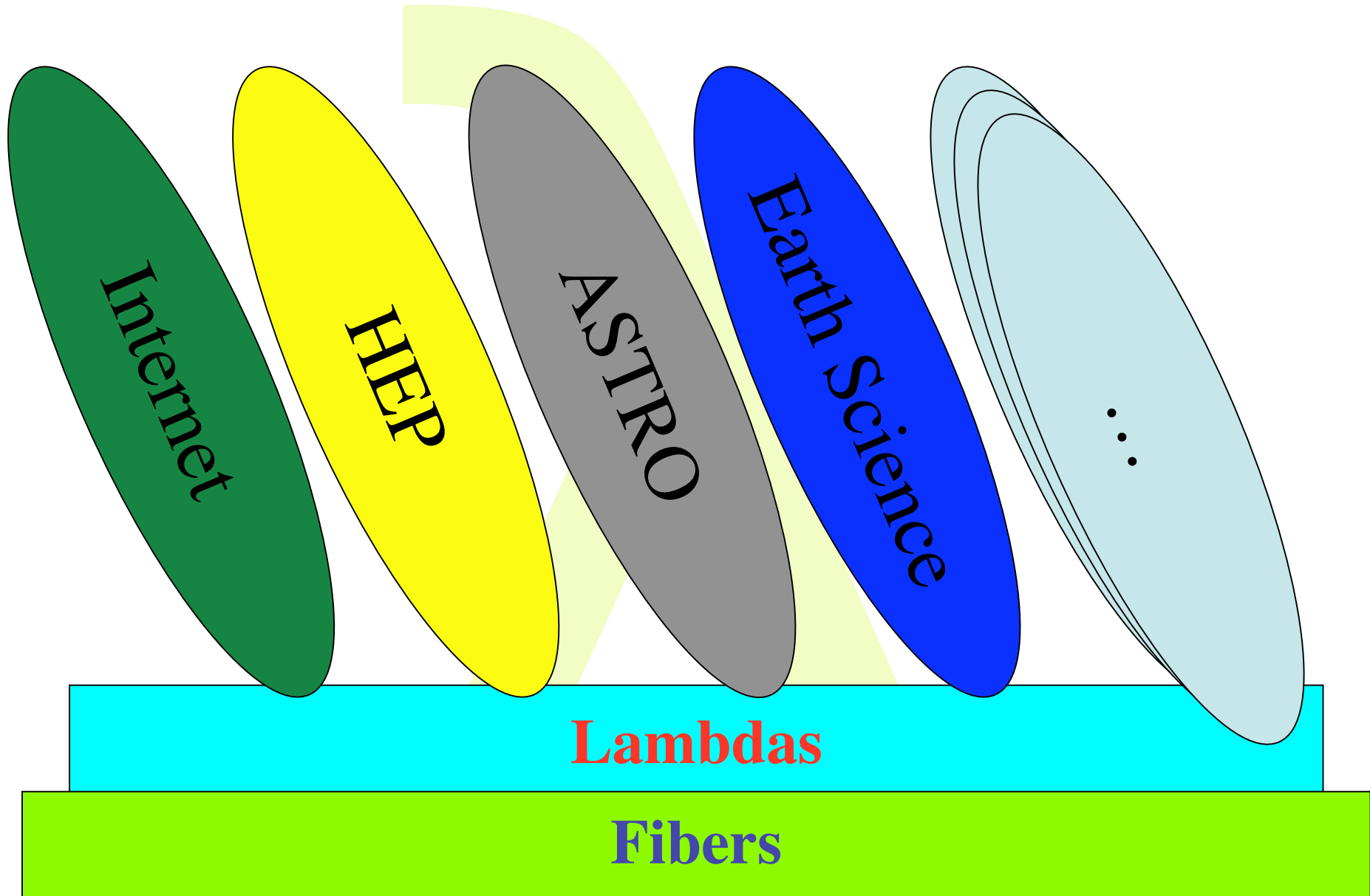
Control Plane and Grid Integration

The GLIF can only function if we agree on the interfaces and protocols that talk to each other in the control plane on the contributed Lambda resources. The main players in this field are already meeting, almost on a bi-monthly schedule. Although not essential, this GLIF meeting could also host a breakout session on control plane middleware.

GLIF - 5 meeting

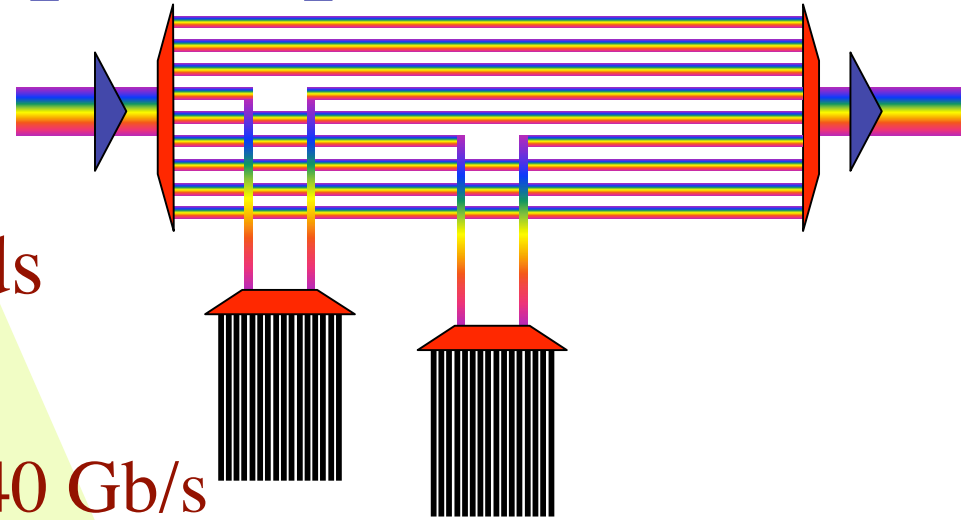
- Collocated with iGrid2005 San Diego
- CAL-(IT)²
- Thursday 29 sept 2005
 - Presentations track
- Friday 30 sept 2005
 - Work group meetings
- **NOT** on invitation only anymore!
 - Open meeting for participants
 - Industry rep's only on workgroup chairs invitation (no marketing!)

Discipline Networks



SURFnet 6 principles

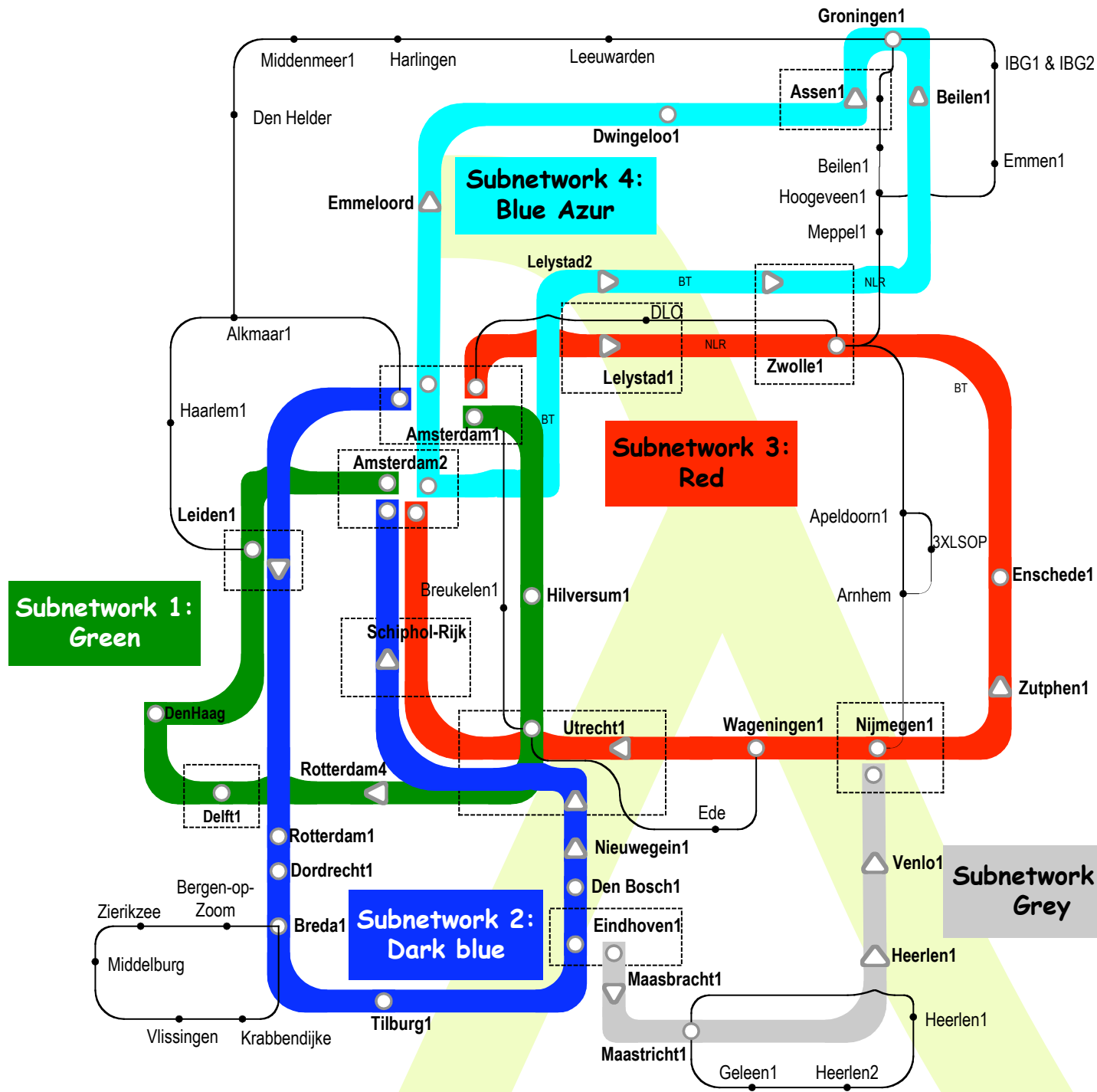
- Based on dark fiber
- 4 DWDM rings of 9 bands
 - each 4, later 8, colors
 - Each capable of 10, later 40 Gb/s
- Universities have POP's on ring, each 1 band
- Connect with 1 or 10 Gb/s Ethernet
- Routing in Amsterdam in 2 core POP's!
- International connectivity in Amsterdam
- Lambda service between ring POP's and to NetherLight





6000 km fiber
Iru for 15 years
3900 km railways

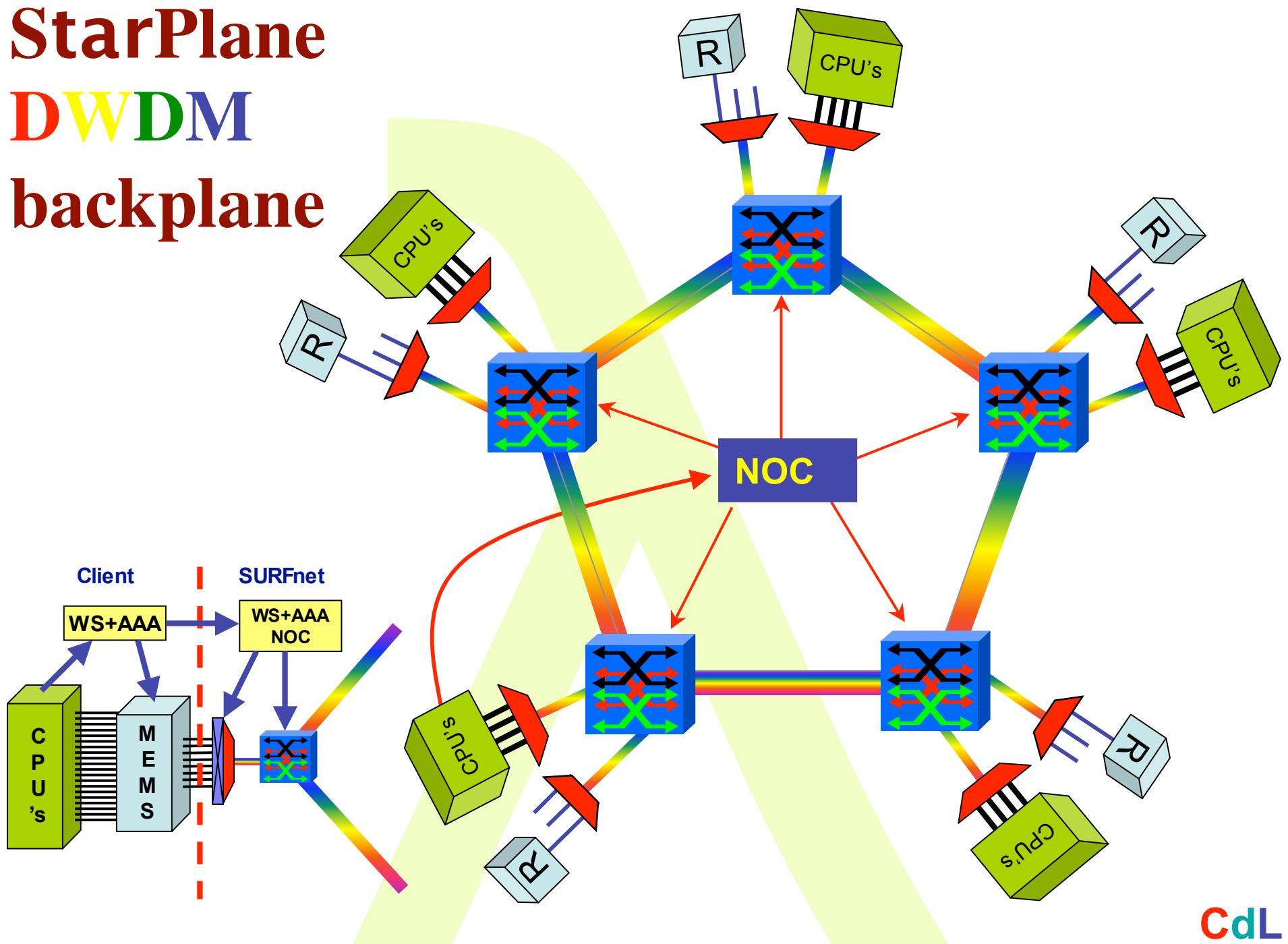
Common Photonic Layer (CPL) in SURFnet6



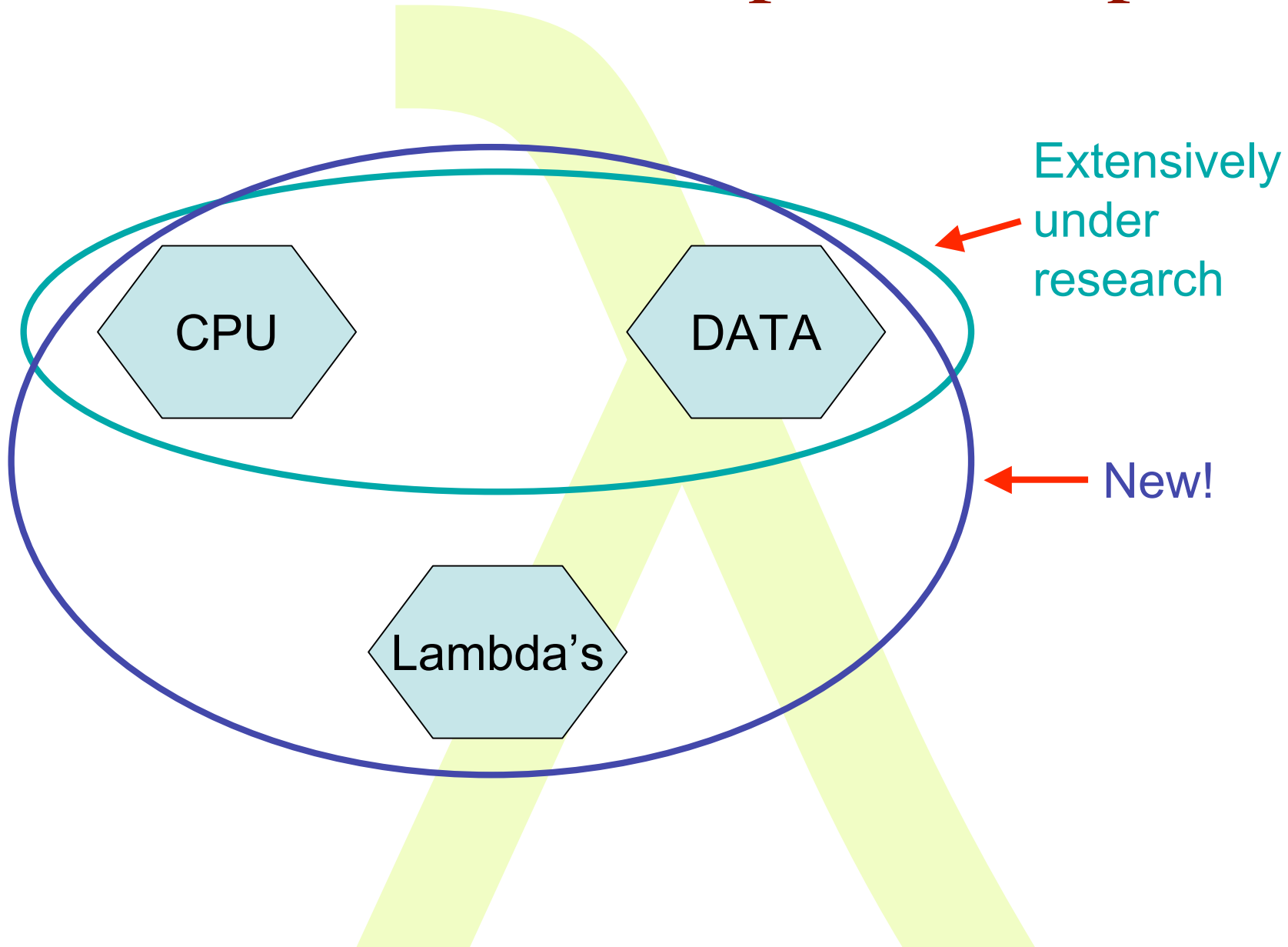
SURFnet on Lambda inspection in Science Park Amsterdam :-)



StarPlane DWDM backplane



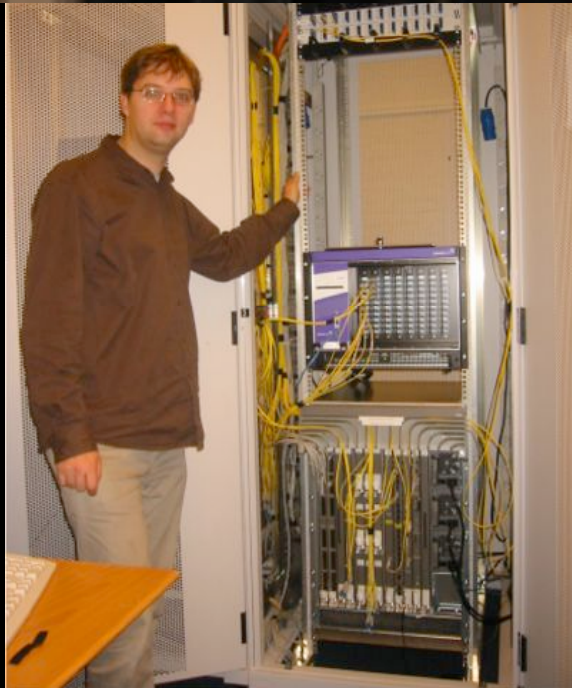
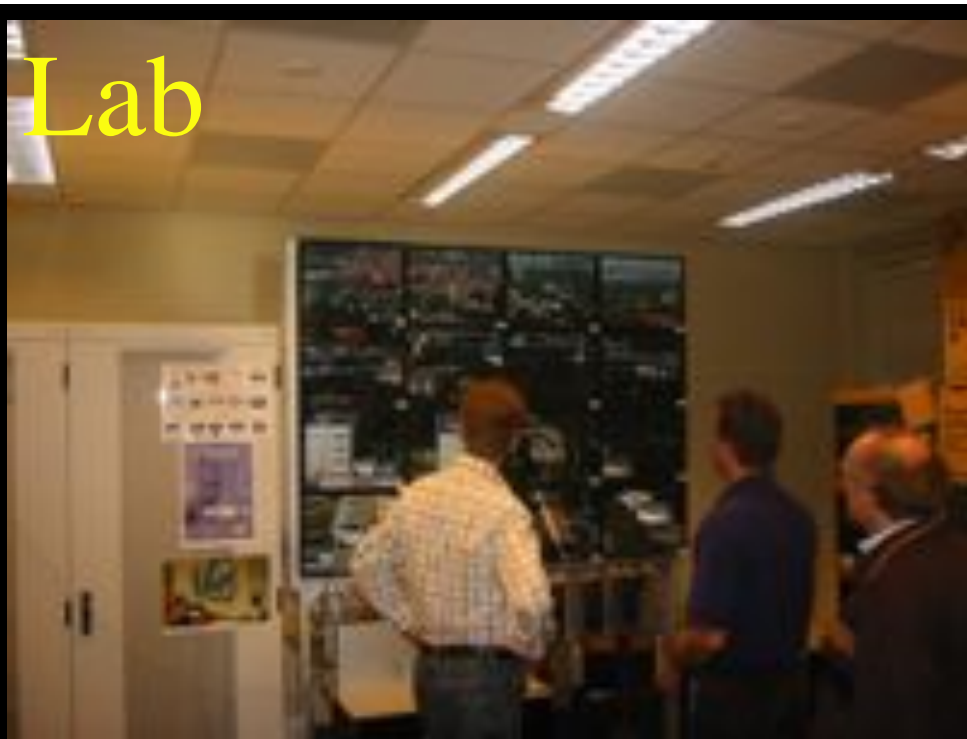
GRID-Colocation problem space



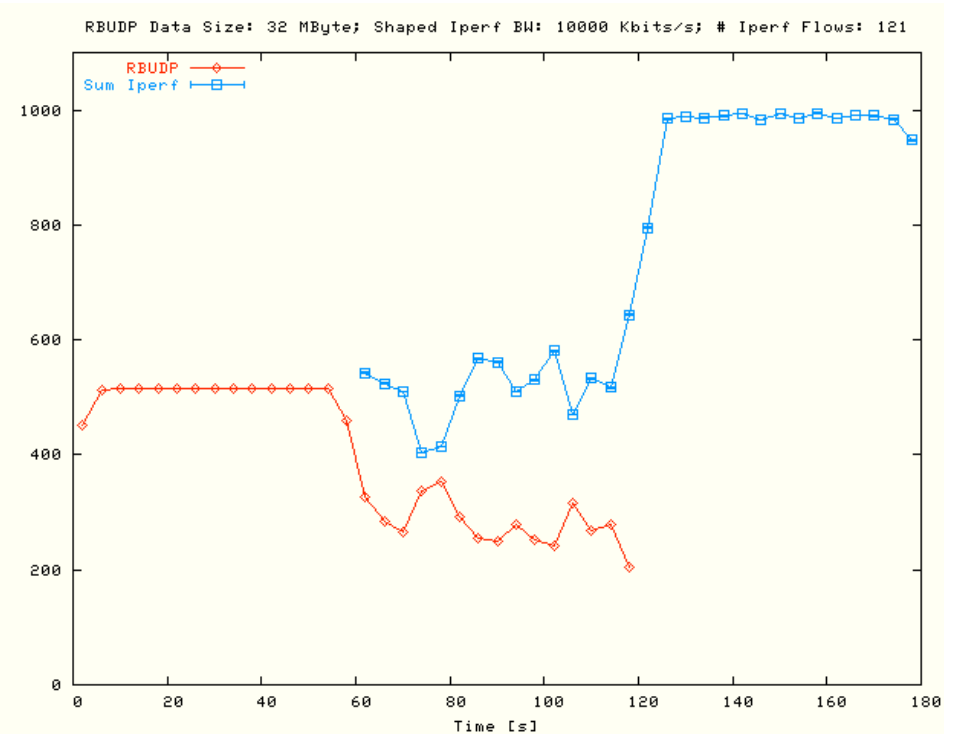
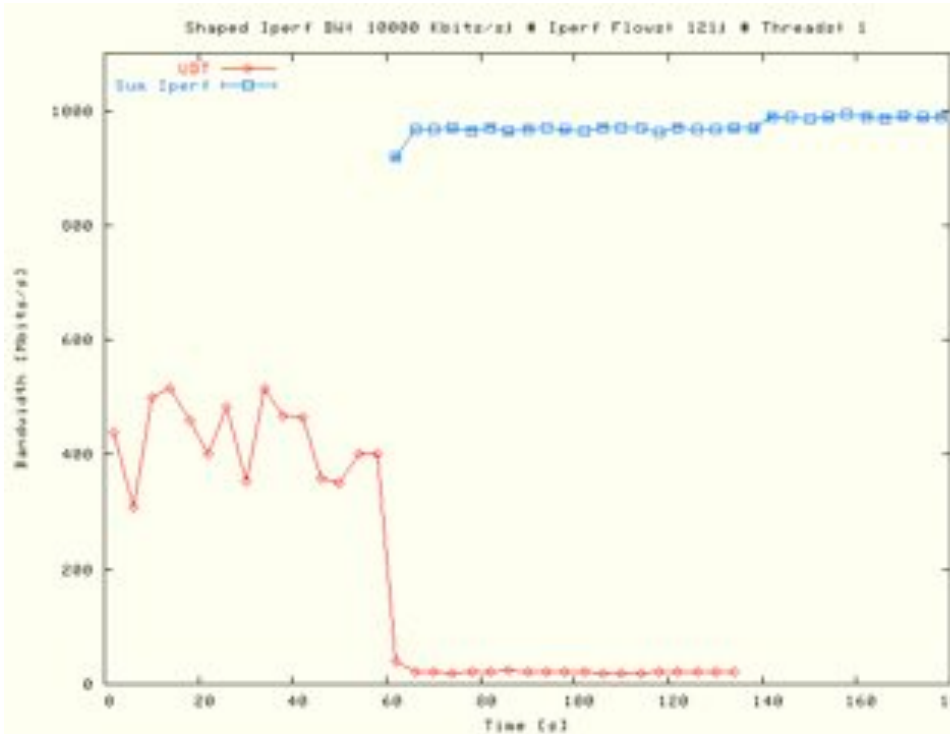
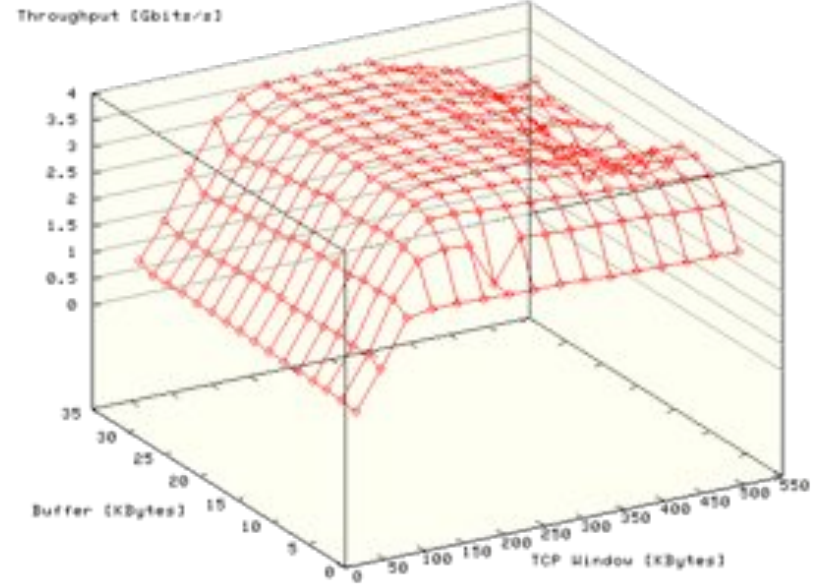
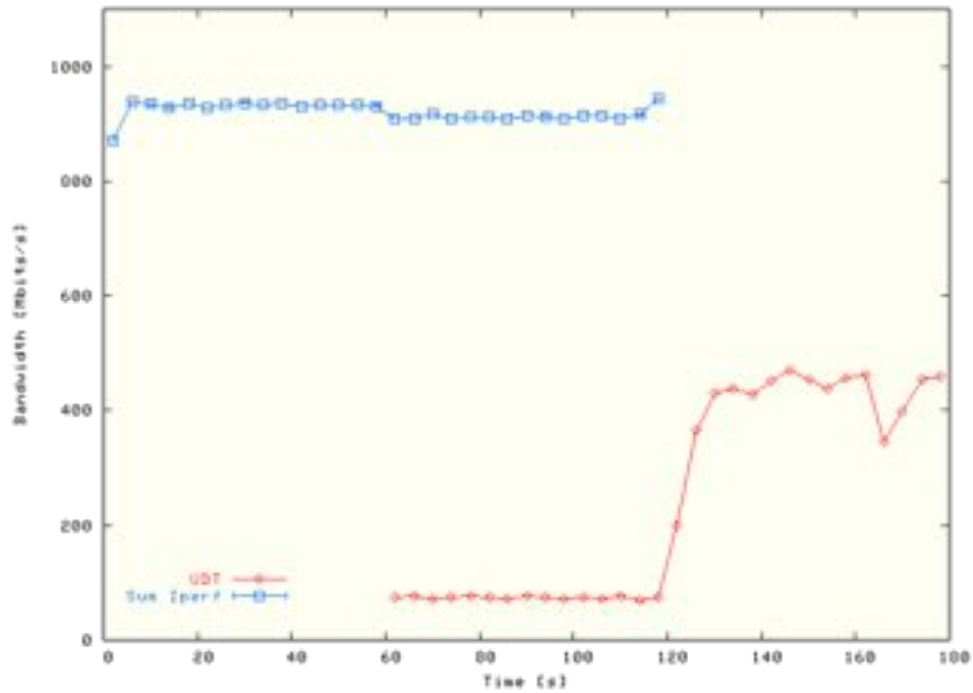
Research @ AIR

- Optical networking architectures and models
 - Optical Internet Exchange architecture
 - Lambda routing and assignment
- IP transport protocols, performances monitoring and measurements
 - With respect to performance
 - Monitoring and reporting
 - Traffic generation with grid infrastructure
- Authorization, Authentication and Accounting
 - Concepts
 - Proof of concepts
 - Application

The Lab



Protocol tests



Layer - 2 requirements from 3/4



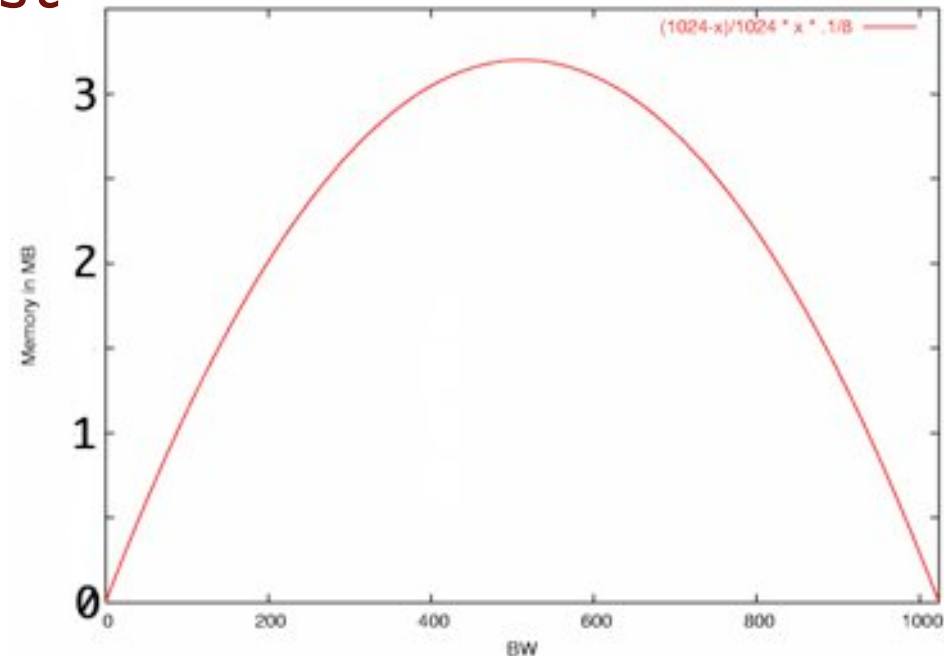
TCP is bursty due to sliding window protocol and slow start algorithm.

$$\text{Window} = \text{BandWidth} * \text{RTT} \quad \& \quad \text{BW} == \text{slow}$$

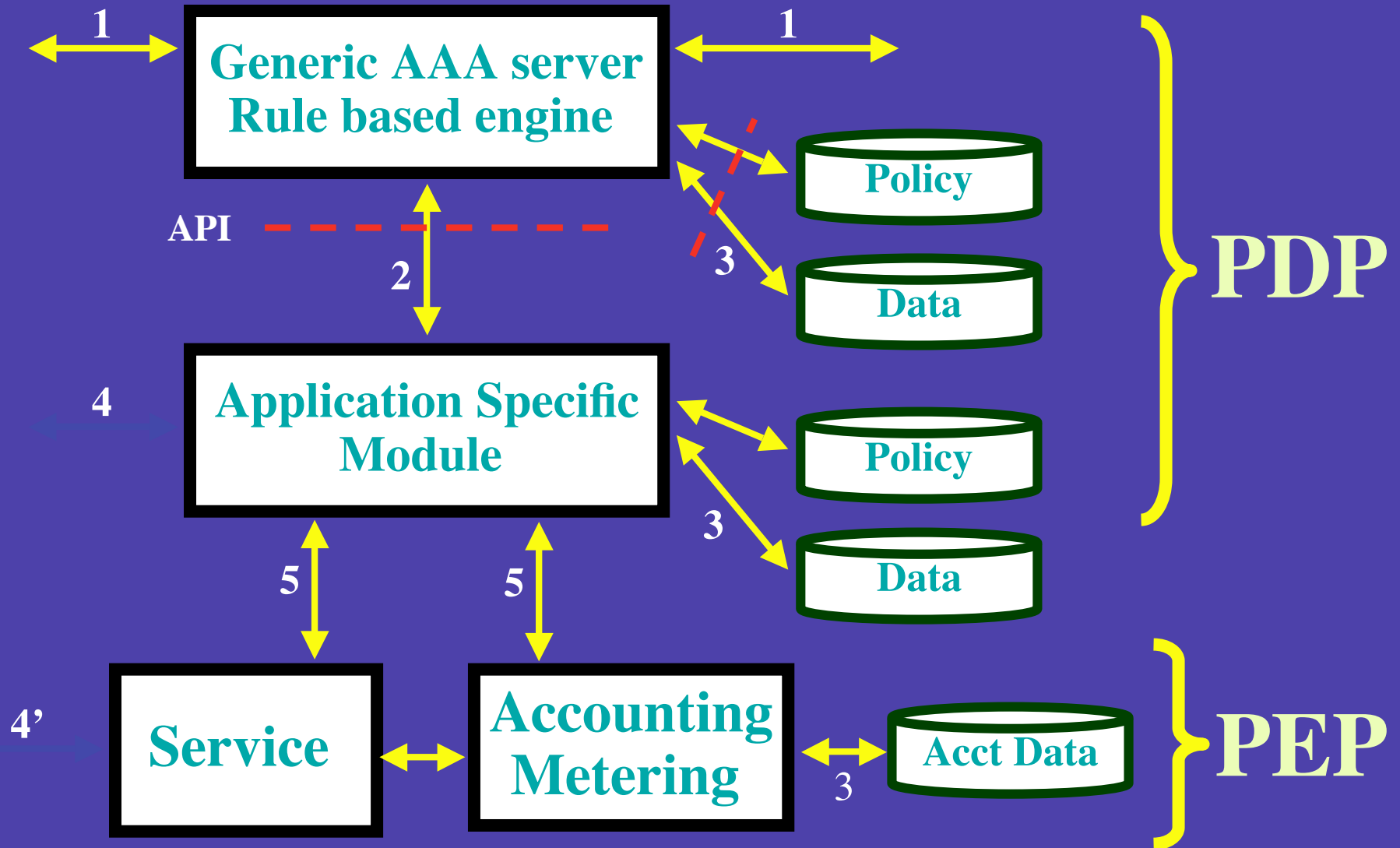
$$\text{Memory-at-bottleneck} = \frac{\text{fast} - \text{slow}}{\text{fast}} * \text{slow} * \text{RTT}$$

So pick from menu:

- ◆ *Flow control*
- ◆ *Traffic Shaping*
- ◆ *RED (Random Early Discard)*
- ◆ *Self clocking in TCP*
- ◆ *Deep memory*

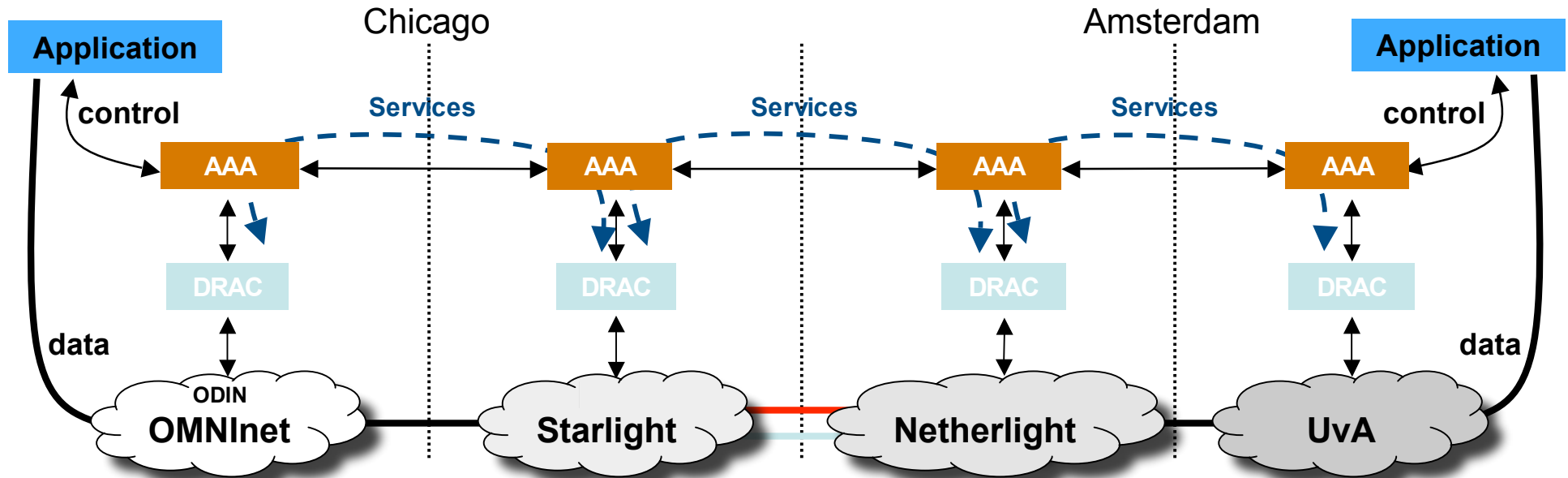


Starting point



RFC 2903 - 2906 , 3334 , policy draft

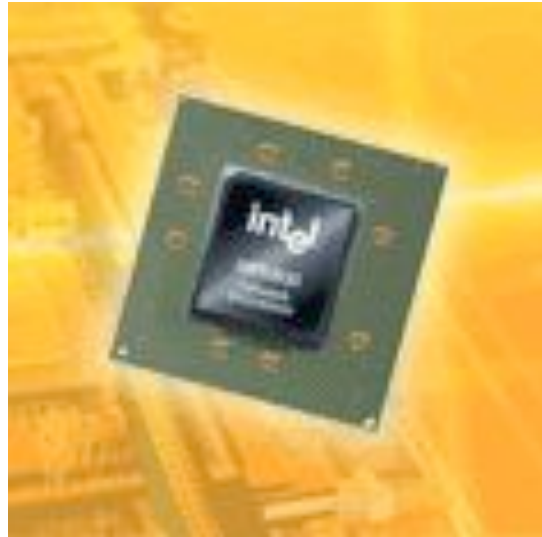
SC2004 CONTROL CHALLENGE



- finesse the control of bandwidth across multiple domains
- while exploiting scalability and intra-, inter-domain fault recovery
- thru layering of a novel SOA upon legacy control planes and NEs



intel. IXP series Network Processor Units



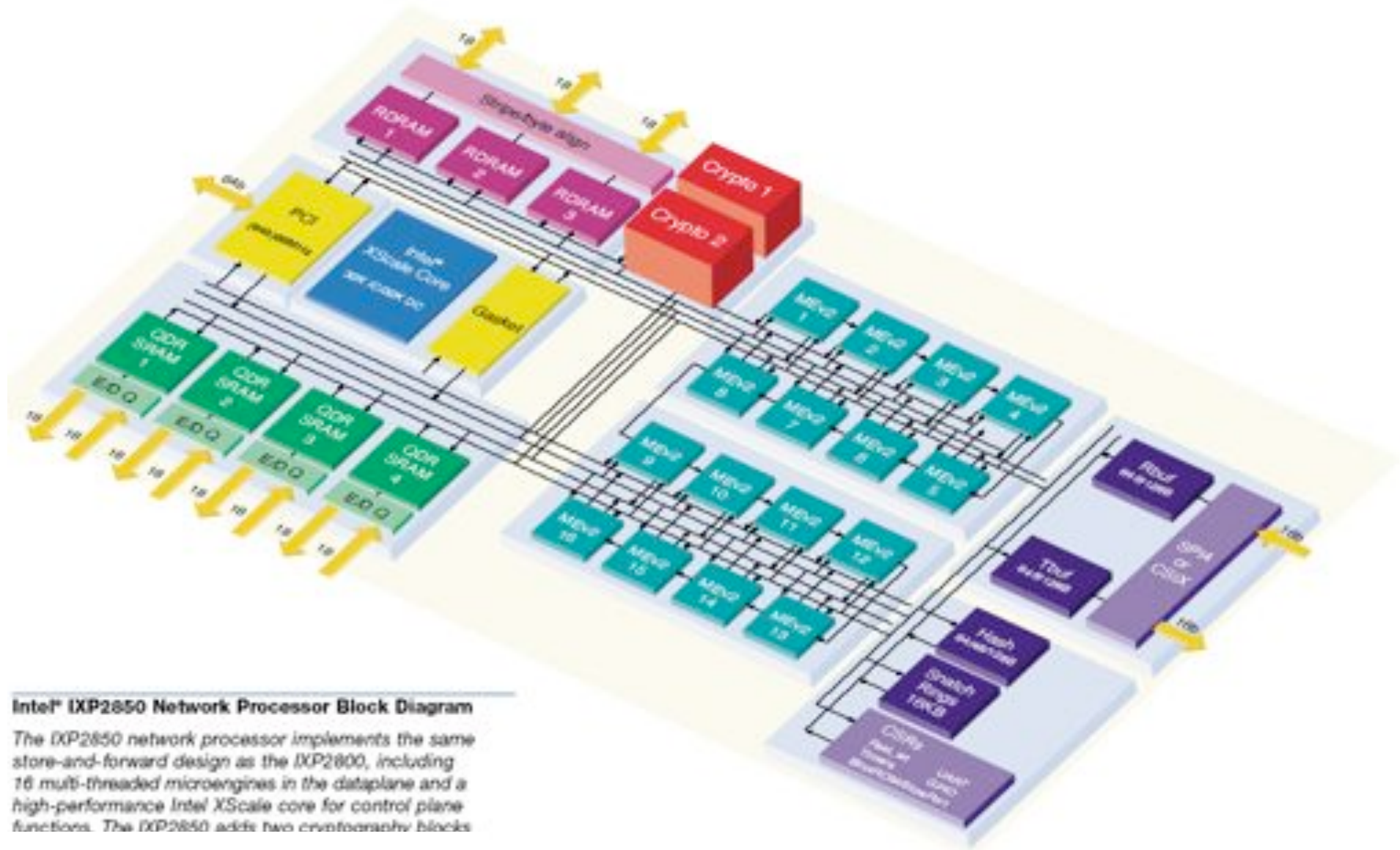
Features:

- The IXP 2850 is able to perform packet functions at 10 gb/s
- 16 programmable Micro Engines to allow parallel dataplane processing.
- Two crypto units support bulk security algorithms (AES, DES, 3DES, SHA1)
- Designed for IPSec, however is general enough to do other things.
- Supports Cypher Block Chaining in combination with MAC.



UNIVERSITEIT VAN AMSTERDAM

GigaPort

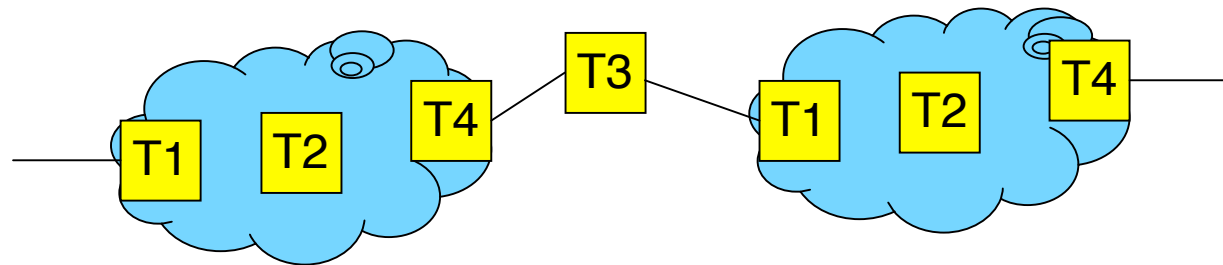


Intel® IXP2850 Network Processor Block Diagram

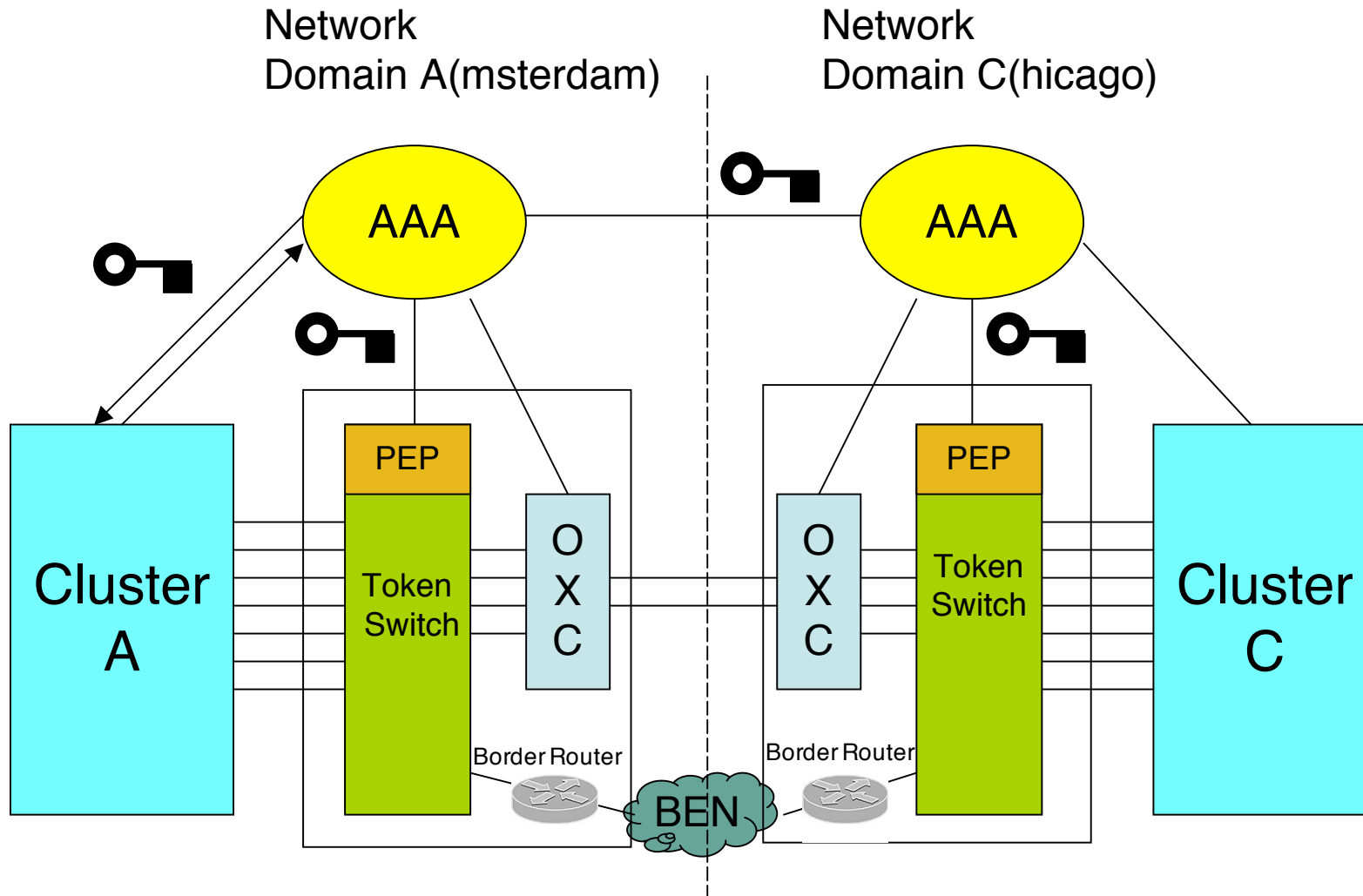
The IXP2850 network processor implements the same store-and-forward design as the IXP2800, including 16 multi-threaded microengines in the dataplane and a high-performance Intel XScale core for control plane functions. The IXP2850 adds two cryptography blocks

Fundamental places for token switches

1. At network ingress point (switch, access server etc.) admitting only valid tokens.
2. Inside a network.
3. Between networks.
4. At network egress point (switch, gateway device)



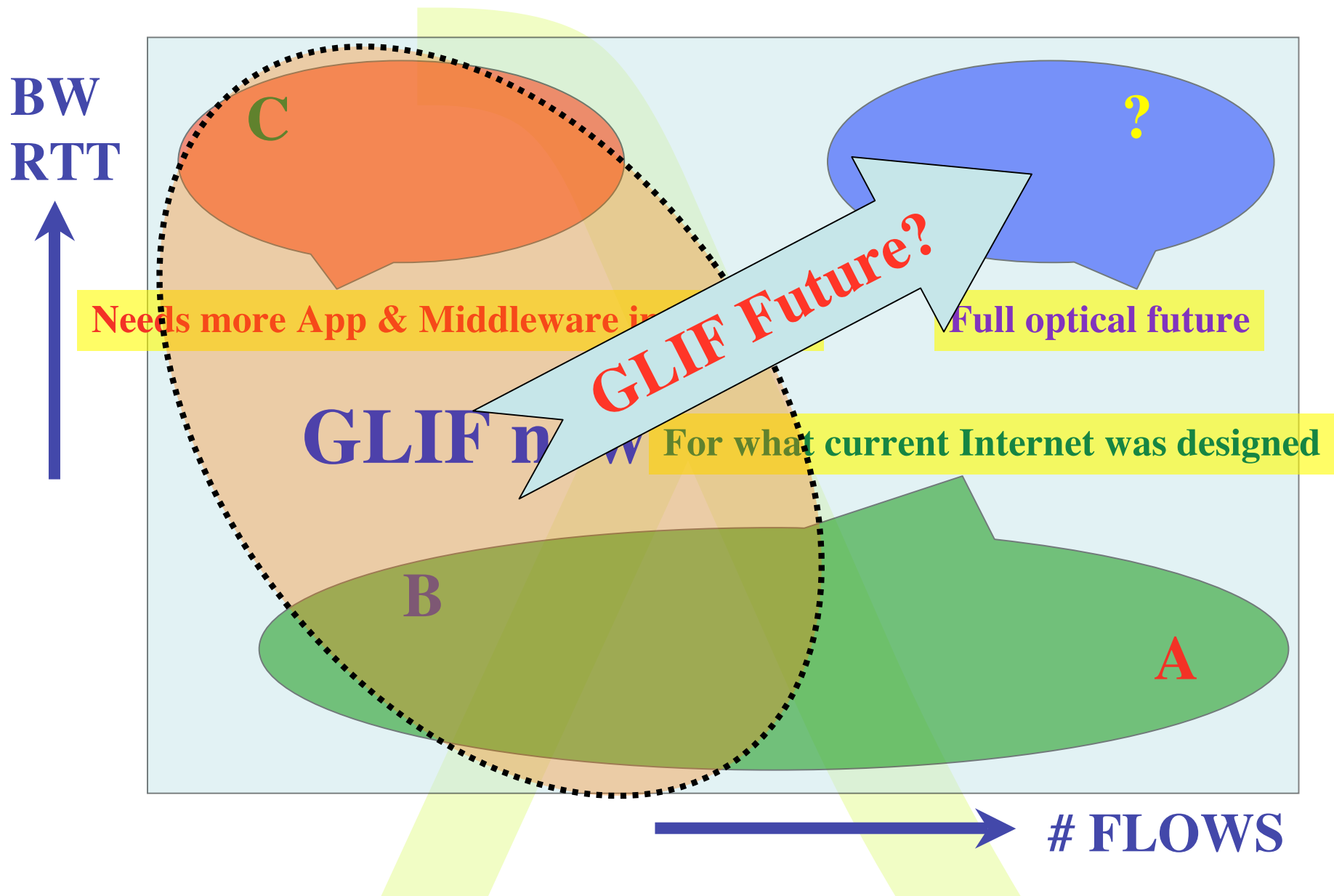
Example experiment agent model



UNIVERSITEIT VAN AMSTERDAM

GigaPort

Transport of flows



Not quite ~~ENDING~~ END

Thanks to

SURFnet: Kees Neggers, UIC&iCAIR: Tom DeFanti, Joel Mambretti, CANARIE: Bill St. Arnaud

Freek Dijkstra, Hans Blom, Leon Gommans, Bas van oudenaarde, Arie Taal, Pieter de Boer, Bert Andree, Fred Wan, Jeroen van der Ham, Karst Koymans, Paola Grosso, Yuri Demchenko, Rob Meijer, VL-team.

Partially complete list:

Caas
Chase
Cess
Kess
Case

