Lambda-Grid developments

www.science.uva.nl/~delaat

Cees de Laat

GigaPort
EU

University of Amsterdam

SARA
NCF

Contents of this talk

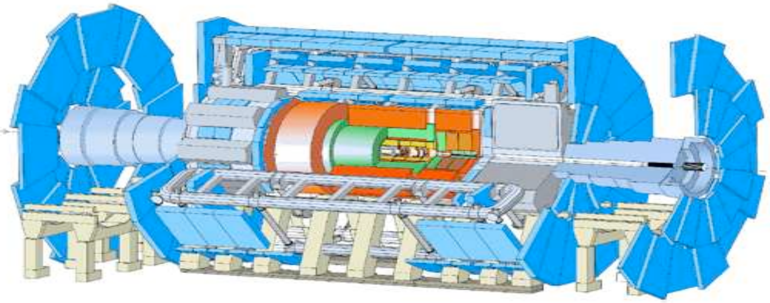
- Demanding applications
- Model of Lambda networking
- Current experiments

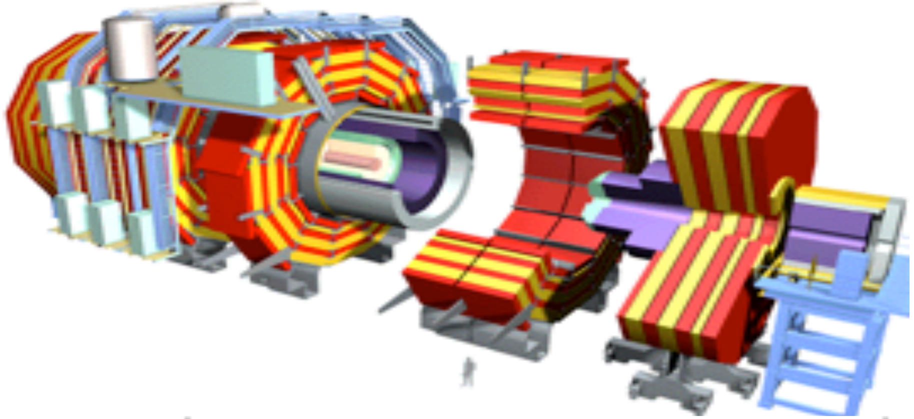
Contents of this talk

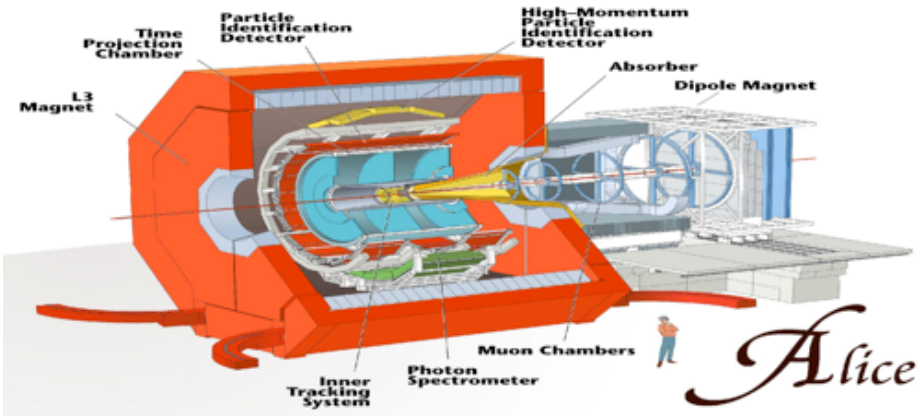
- Demanding applications
- Model of Lambda networking
- Current experiments

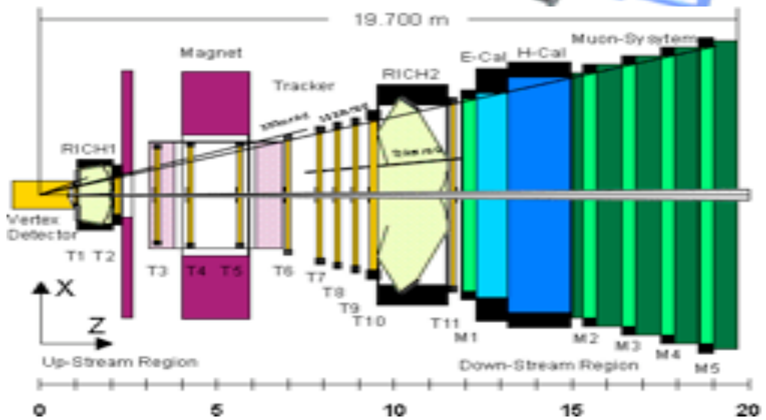
Four LHC Experiments: The Petabyte to Exabyte Challenge

- **ATLAS, CMS, ALICE, LHCB**









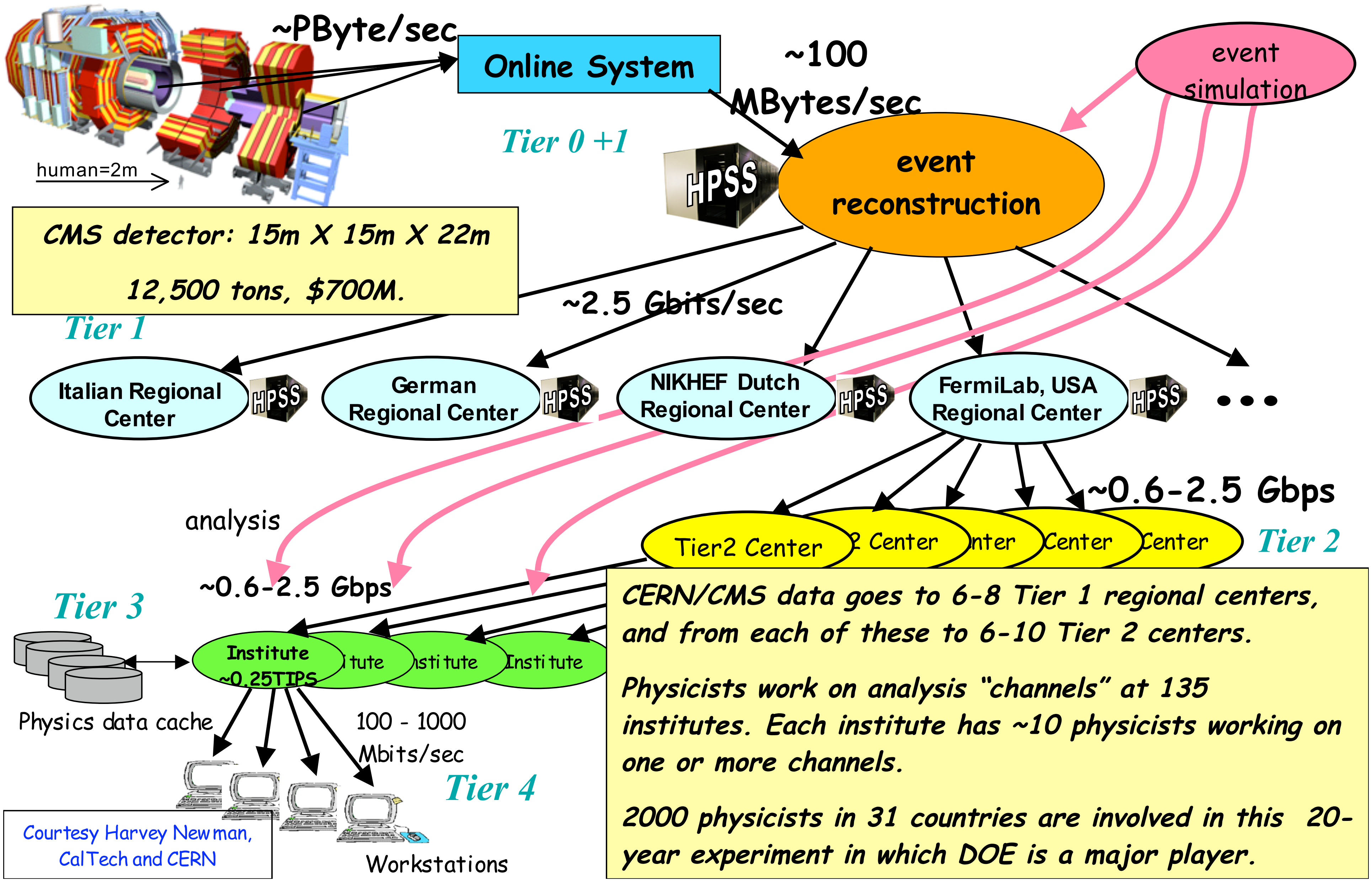
6000+ Physicists & Engineers; 60+ Countries; 250 Institutions

Tens of PB 2008; To 1 EB by ~2015

Hundreds of TFlops To PetaFlops

LHC Data Grid Hierarchy

CMS as example, Atlas is similar



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

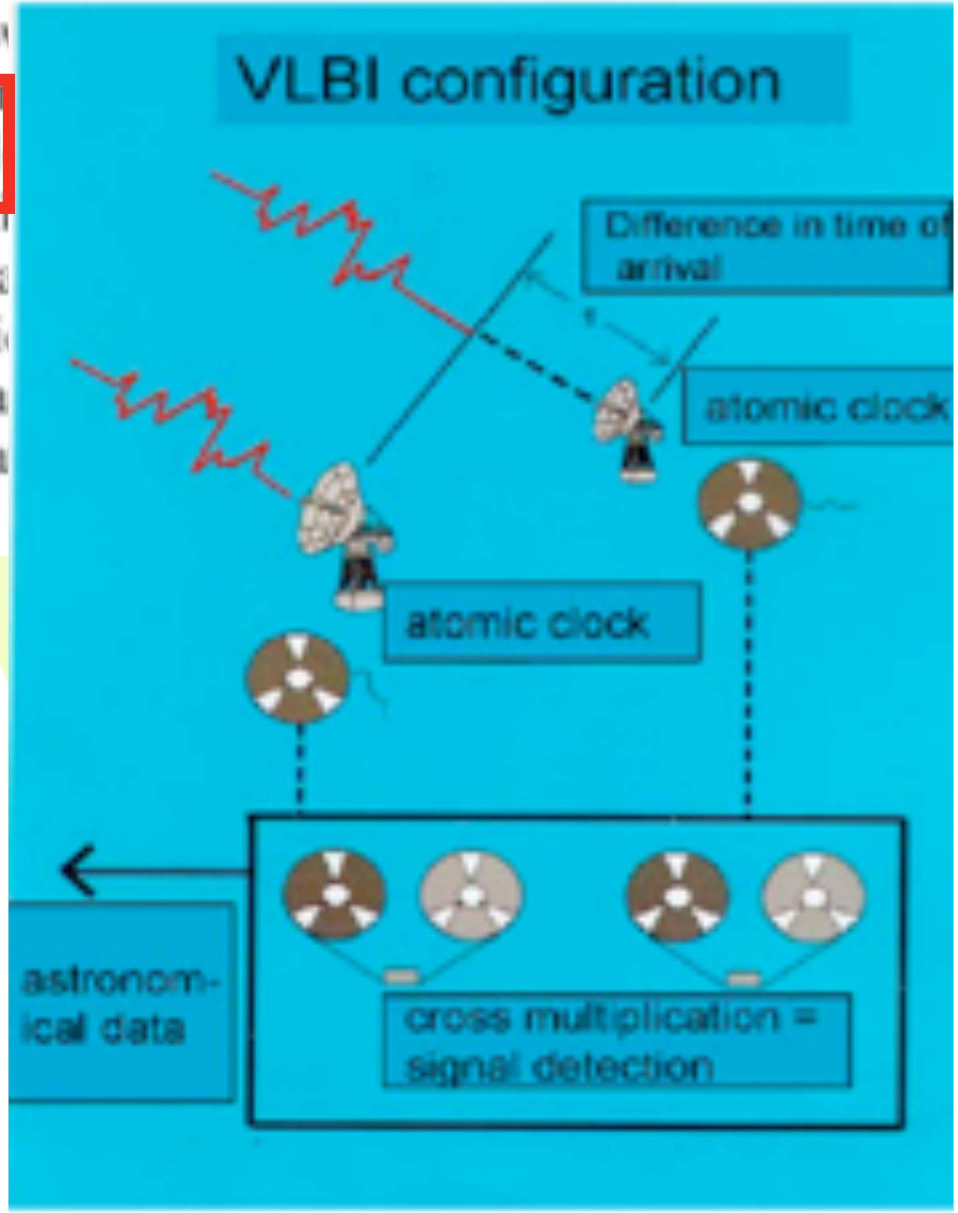
Physicists work on analysis "channels" at 135 institutes. Each institute has ~10 physicists working on one or more channels.

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

VLBI

ger term VLBI is easily capable of generating many Gb of data per
The sensitivity of the VLBI array scales w
(√data-rate) and there is a strong push to
. Rates of 8Gb/s or more are entirely feasibl
nder development. It is expected that parall
orrelator will remain the most efficient approa
s distributed processing may have an applic
lti-gigabit data streams will aggregate into la
or and the capacity of the final link to the da
tor.

Westerbork Synthesis Radio Telescope - Netherlands



Lambdas as part of instruments

www.lofar.org

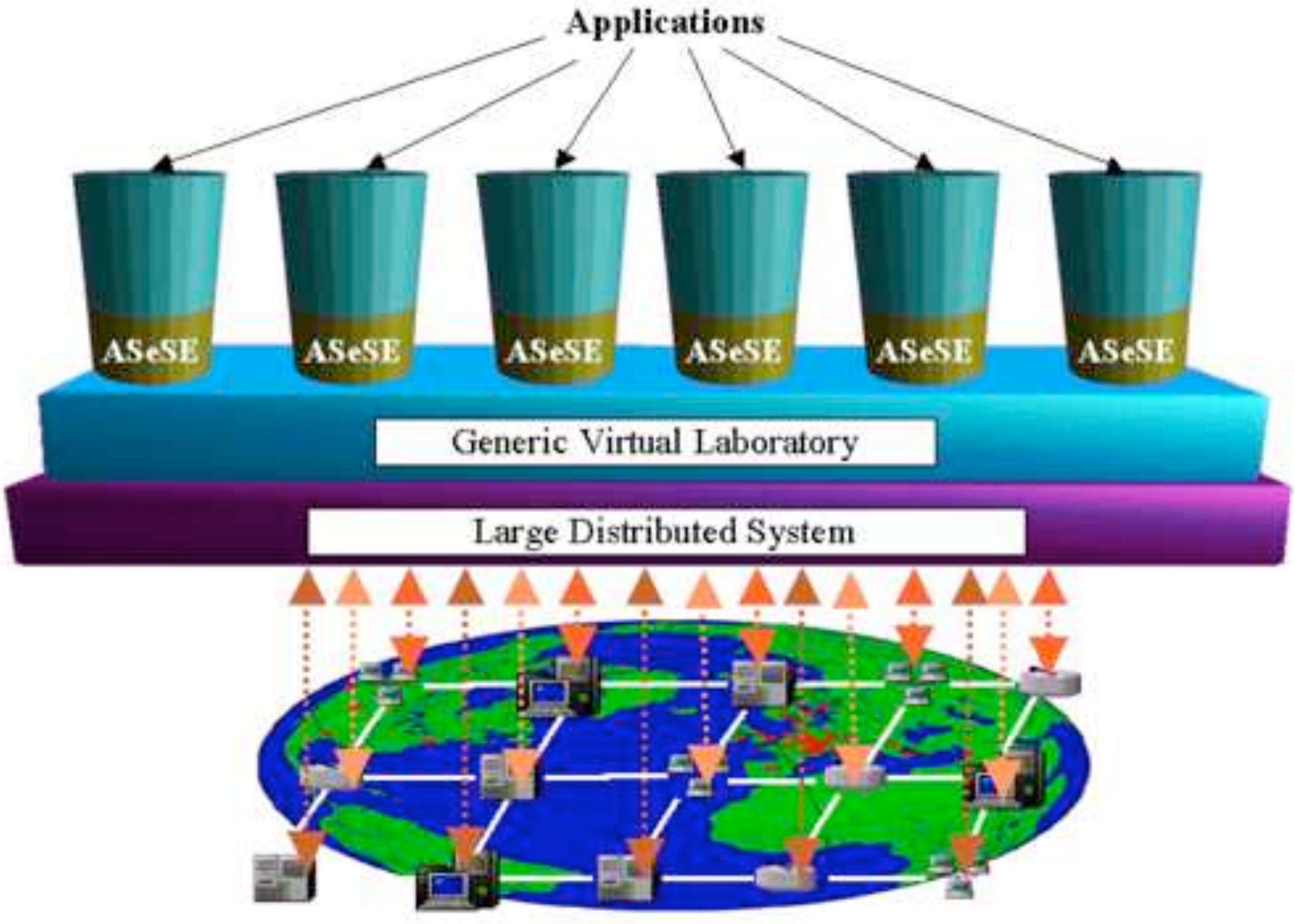
OptIPuter Project Goal: Scaling to 100 Million Pixels

- **JuxtaView (UIC EVL) for PerspecTile LCD Wall**
  - Digital Montage Viewer
  - 8000x3600 Pixel Resolution~30M Pixels
- **Display Is Powered By**
  - 16 PCs with Graphics Cards
  - 2 Gigabit Networking per PC

Source: Jason Leigh, EVL, UIC; USGS EROS

UCSD UCIrvine

VLE middleware

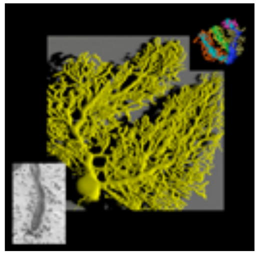


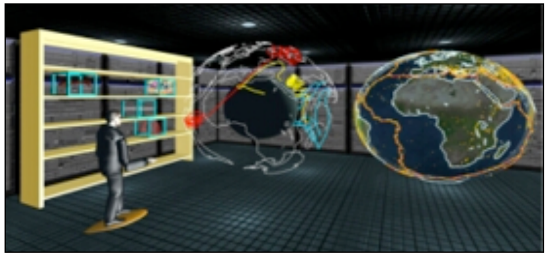
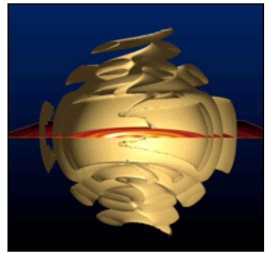
iGrid 2002

September 24-26, 2002, Amsterdam, The Netherlands

- 28 demonstrations from 16 countries: Australia, Canada, CERN, France, Finland, Germany, Greece, Italy, Japan, The Netherlands, Singapore, Spain, Sweden, Taiwan, United Kingdom, United States
- Applications demonstrated: art, bioinformatics, chemistry, cosmology, cultural heritage, education, high-definition media streaming, manufacturing, medicine, neuroscience, physics, tele-science





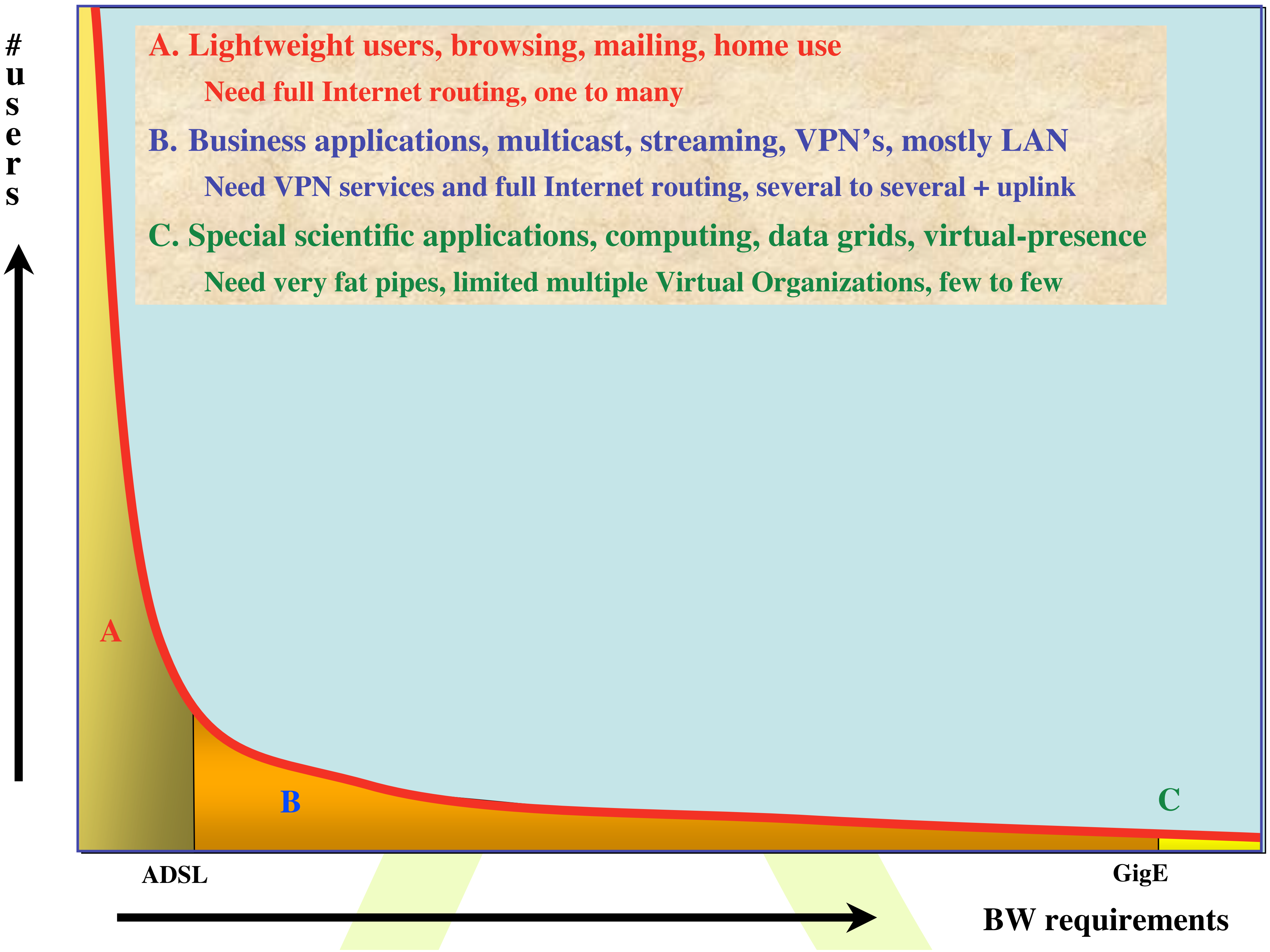
- Grid technologies demonstrated: Major emphasis on grid middleware, data management grids, data replication grids, visualization grids, data/visualization grids, computational grids, access grids, grid portals
- 25Gb transatlantic bandwidth (100Mb/attendee, 250x iGrid2000!)

www.igrid2002.org

Note: iGrid2005 @ San Diego sept 2005

Contents of this talk

- Demanding applications
- Model of Lambda networking
- Current experiments

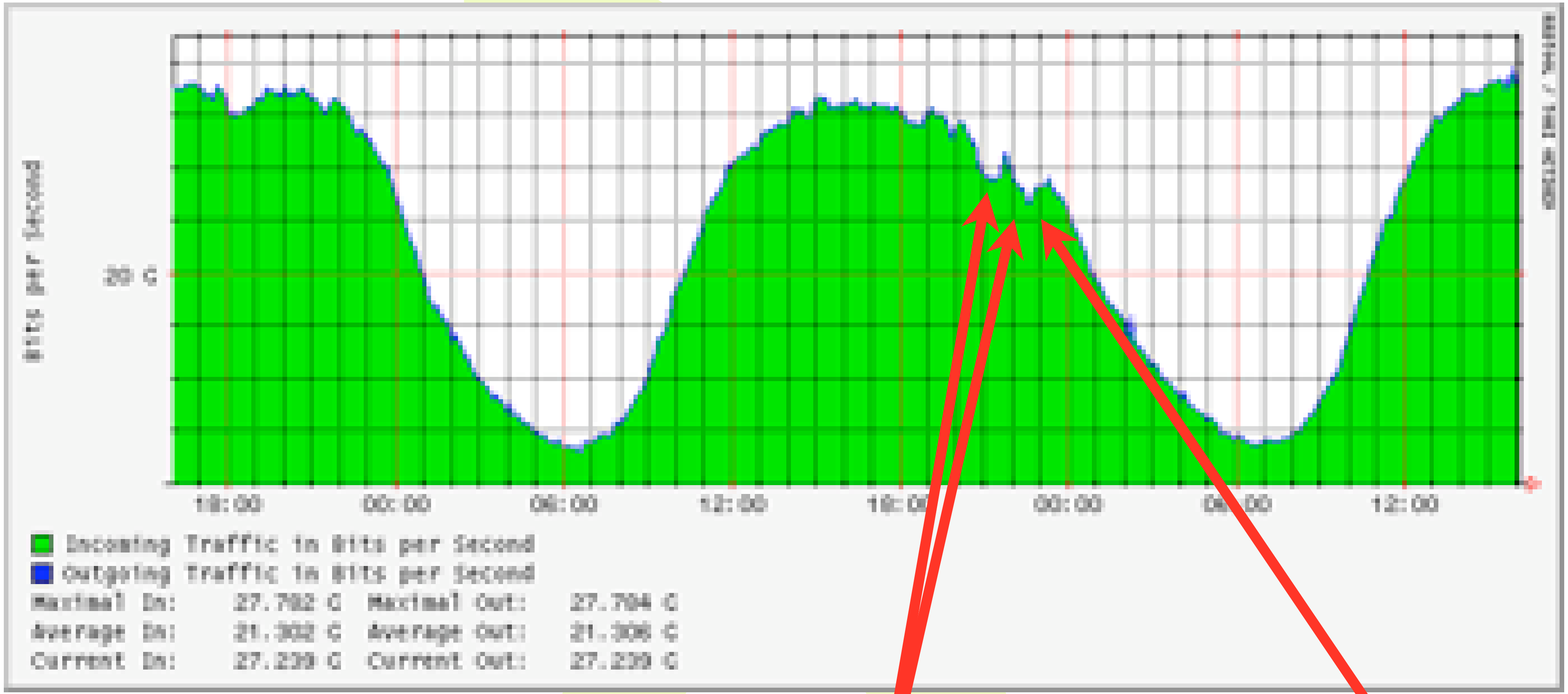


A. Lightweight users, browsing, mailing, home use
Need full Internet routing, one to many
B. Business applications, multicast, streaming, VPN's, mostly LAN
Need VPN services and full Internet routing, several to several + uplink
C. Special scientific applications, computing, data grids, virtual-presence
Need very fat pipes, limited multiple Virtual Organizations, few to few
users
A
B
C
ADSL
GigE
BW requirements

The Dutch Situation

- **Estimate A**
  - **17 M people, 6.4 M households, 25 % penetration of 0.5-2.0 Mb/s ADSL, 40 times under-provisioning ==> 20 Gb/s**

AMS-IX



June 19th 2004

Lost :-(

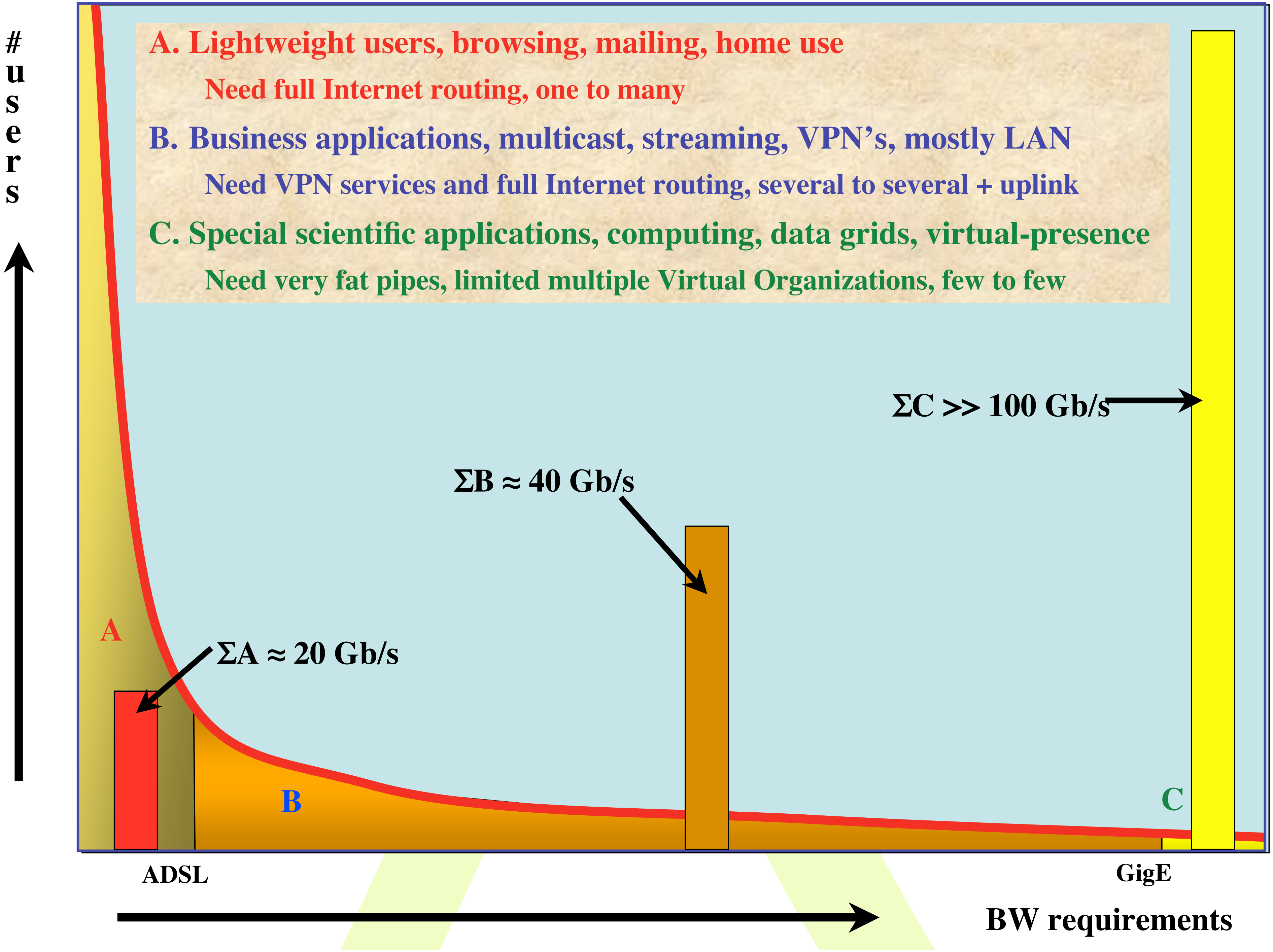
European championship football **Holland -- Czech Republic**

The Dutch Situation

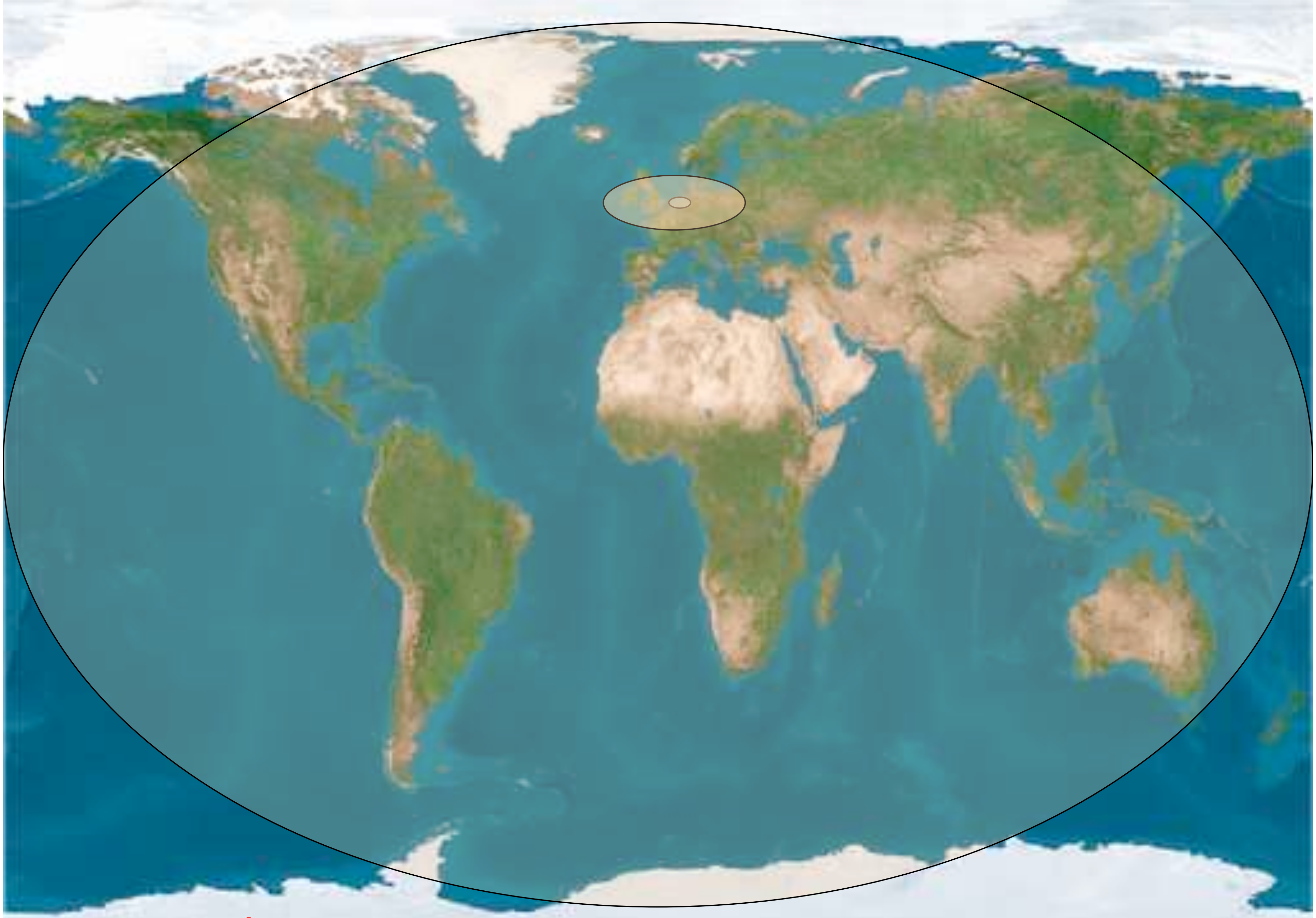
- **Estimate A**
  - **17 M people, 6.4 M households, 25 % penetration of 0.5-2.0 Mb/s ADSL, 40 times under-provisioning ==> 20 Gb/s**
- **Estimate B**
  - **SURFnet has 10 Gb/s to about 12 institutes and 0.1 to 1 Gb/s to 180 customers, estimate same for industry (overestimation) ==> 20-40 Gb/s**

The Dutch Situation

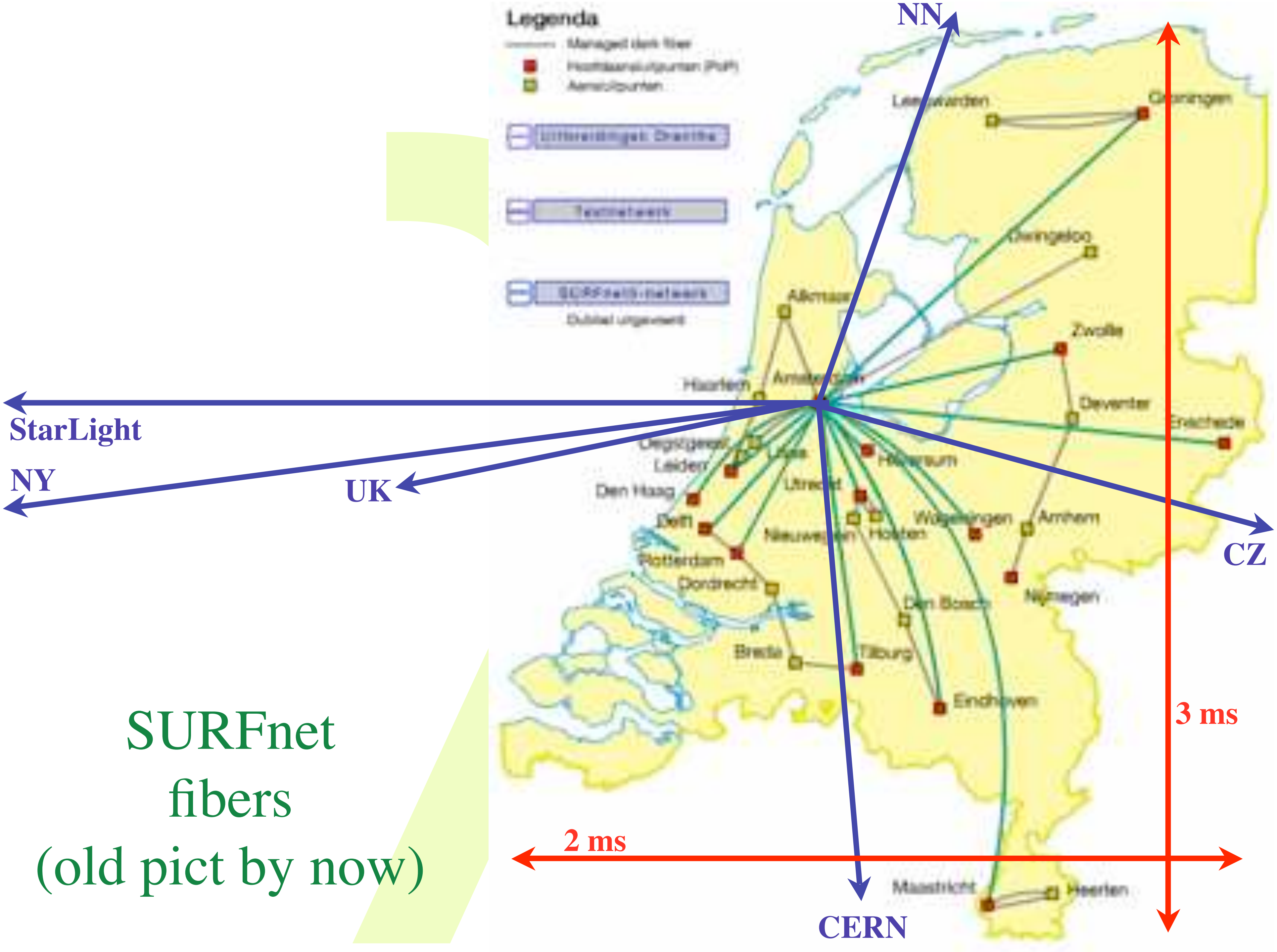
- **Estimate A**
  - **17 M people, 6.4 M households, 25 % penetration of 0.5-2.0 Mb/s ADSL, 40 times under-provisioning ==> 20 Gb/s**
- **Estimate B**
  - **SURFnet has 10 Gb/s to about 12 institutes and 0.1 to 1 Gb/s to 180 customers, estimate same for industry (overestimation) ==> 20-40 Gb/s**
- **Estimate C**
  - **Leading HEF and ASTRO + rest ==> 80-120 Gb/s**
  - **LOFAR ==> ≈ 26 Tbit/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. Special scientific applications, computing, data grids, virtual-presence
Need very fat pipes, limited multiple Virtual Organizations, few to few
users
ΣC >> 100 Gb/s
ΣB ≈ 40 Gb/s
ΣA ≈ 20 Gb/s
A
B
C
ADSL
GigE
BW requirements



λ's on scale 2-20-200 ms rtt



SURFnet
fibers
(old pict by now)

The only formula

$$\#\,\lambda(rtt,t) \approx \frac{200 * e^{(t-2002)}}{rtt}$$

Compares very well with SURFnet's resources and Lambda's @ NetherLight

- **1 Transatlantic Lambda in 2001, now ∼10 from EU+US**
- **4200 km dark fiber in Holland ≈ railway net**

So what are facts

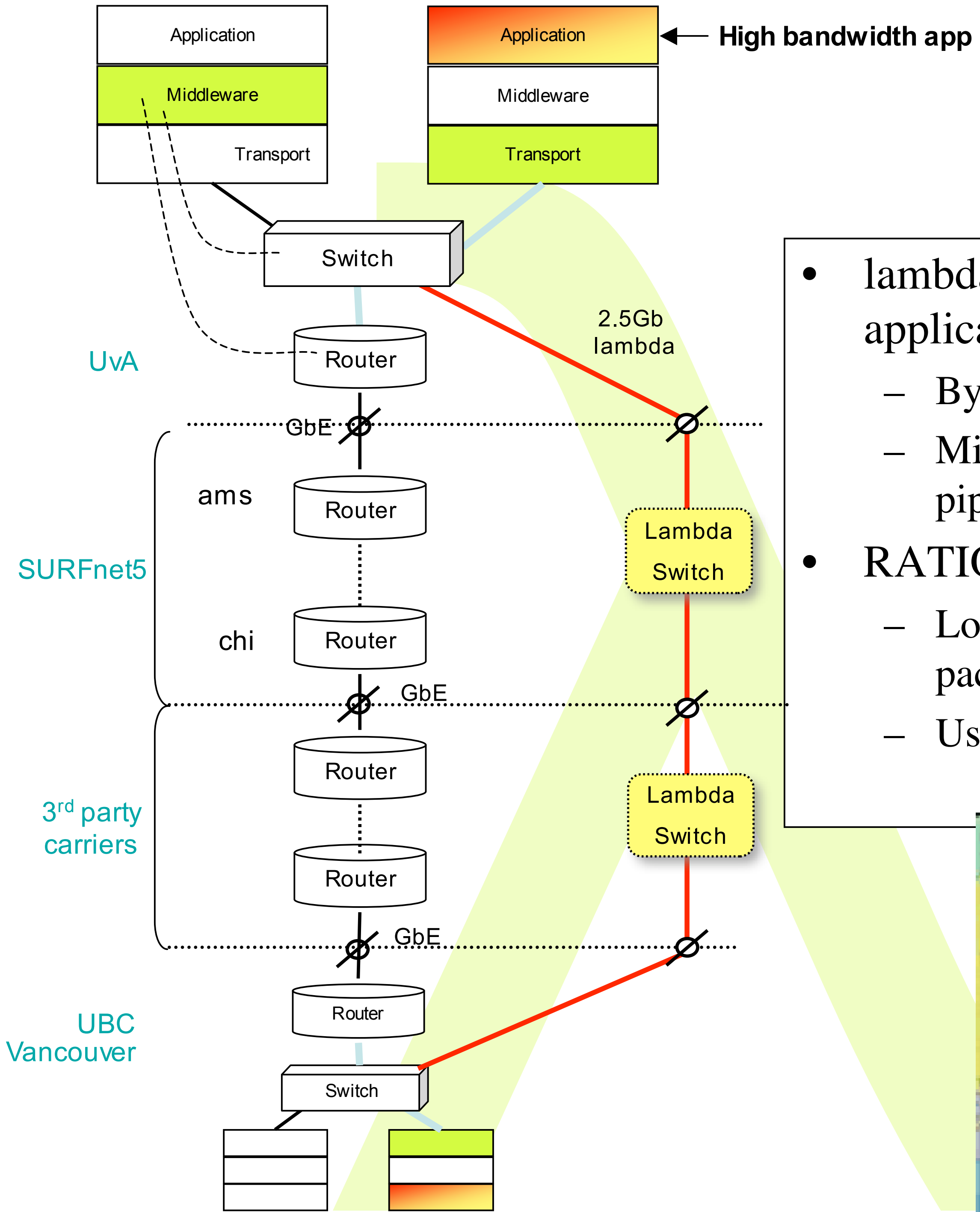
- Costs of fat pipes (fibers) are one/third of cost of equipment to light them up
  - Is what Lambda salesmen tell me
- Costs of optical equipment 10% of switching 10 % of full routing equipment for same throughput
  - 10G routerblade -> 200 k$, 10G switch port -> 20 k$, Mems dev port -> 1 k$
  - 100 Byte packet @ 40 Gb/s -> 20 ns -> time to look up destination in 140 kEntries routing table (light speed from me to you!)
- Bottom line: look for a hybrid architecture which serves all classes in a cost effective way (A -> L3 , B -> L2 , C -> L1)
- Look at worldwide ethernet infrastructure:
  - Tested 10 gbps Ethernet WANPHY Amsterdam-CERN
  - http://www.surfnet.nl/en/publications/pressreleases/021003.html

UVA/EVL's
64*64
Optical Switch
@ NetherLight
in SURFnet POP @
SARA

Costs 1/100th of a
similar throughput
router
or 1/10th of an
Ethernet switch but
with specific services!

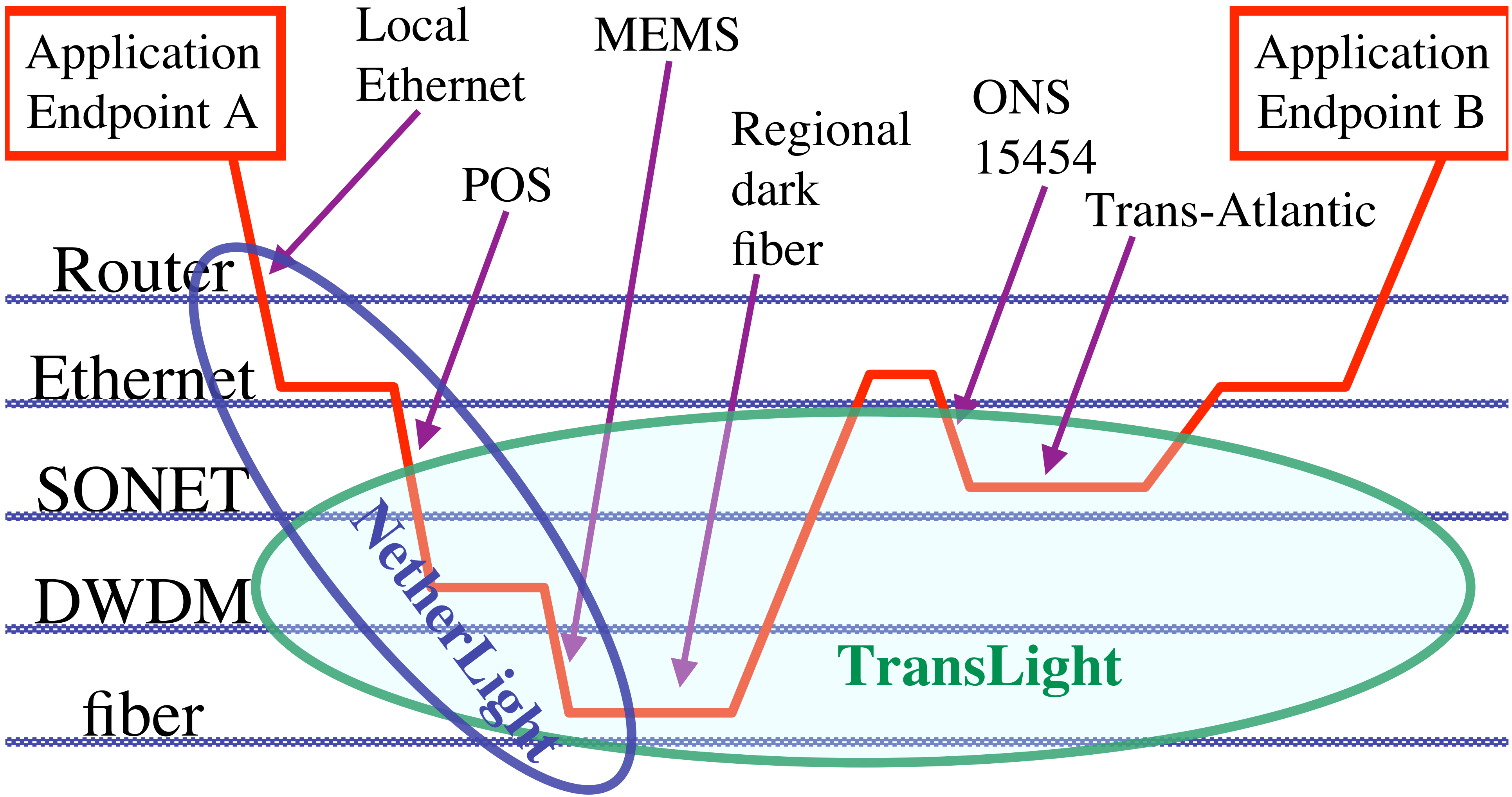
Services

SCALE / CLASS	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

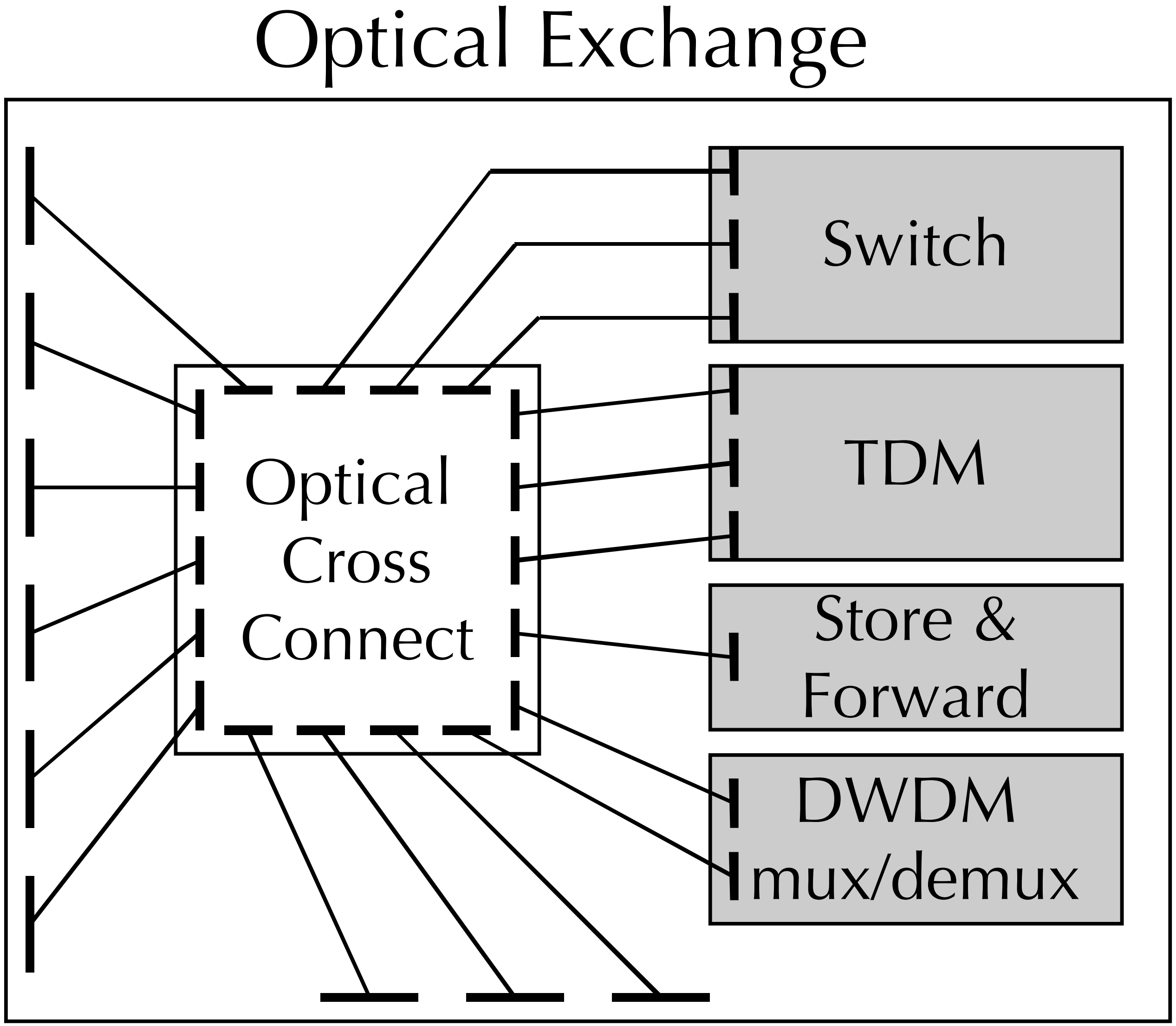


- lambda for high bandwidth applications
  - Bypass of production network
  - Middleware may request (optical) pipe
- RATIONALE:
  - Lower the cost of transport per packet
  - Use Internet as controlplane!

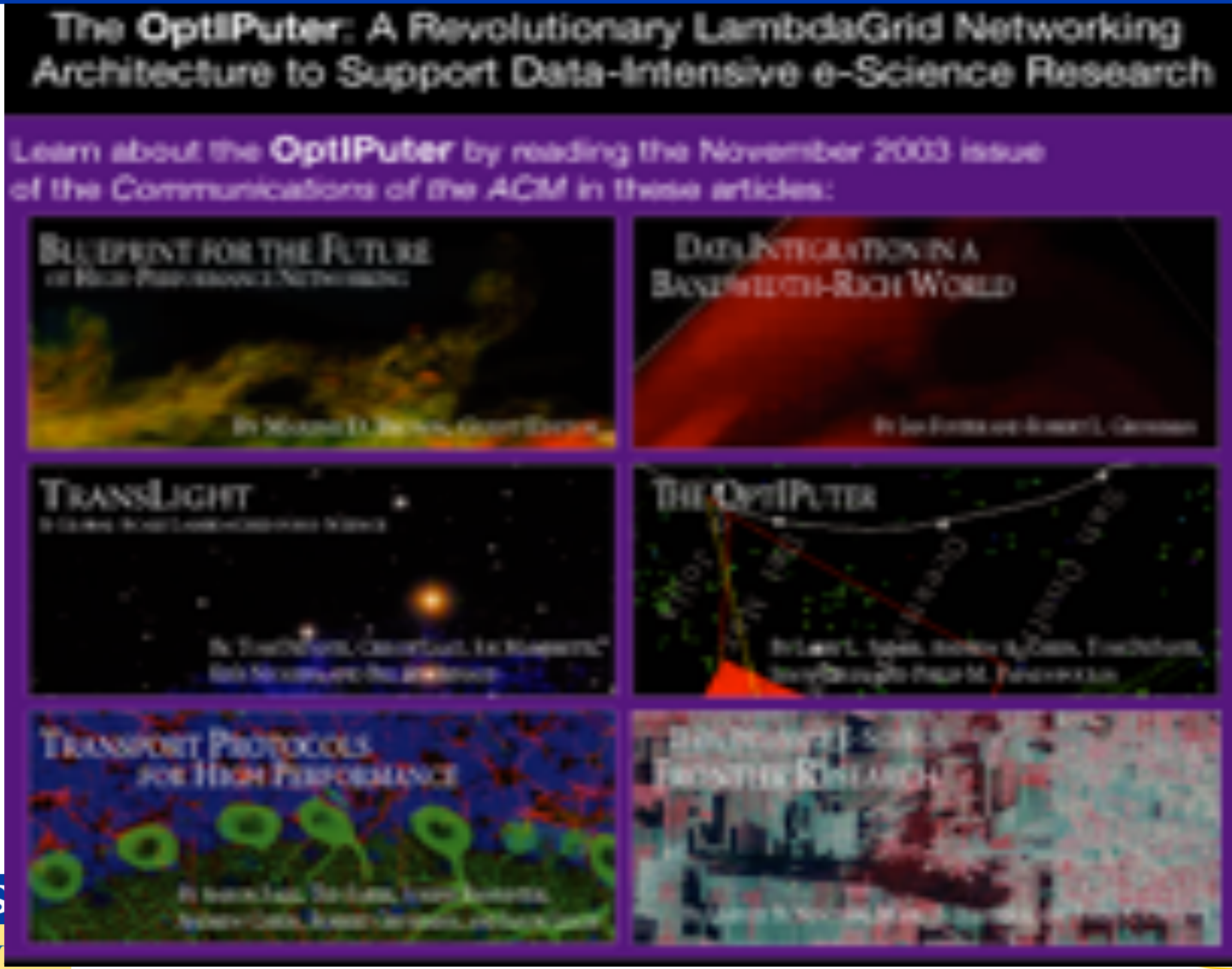
How low can you go?



Optical Exchange as Black Box



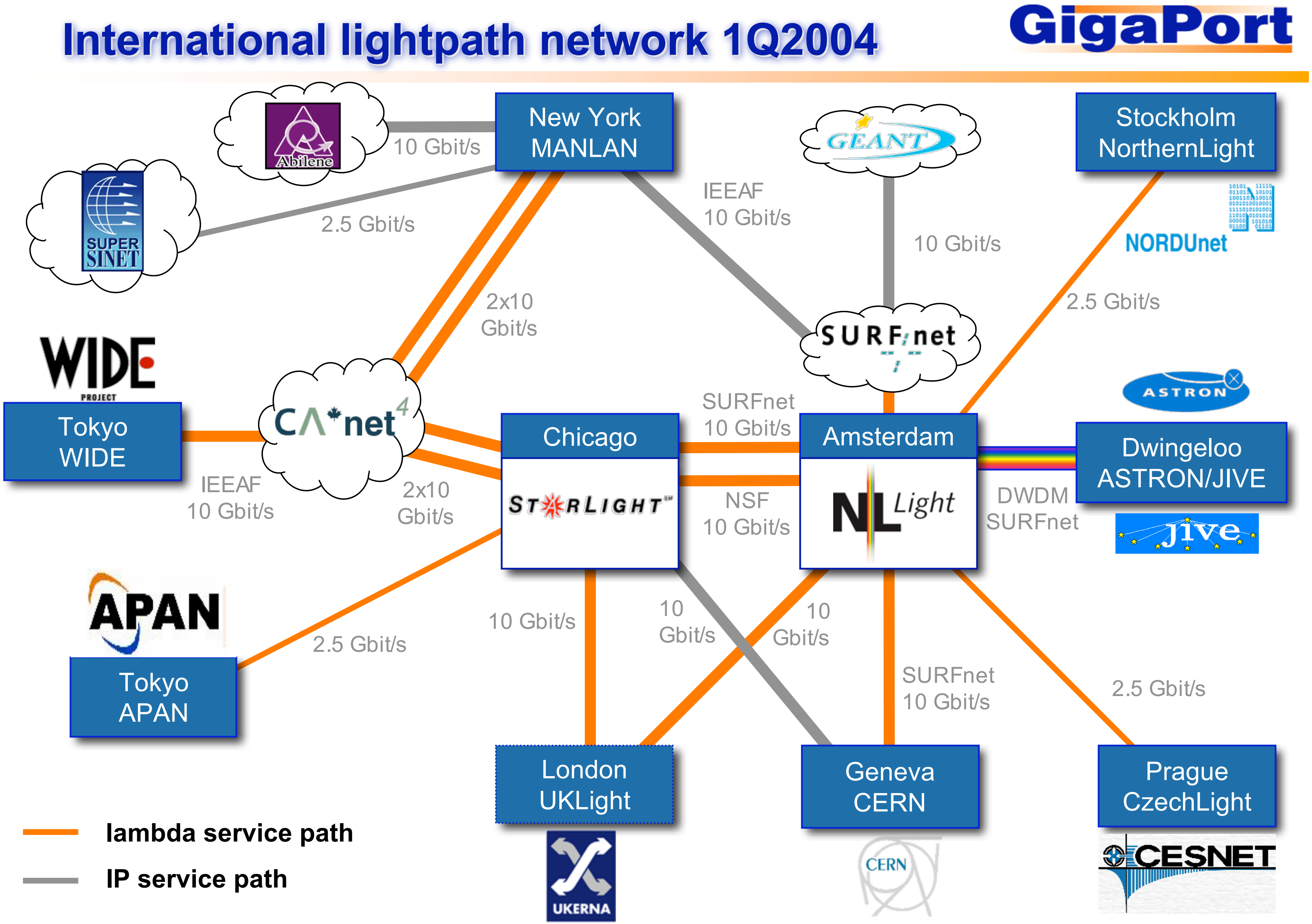
See Nov 2003 CACM
For Articles on OptIPuter Technologies



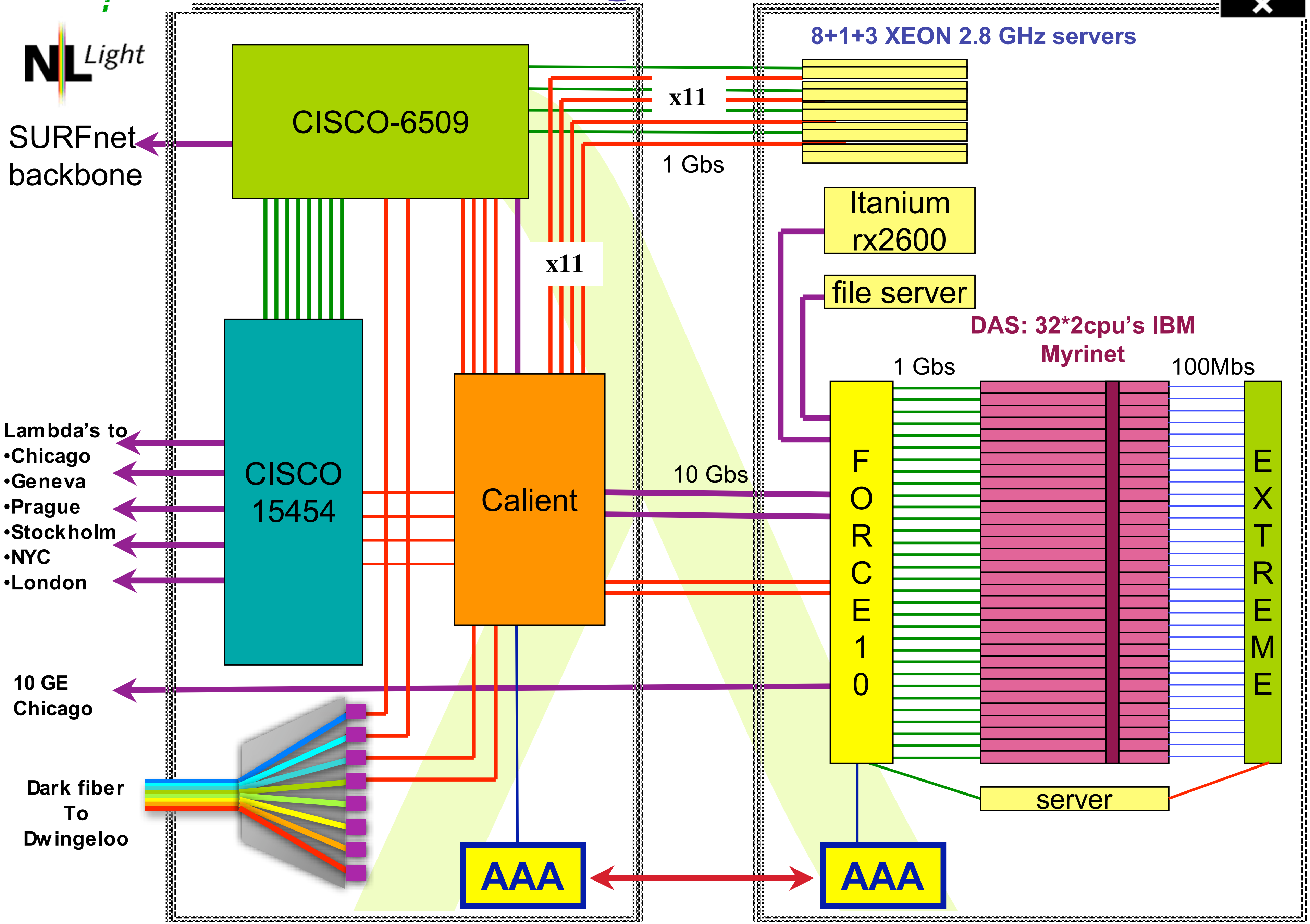
Contents of this talk

- Demanding applications
- Model of Lambda networking
- Current experiments

International lightpath network 1Q2004



NetherLight <-> UvA



Little GLORIAD

http://www.nsf.gov/od/lpa/news/03/pr03151.htm

Research on Networks (CdL)

GigaPort

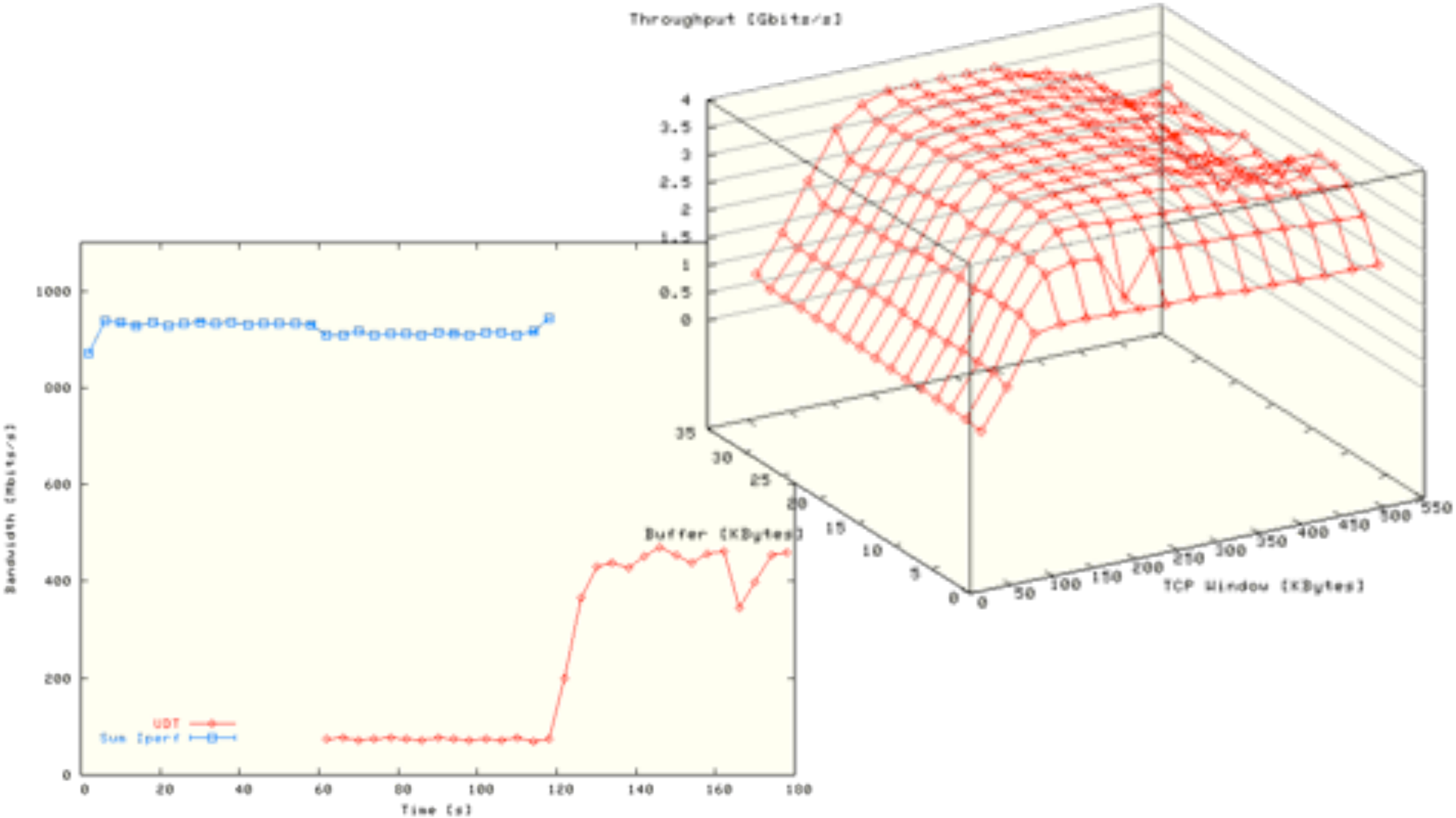
- **Optical Networking:** What innovation in architectural models, components, control and light path provisioning are needed to integrate dynamically configurable optical transport networks and traditional IP networks to a generic data transport platform that provides end-to-end IP connectivity as well as light path (lambda and sub-lambda) services?
- **High performance routing and switching:** What developments need to be made in the Internet Protocol Suite to support data intensive applications, and scale the routing and addressing capabilities to meet the demands of the research and higher education communities in the forthcoming 5 years?
- **Management and monitoring:** What management and monitoring models on the dynamic hybrid network infrastructure are suited to provide the necessary high level information to support network planning, network security and network management?
- **Grids and access; reaching out to the user:** What new models, interfaces and protocols are capable of empowering the (grid) user to access, and the provider to offer, the network and grid resources in a uniform manner as tools for scientific research?
- **Testing methodology:** What are efficient and effective methods and setups to test the capabilities and performance of the new building blocks and their interworking, needed for a correct functioning of a next generation network?

SURFnet

Research topics

- Optical networking architectures and models for usage
- Transport protocols for massive amounts of data
- Authorization of complex resources in multiple domains
- Embedding in Grid environments

Example Measurements



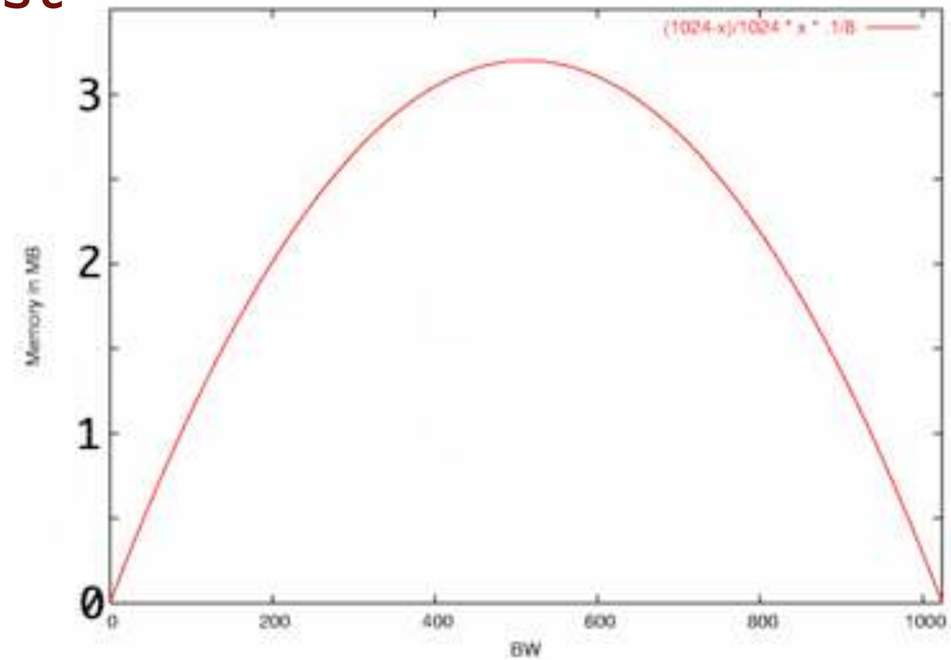
Layer - 2 requirements from 3/4

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

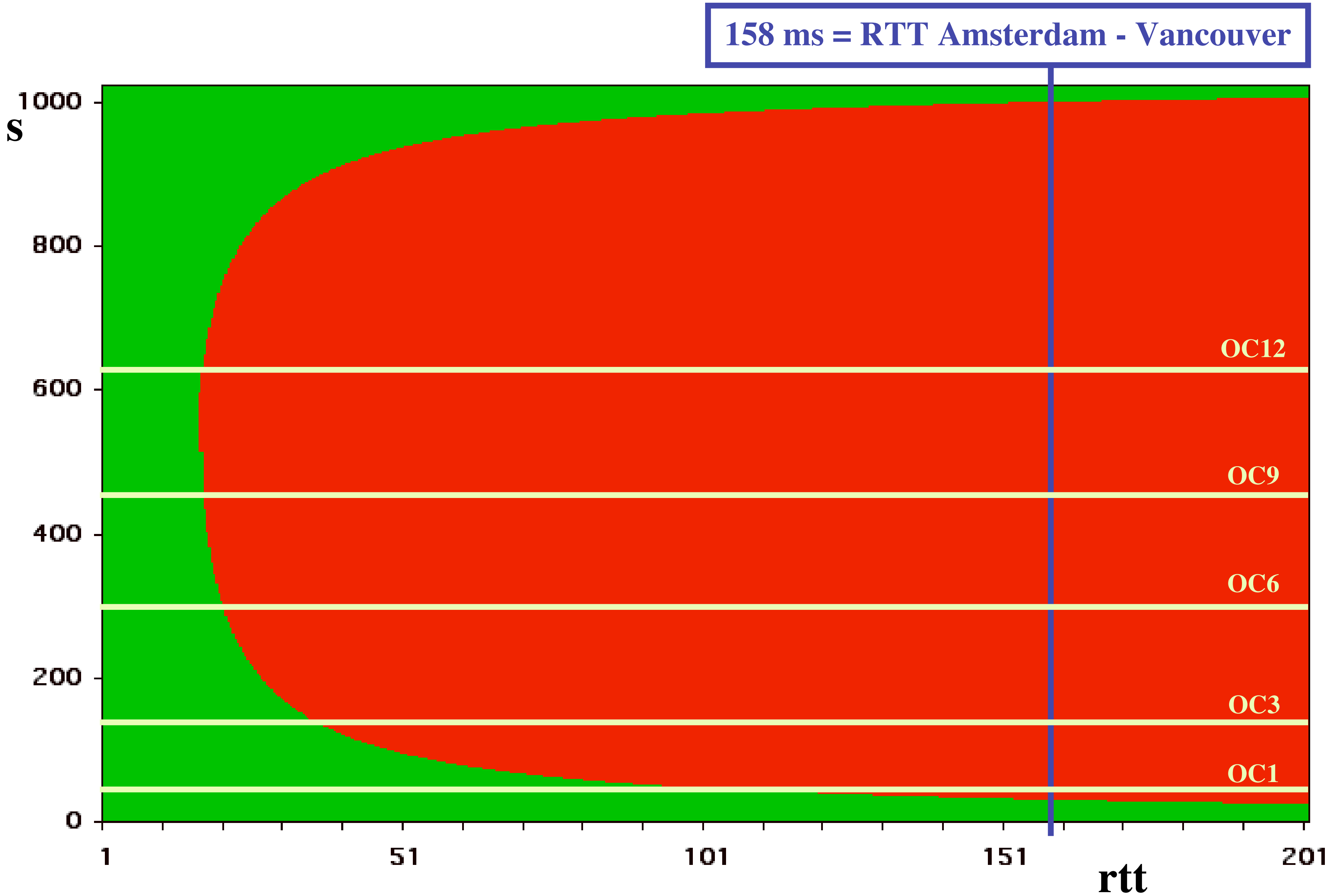
```
Window = BandWidth * RTT    &    BW == slow

                          fast - slow
Memory-at-bottleneck = ----------- * slow * RTT
                              fast
```

So pick from menu:

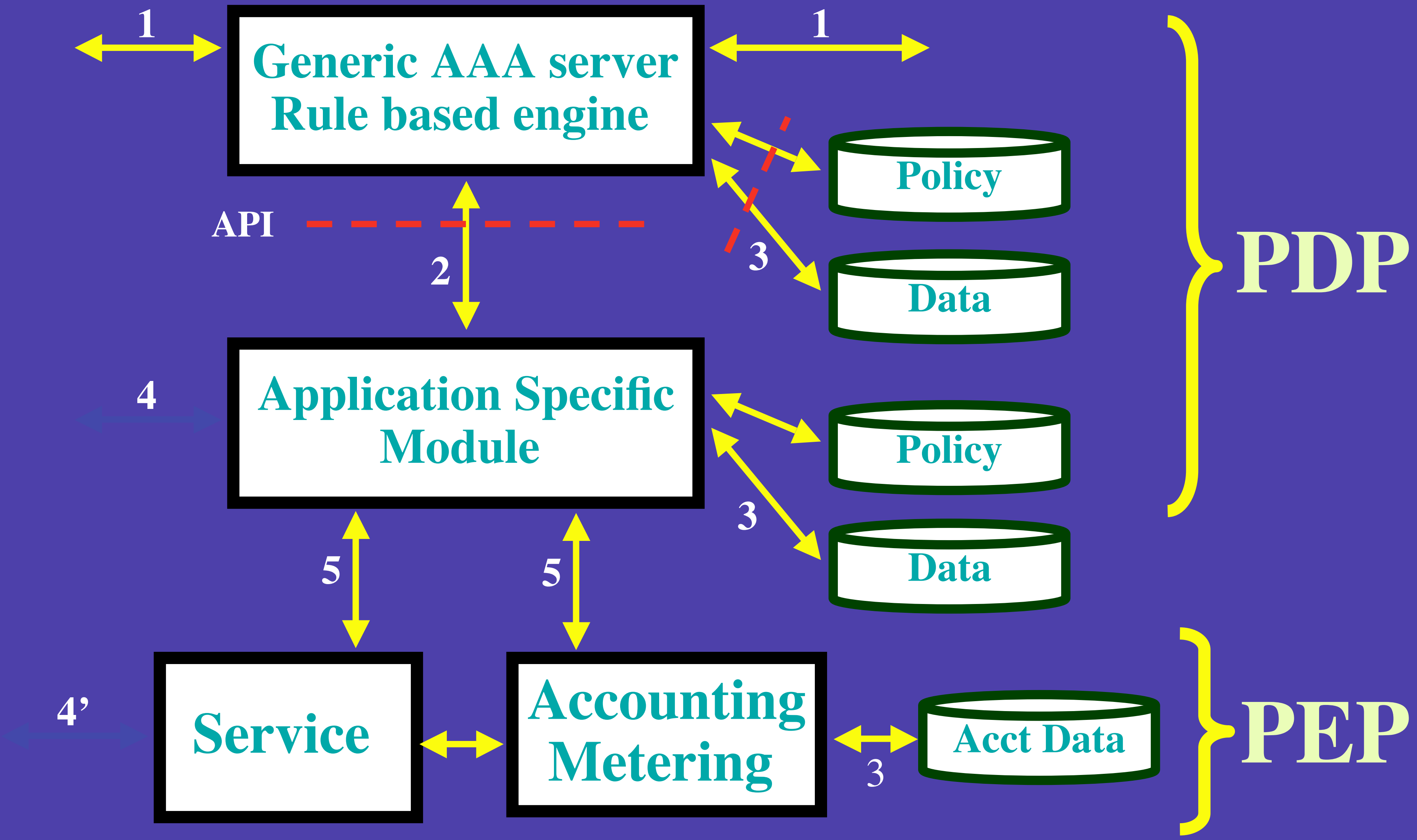
- *Flow control*
- *Traffic Shaping*
- *RED (Random Early Discard)*
- *Self clocking in TCP*
- *Deep memory*



Forbidden area, solutions for s when f = 1 Gb/s, M = 0.5 Mbyte AND NOT USING FLOWCONTROL

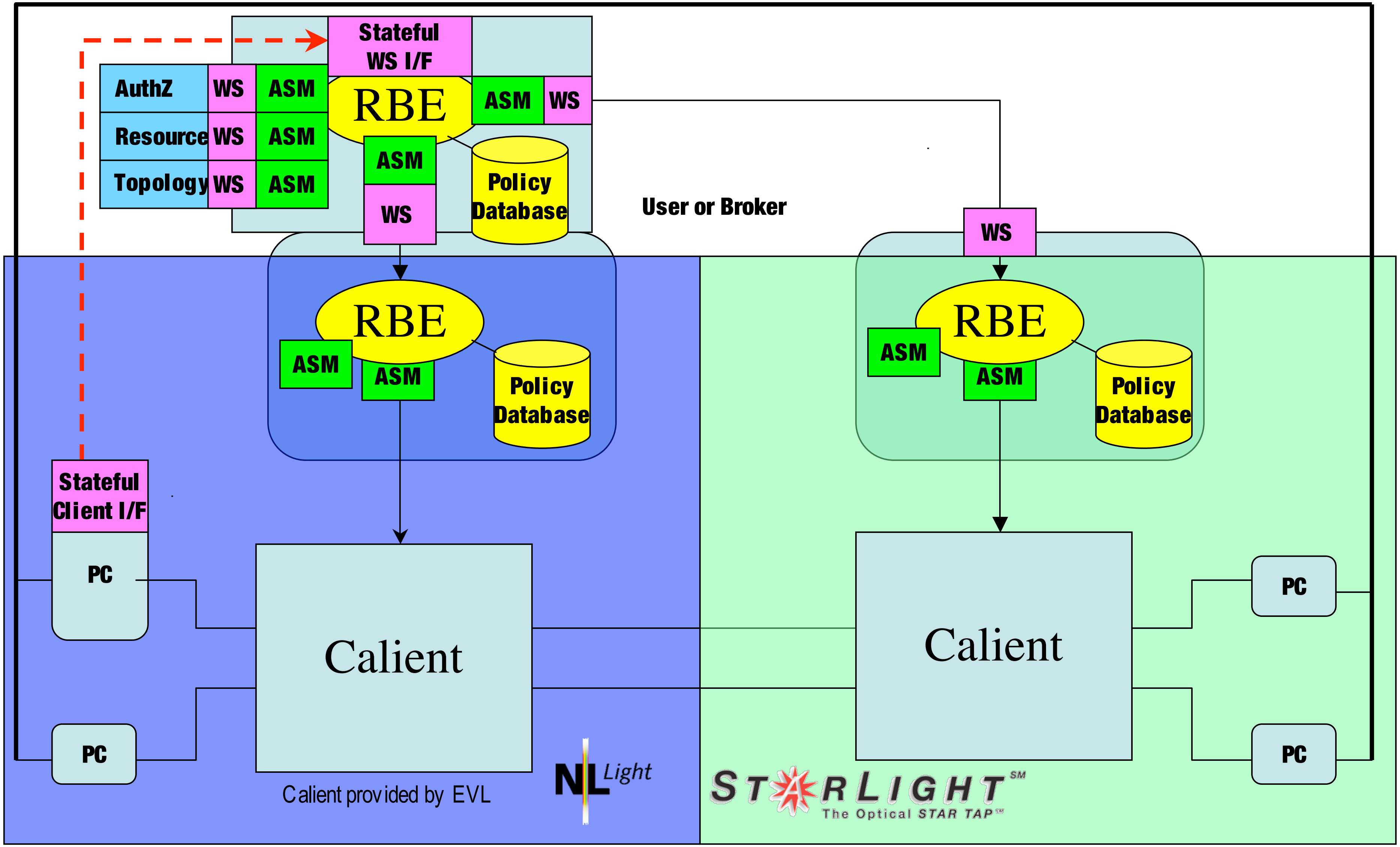


Starting point



RFC 2903 - 2906 , 3334 , policy draft

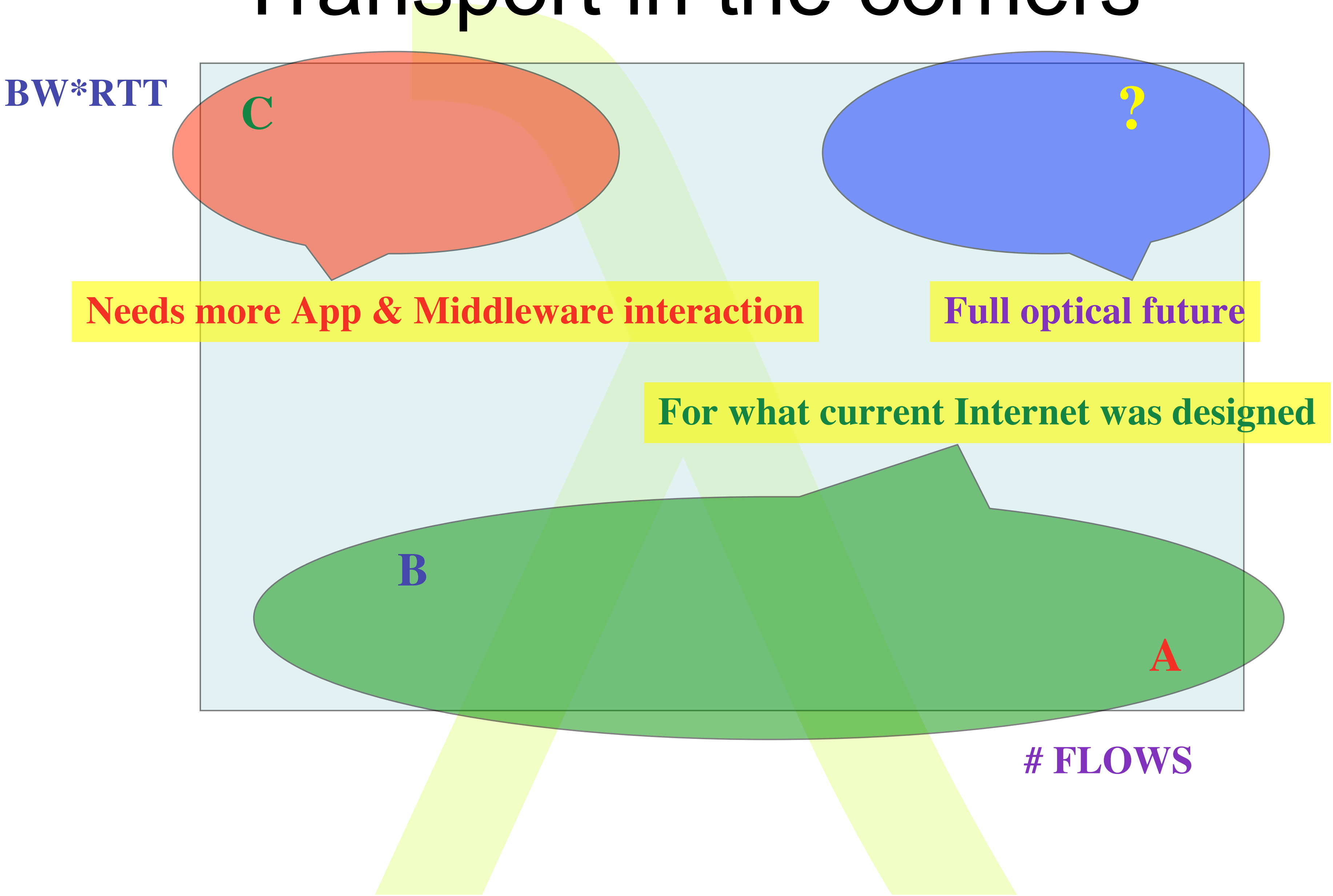
AAA based demo at SC2003



Conlusions

- Demanding applications
  - (Science) data repositories mirroring
  - Instrumentation grids
  - Visualisation and collaboration support
- Model of Lambda networking
  - Identify traffic types
  - Scales of infrastructure
  - Map efficiently to lower the cost/packet
- Current experiments
  - NetherLight
  - VLE/eScience Amsterdam
  - Networking research
    (control plane, transport protocols, optical net models)

Transport in the corners



The 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, Martijn de Munnik, Antony Antony, Rob Meijer, VL-team.