The Lambda Grid

www.science.uva.nl/~delaat

Cees de Laat

SURFnet

EU

University of Amsterdam

SARA

NIKHEF

NCF

The Lambda Grid

www.science.uva.nl/~delaat

Cees de Laat

SURFnet

EU

University of Amsterdam

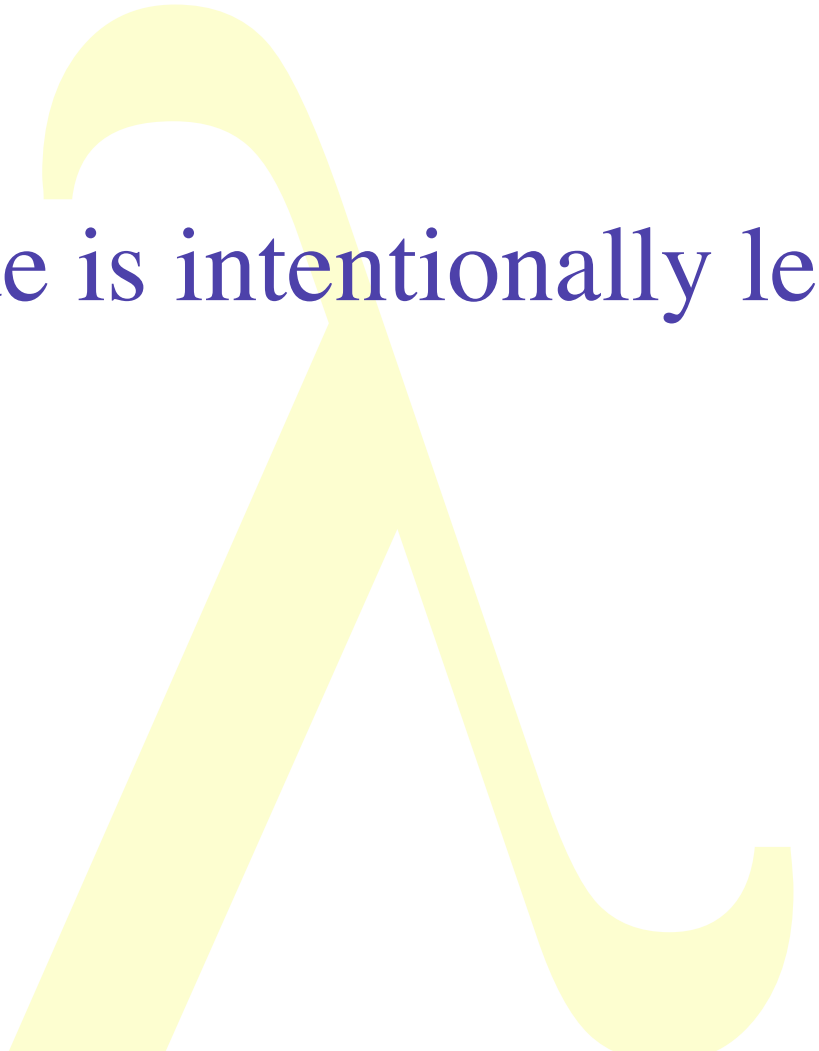
SARA

NIKHEF

NCF

Contents of this talk

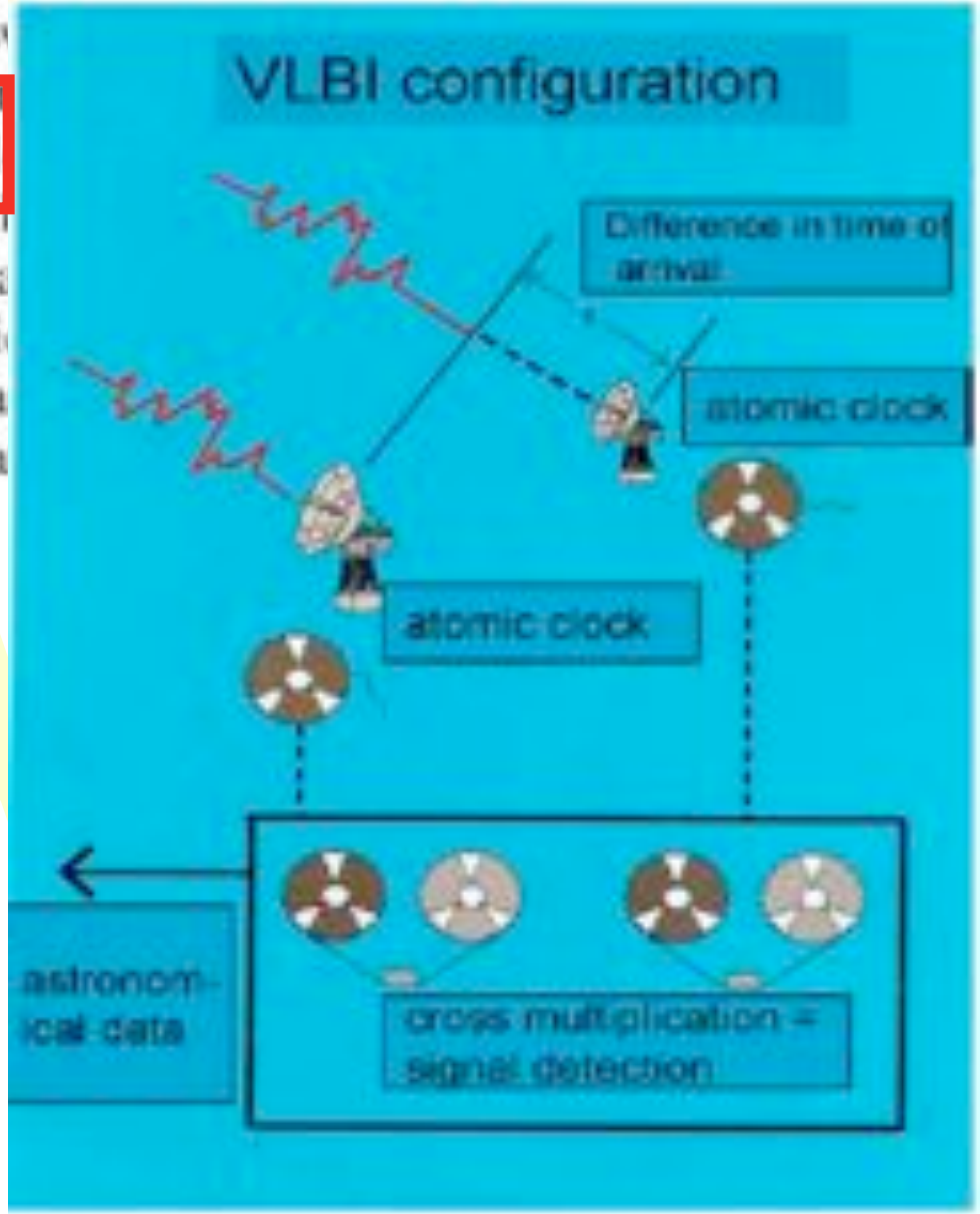
This slide is intentionally left blank



VLBI

he longer term VLBI is easily capable of generating many Gb of data per
scope. The sensitivity of the VLBI array scales w
dwidth (=data-rate) and there is a strong push to
dwidths. Rates of 8Gb/s or more are entirely feasibl
also under development. It is expected that paral
cated correlator will remain the most efficient approa
evolves distributed processing may have an appli
eral, multi-gigabit data streams will aggregate into la
correlator and the capacity of the final link to the da
ting factor.

Westerbork Synthesis Radio Telescope - Netherlands



eVLBI

Lambdas as part of instruments

LOFAR

www.lofar.org

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!)

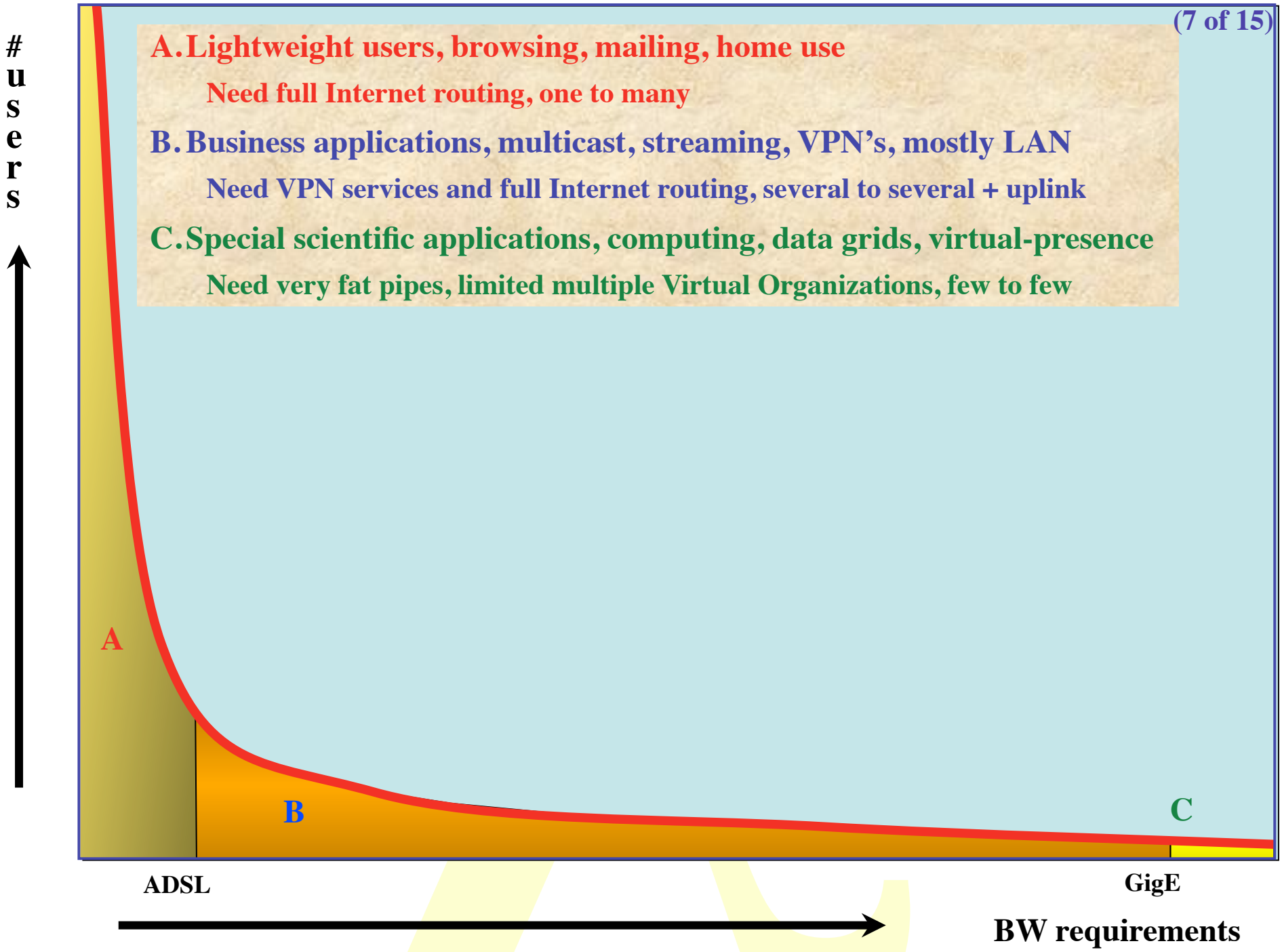
www.igrid2002.org

iGrid 2002

Sept 24-26, 2002, Amsterdam, The Netherlands

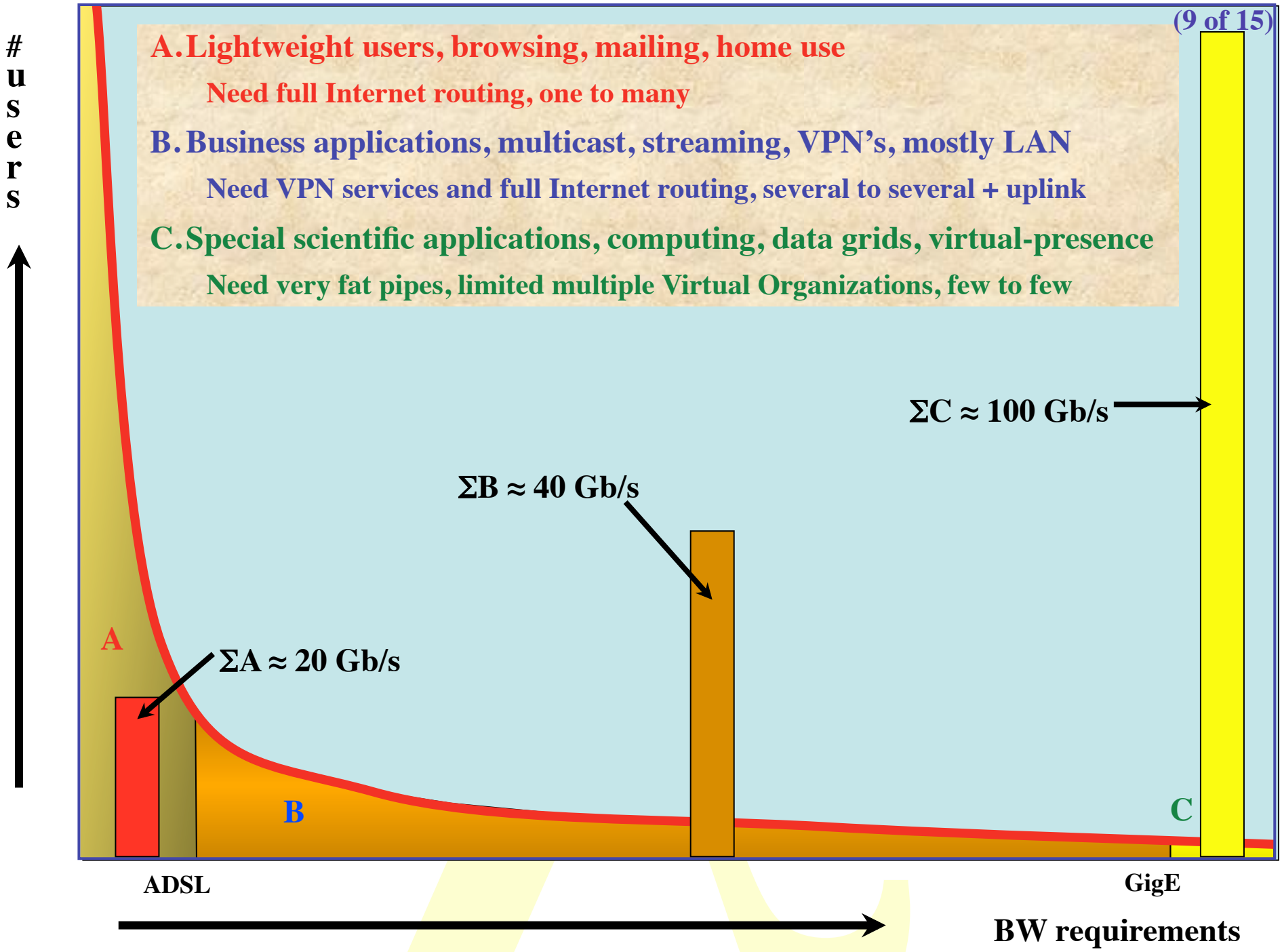
Conference issue
FGCS
Volume 19 (2003)
Number 6 august
22 refereed papers!

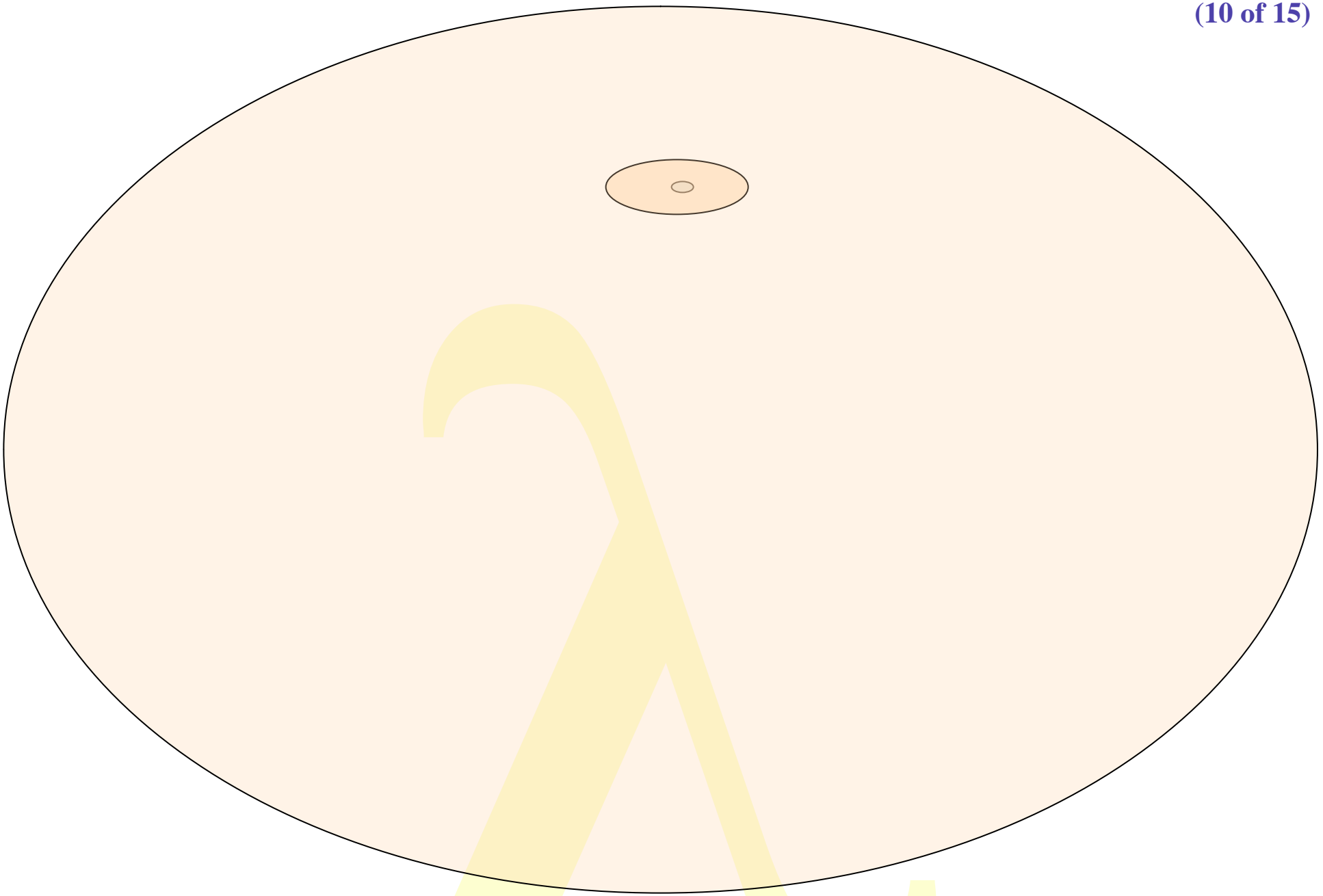
THESE
ARE
THE
APPLICATIONS!


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

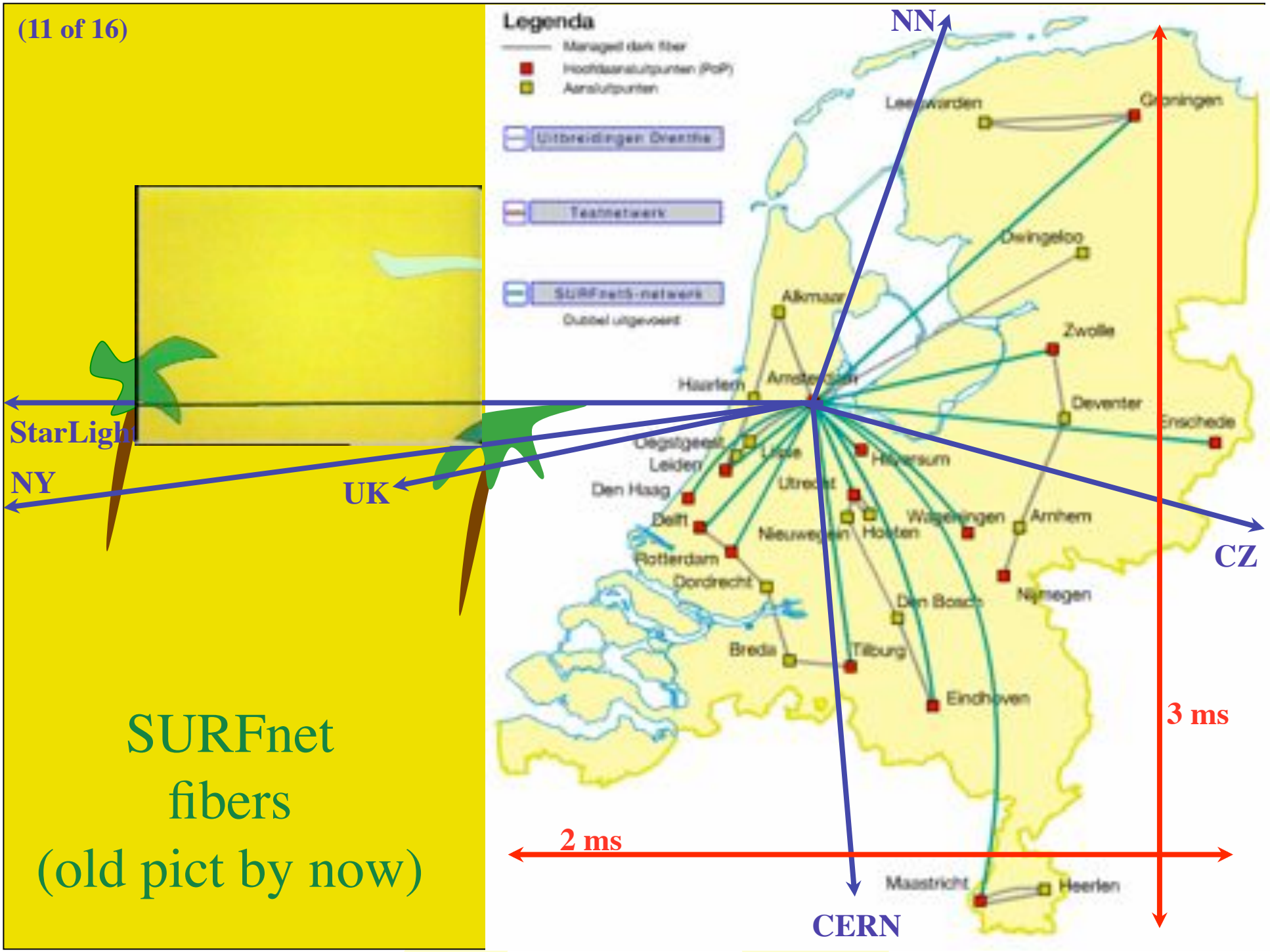
The Dutch Situation

- **Estimate A**
  - **17 M people, 6.4 M households, 25 % penetration of 0.5 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 ==> 20 TByte/s**


#users
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
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)

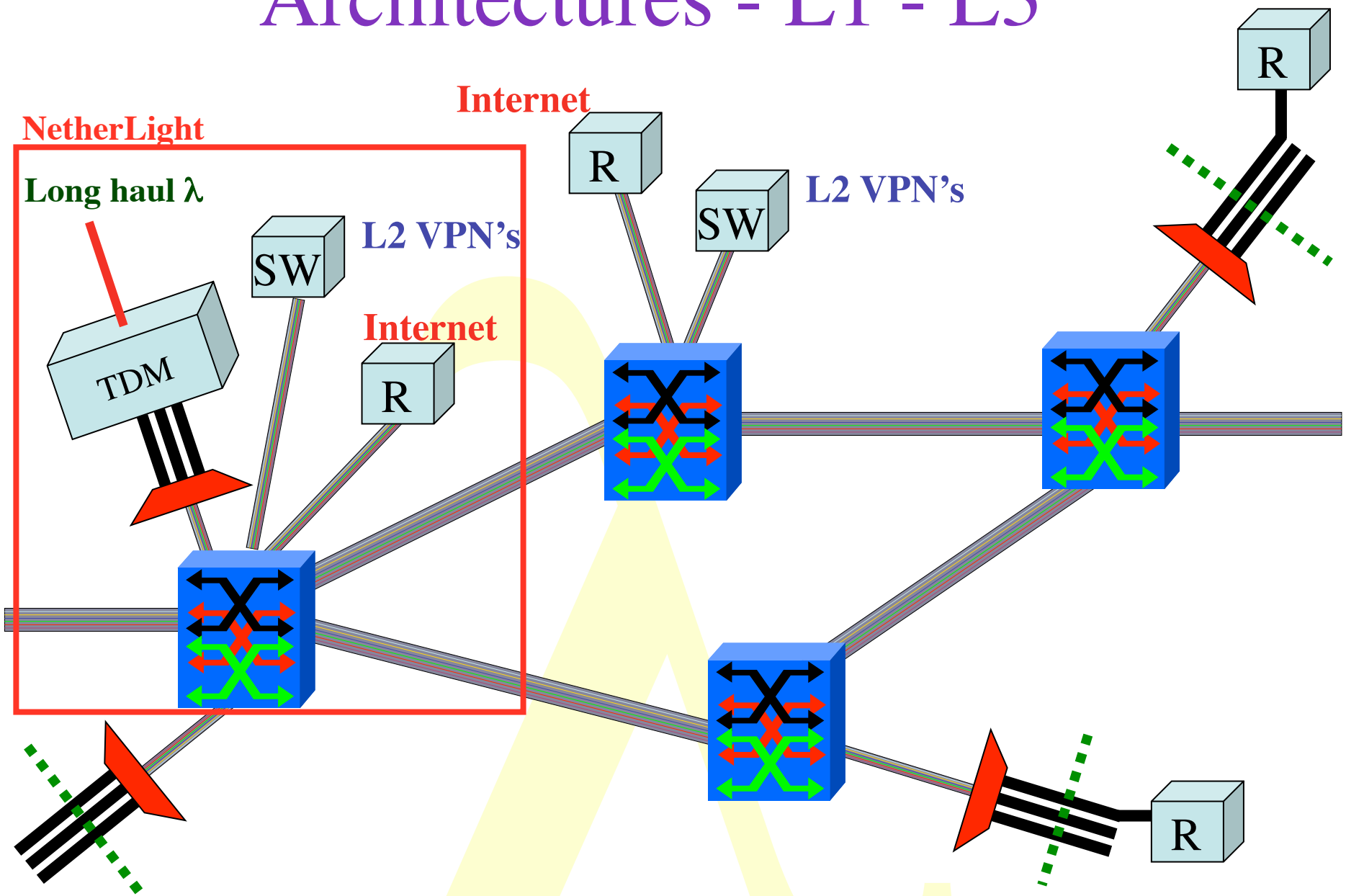
So what are the facts

- 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
  - 100 Byte packet @ 40 Gb/s -> 20 ns to look up in 140 kEntries routing table (light speed from me to you!)
- Big sciences need fat pipes
- Bottom line: look for a hybrid architecture which serves all classes in a cost effective way (A -> L3 , B -> L2 , C -> L1)
- Tested 10 gbps Ethernet WANPHY Amsterdam-CERN (ATLAS)
  - http://www.surfnet.nl/en/publications/pressreleases/021003.html

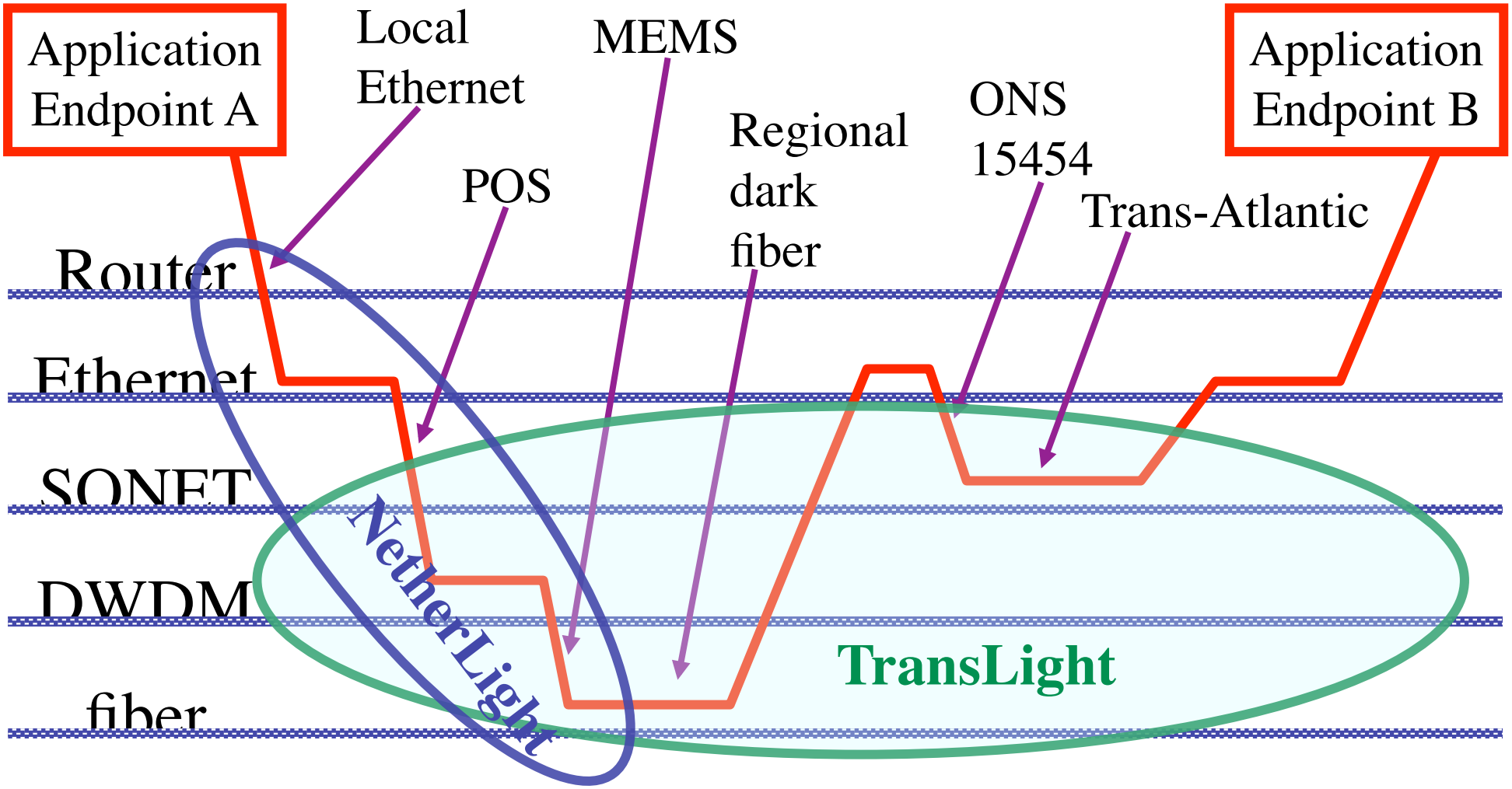
Services

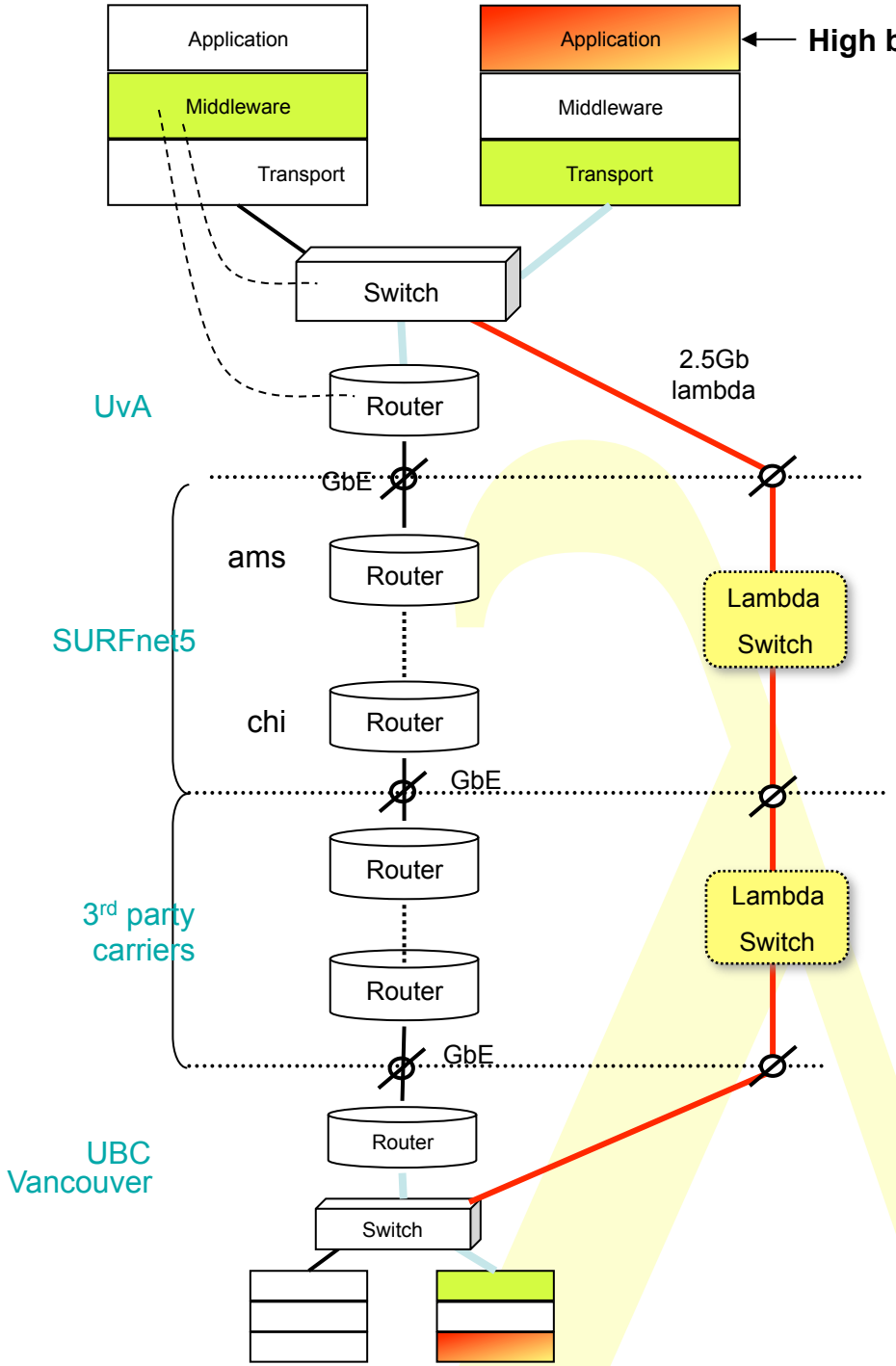
SCALE / CLASS	2 Metro	20 National/ regional	200 World
A	Switching/ routing	Routing	ROUTER$
B	Switches + E-WANPHY VPN's,	Switches + E-WANPHY (G)MPLS	ROUTER$
C	dark fiber Optical switching	Lambda switching	Sub-lambdas, ethernet-sdh

Architectures - L1 - L3



How low can you go?

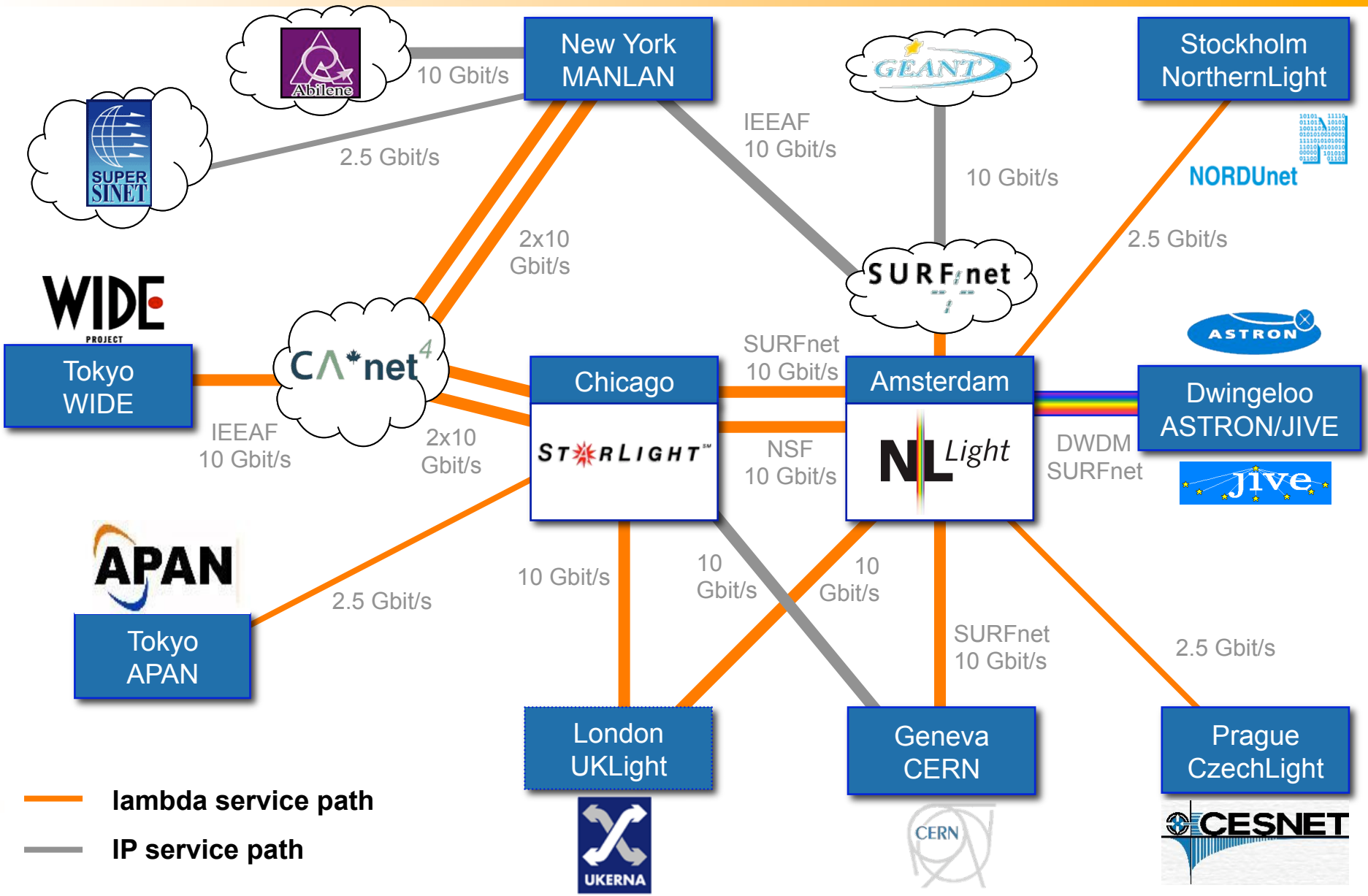




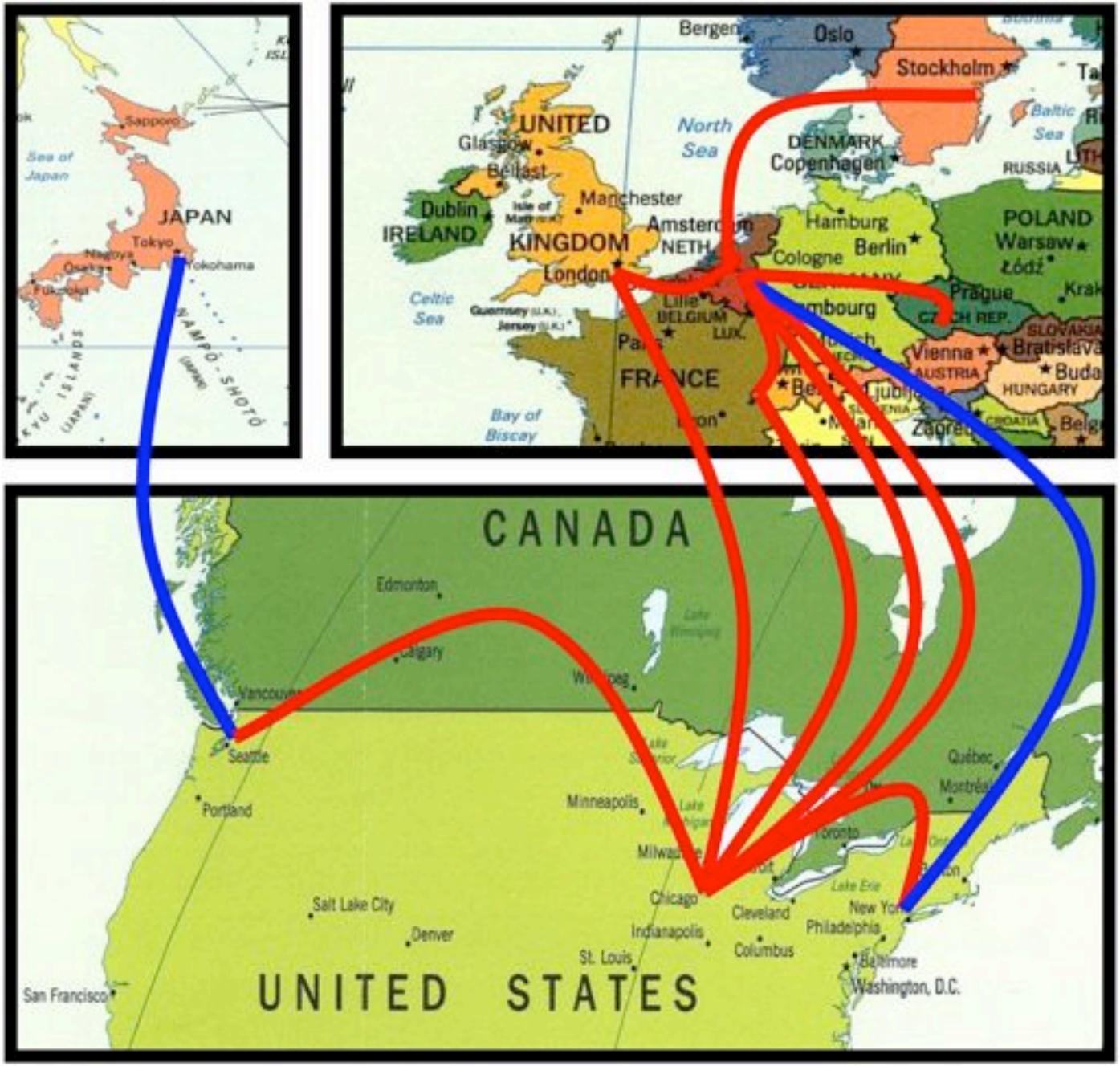
- lambda for high bandwidth applications
  - Bypass of production network
  - Middleware may request (optical) pipe
- RATIONALE:
  - Lower the cost of transport per packet
  - Use Internet as controlplane!

International lightpath network 1Q2004

GigaPort



TransLight Lambdas



European lambdas to US
–6 GigEs Amsterdam—Chicago
–2 GigEs CERN—Chicago
–8 GigEs London—Chicago

Canadian lambdas to US
–8 GigEs Chicago—Canada—NYC

–8 GigEs Chicago—Canada—Seattle

US lambdas to Europe
–4 GigEs Chicago—Amsterdam
–2 GigEs Chicago—CERN

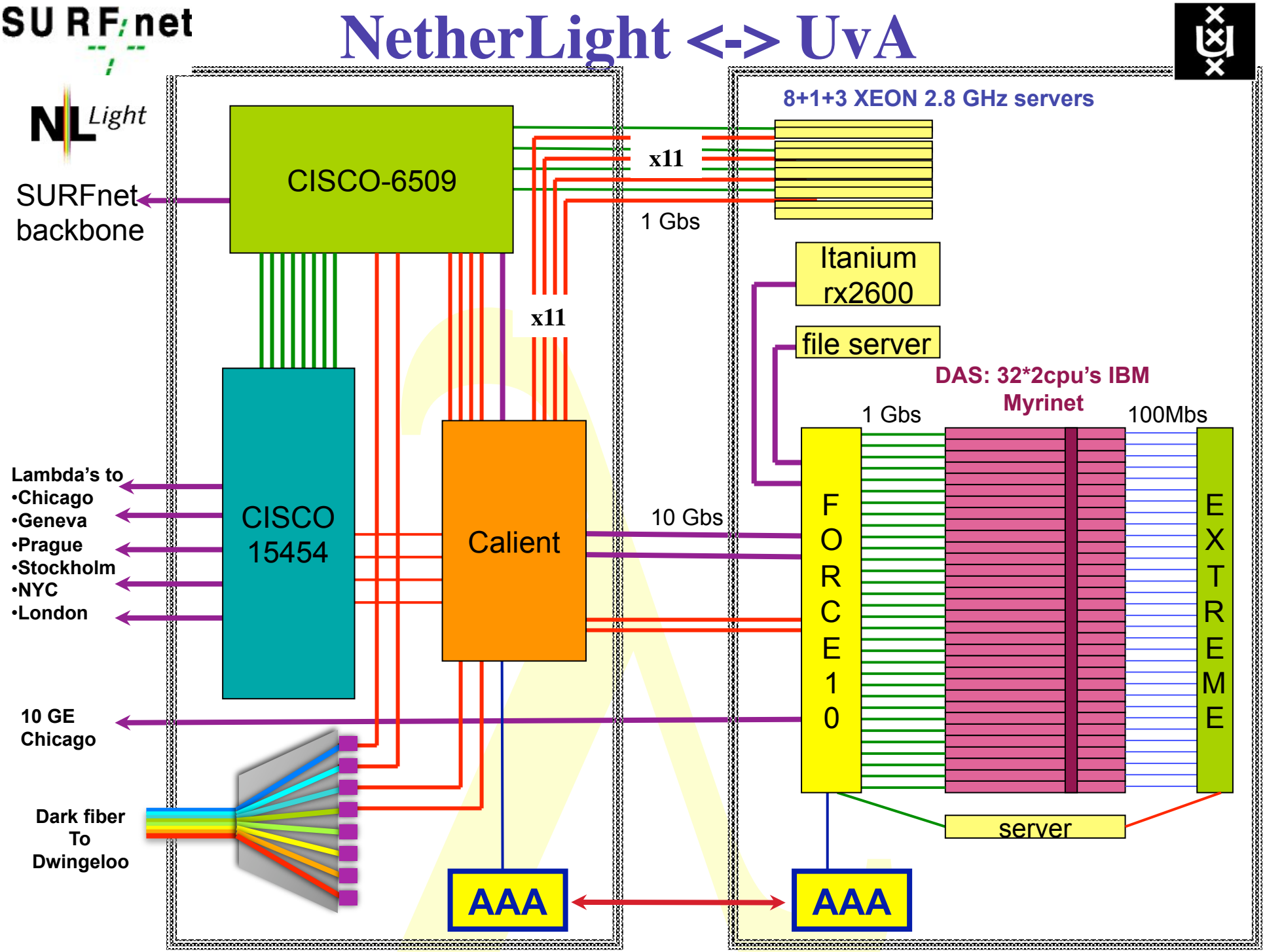
European lambdas
–8 GigEs Amsterdam—CERN
–2 GigEs Prague—Amsterdam
–2 GigEs Stockholm—Amsterdam
–8 GigEs London—Amsterdam

IEEAF lambdas (blue)
–8 GigEs Seattle—Tokyo
–8 GigEs NYC—Amsterdam

Little GLORIAD

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

NetherLight <-> UvA

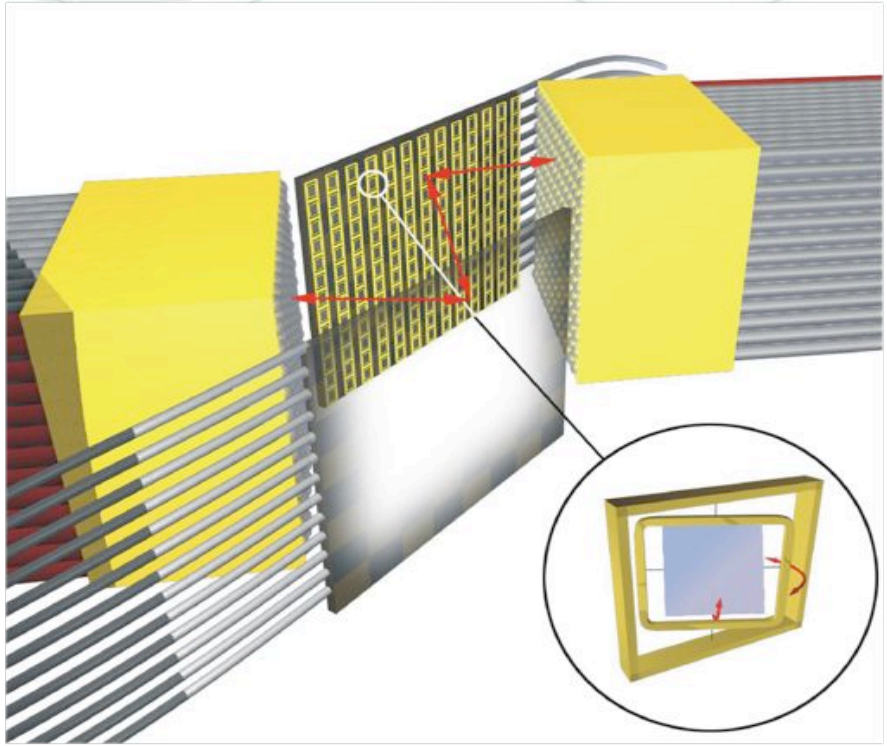


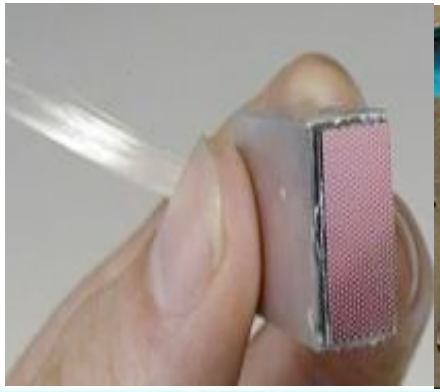
UVA/EVL's
64*64
Optical Switch
@ NetherLight
in SURFnet POP
@ SARA
Costs 1/100th of
a similar
throughput router
but with specific
services!

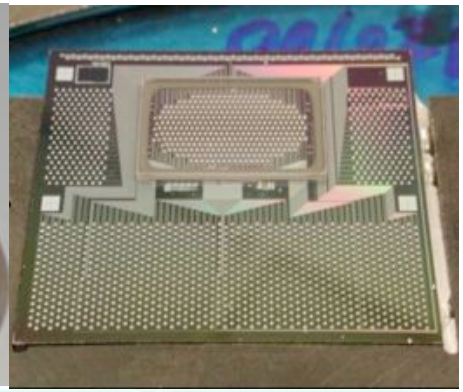
Core Switch Technology

(Intermezzo-2)

Calient NETWORKS







- **3D MEMS structure**
  - Bulk MEMS – High Density Chips
  - Electrostatic actuation
  - Short path length (~4cm)
  - <1.5 dB median loss
- **Completely Non-blocking**
  - Single-stage up to 1Kx1K
  - 10 ms switching time
- **Excellent Transparency**
  - Polarization
  - Bit rate
  - Wavelength

Layer - 2 requirements from 3/4

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

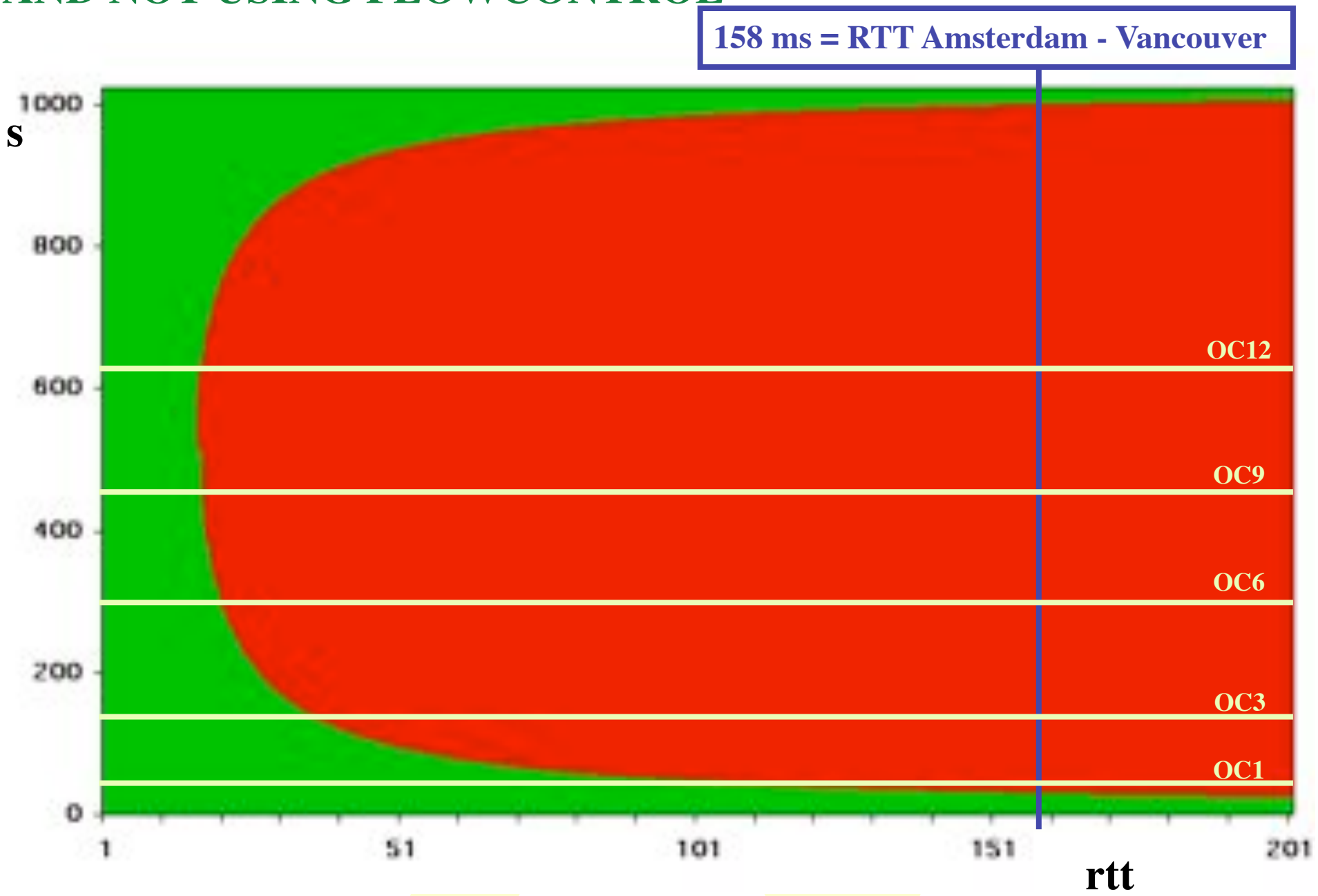
```
Window = BandWidth * RTT     &    BW == slow

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

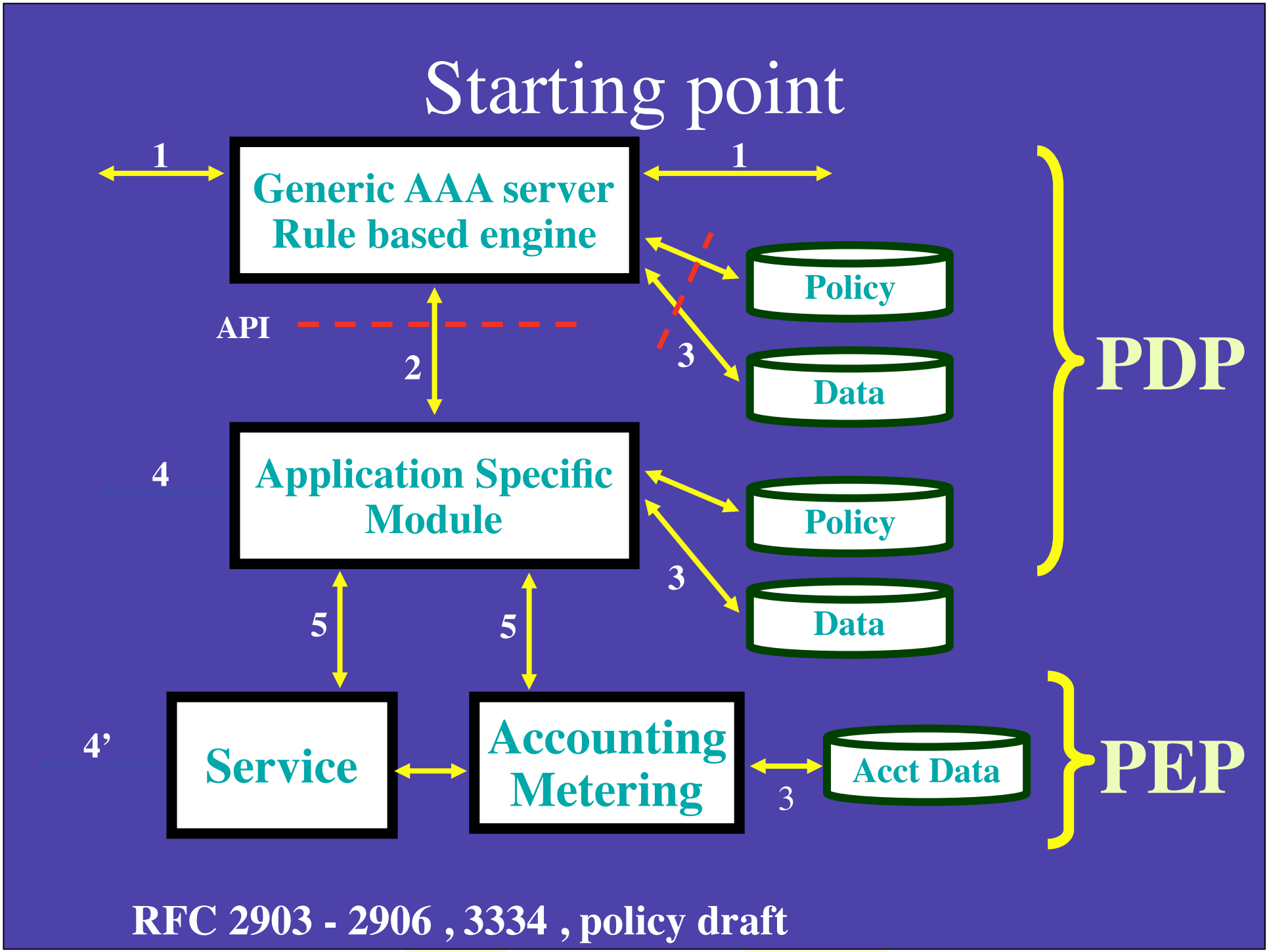
So pick from menu:

- *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 (20 of 22)
AND NOT USING FLOWCONTROL



Starting point

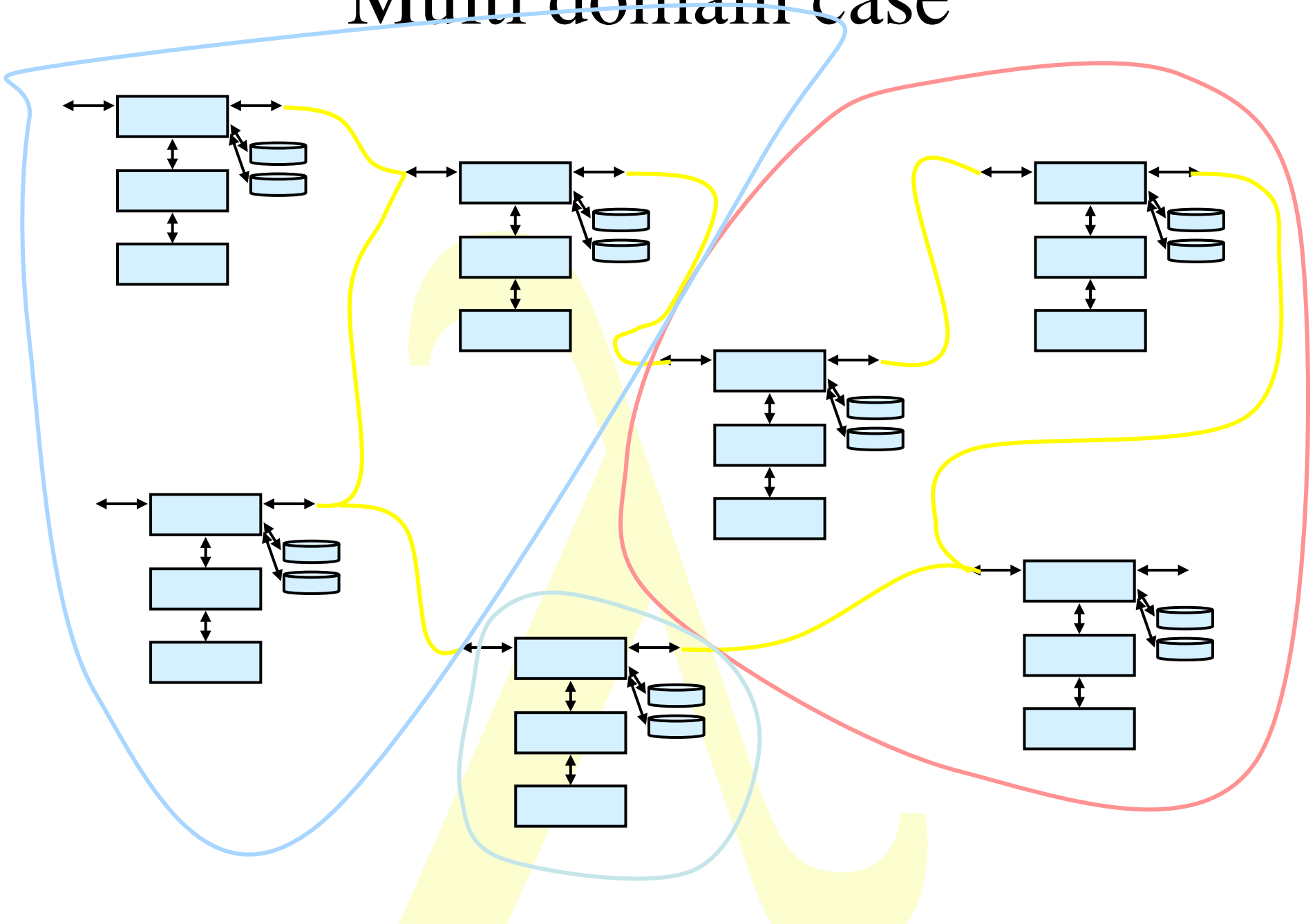


RFC 2903 - 2906 , 3334 , policy draft

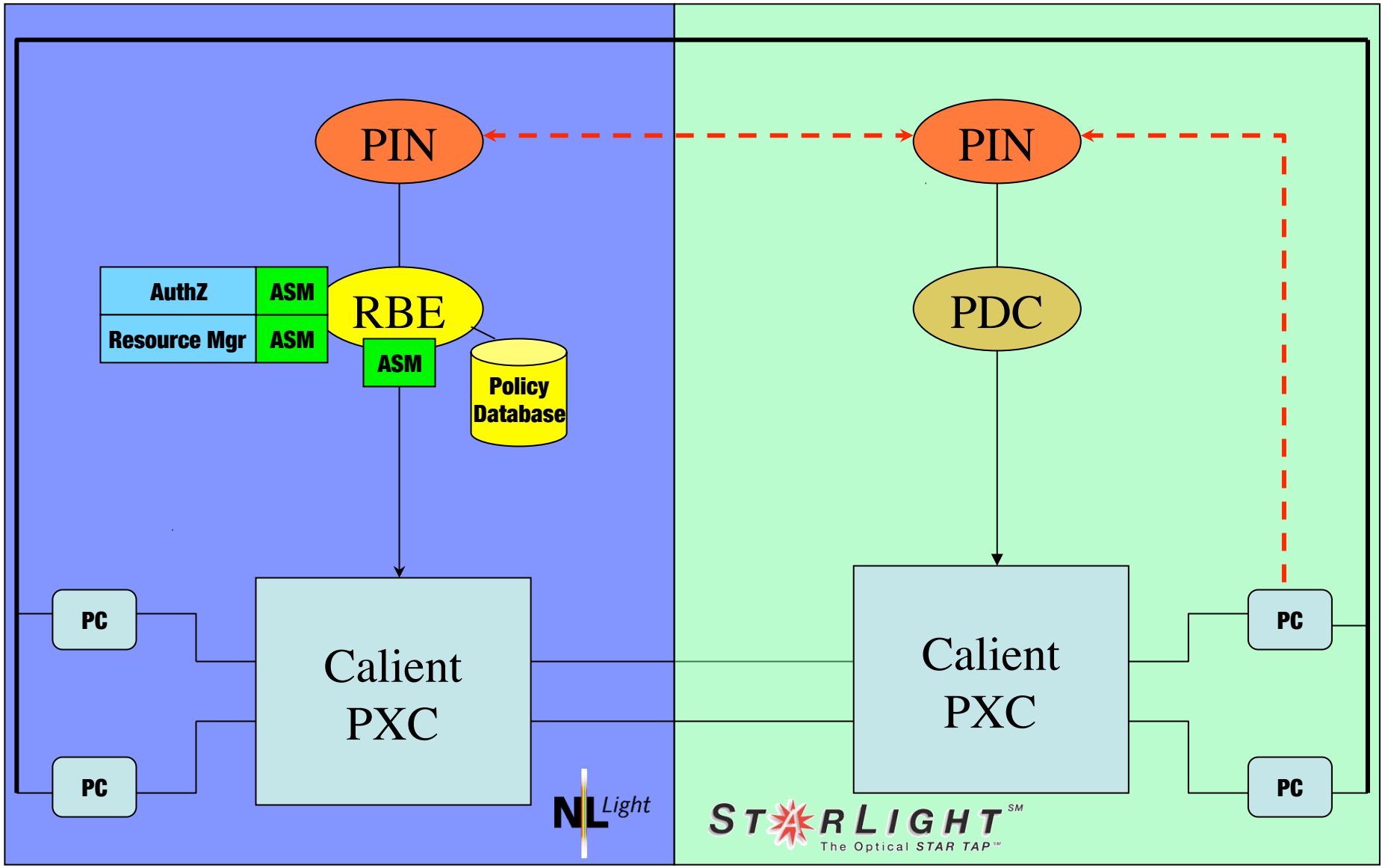
Multi Domain Lambda setup

- AAA based on RFC 2903-2906
- OGSI web services wrapper
- Interface to CALIENT optical switch, layer 2 switches
- Interface to PDC
- Broker for path searching, selection
- Web and application interface
- Demonstration on SC2003

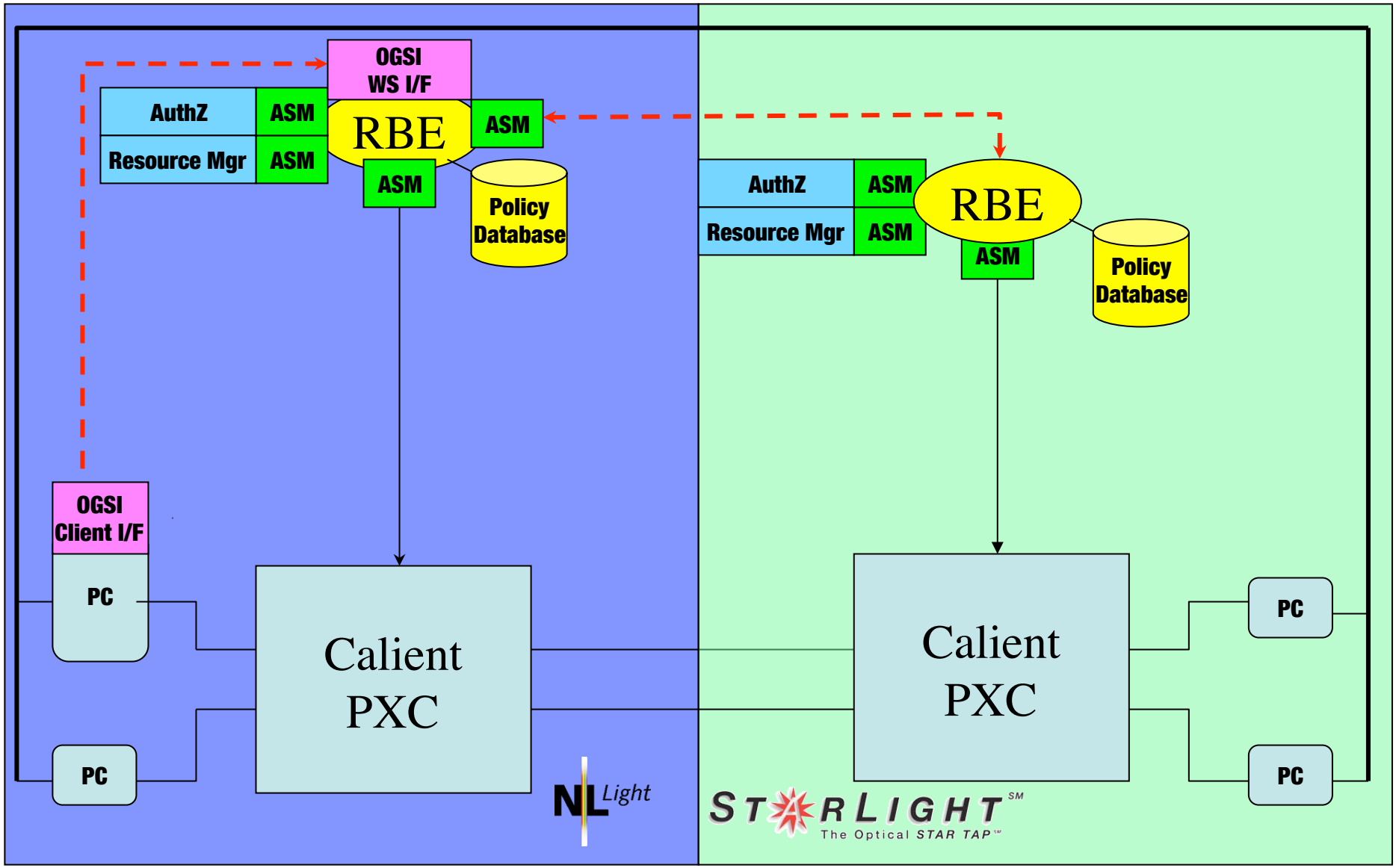
Multi domain case



Multi-domain experiment 1 at SC2003



Multi-domain experiment 2 at SC2003

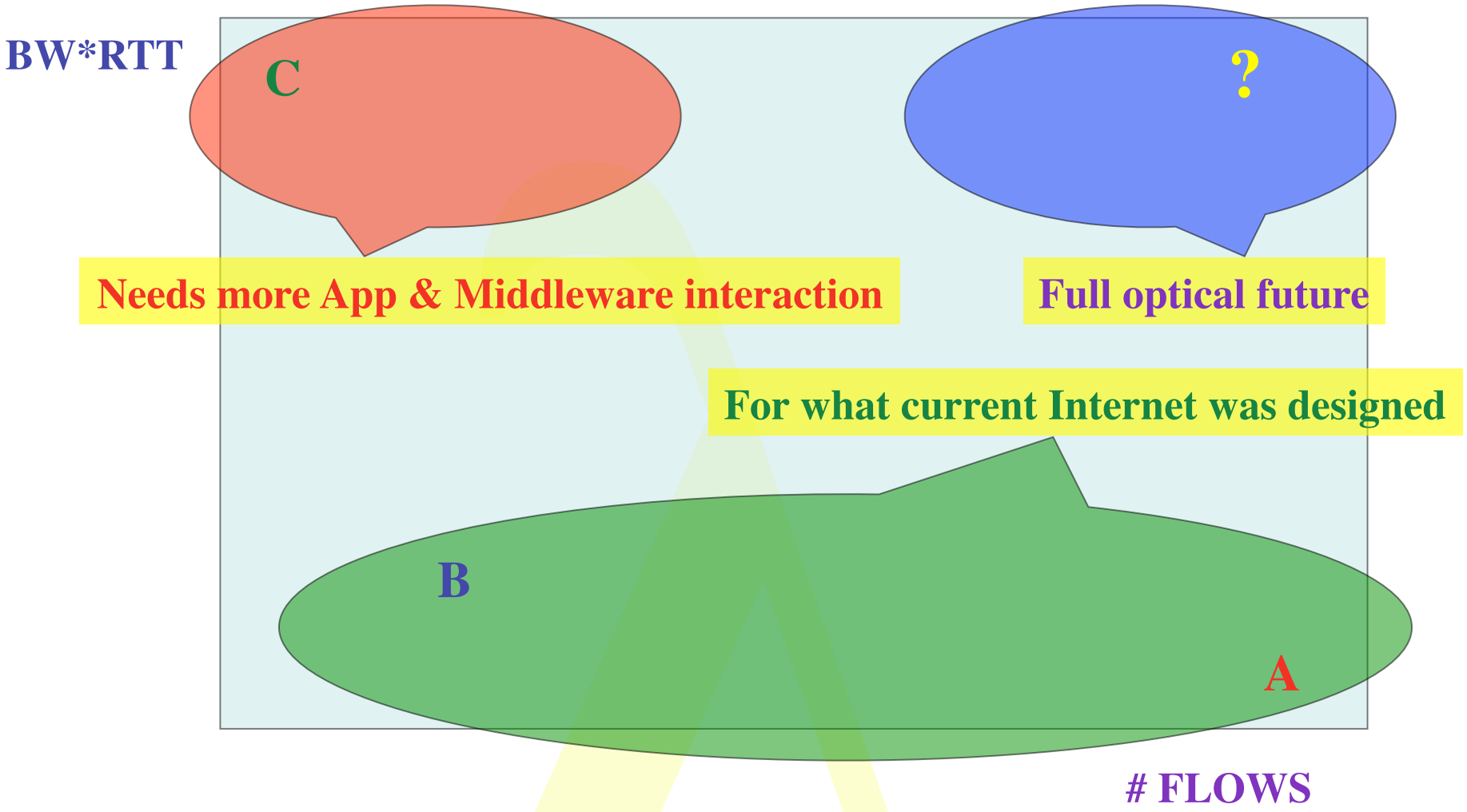


Lambda workshop

- **Amsterdam - Terena**
  - **Concepts**
  - **Initial testbed (SURFnet Lambda to StarLight)**
- **Amsterdam - iGrid2002**
  - **Rechecking concepts models**
  - **Initial experiences and measurements**
  - **Expansion of Lambda testbed**
- **Reykjavik - NORDUnet**
  - **Towards persistent demonstrations and applications**
- **Next one in UK sept 3th 2004 (tentative)**

3th Lambda workshop @ NORDUnet 2003

Transport in the corners



Revisiting the truck of tapes

Consider one fiber

- **Current technology allows 320 λ in one of the frequency bands**
- **Each λ has a bandwidth of 40 Gbit/s**
- **Transport: 320 * $40*10^9$ / 8 = 1600 GByte/sec**

- **Take a 10 metric ton truck**
  - **One tape contains 50 Gbyte, weights 100 gr**
  - **Truck contains (10000 / 0.1) * 50 Gbyte = 5 PByte**

- **Truck / fiber = 5 PByte / 1600 GByte/sec = 3125 s ≈ one hour**

- **For distances further away than a truck drives in one hour (50 km) minus loading and handling 100000 tapes the fiber wins!!!**

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.