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

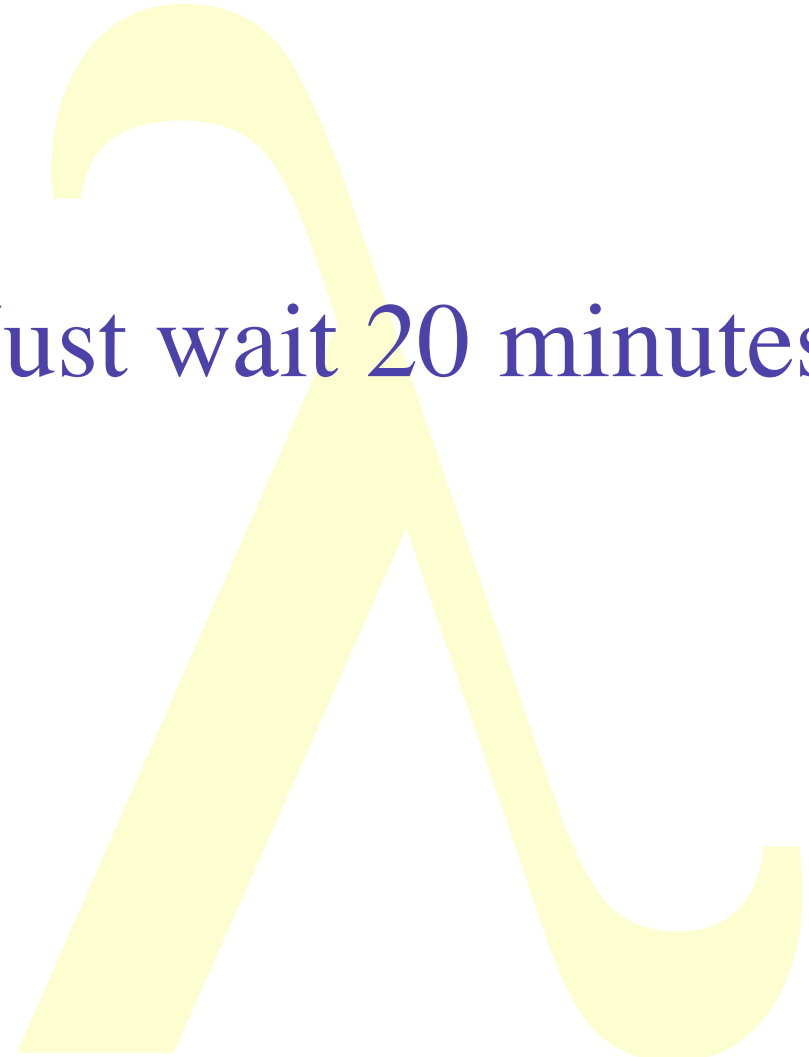
University of Amsterdam

SARA

NIKHEF

NCF

Contents of this talk

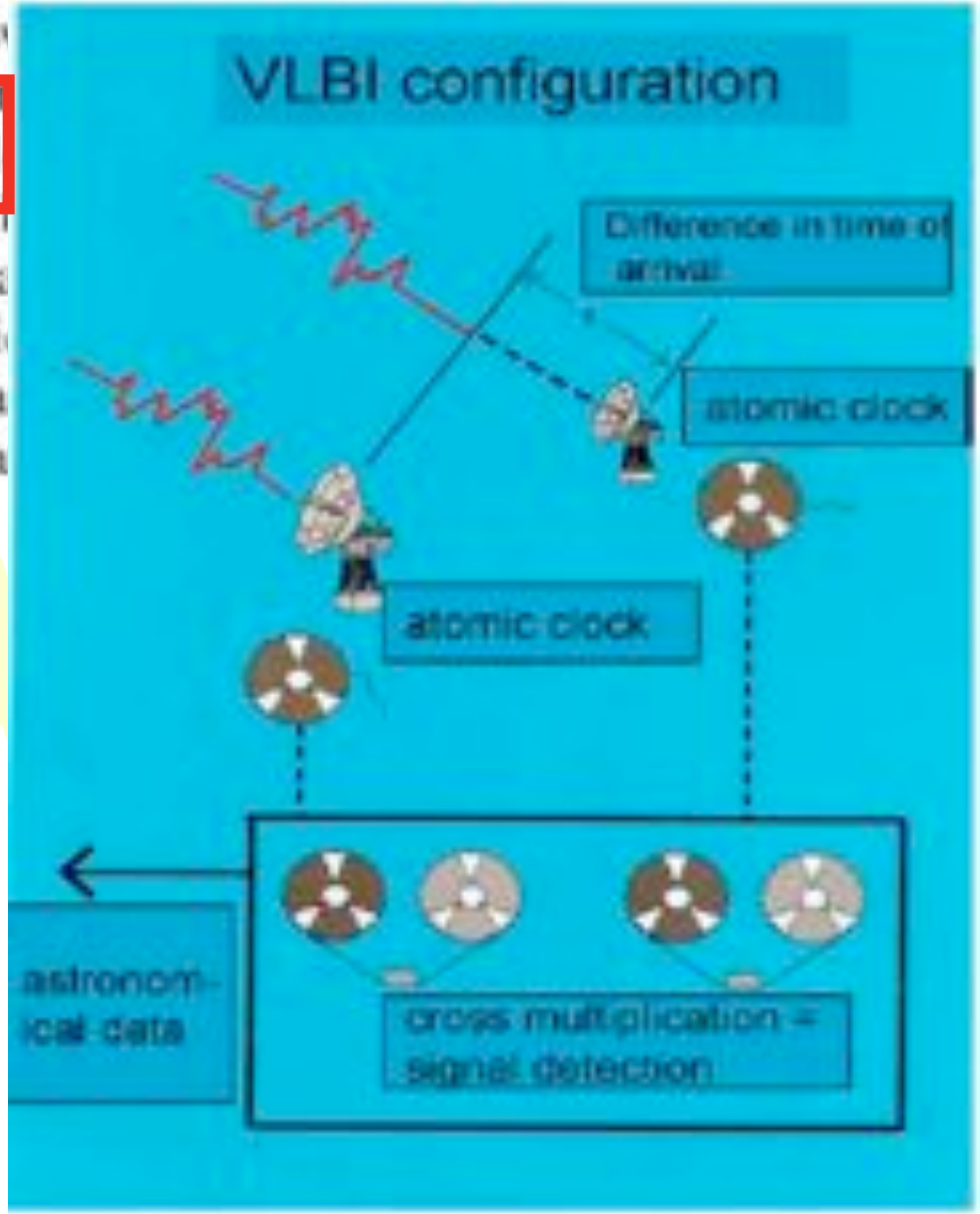


Just wait 20 minutes

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



Lambdas as part of instruments

GigaPort

LOFAR

www.lofar.org

SURFnet

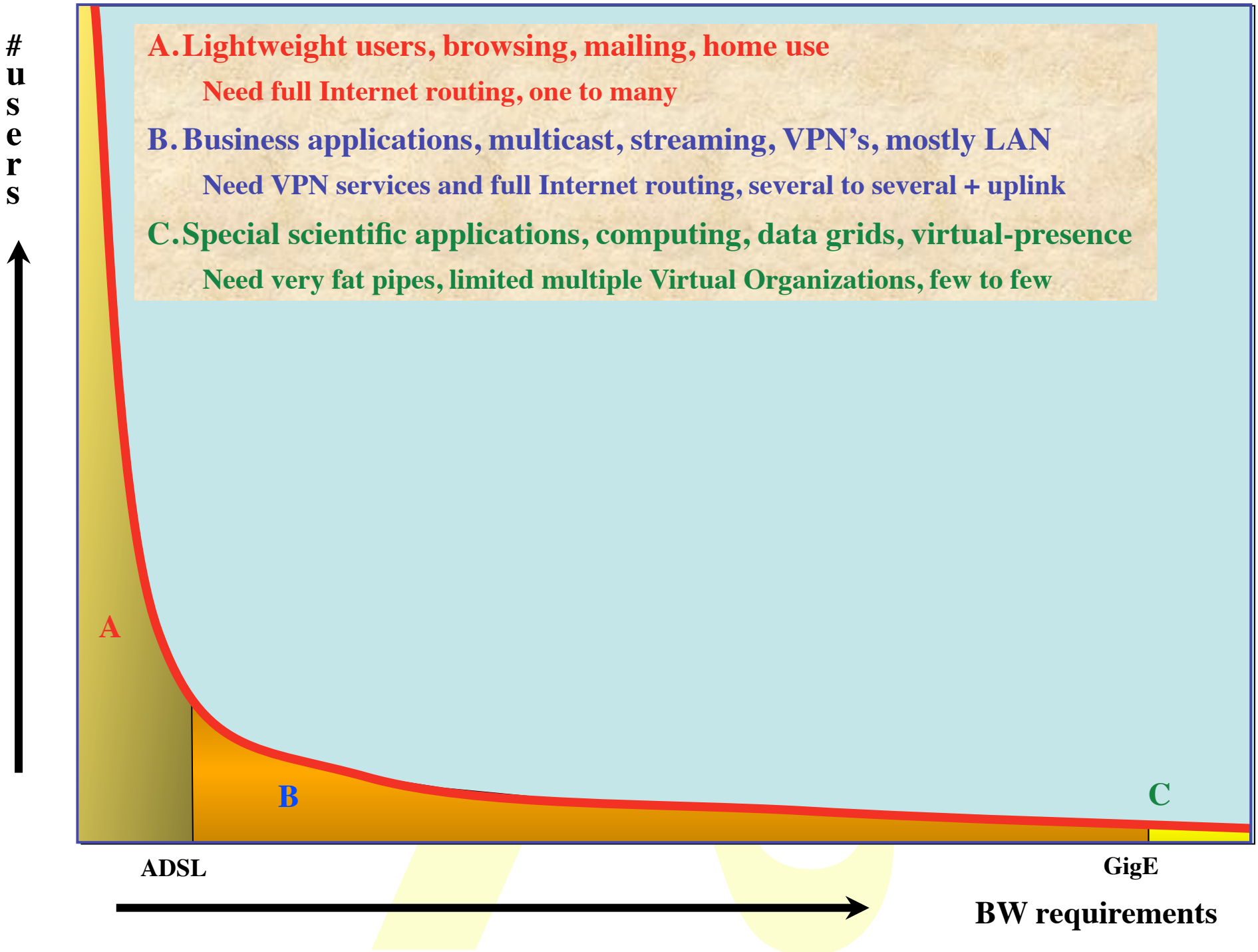
iGrid 2002

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

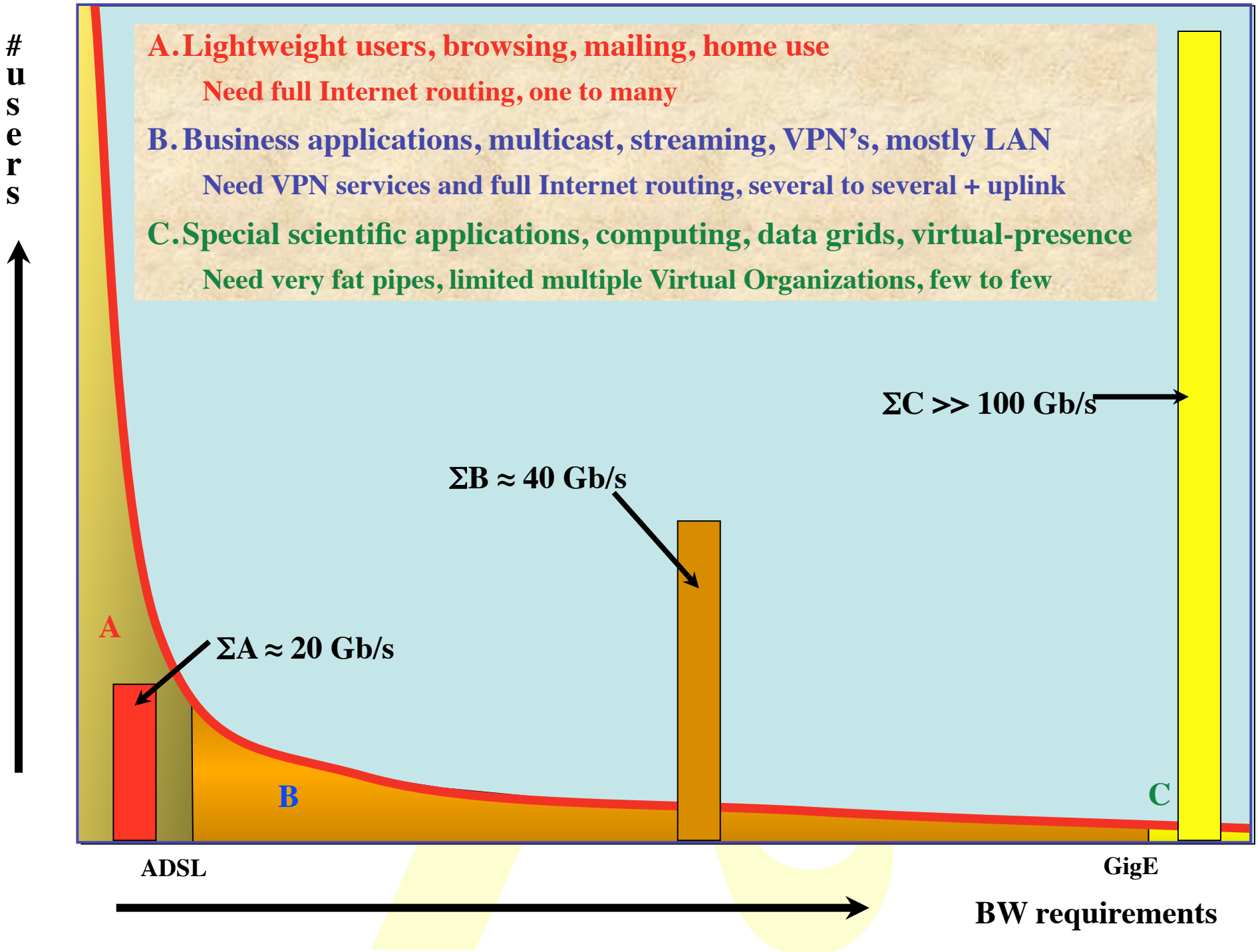
- Grid technologies demonstrated: Major emphasis on grid middleware, data management grids, data replication grids, visualization grids, data/visualization grids, computational grids, access grids, grid portals
- 25Gb transatlantic bandwidth (100Mb/attendee, 250x iGrid2000!)

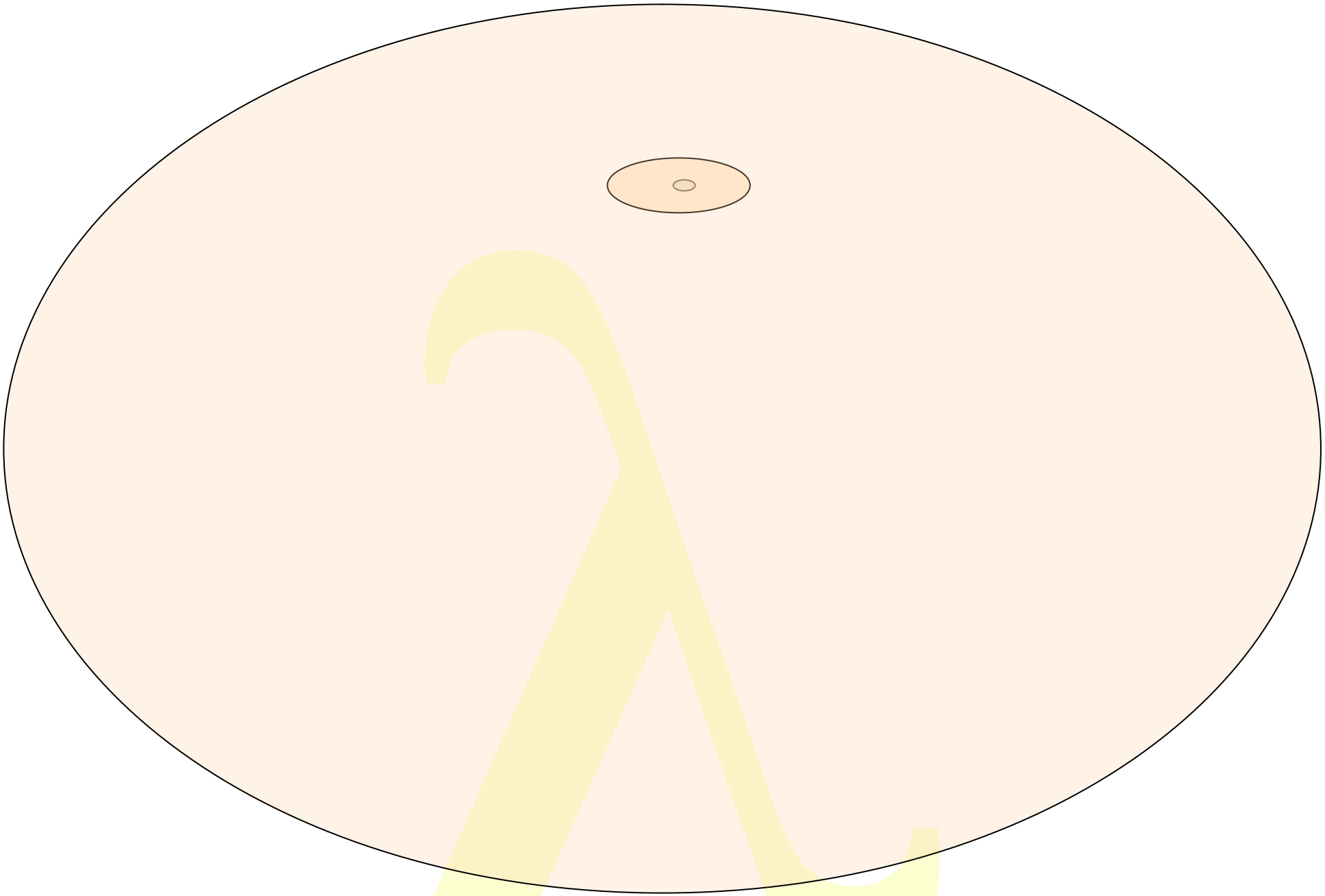
www.igrid2002.org

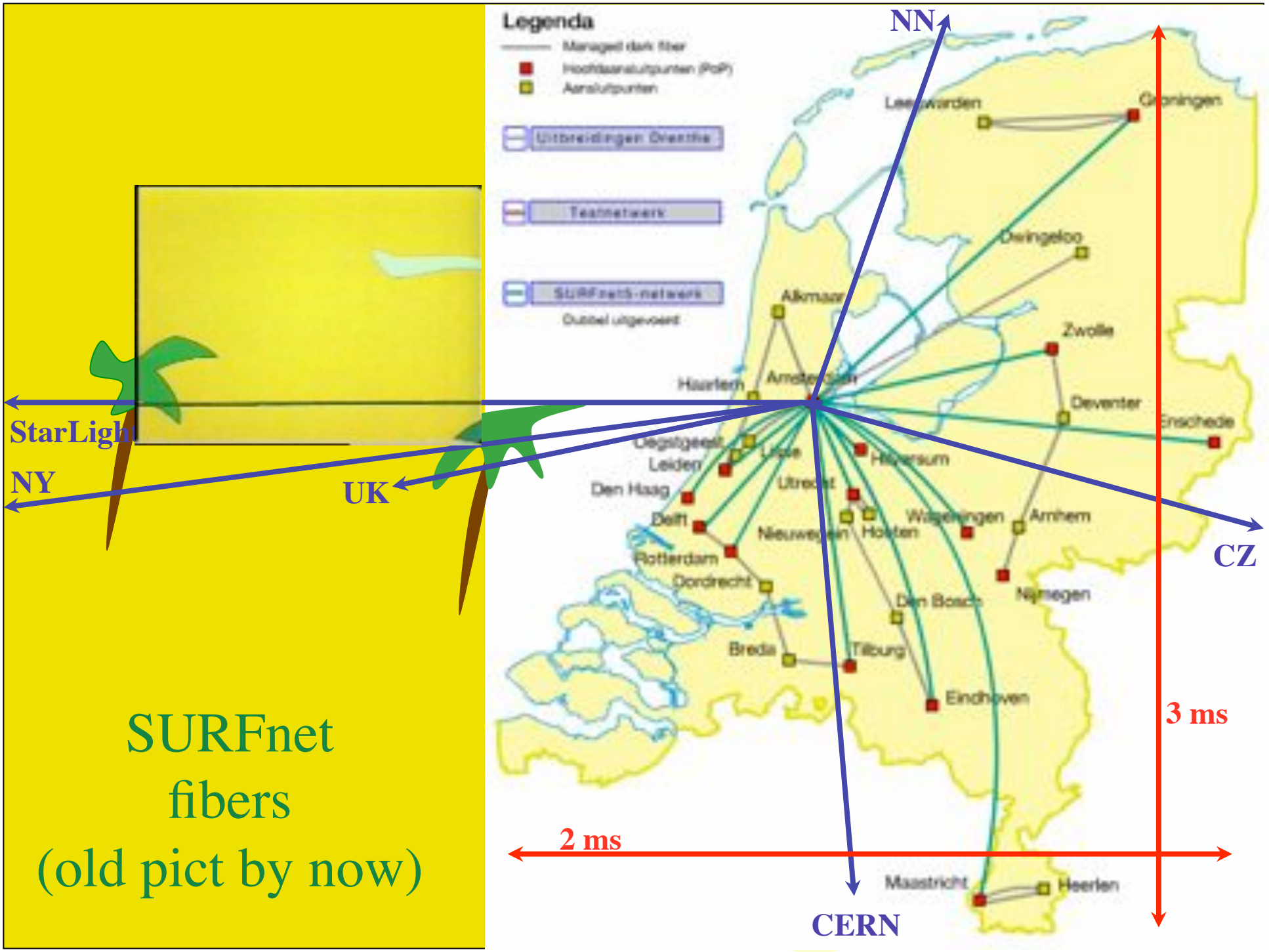

#users
A. Lightweight users, browsing, mailing, home use
Need full Internet routing, one to many
B. Business applications, multicast, streaming, VPN's, mostly LAN
Need VPN services and full Internet routing, several to several + uplink
C. Special scientific applications, computing, data grids, virtual-presence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



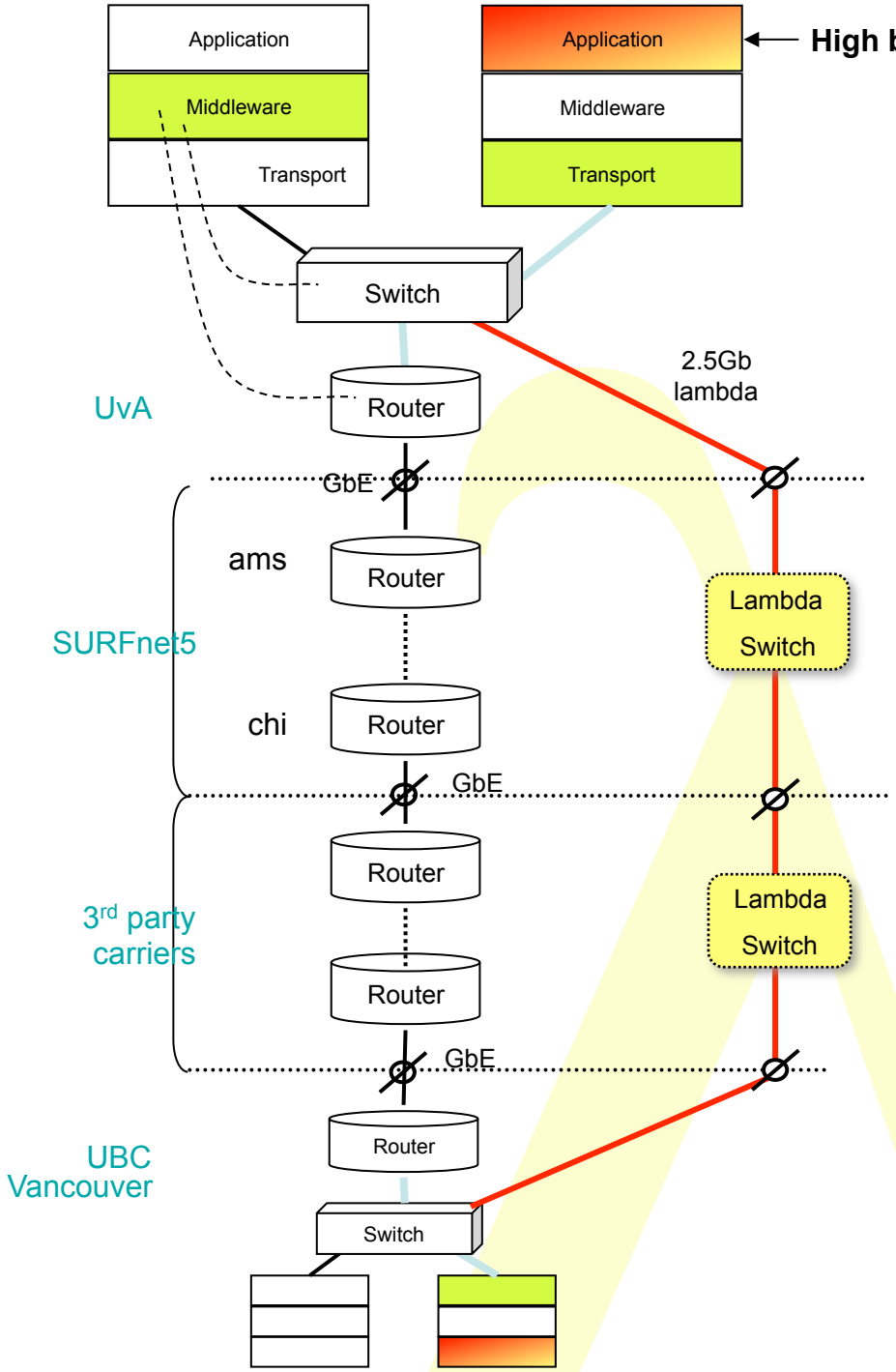
SURFnet
fibers
(old pict by now)

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!

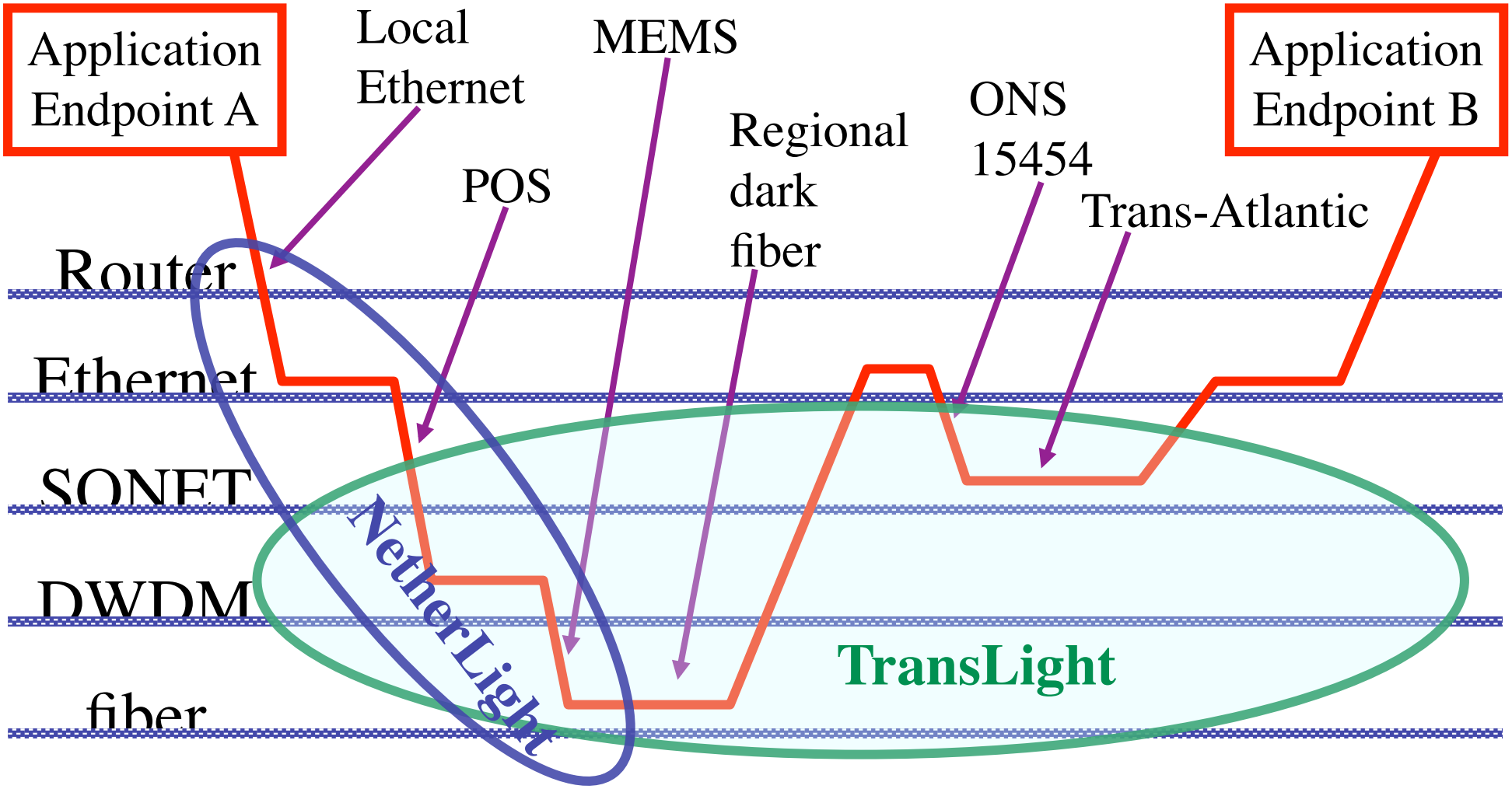
Services

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 Opt- switch	Lambda switching	Sub-lambdas, sonet/sdh, ethernet



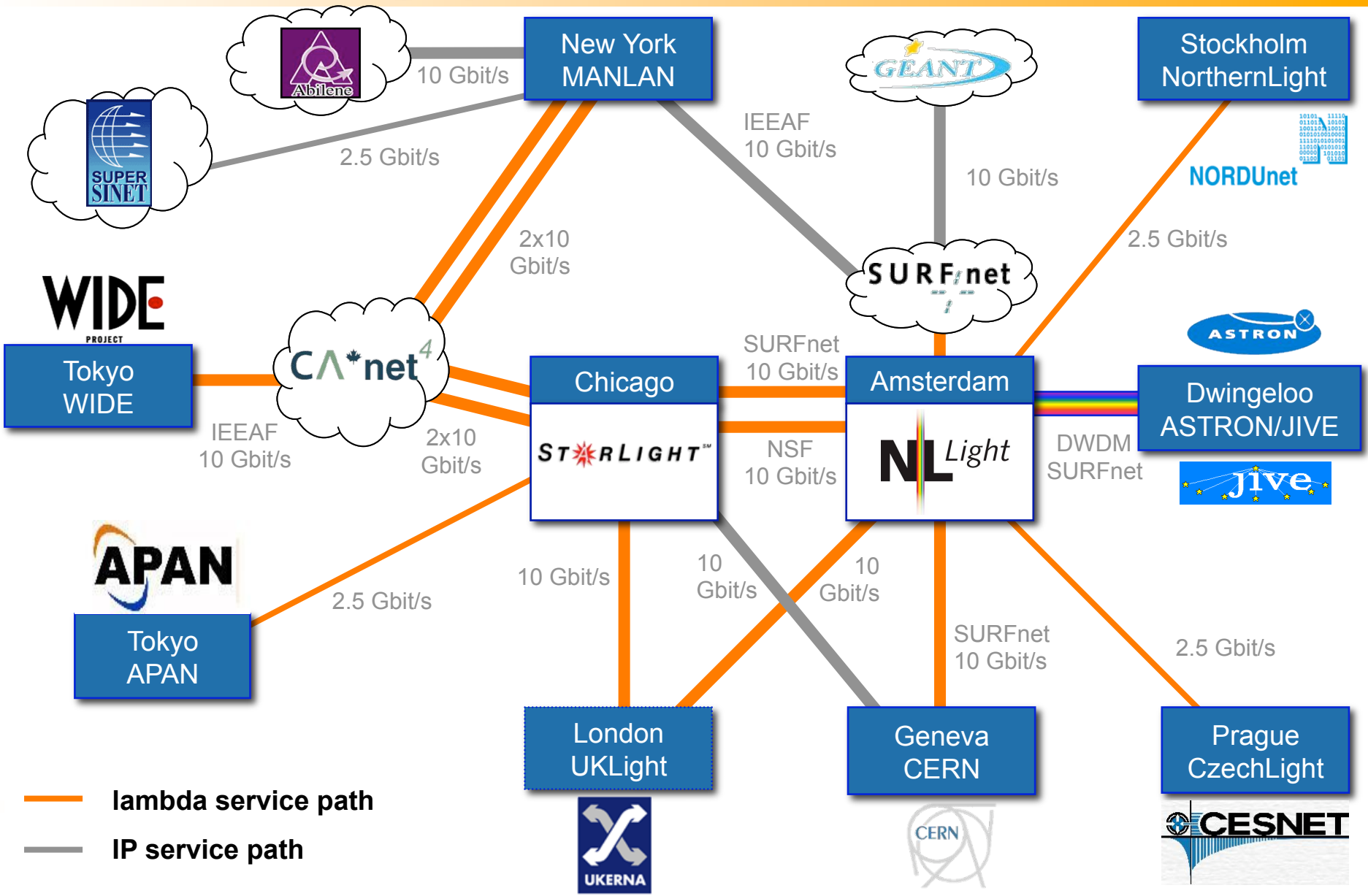
- lambda for high bandwidth applications
  - Bypass of production network
  - Middleware may request (optical) pipe
- RATIONALE:
  - Lower the cost of transport per packet
  - Use Internet as controlplane!

How low can you go?

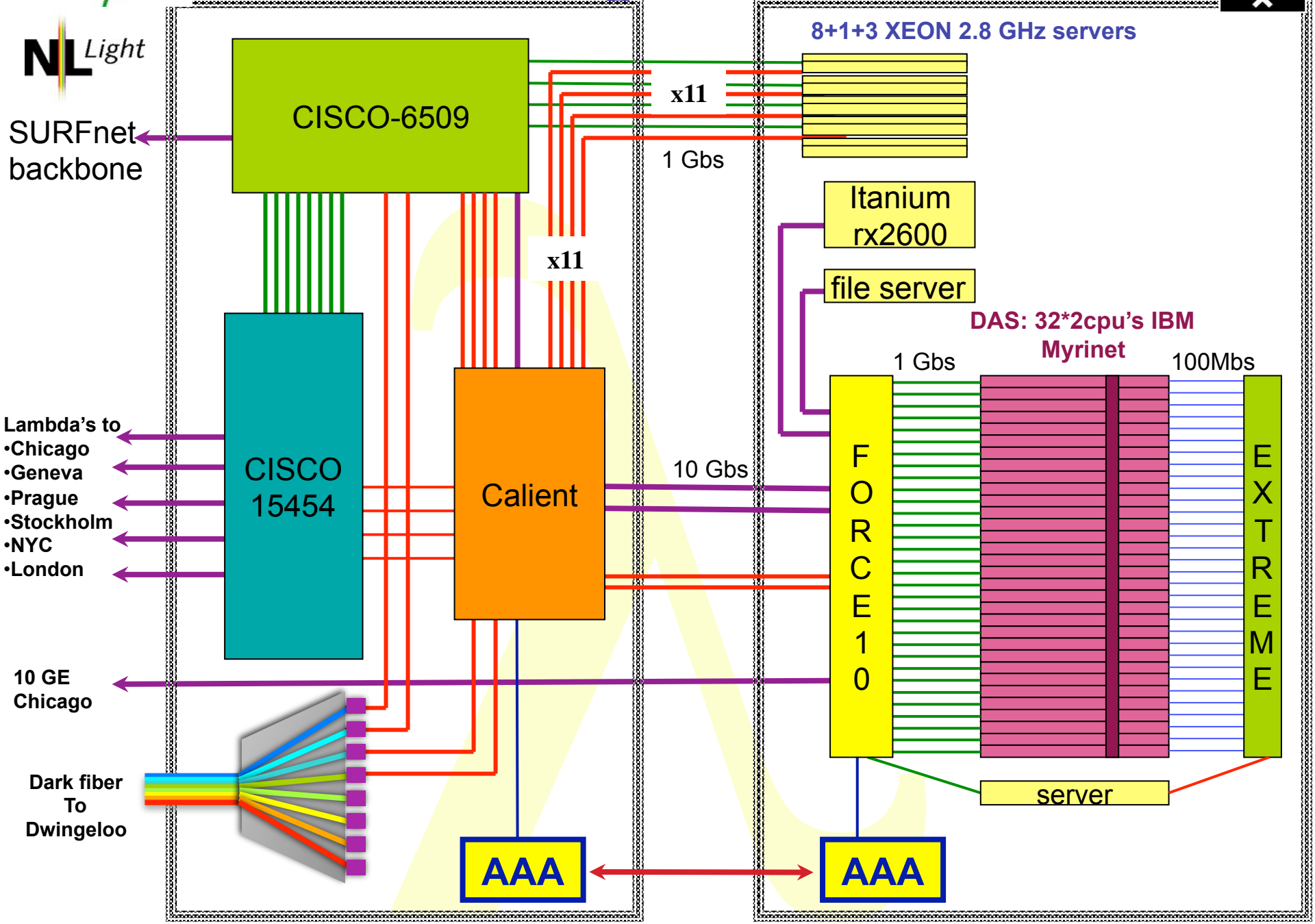


International lightpath network 1Q2004

GigaPort



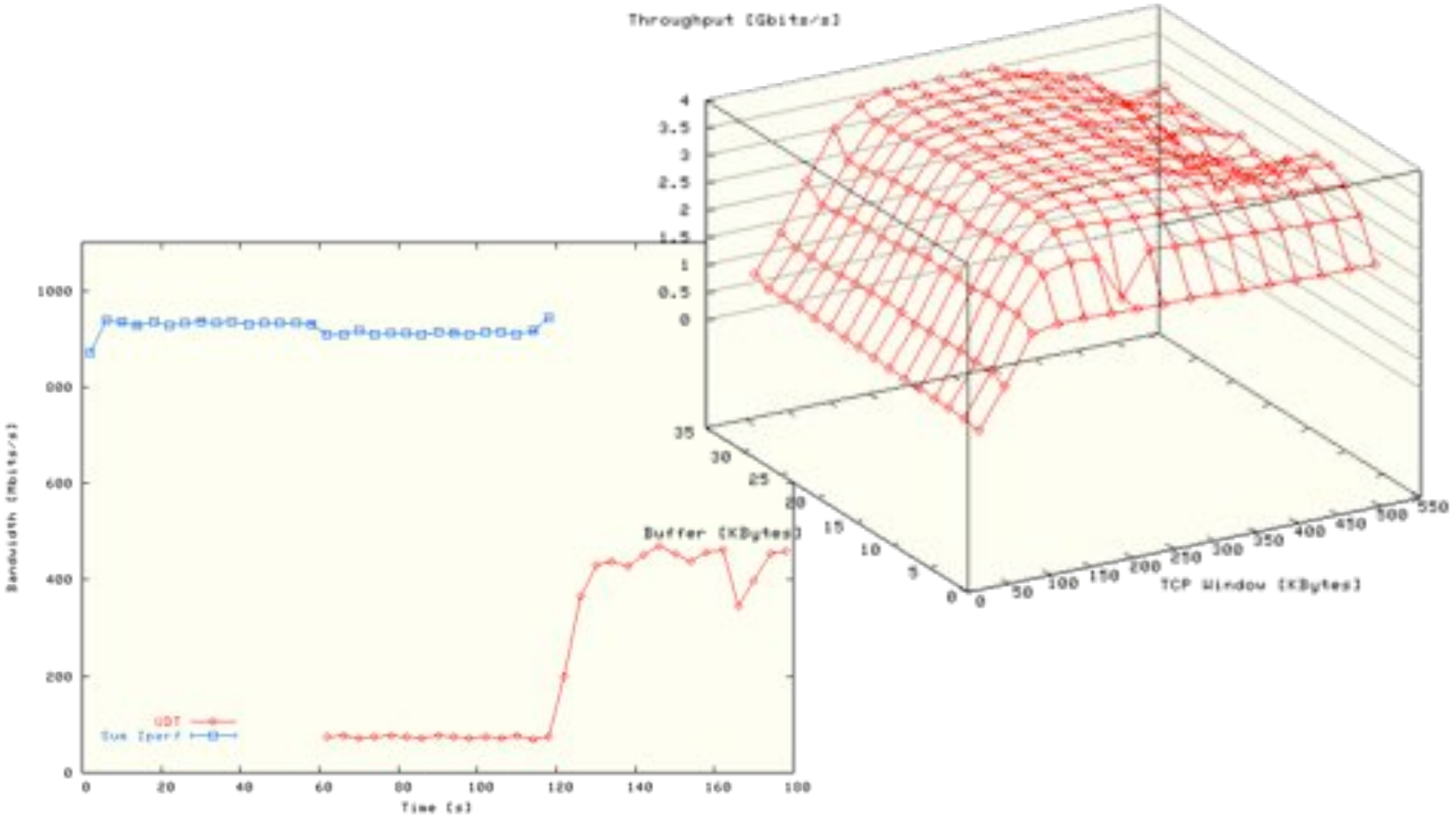
NetherLight <-> UvA



Research topics

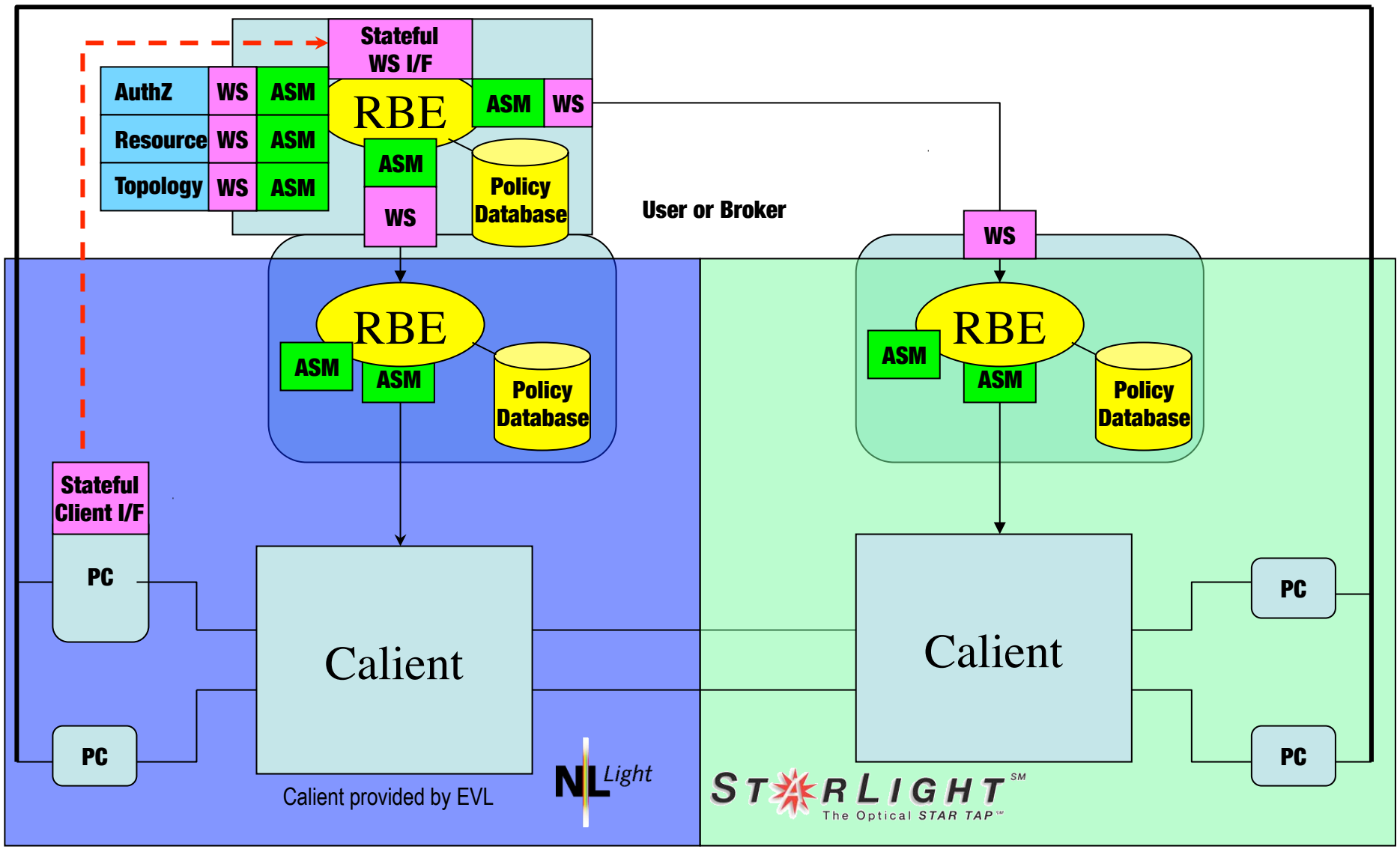
- Optical networking architectures and models for usage
- Transport protocols for massive amounts of data
- Authorization of complex resources in multiple domains
- Embedding in Grid environments

Example Measurements



UvA Universiteit van Amsterdam

AAA based demo at SC2003



Research on Networks (CdL)

GigaPort

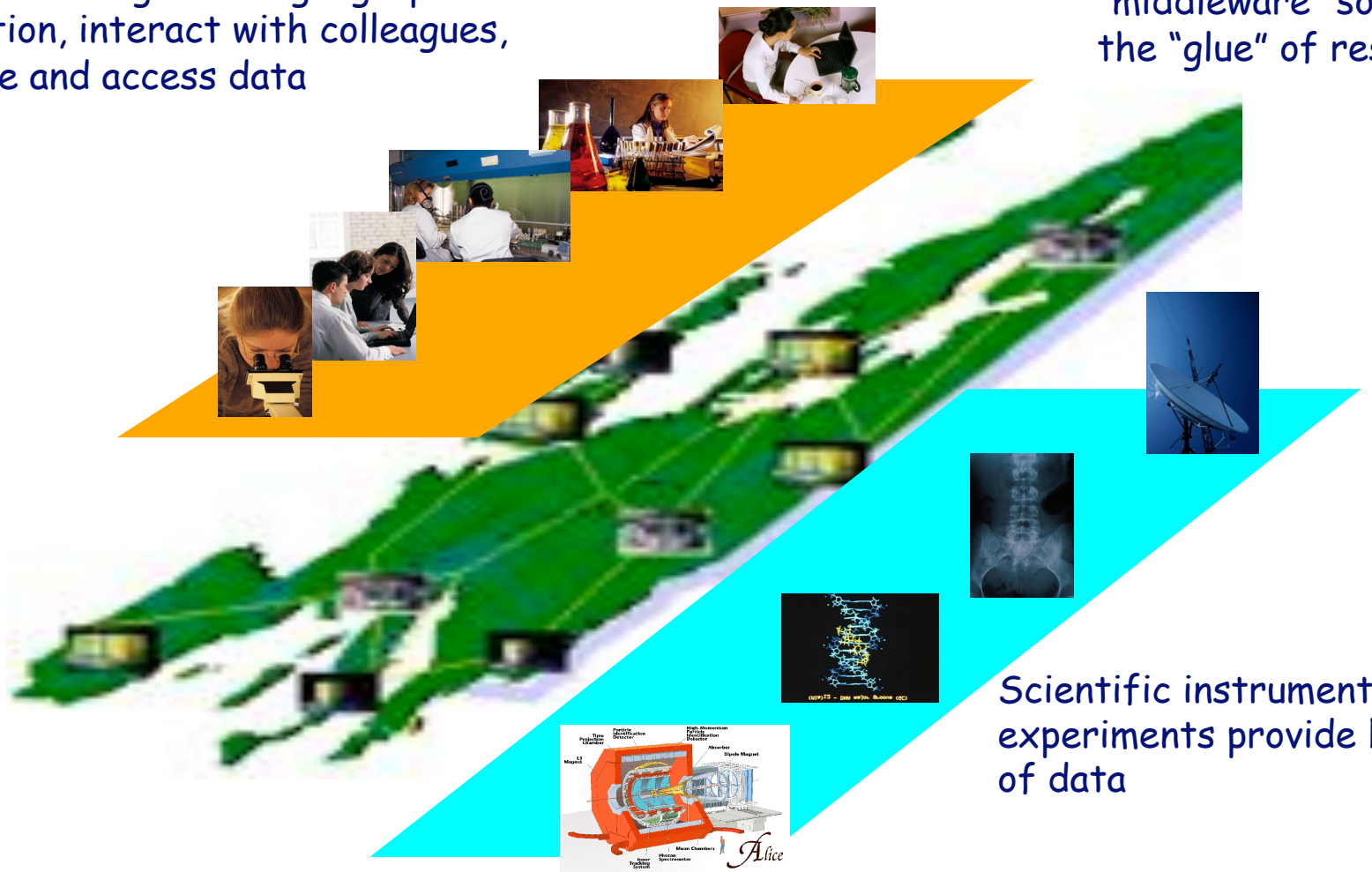
- **Optical Networking:** What innovation in architectural models, components, control and light path provisioning are needed to integrate dynamically configurable optical transport networks and traditional IP networks to a generic data transport platform that provides end-to-end IP connectivity as well 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

Grid - a Vision

Researchers perform their activities regardless geographical location, interact with colleagues, share and access data

The GRID: networked data processing centres and "middleware" software as the "glue" of resources.

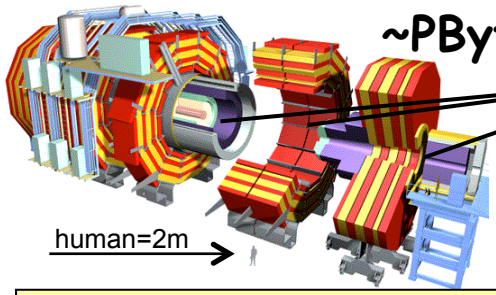


Scientific instruments and experiments provide huge amounts of data

Ref: EDG

LHC Data Grid Hierarchy

CMS as example, Atlas is similar



~PByte/sec

Online System

~100 MBytes/sec

event simulation

Tier 0 +1

HPSS

event reconstruction

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

HPSS

• • •

~0.6-2.5 Gbps

analysis

Tier2 Center

Tier 2

~0.6-2.5 Gbps

Tier 3

Institute ~0.25TIPS

Physics data cache

100 - 1000 Mbits/sec

Tier 4

Workstations

Courtesy Harvey Newman, CalTech and CERN

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.

Grid Projects Participation

- European Data Grid – finished 01/04/2004
  - NIKHEF: HEP integration, fabric management, core site operation
  - SARA: Grid Mass Storage, testbed site operation
- VL-E (started 01/04/2004, ends 01/04/2009)
  - UvA: PI, workflow, knowledge management, AAA, networking,
  - NIKHEF: Data Intensive Sciences, Scaling & Validation
  - SARA: Scaling & Validation
  - Many others in NL!
- EGEE (started 01/04/2004, ends 01/04/2006 + 2)
  - UvA: Security
  - SARA: Regional Operations Center
  - NIKHEF: Applications, Operations, Security
- LCG (started 2003, end indefinite)
  - NIKHEF/SARA is a core site (Tier-1 candidate)
- NEXTGRID, GRIDCOORD, COSSGRID, EU-INFRA, Expert group, GEANT-NG, CROSSGRID, etc., etc., etc.
- GGF, GridForum.nl)

Grid Resources in the Netherlands

- NIKHEF
  - Approx 280 CPUs, ranging from 800 MHz PIIIs to 2.4 GHz Xeon
  - Approx 3.5 terabyte grid-accessible storage
  - Various main services (data, VO, and workload management)
- SARA
  - Matrix grid cluster
    - 72 CPUs 3.06 GHz Xeon
    - 1 terabyte fast grid-enabled storage
  - TERAS supercomputer
    - Approx 12 terabyte grid-enabled storage
- NIKHEF/SARA/UVA/SURFnet cooperation towards Tier 1 LHC Grid site
  - Implies > 2k CPUs and > 1 petabyte of storage, connectivity

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

it²

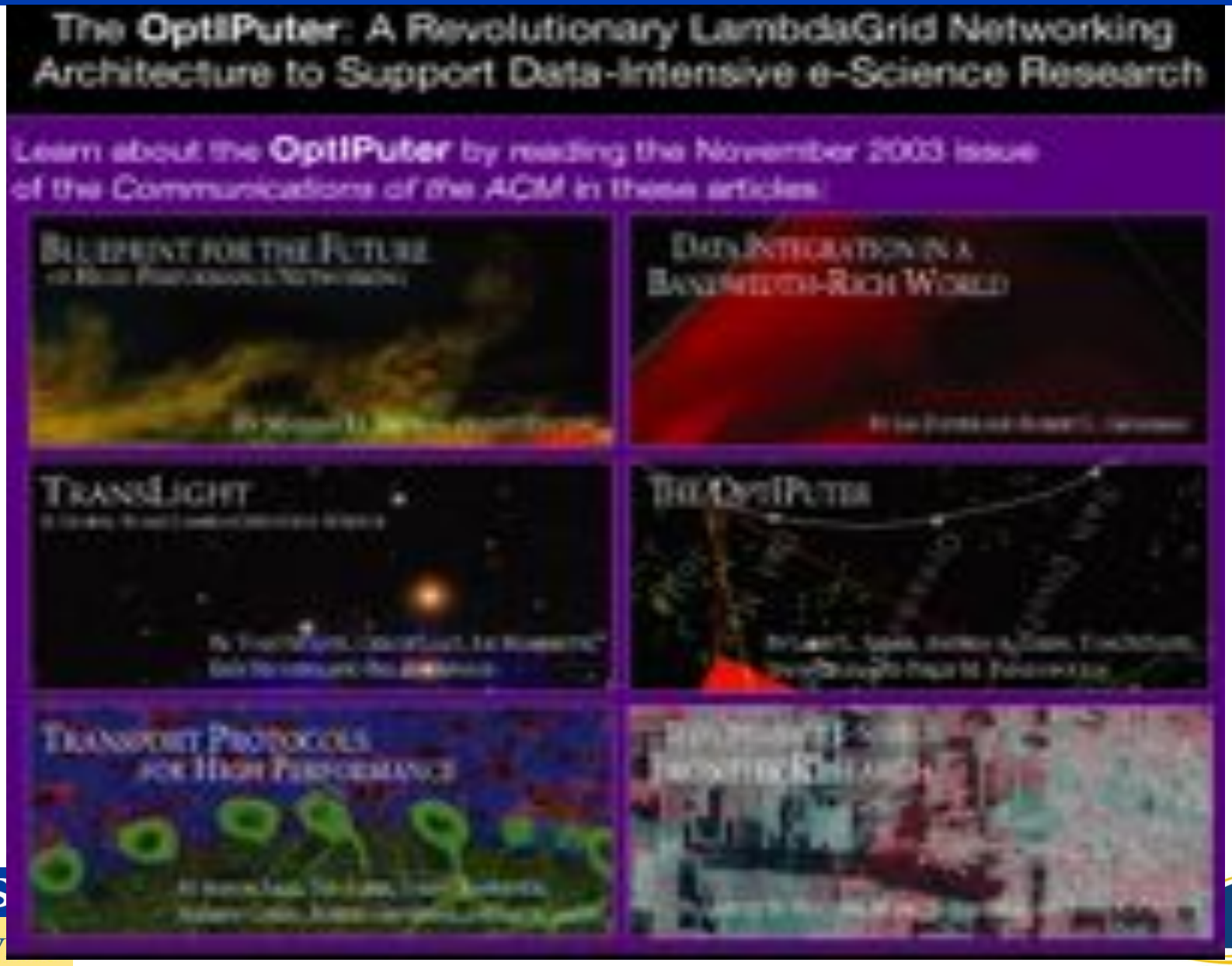
OptIPuter Overview

- Motivation from e-Science Distributed Cyberinfrastructure
- What are the Science Barriers We are Trying to Overcome?
  - Gigabyte Data Objects Need Interactive Visualization
  - Shared Internet Limits Speed of File Transfers
  - Inadequate Campus Grid Infrastructure
- Creating a Multi-Latency OptIPuter Laboratory
- System Software: From Grid to LambdaGrid
- Education and Outreach
- Project Management

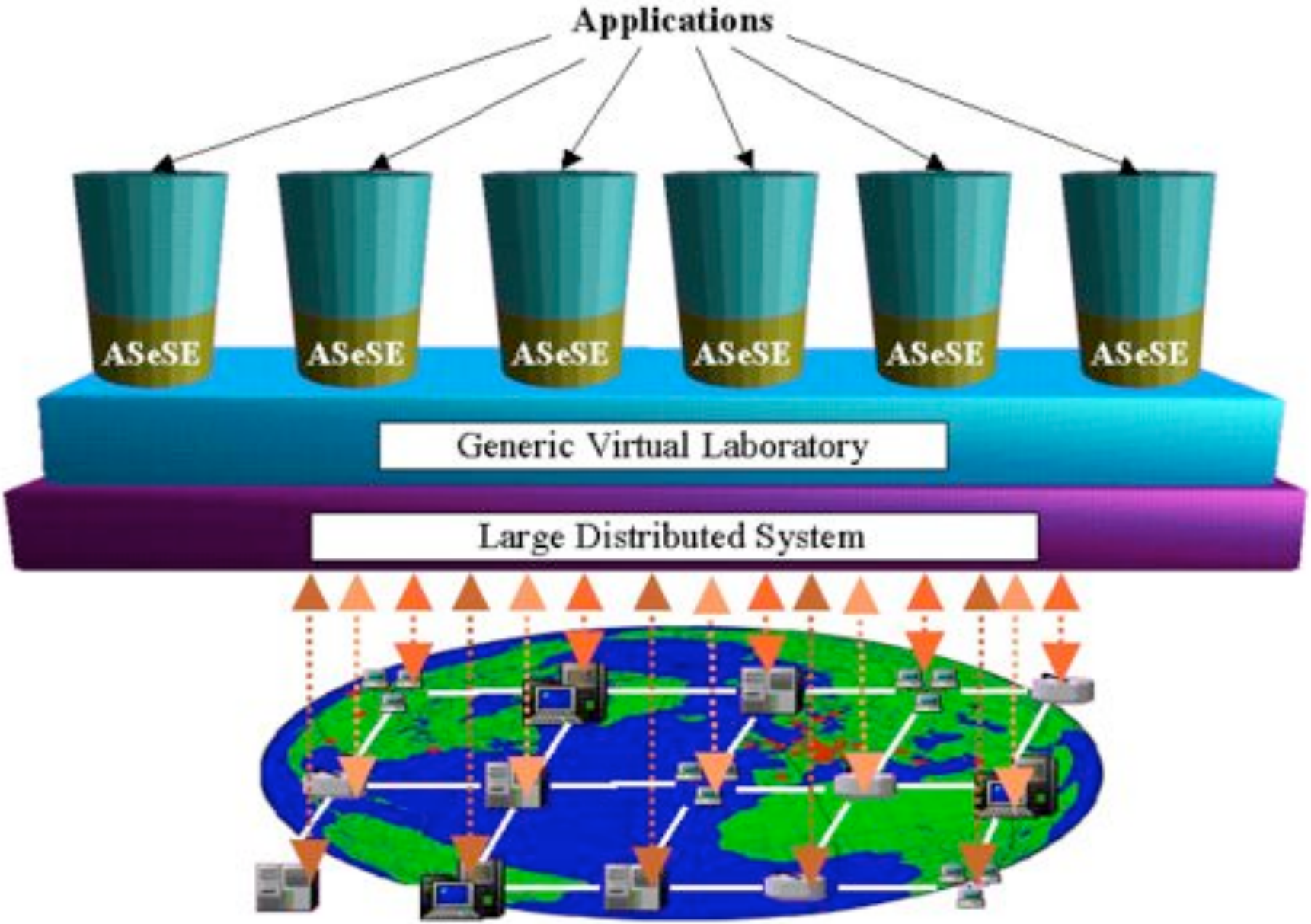
UCSD
UCIrvine

it²

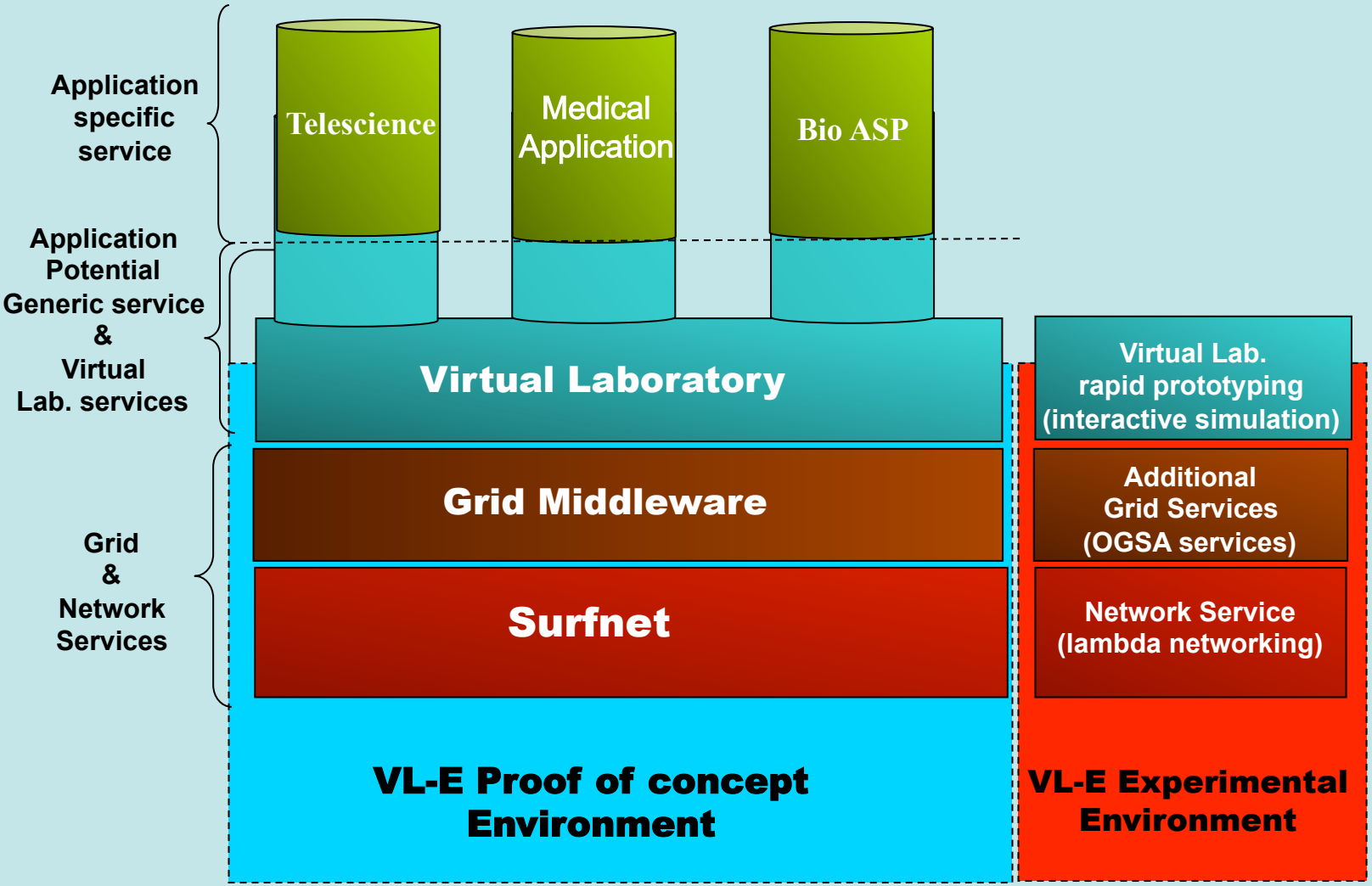
See Nov 2003 CACM
For Articles on OptIPuter Technologies



VLE middleware



Virtual Laboratory Amsterdam



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 = 1.6 TByte/s**

- **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 / 1.6 TByte/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.