CineGrid Networking

CG-2009

Cees de Laat

University of Amsterdam

Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

CineGrid Mission

To build an interdisciplinary community that is focused on the research, development, and demonstration of networked collaborative tools to enable the production, use and exchange of very-high-quality digital media over photonic networks.

http://www.cinegrid.org/

Keio/Calit2 Collaboration: Trans-Pacific 4K Teleconference

Like High-Def? Here Comes the Next Level

By JOHN MARKOFF
Published: September 26, 2005

The New York Times
ON THE WEB

Keio University
President Anzai

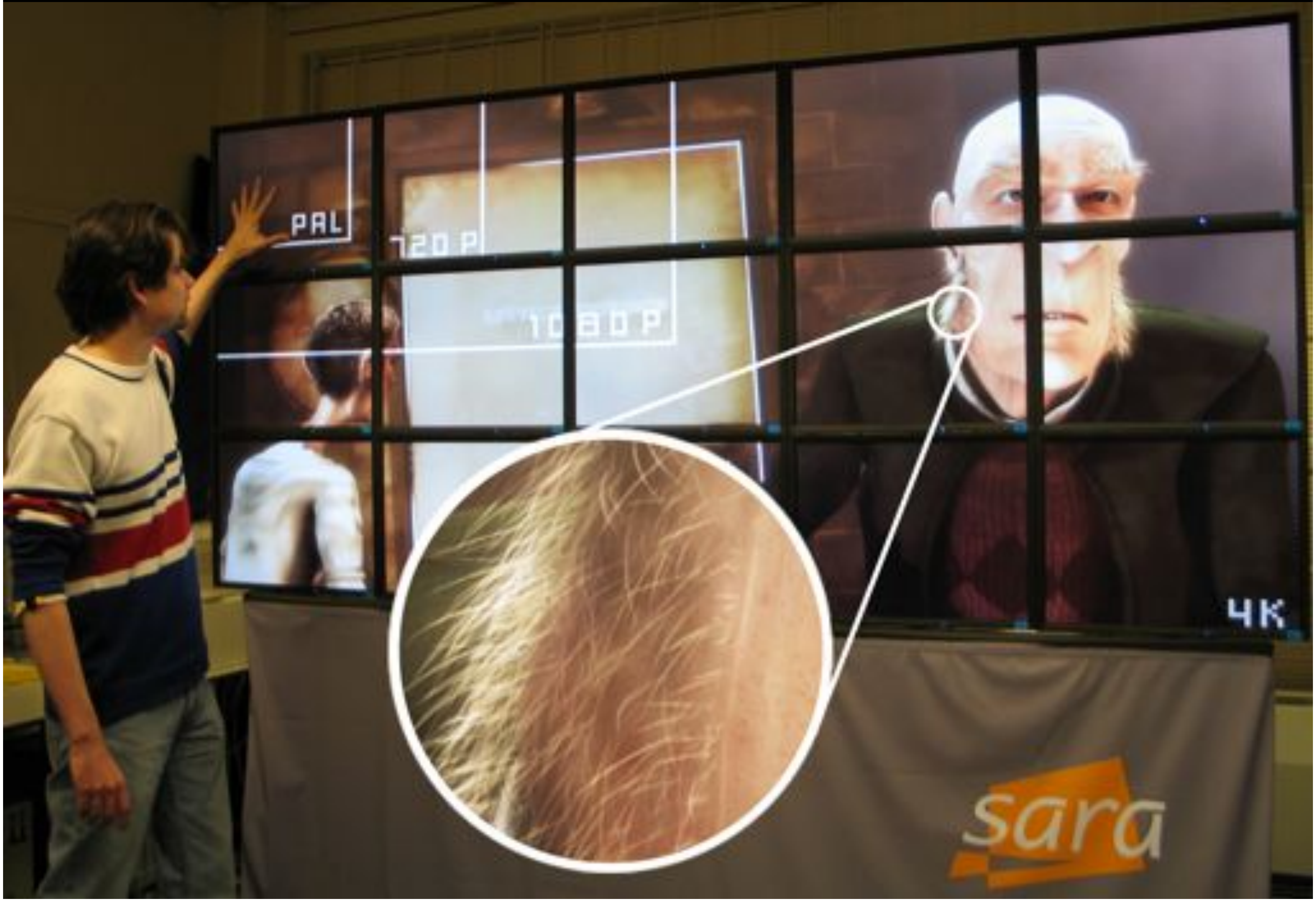
UCSD
Chancellor Fox

Used
1Gbps
Dedicated

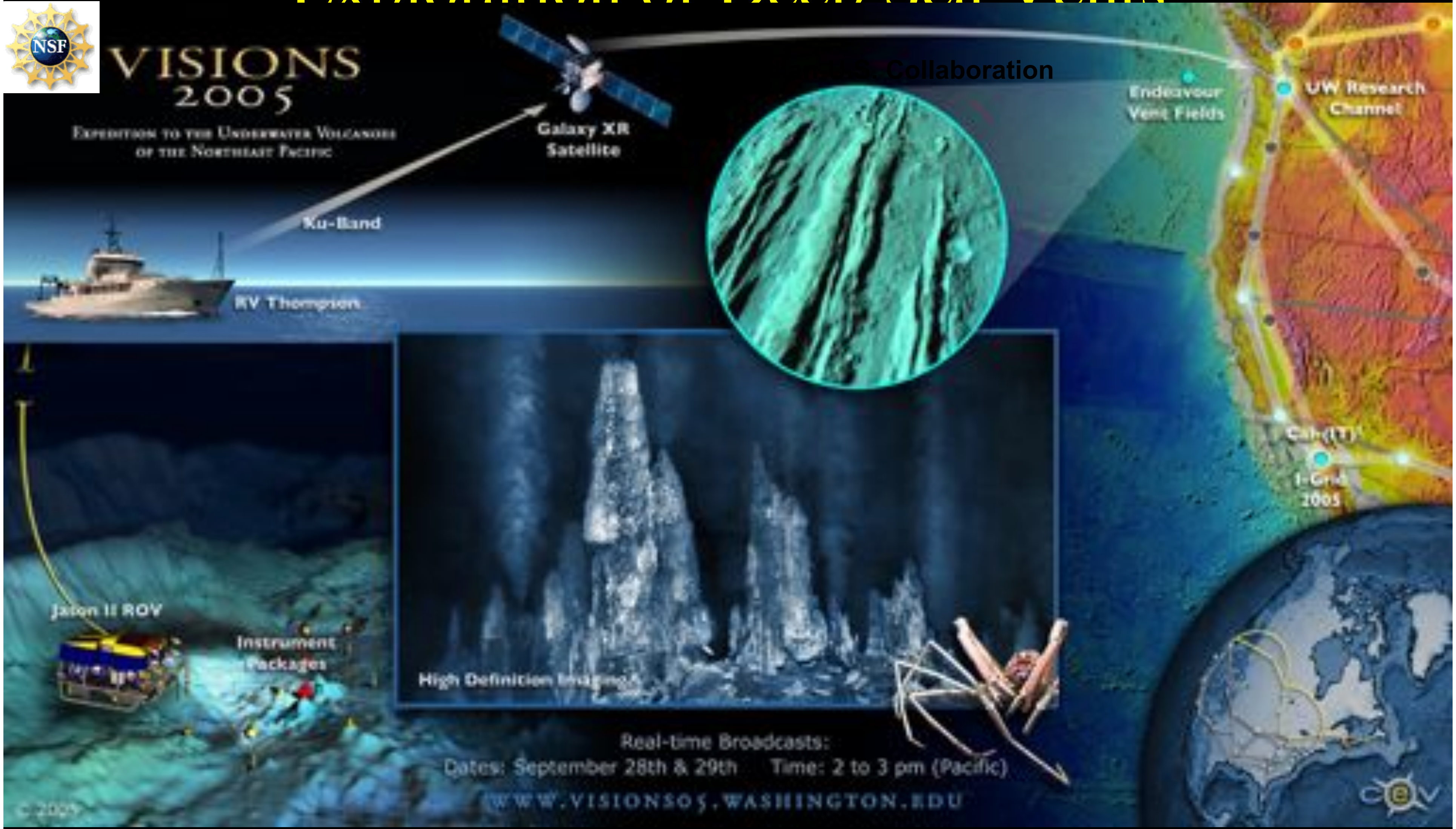
Sony
NTT
SGI

iGrid 2005

CineGrid@SARA



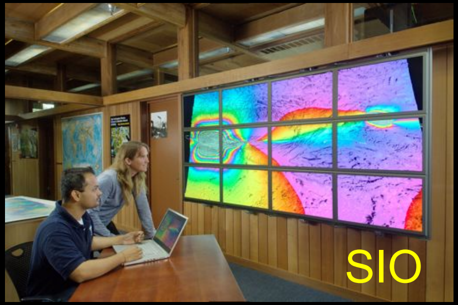
First Remote Interactive High Definition Video Exploration of Deep Sea Vents



Collaboration

Source John Delaney & Deborah Kelley, UWash

US and International OptIPortal Sites

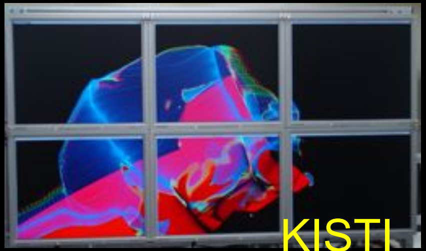

SIO

NCMIR

USGS EDC

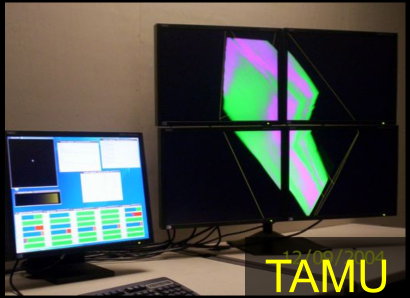
NCSA & TRECC

SARA


KISTI

AIST

RINCON & Nortel


TAMU

UCI

UIC

CALIT2

The "Dead Cat" demo

SC2004 & iGrid2005

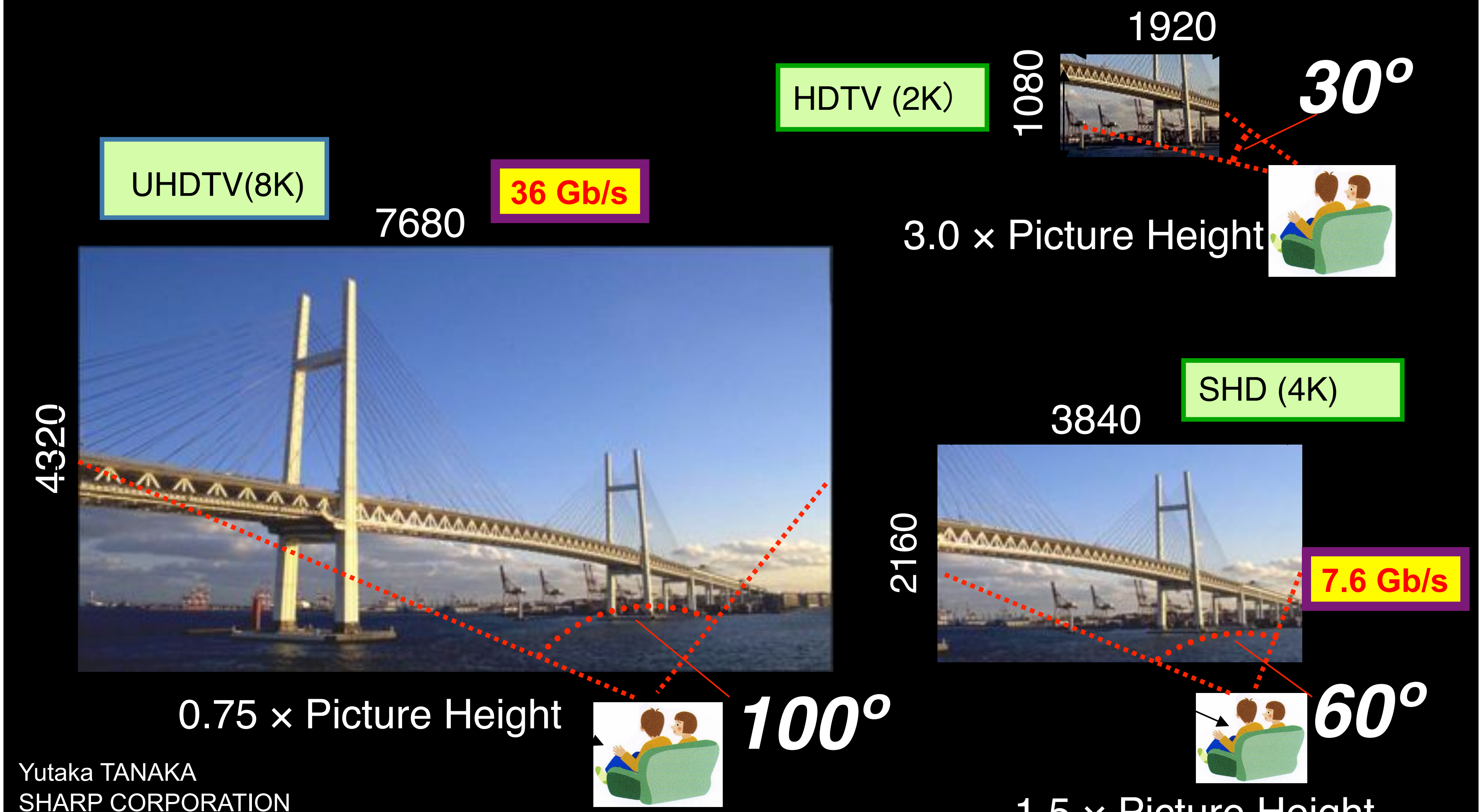
SC2004,
Pittsburgh,
Nov. 6 to 12, 2004
iGrid2005,
San Diego,
sept. 2005

Produced by:
Michael Scarpa
Robert Belleman
Peter Sloot

Many thanks to:
AMC
SARA
GigaPort
UvA/AIR
Silicon Graphics, Inc.
Zoölogisch Museum

Why is more resolution is better?

1. More Resolution Allows Closer Viewing of Larger Image
2. Closer Viewing of Larger Image Increases Viewing Angle
3. Increased Viewing Angle Produces Stronger Emotional Response



HDTV (2K)

1920

1080

30º

3.0 × Picture Height

UHDTV(8K)

36 Gb/s

7680

4320

0.75 × Picture Height

100º

SHD (4K)

3840

2160

7.6 Gb/s

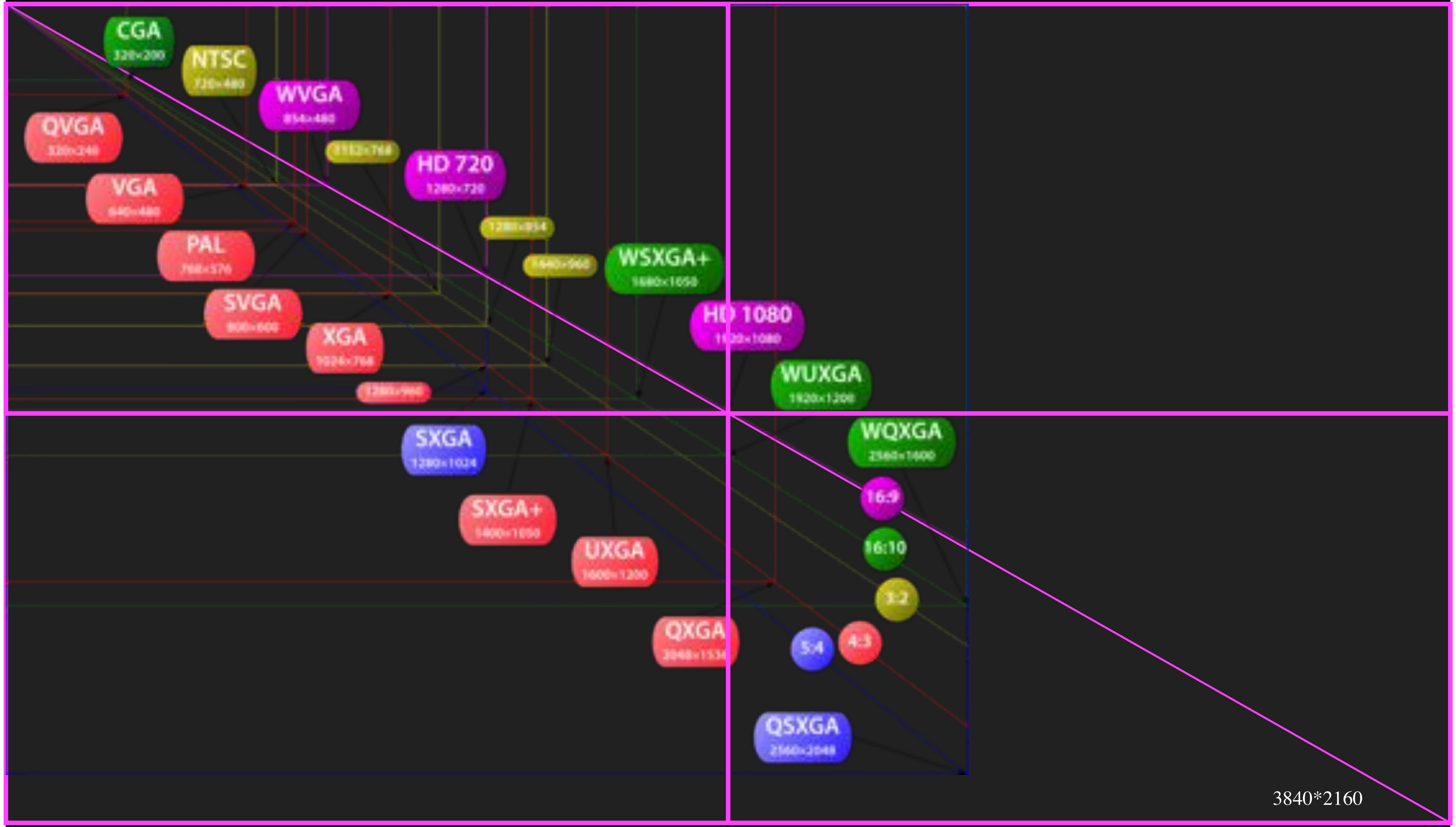
60º

1.5 × Picture Height

Yutaka TANAKA
SHARP CORPORATION
Advanced Image Research Laboratories

Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A



Formats - Numbers - Bits

Format - Numbers - Bits (examples!)

Format	X	Y	Rate	Color	Frame	Frame	Flow	Stream
			/s	bits/pix	pix	MByte	MByt/s	Gbit/s
720p	1280	720	60	24	921.600	2.8	170	1.3
1080p	1920	1080	30	24	2.073.600	6.2	190	1.5
2k (24)	2048	1080	24	36	2.211.840	10	240	1.2
2k (48)	2048	1080	48	36	2.211.840	10	480	2.4
SHD	3840	2160	30	24	8.294.400	25	750	6.0
4k	4096	2160	24	36	8.847.360	~ 40	960	7.6
8k	7680	4320	24	36	33.177.600	~ 150	4478	36

Note: this is excluding sound!

Note: these are raw uncompressed data rates ex overhead!

Formats - Numbers - Bits

- Formats:

– uncompressed unreadable (UMF)	3/4 GBytesec
– compressed unreadable (jpeg2000)	300 - 700 Mbit/s
– uncompressed readable (eg TIFF)	1.2 GB/s, 4.3 TB/h
– compressed readable (eg DXT)	300 - 800 Mbit/s

- Do not compress away the science!
- Storage
  - Holland festival taking uncompressed about 12 TByte

Number, numbers and more numbers!

- **Digital Motion Picture for Audio Post-Production**
  - 1 TV Episode Dubbing Reference 1 GB
  - 1 Theatrical 5.1 Final Mix 8 GB
  - 1 Theatrical Feature Dubbing reference 30 GB
- **Digital Motion Picture Acquisition**
  - 6:1 up to 20:1 shooting ratios
  - 4k @ 24 FPS @ 10bit/color: ~48MB/Frame uncompressed
  - ~8TB for Finished 2 Hr Feature
- **Digital Dailies**
  - HD compressed MPEG-2 @ 25Mb/s
  - Data Size: ~22GB for 2 Hours
- **Digital Post-production and Visual Effects**
  - Terabytes, Gigabytes, Megabytes To Select Sites Depending on Project
- **Digital Motion Picture Distribution**
  - Film Printing in Regions
    - Features ~8TB
    - Trailers ~200GB
  - Digital Cinema to Theatres
    - Features ~200 - 300GB DCP
    - Trailers ~2 - 4GB DCP
- Online Download
  - Features ~1.3GB
  - TV Shows ~600MB

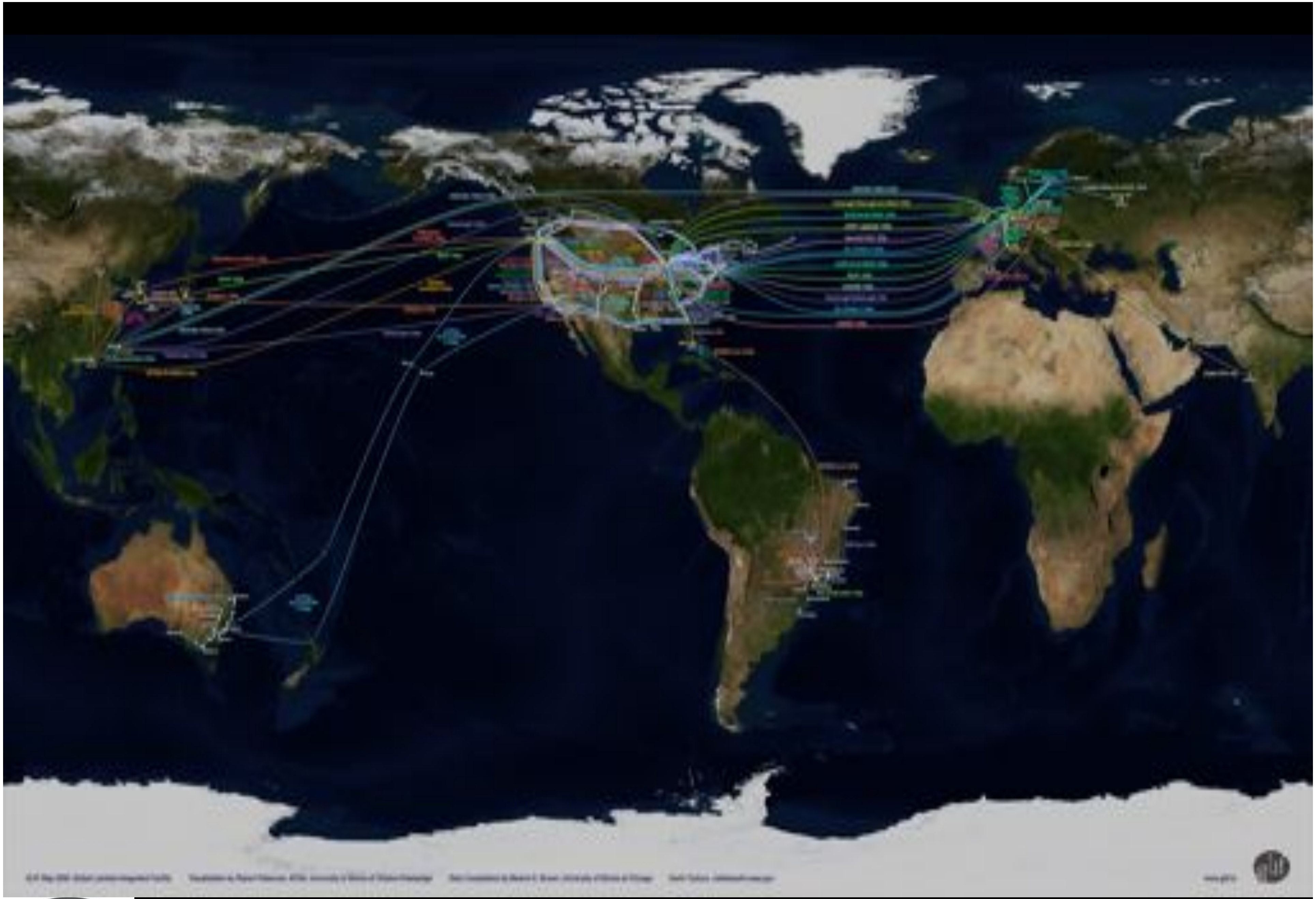
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

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**



GLIF 2008

Visualization courtesy of Bob Patterson, NCSA
Data collection by Maxine Brown.

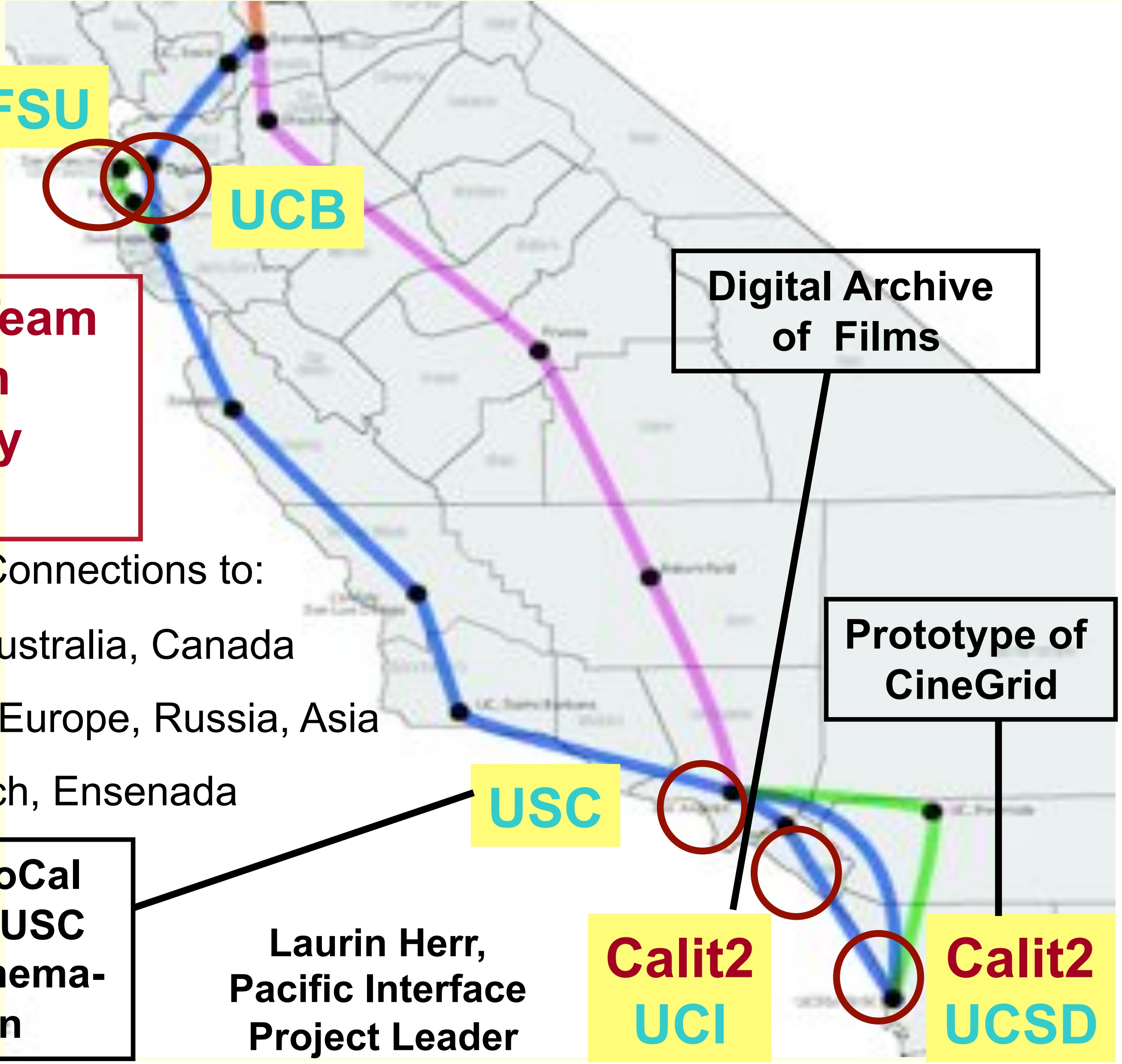
Calit2 is Partnering with CENIC to Connect California Industries and Researchers Into CineGrid

Partnering with SFSU's Institute for Next Generation Internet

Calit2's CineGrid Team is Working with Cinema Industry in LA and SF

In addition, 1Gb and 10Gb Connections to:

- Seattle then to Asia, Australia, Canada
- Chicago, Amsterdam, Europe, Russia, Asia
- Tijuana, Rosarita Beach, Ensenada



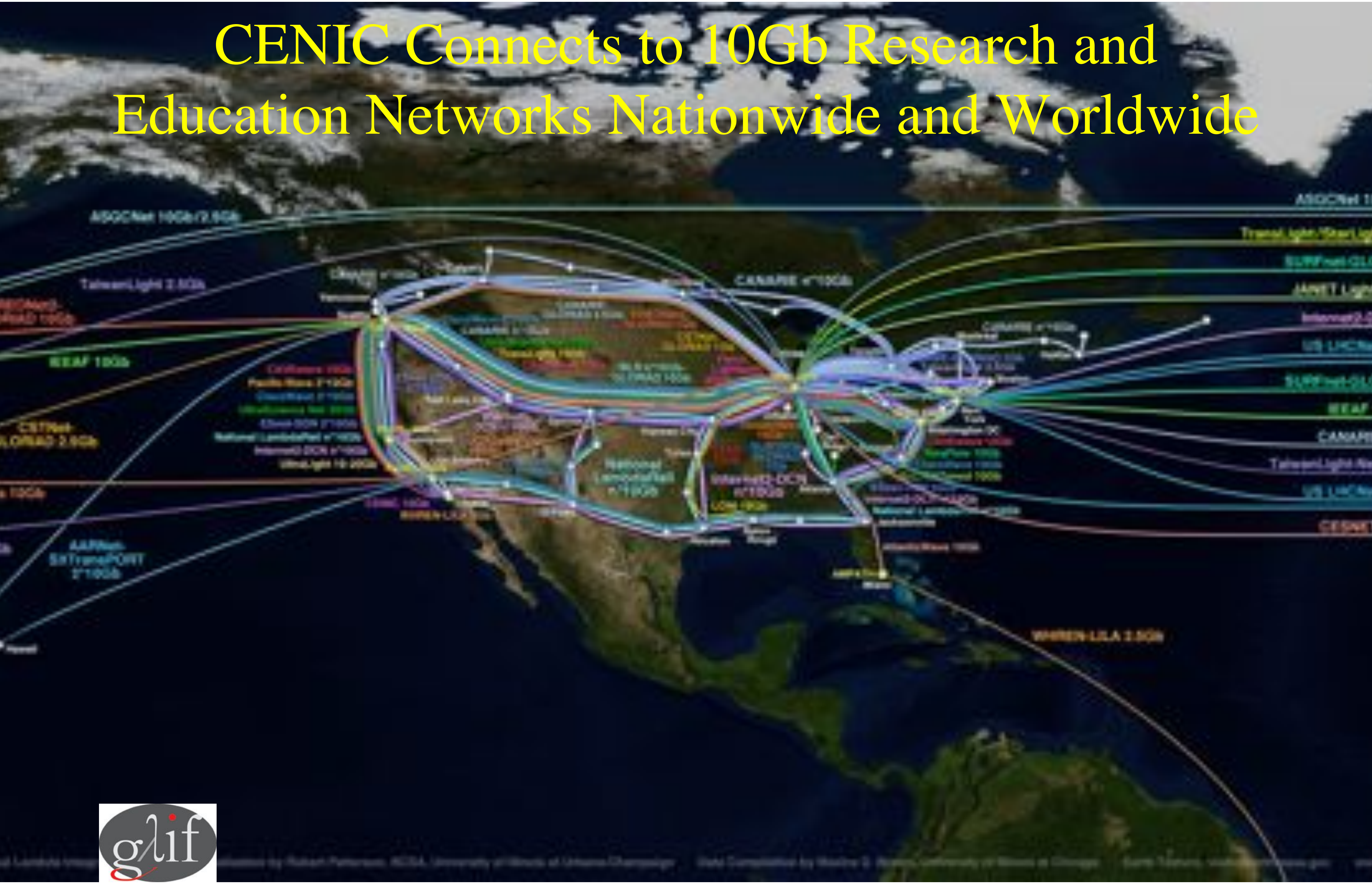
Digital Archive of Films

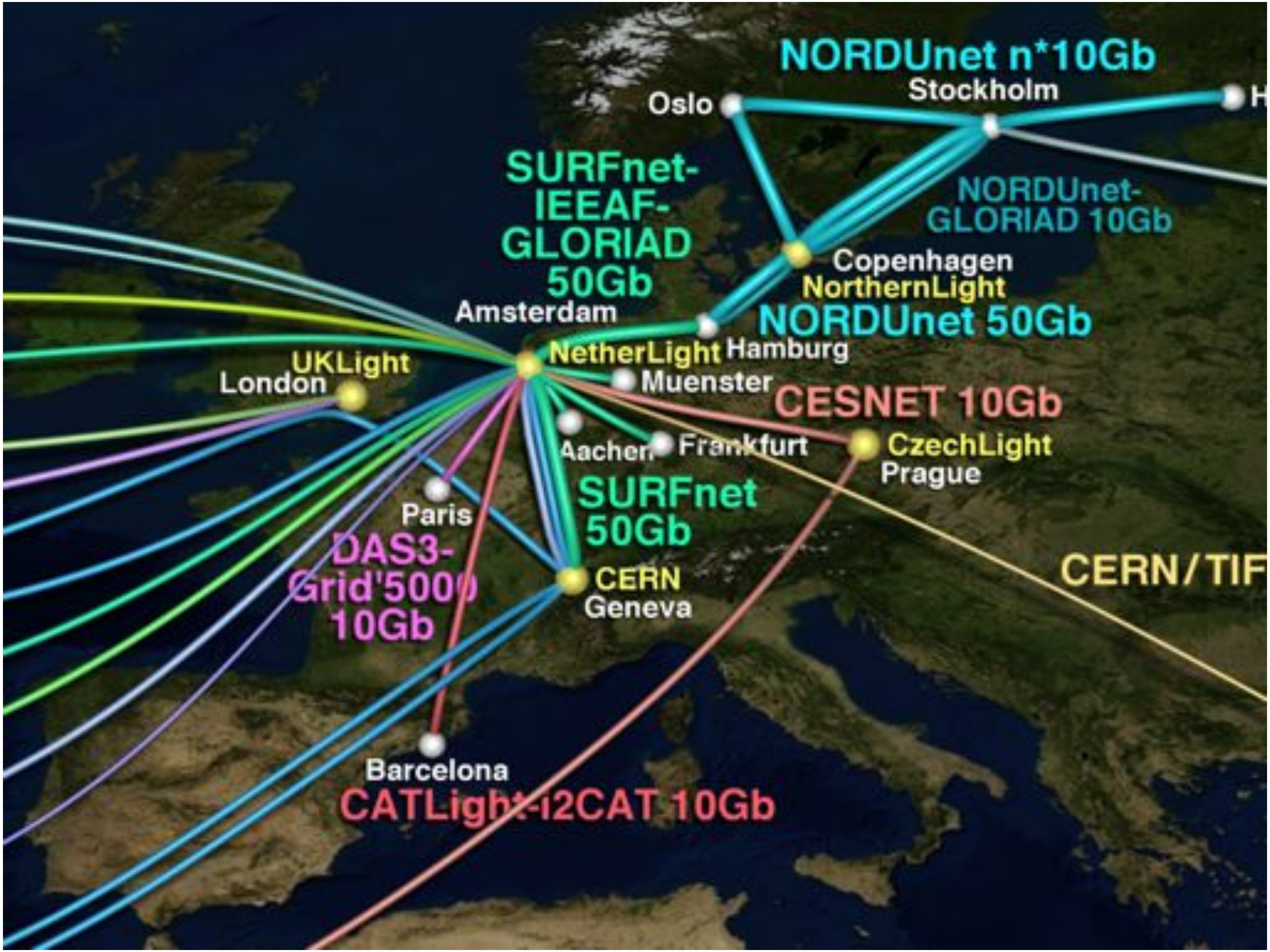
Prototype of CineGrid

Extending SoCal OptIPuter to USC School of Cinema-Television

Laurin Herr,
Pacific Interface
Project Leader

CENIC Connects to 10Gb Research and Education Networks Nationwide and Worldwide



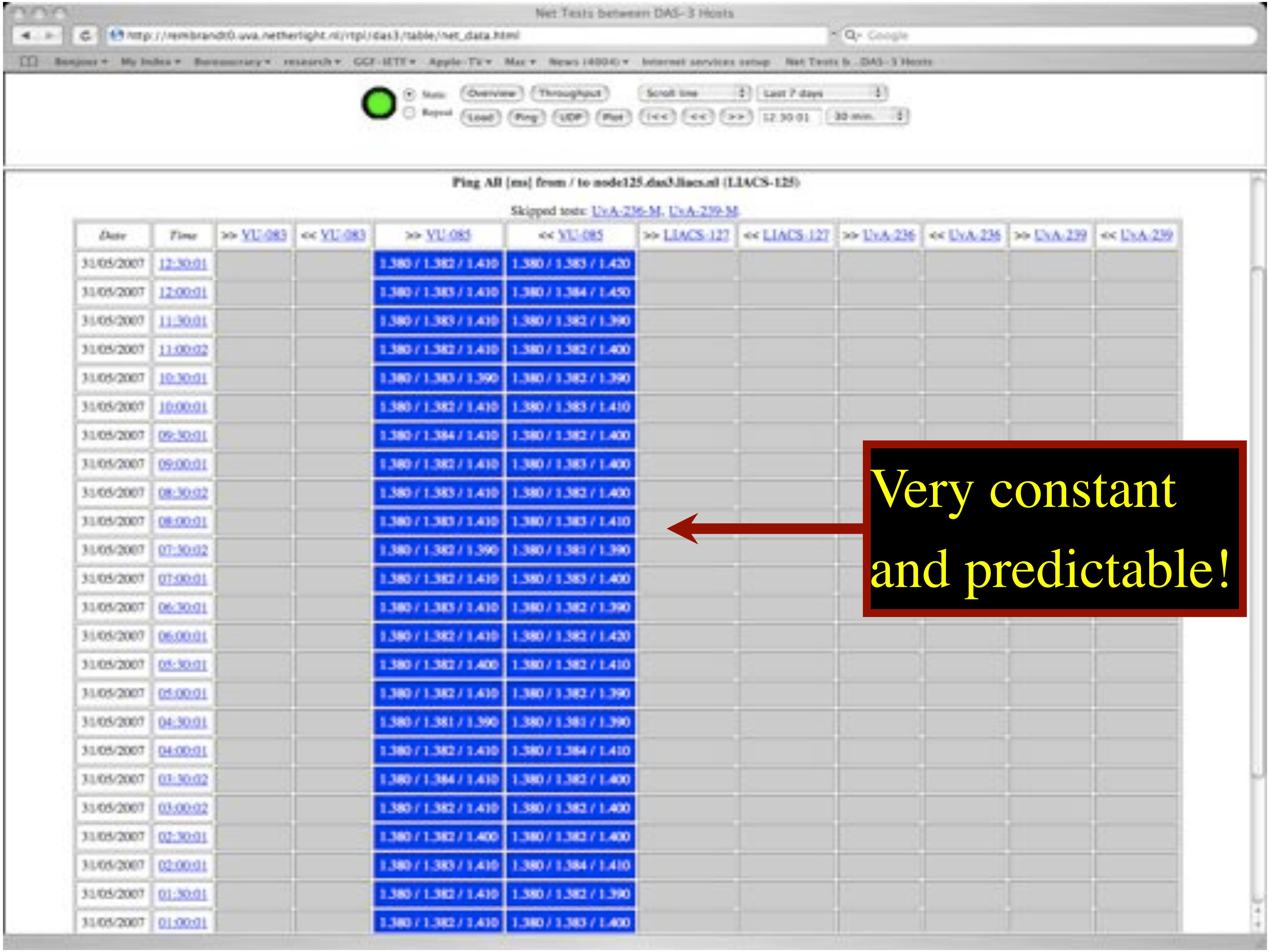

NORDUnet n*10Gb
Stockholm
Oslo
H
SURFnet-
IEEAF-
GLORIAD
50Gb
NORDUnet-
GLORIAD 10Gb
Copenhagen
NorthernLight
Amsterdam
NORDUnet 50Gb
NetherLight
Hamburg
UKLight
London
Muenster
CESNET 10Gb
Aachen
Frankfurt
CzechLight
Prague
Paris
SURFnet
50Gb
DAS3-
Grid'5000
10Gb
CERN
Geneva
CERN/TIF
Barcelona
CATLight-i2CAT 10Gb

Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

What is a LightPath

- A LightPath is a circuit like connection that connects end systems to each other. This uses usually the same infrastructure as the Internet, but a LightPath gets dedicated resources next to Internet.
- A LightPath can be a combination of:
  - A color in a fiber (Lambda)
  - Sonet/sdh circuit in a sonet infrastructure
  - Vlans and dedicated ports in an ethernet switch
  - Etc.
- Aim is to get predictable and knowable connection characteristics
- Let us look at examples setups used recently!



Ping All [ms] from / to node125.das3.liacs.nl (LIACS-125)

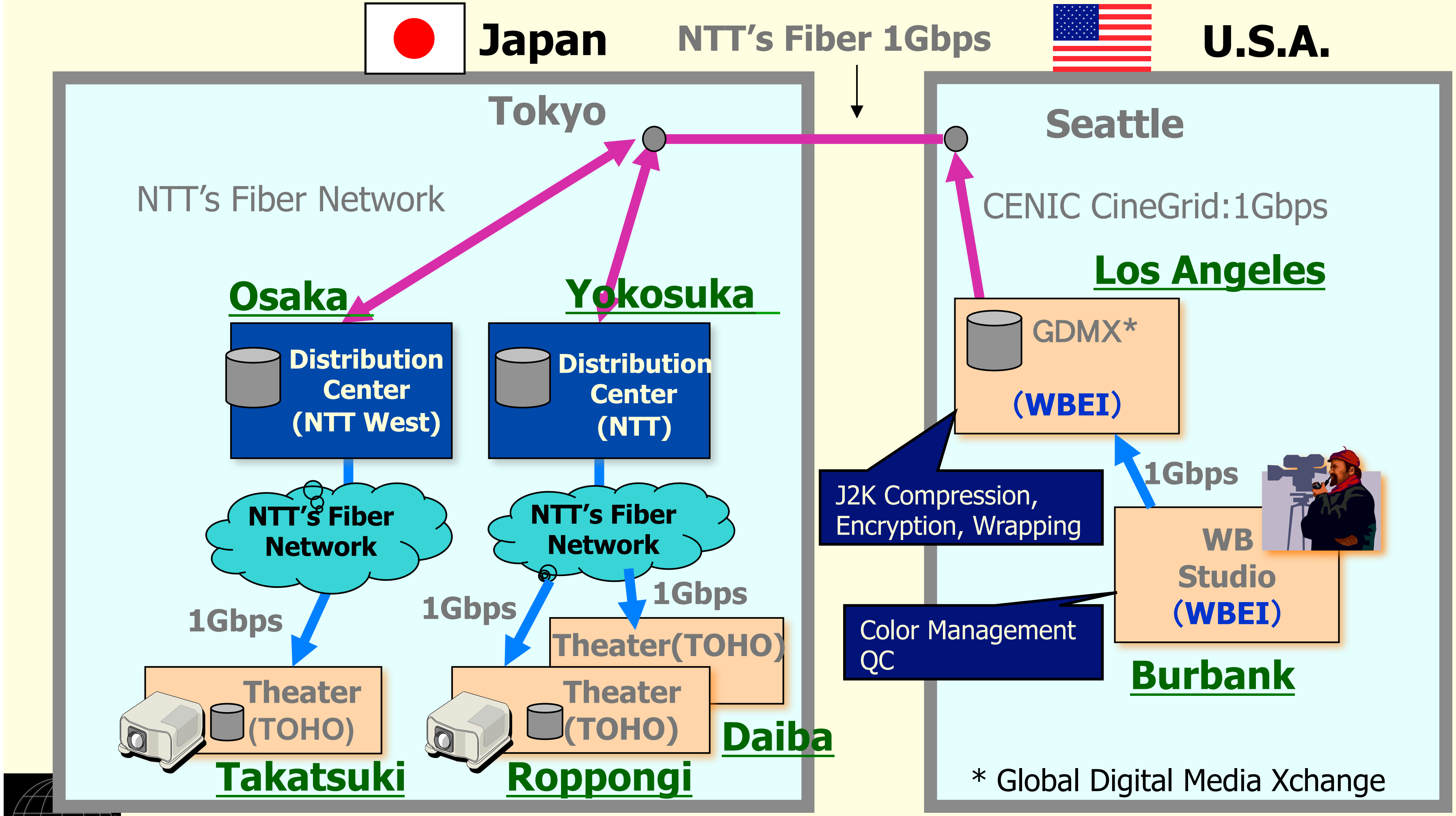
Skipped tests: UvA-236-M, UvA-239-M.

Date	Time	>> VU-083	<< VU-083	>> VU-085	<< VU-085	>> LIACS-127	<< LIACS-127	>> UvA-236	<< UvA-236	>> UvA-239	<< UvA-239
31/05/2007	12:30:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.420						
31/05/2007	12:00:01			1.380 / 1.383 / 1.410	1.380 / 1.384 / 1.450						
31/05/2007	11:30:01			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	11:00:02			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	10:30:01			1.380 / 1.383 / 1.390	1.380 / 1.382 / 1.390						
31/05/2007	10:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.410						
31/05/2007	09:30:01			1.380 / 1.384 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	09:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						
31/05/2007	08:30:02			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	08:00:01			1.380 / 1.383 / 1.410	1.380 / 1.383 / 1.410						
31/05/2007	07:30:02			1.380 / 1.382 / 1.390	1.380 / 1.381 / 1.390						
31/05/2007	07:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						
31/05/2007	06:30:01			1.380 / 1.383 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	06:00:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.420						
31/05/2007	05:30:01			1.380 / 1.382 / 1.400	1.380 / 1.382 / 1.410						
31/05/2007	05:00:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	04:30:01			1.380 / 1.381 / 1.390	1.380 / 1.381 / 1.390						
31/05/2007	04:00:01			1.380 / 1.382 / 1.410	1.380 / 1.384 / 1.410						
31/05/2007	03:30:02			1.380 / 1.384 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	03:00:02			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.400						
31/05/2007	02:30:01			1.380 / 1.382 / 1.400	1.380 / 1.382 / 1.400						
31/05/2007	02:00:01			1.380 / 1.383 / 1.410	1.380 / 1.384 / 1.410						
31/05/2007	01:30:01			1.380 / 1.382 / 1.410	1.380 / 1.382 / 1.390						
31/05/2007	01:00:01			1.380 / 1.382 / 1.410	1.380 / 1.383 / 1.400						

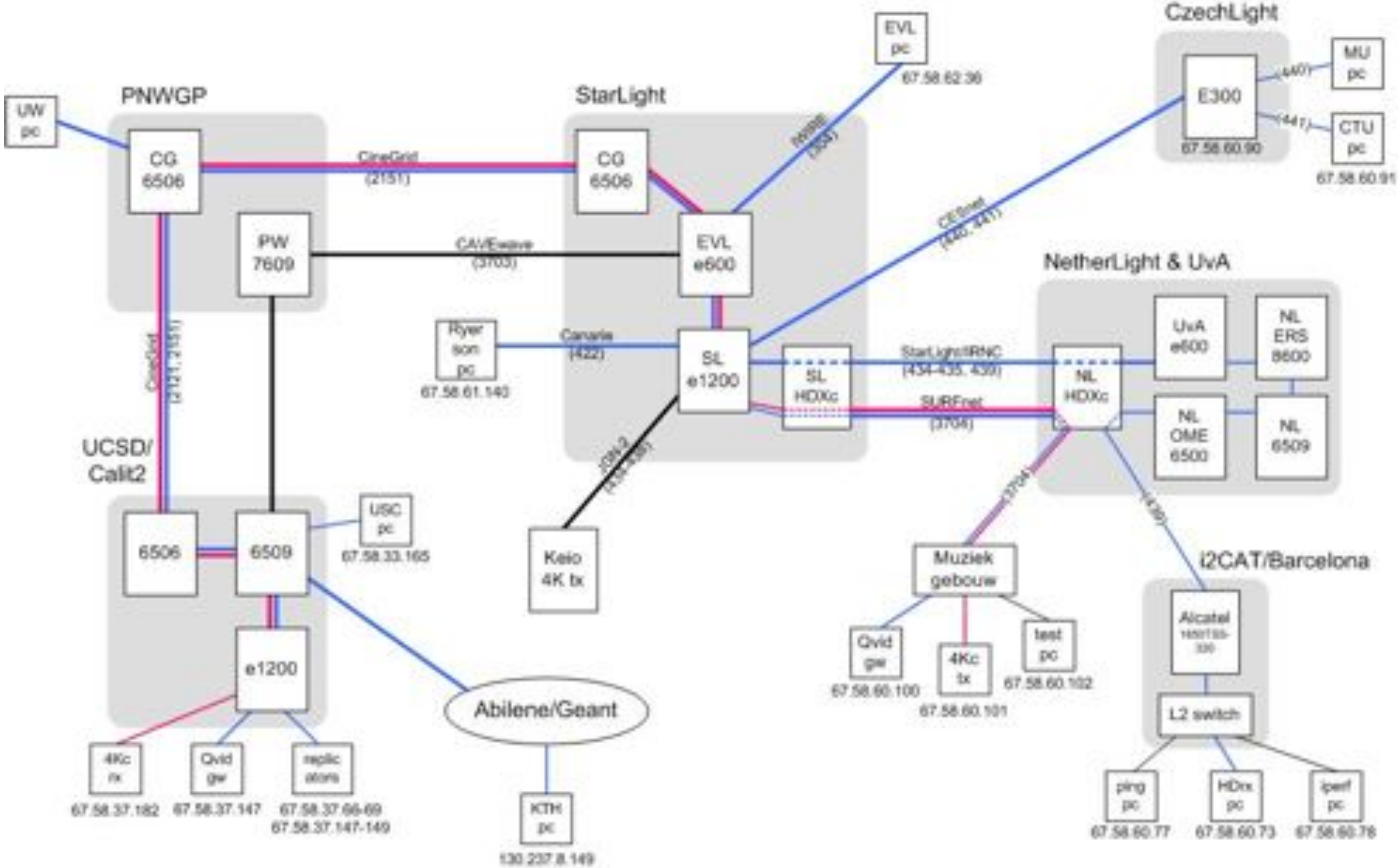
Very constant and predictable!

Network for “4K Pure Cinema” Trial

DCP is directly transferred from GDMX in LA to distribution centers in Japan via fiber network. Within Japan, DCP is distributed from the distribution centers to TOHO theaters. Key is distributed from Osaka center, based on the contract between WB Japan and TOHO cinemas.



* Global Digital Media Xchange

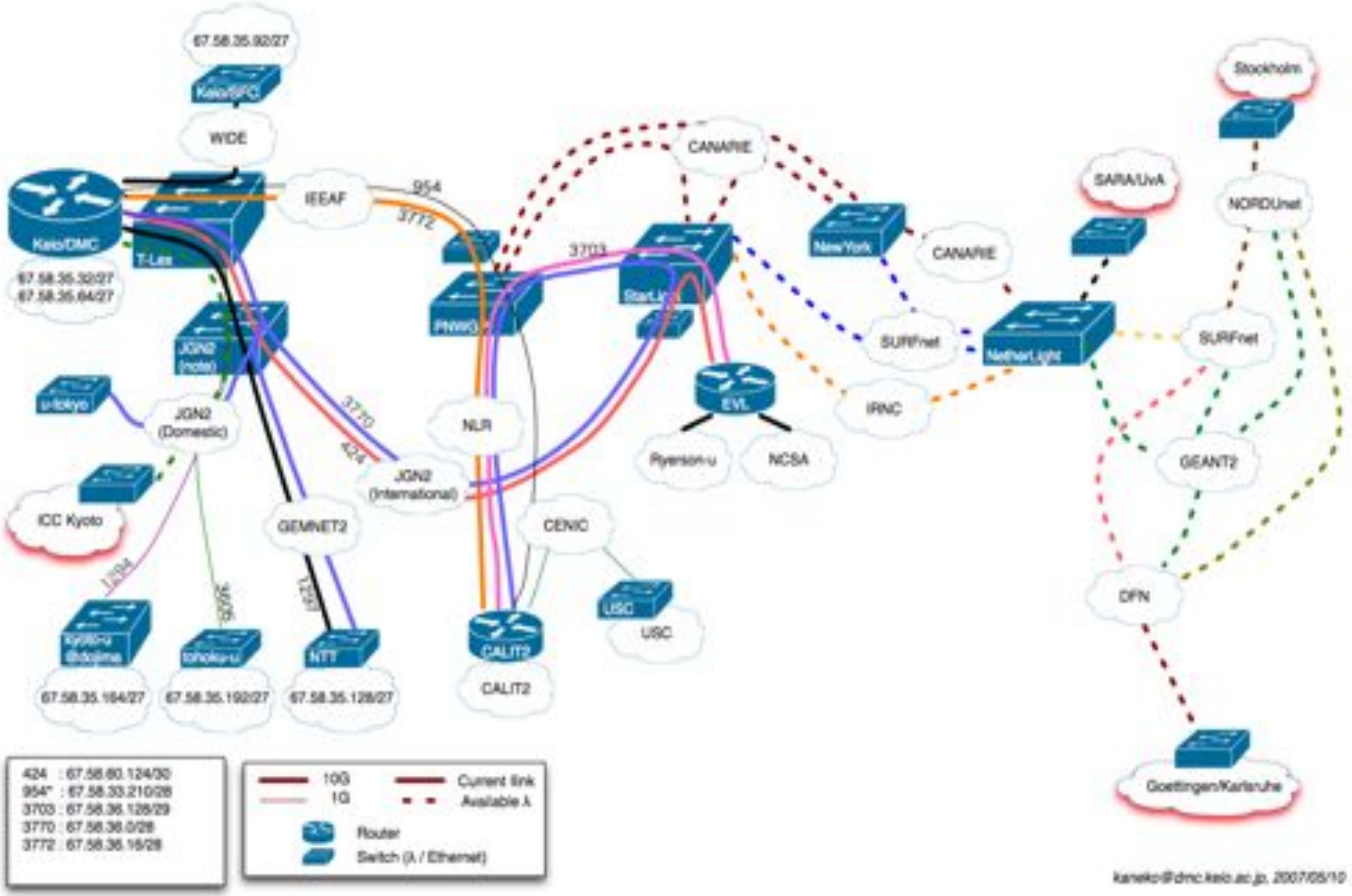


🟥 AMS -> Calit2
4K compressed

🟦 AMS -> Calit2 -> sites
HD compressed

Holland Festival
CineGrid 2007
19-21 June 2007
Drawing by Alan Verlo, et al.

Current Links & Available Links for Kyoto Prize Events



424 : 67.58.60.124/30
954* : 67.58.33.210/28
3703 : 67.58.36.128/29
3770 : 67.58.36.0/28
3772 : 67.58.36.16/28

kaneko@dmc.keio.ac.jp, 2007/05/10

Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

Internet Transport Protocols

- IP = Internet Protocol
  - Connectionless packet transport service
  - Datagrams of max 64 kByte
  - Can be fragmented down the way
  - Packets can get lost, duplicated or out of order!
- TCP/IP = Transmission Control Protocol
  - Reliable byte-stream over potentially unreliable packet service
  - Connection oriented, exactly once and in order, end to end duplex
- UDP = User Datagram Protocol
  - Packet service up to 64 kByte
  - Connectionless, unidirectional, L2 switches may start flooding
  - Unreliable delivery, can get out of order, duplicated, lost

Flow control vs Congestion control

- Flow control
  - To prevent a fast sender overflowing a slow receiver
  - Receiver signals sender so it can adapt
- Congestion control
  - Traffic jams in the Internet: packets may get lost
  - For TCP protocol control loops via ack’s and ICMP packets
  - TCP is friendly protocol, can adapt but performance usually takes severe hit
  - RTT is reaction and recovery time

Windows and buffering for reliable protocols

- Round Trip Time (rtt) is time it takes to send a shortest message and get the answer back (unix tool ping)
- That is the shortest time the sender can know that traffic arrived at the other end
- Sender can only discard old data after receiving ack’s
- Lightspeed in fiber = 200000 km/s
- 100 km = 200 km round trip = 1/1000 sec = 1 ms rtt
  - Amsterdam - Geneve ≈ 20 ms
  - Amsterdam - Chicago ≈ 90 ms
  - Amsterdam - San Diego ≈ 160 ms
  - Amsterdam - Tokyo ≈ 250 ms
  - Amsterdam - Sydney ≈ 300 ms

Buffer space

Window = RTT * BW

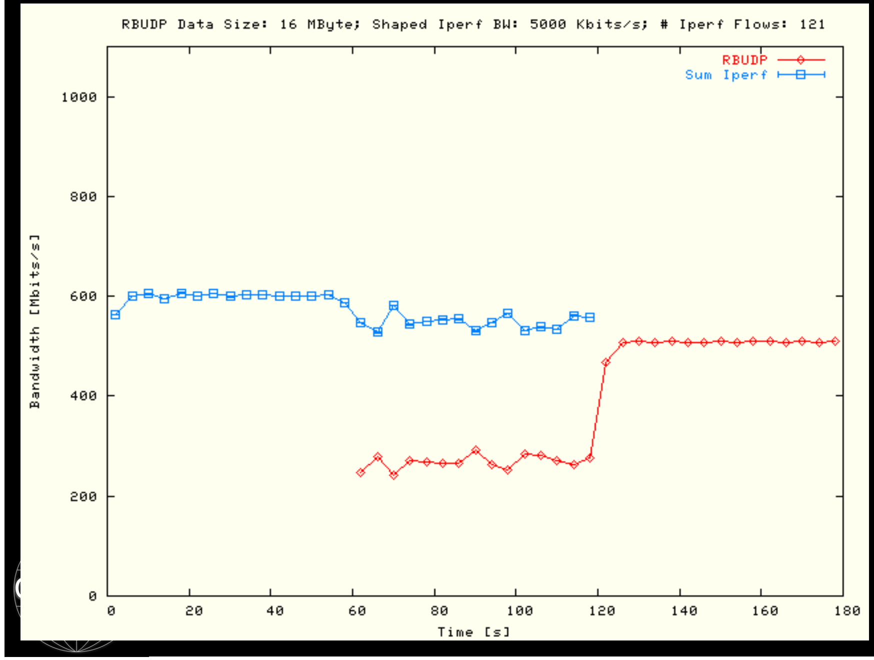
RTT	100 Mbit/s	1 Gbit/s	10 Gbit/s
1	12.5 kB	125 kB	1.25 MB
2	25 kB	250 kB	2.5 MB
5	62.5 kB	615 kB	6.15 MB
10	125 kB	1.25 MB	12.5 MB
20	250 kB	2.5 MB	25 MB
50	625 kB	6.25 MB	62.5 MB
100	1.25 MB	12.5 MB	125 MB
200	2.5 MB	25 MB	250 MB
500	6.25 MB	62.5 MB	625 MB
1000	12.5 MB	125 MB	1250 MB

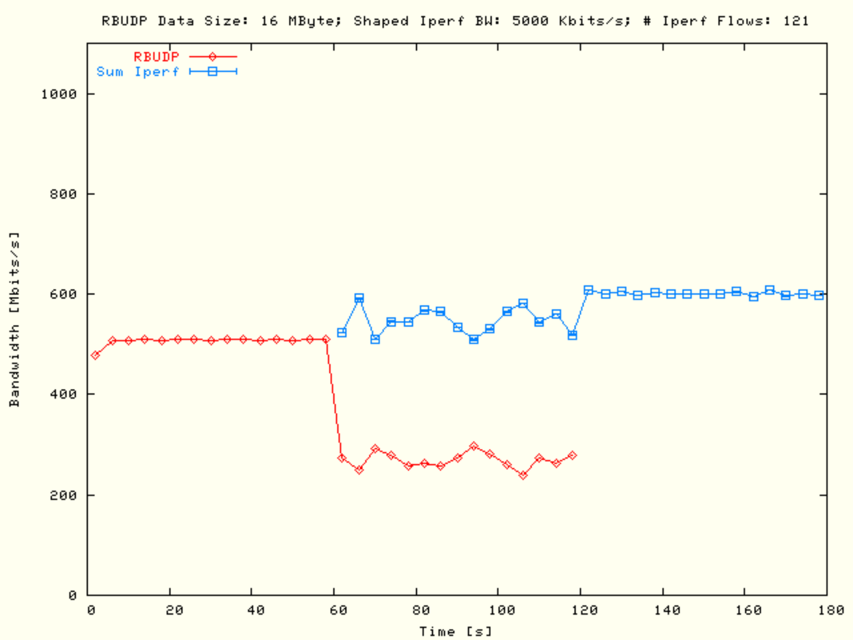
TCP Tuning (if not auto-tuning)

- 1 Gbit/s on 160 ms RTT (= Amsterdam - San Diego) :
  - sysctl -w kern.ipc.maxsockbuf=50000000
  - sysctl -w net.inet.tcp.sendspace=21000000
  - sysctl -w net.inet.tcp.recvspace=21000000
  - sysctl -w net.inet.udp.maxdgram=57344
  - sysctl -w net.inet.udp.recvspace=74848
  - sysctl -w net.local.stream.sendspace=32768
  - sysctl -w net.local.stream.recvspace=32768
  - sysctl -w kern.ipc.somaxconn=512
  - sysctl -w net.inet.tcp.mssdflt=1460
  - sysctl -w net.inet.tcp.delayed_ack=2
  - sysctl -w net.inet.tcp.rfc1323=1
  - sysctl -w net.inet.tcp.rfc1644=1
  - sysctl -w net.inet.tcp.newreno=1

Other issues & protocols

- When using UDP, watch for bottleneck!
- About 10 other non standard protocols
- FAST TCP
  - Modified receiver algorithms
- RBUDP
  - Runs on top of UDP, simple back-off and retransmission scheme





Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

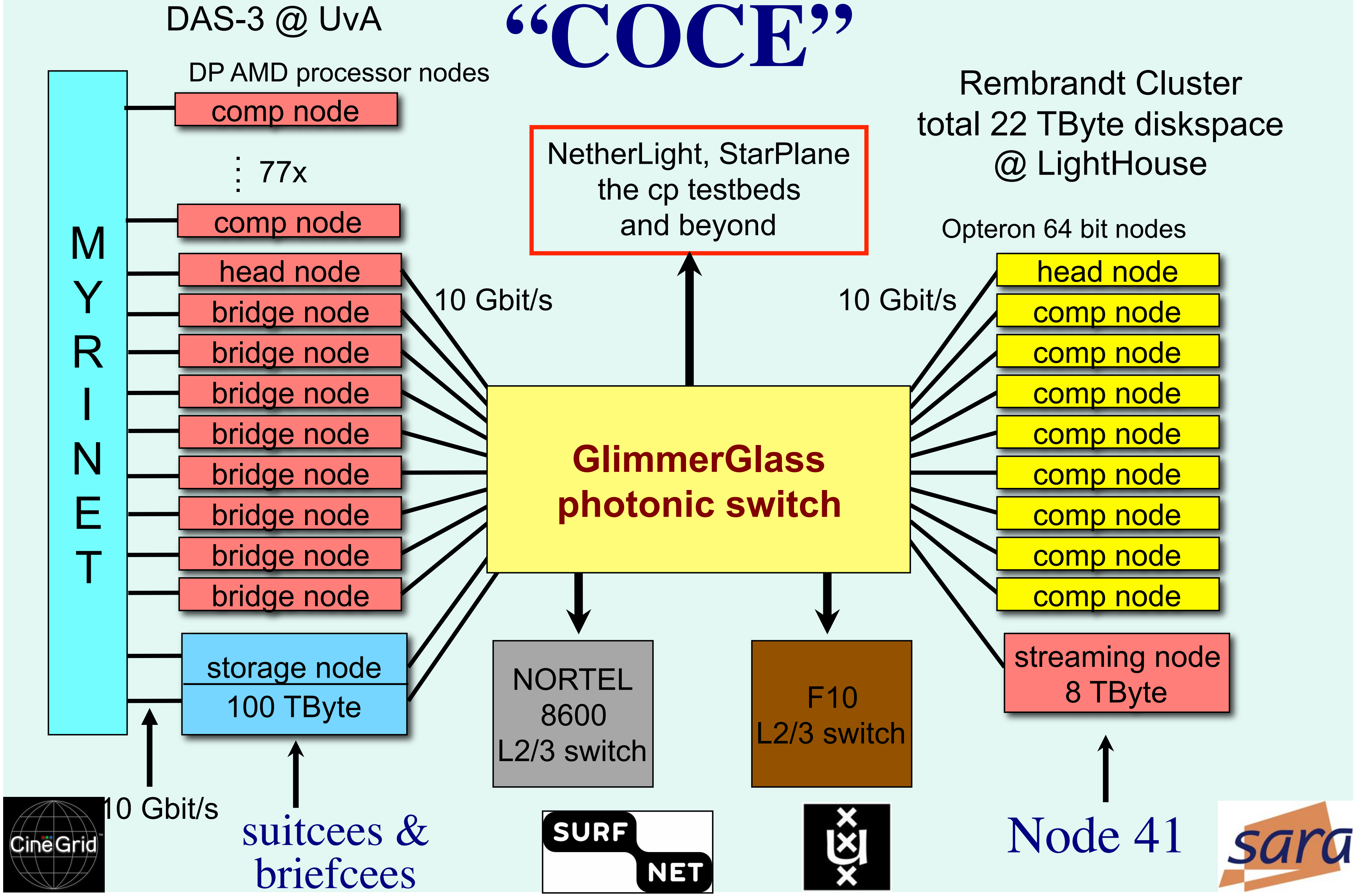
End System Issues

- Ethernet card interface to computer bus system
  - PCI-X
    - 32/64 bit 66/133/266 MHZ -> about 8 Gbit/s max in 133 MHZ mode
  - PCI-Express
    - 2.5 Gbit/s per lane, 4, 8, 16 lanes
- Memory organization
- CPU cache
  - Effect when things go out of cache (small windows, etc.)
- CPU core
  - Takes 1 core to handle network (affinity may help)
- Disk raid subsystem
  - raid0 twice as fast as raid5
  - One disk does typically 40 MB/s write, 60 MB/s read

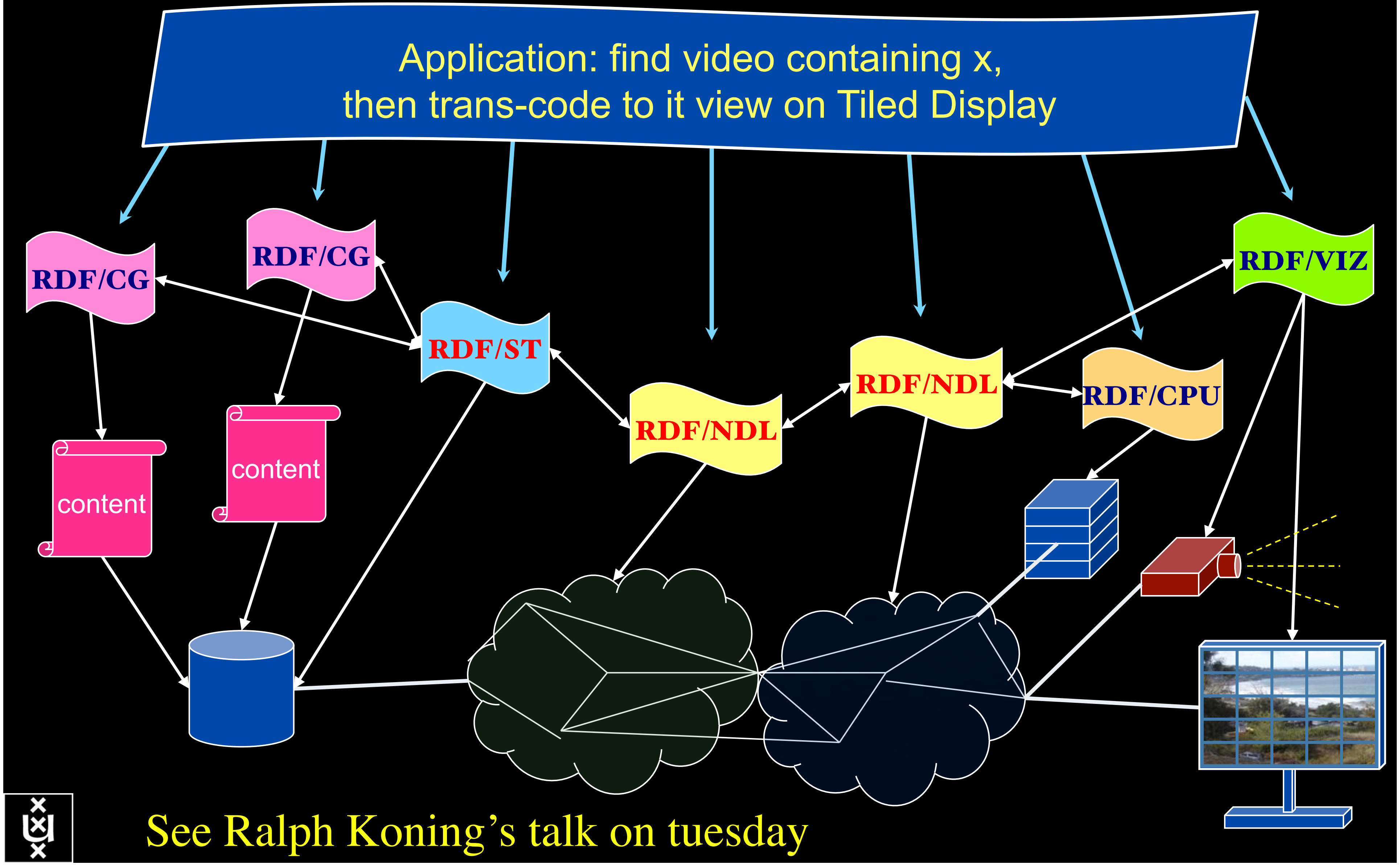
Contents

1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

Amsterdam CineGrid S/F node "COCE"



RDF describing Infrastructure



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

RDF/CG

RDF/CG

RDF/ST

RDF/NDL

RDF/NDL

RDF/CPU

RDF/VIZ

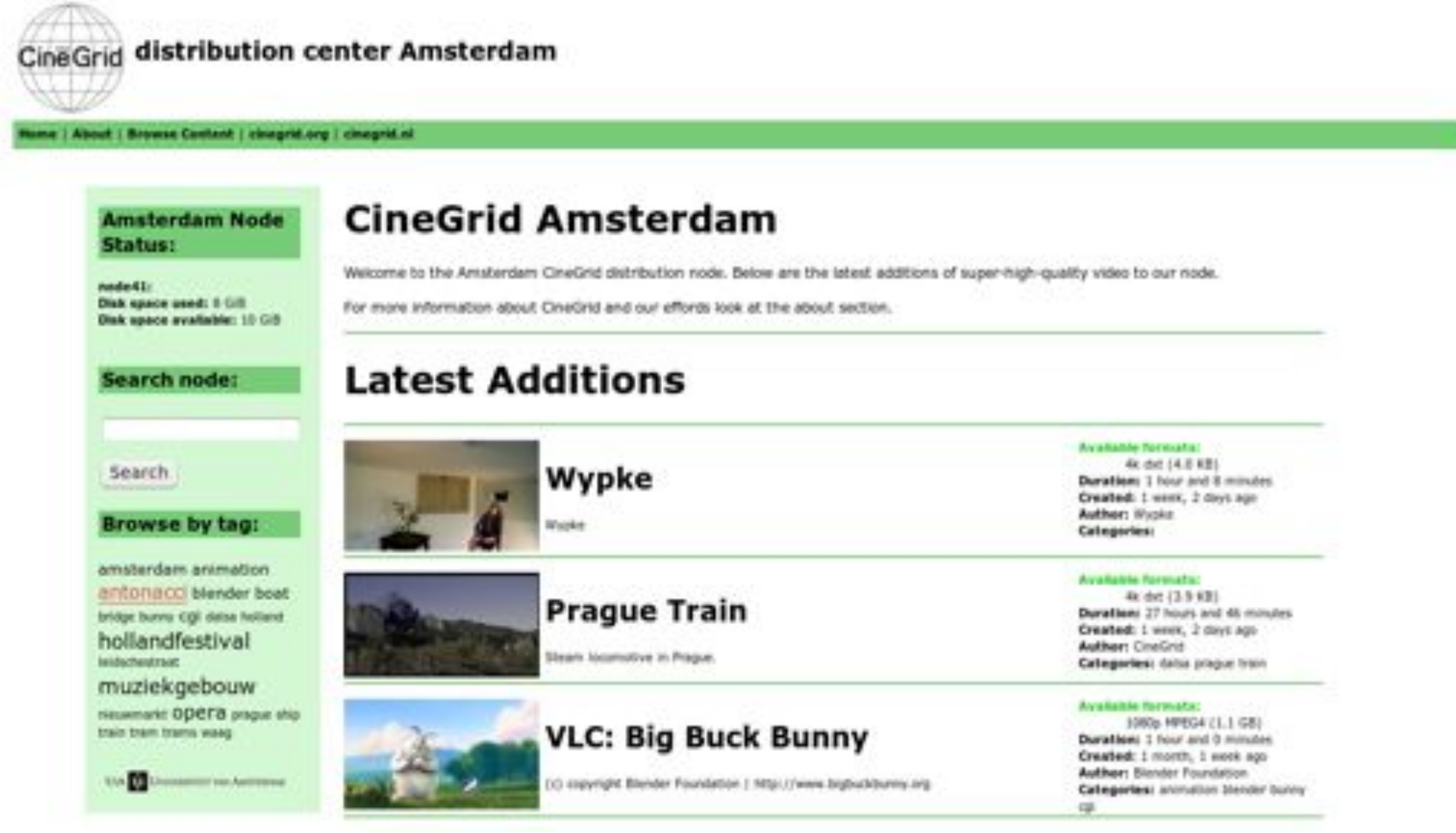
content

content

See Ralph Koning’s talk on tuesday

CineGrid portal

100 Tbyte
Cache & Store & Forward



CineGrid distribution center Amsterdam

Home | About | Browse Content | cinegrid.org | cinegrid.nl

Amsterdam Node Status:

node41:
Disk space used: 8 GiB
Disk space available: 10 GiB

Search node:

Search

Browse by tag:

amsterdam animation antonacci blender boat bridge bunny cgi dalsa holland hollandfestival leidschestraat muziekgebouw nieuwmarkt opera prague ship train tram trams waag

UvA Universiteit van Amsterdam

CineGrid Amsterdam

Welcome to the Amsterdam CineGrid distribution node. Below are the latest additions of super-high-quality video to our node.

For more information about CineGrid and our effords look at the about section.

Latest Additions

Wypke

Wypke

Available formats:
4k dxt (4.8 KB)
Duration: 1 hour and 8 minutes
Created: 1 week, 2 days ago
Author: Wypke
Categories:

Prague Train

Steam locomotive in Prague.

Available formats:
4k dxt (3.9 KB)
Duration: 27 hours and 46 minutes
Created: 1 week, 2 days ago
Author: CineGrid
Categories: dalsa prague train

VLC: Big Buck Bunny

(c) copyright Blender Foundation | http://www.bigbuckbunny.org

Available formats:
1080p MPEG4 (1.1 GB)
Duration: 1 hour and 0 minutes
Created: 1 month, 1 week ago
Author: Blender Foundation
Categories: animation blender bunny cgi

www.cinegrid.org www.cinegrid.nl

www.science.uva.nl/~delaat

Questions?