# CineGrid Networking

CG-2010


## Cees de Laat & Brian Dunne

University of Amsterdam
CALIT2

# 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. 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

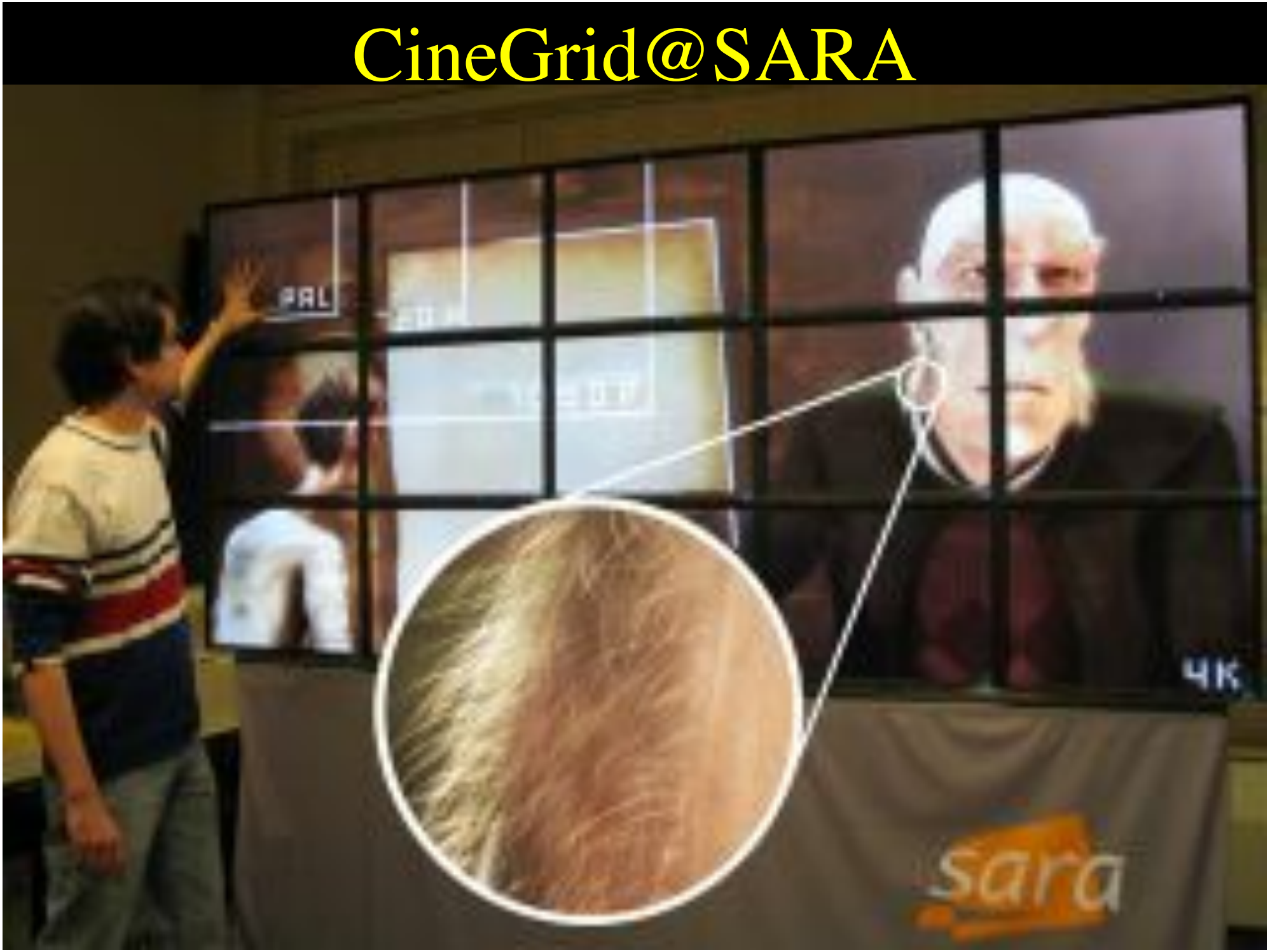
UCSD Chancellor Fox

**Used 1Gbps Dedicated**

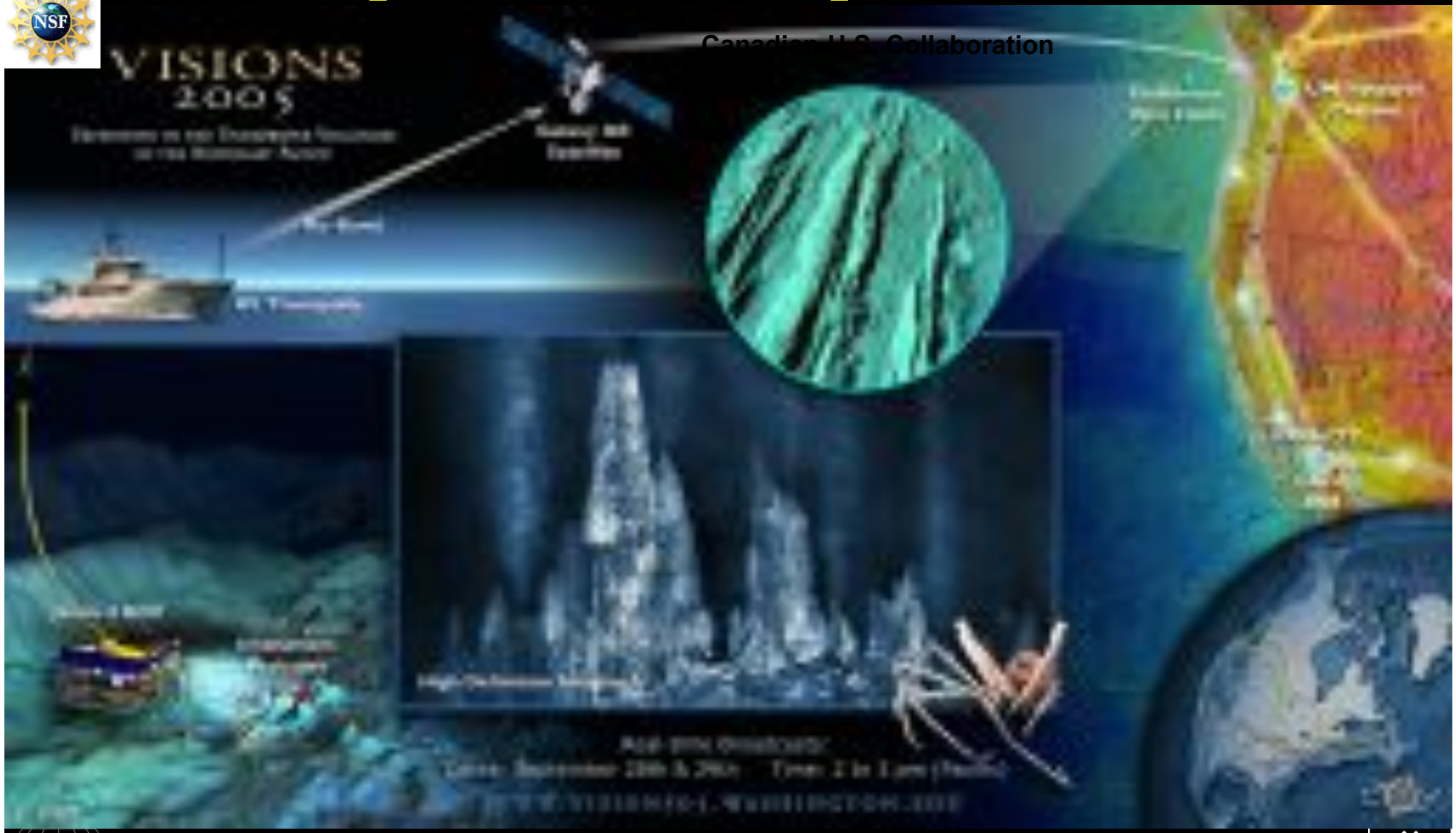
**Sony NTT SGI**

iGrid 2005

# CineGrid@SARA

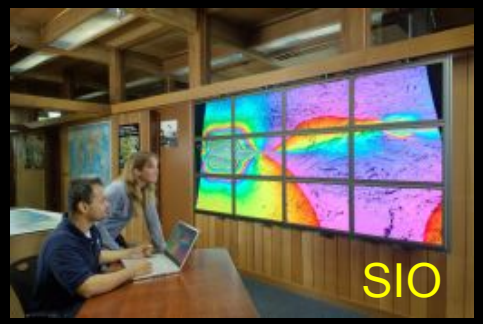


# First Remote Interactive High Definition Video Exploration of Deep Sea Vents




Source John Delaney & Deborah Kelley, UWash

# US and International OptIPortal Sites

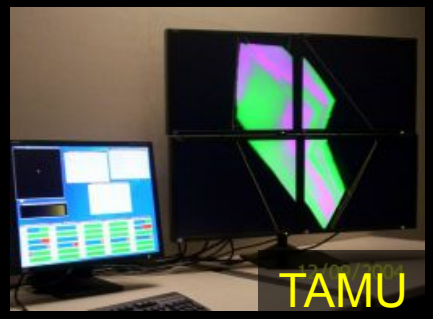
















CineGrid

# The “Dead Cat” demo

## SC2004 & iGrid2005

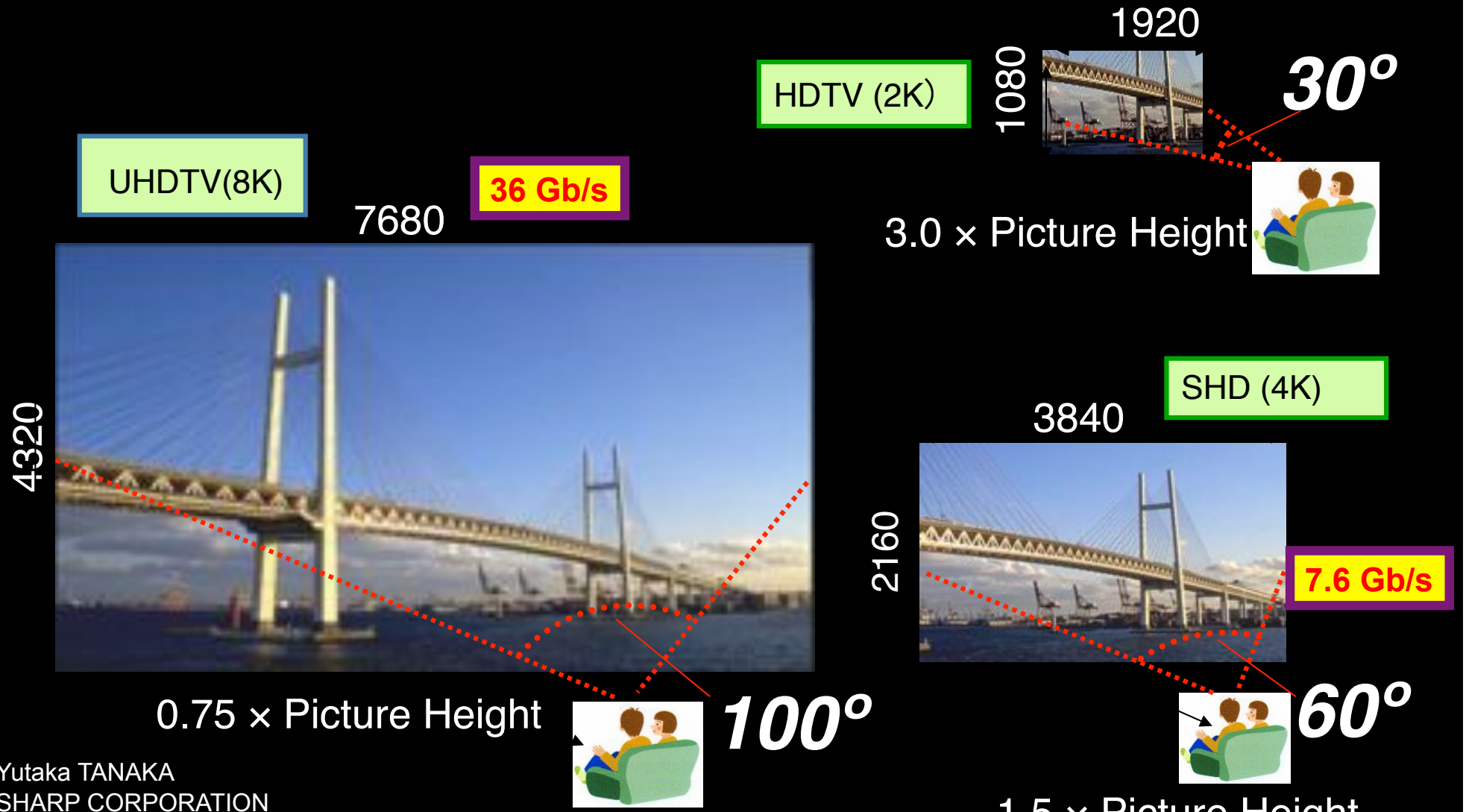
SC2004,
Pittsburgh,
Nov. 6 to 12, 2004
iGrid2005,
San Diego,
sept. 2005

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?

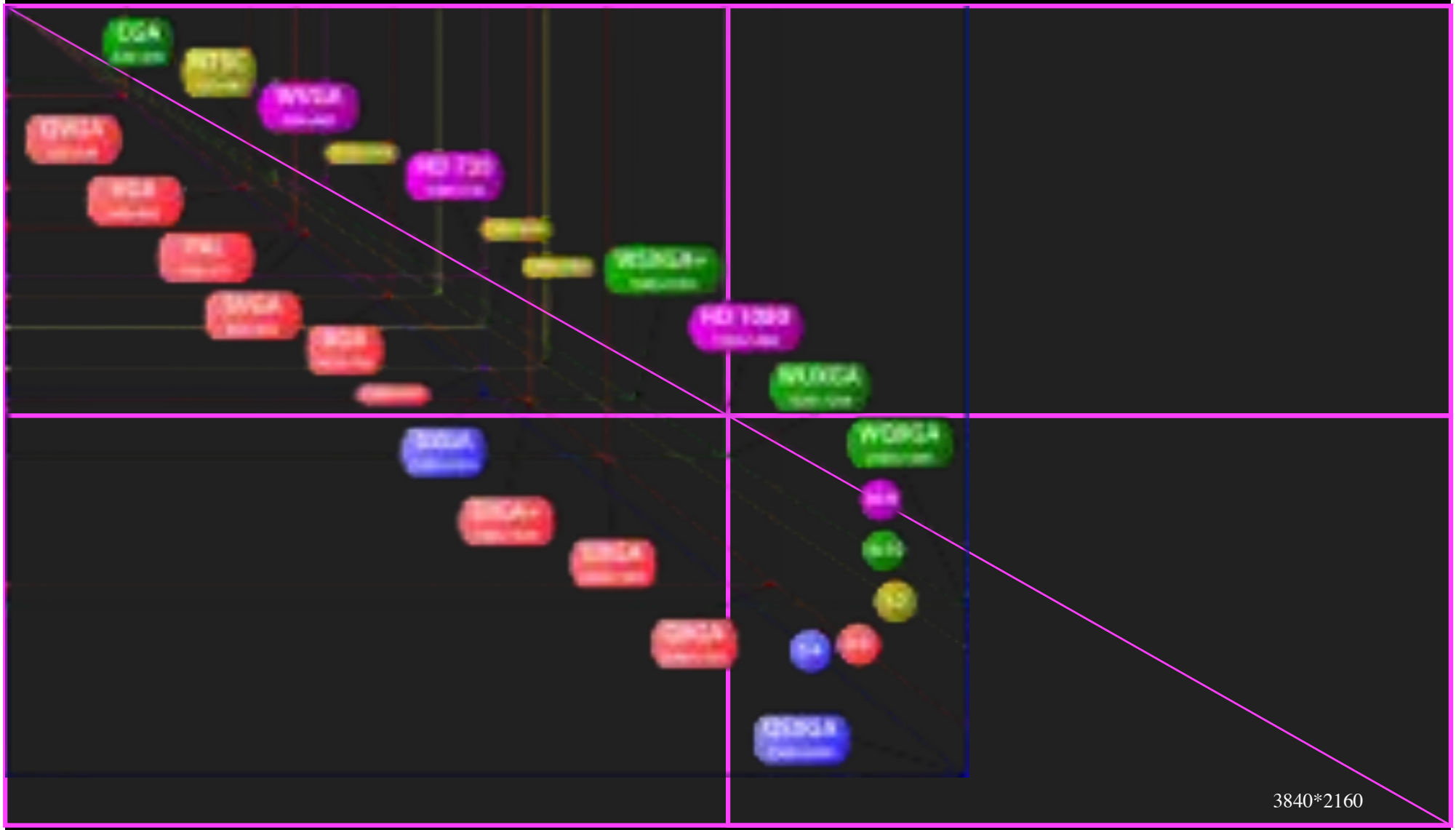
1. More Resolution Allows Closer Viewing of Larger Image
2. Closer Viewing of Larger Image Increases Viewing Angle
3. Increased Viewing Angle Produces Stronger Emotional Response




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 /s | Color bits/pix | Frame pix | Frame MByte | Flow MByt/s | Stream 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 (UMF) 3/4 GBytesec
  - Compressed (jpeg2000) 300 - 700 Mbit/s
  - Uncompressed (eg TIFF) 1.2 GB/s, 4.3 TB/h
  - Compressed (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

CineGrid

# 3D

# FOR 3D MULTIPLY EVERYTHING BY 2!

# 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

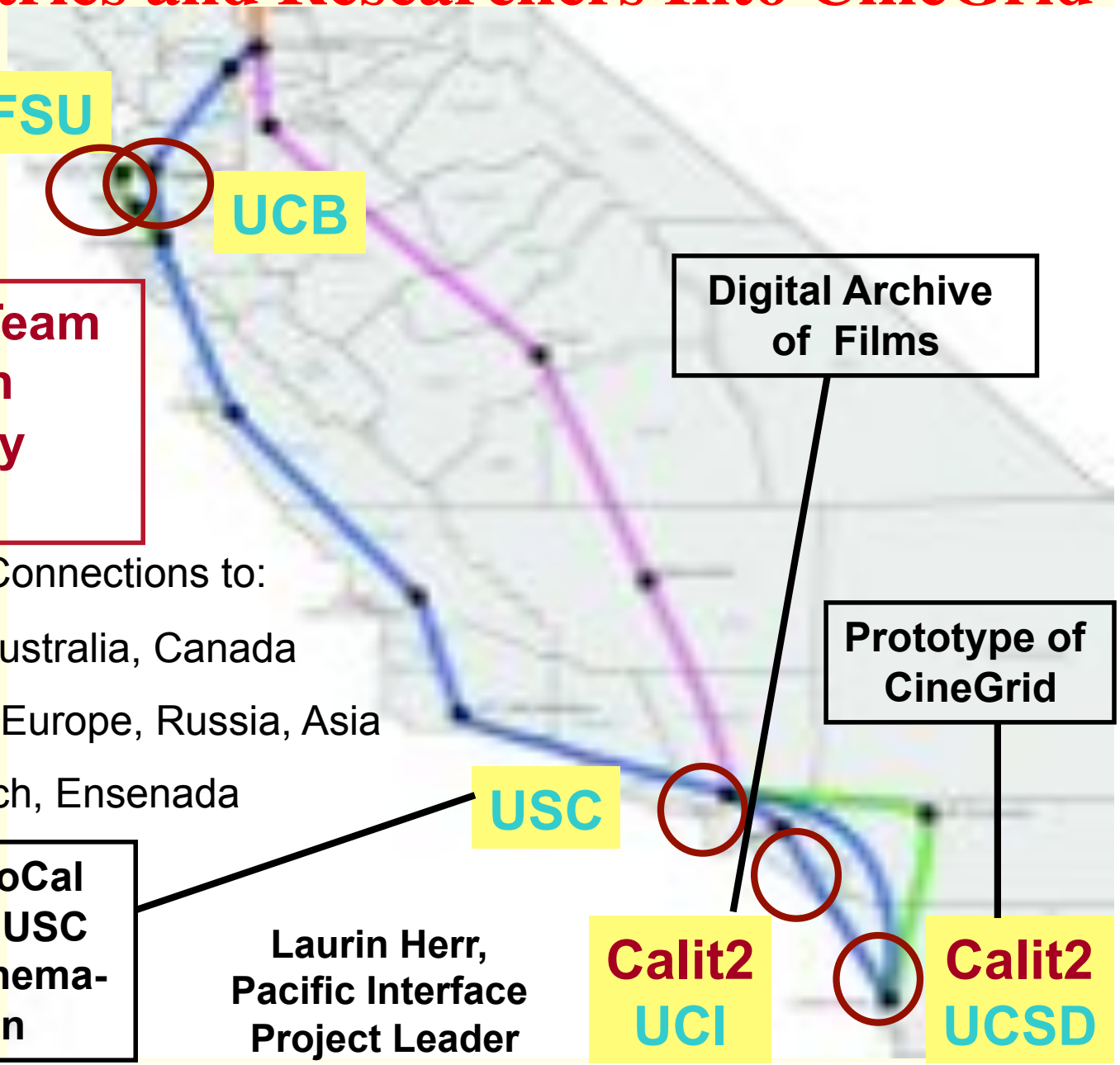
- **GLIF is a world-scale Lambda-based Laboratory for application and middleware development on emerging LambdaGrids, where applications rely on dynamically configured networks based on optical wavelengths**

- **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

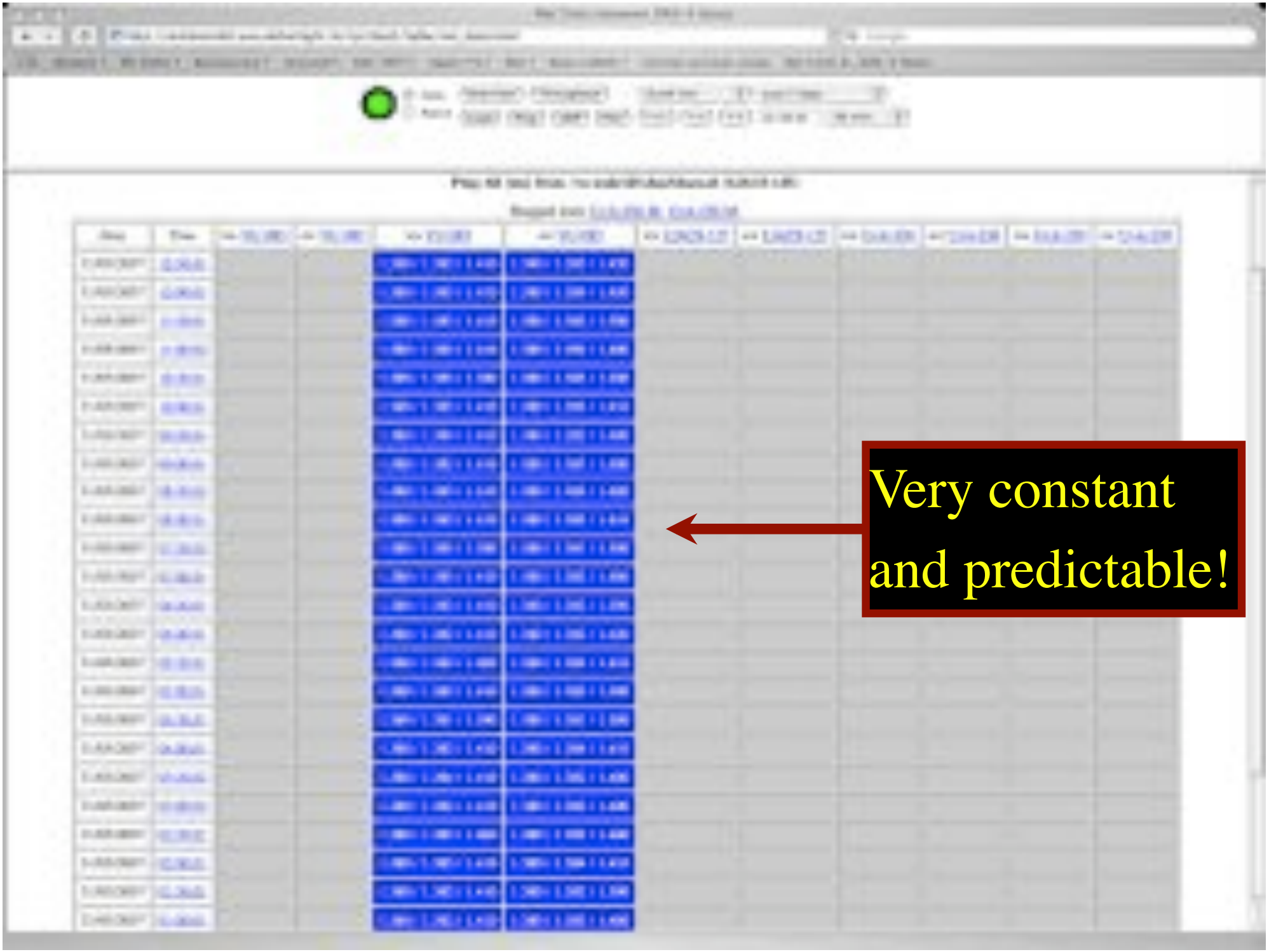

NORDUnet n*10Gb
Stockholm
Oslo
SURFnet-
IEEAF-
GLORIAD
50Gb
NORDUnet-
GLORIAD 10Gb
Copenhagen
NorthernLight
NORDUnet 50Gb
Amsterdam
NetherLight Hamburg
UKLight
London
Munster
CESNET 10Gb
CzechLight
Prague
SURFnet
50Gb
Paris
DASS-
Grid'5000
10Gb
CERN
Geneva
CERN/TIF
Barcelona
CATLight-i2CAT 10Gb

# 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

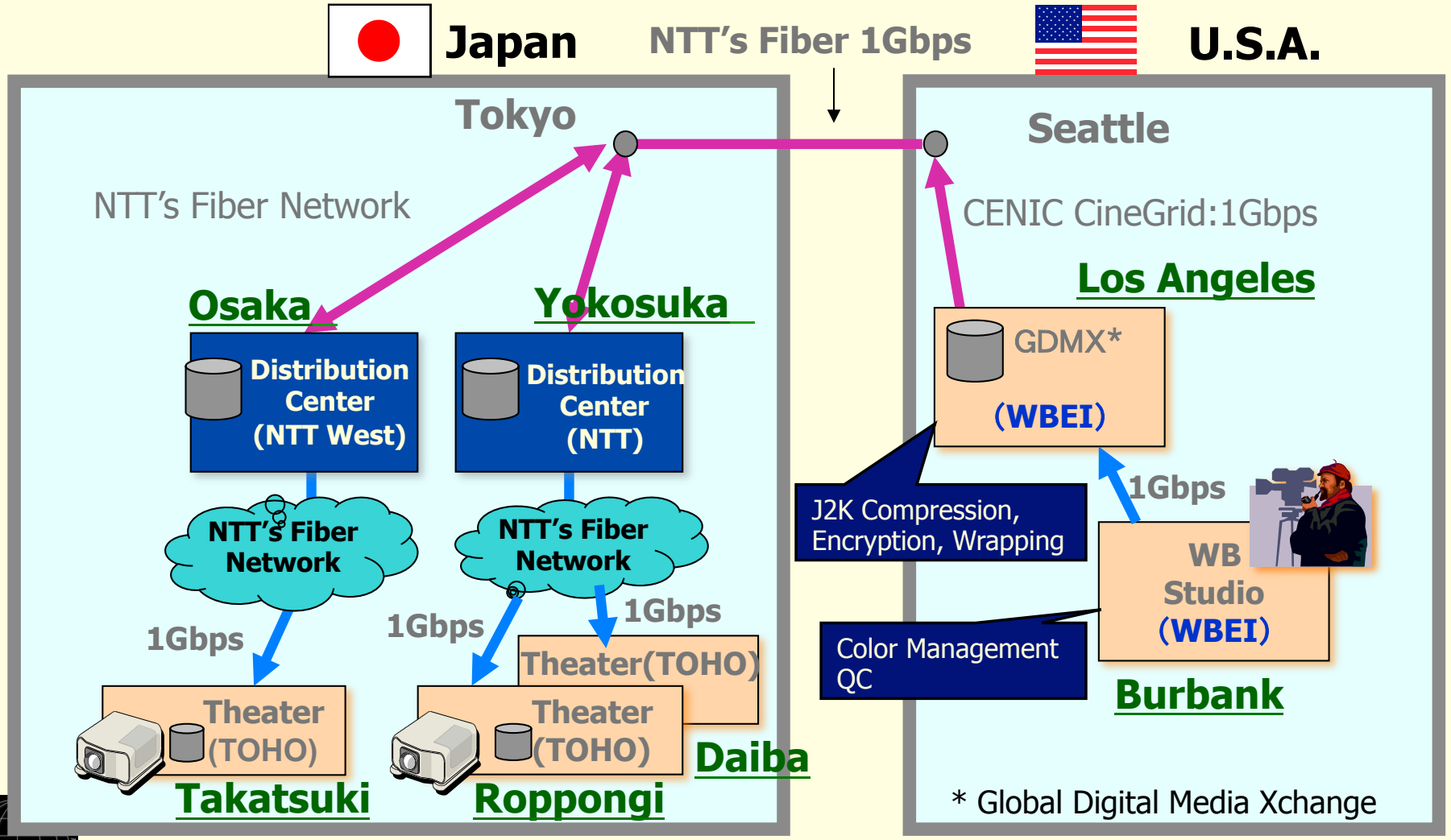
# What is a LightPath

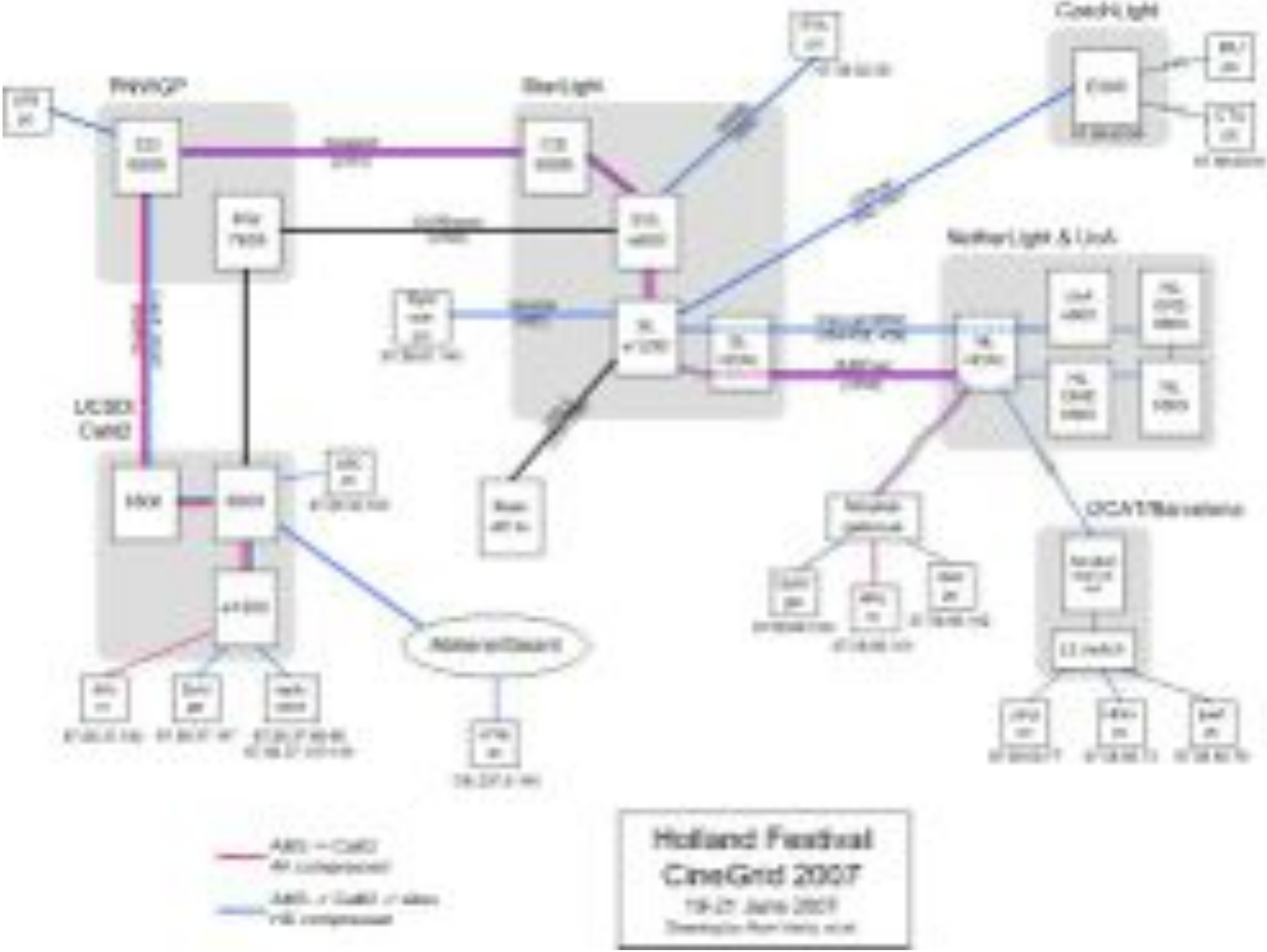
- A LightPath is a circuit like connection that connects end systems to each other. This uses partly the same infrastructure as the Internet, but a LightPath gets dedicated resources next to Internet.
- A LightPath can be a concatenation of:
  - A color in a fiber (true Lambda - λ)
  - SONET/SDH circuit in a SONET infrastructure
  - VLAN's and dedicated ports in an ethernet infrastructure
  - Etc.
- Aim is to get predictable and knowable connection characteristics
- Let us look at examples setups used in CineGrid!


Very constant
and predictable!

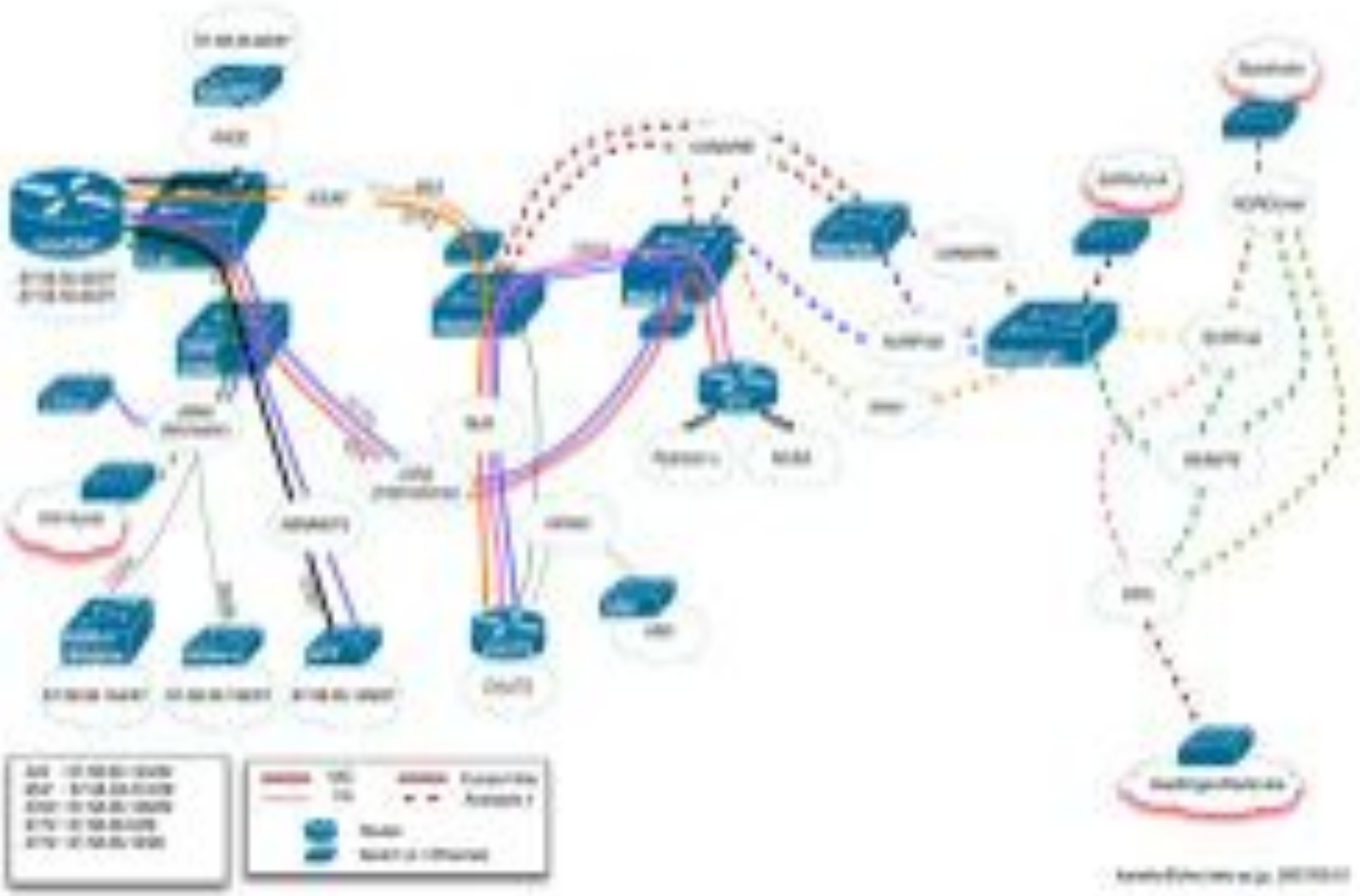
# Network for “4K Pure Cinema” Trial

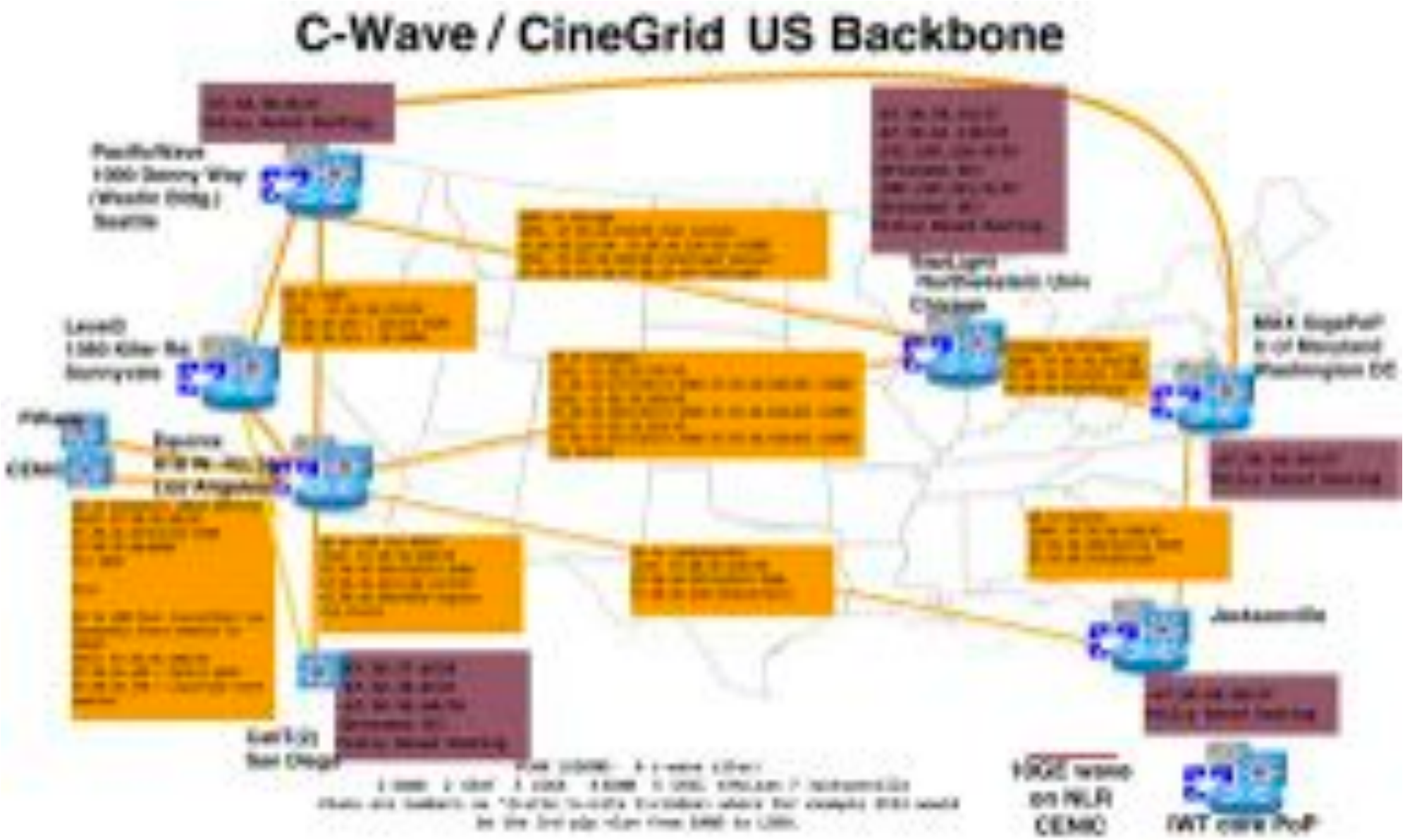
DCP is directly transferred from GDMX in LA to distribution centers in Japan via fiber network. Within Japan, DCP is distributed from the distribution centers to TOHO theaters. Key is distributed from Osaka center, based on the contract between WB Japan and TOHO cinemas.



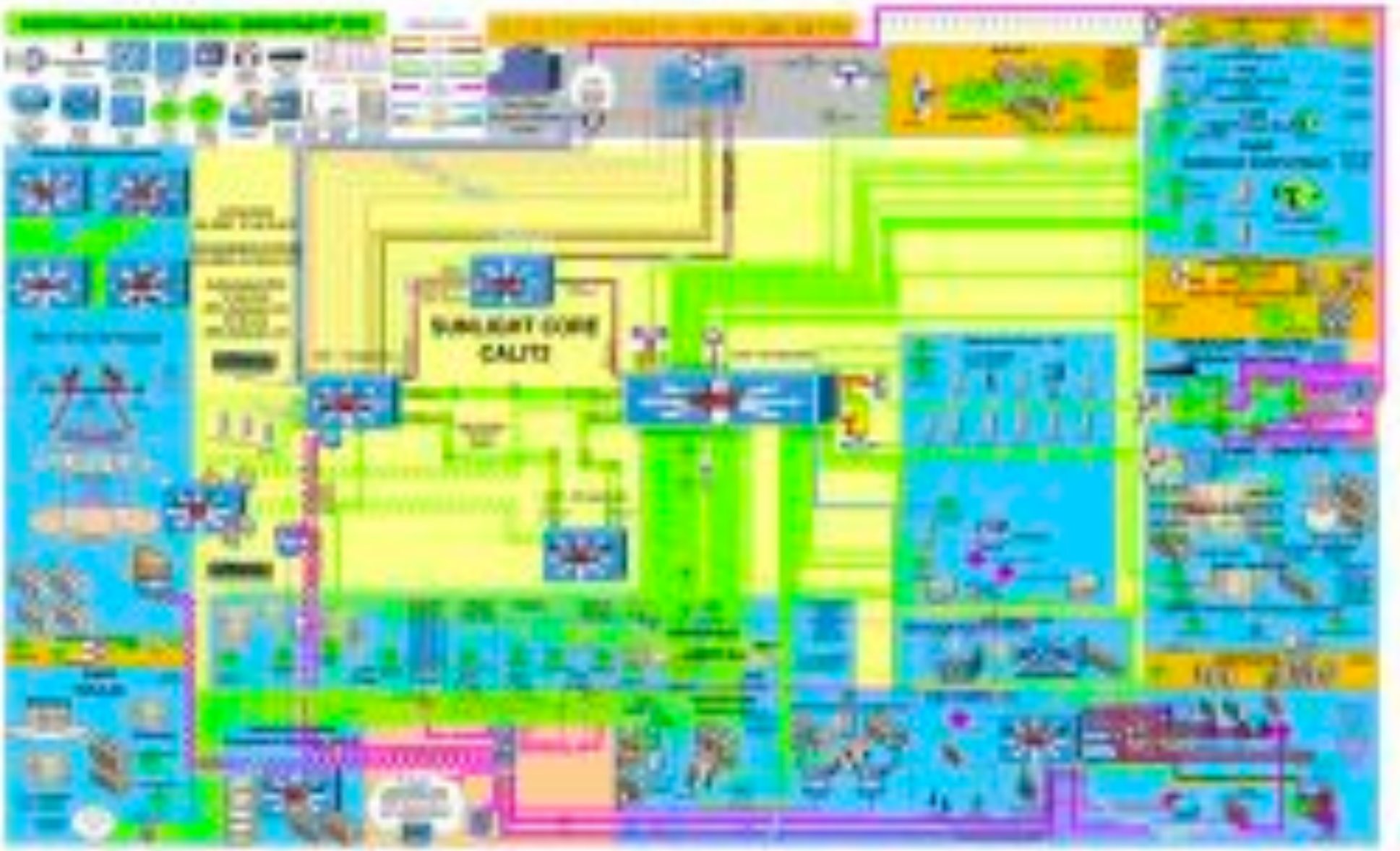

Holland Festival
CineGrid 2007

# Current Links & Available Links for Kyoto Prize Events

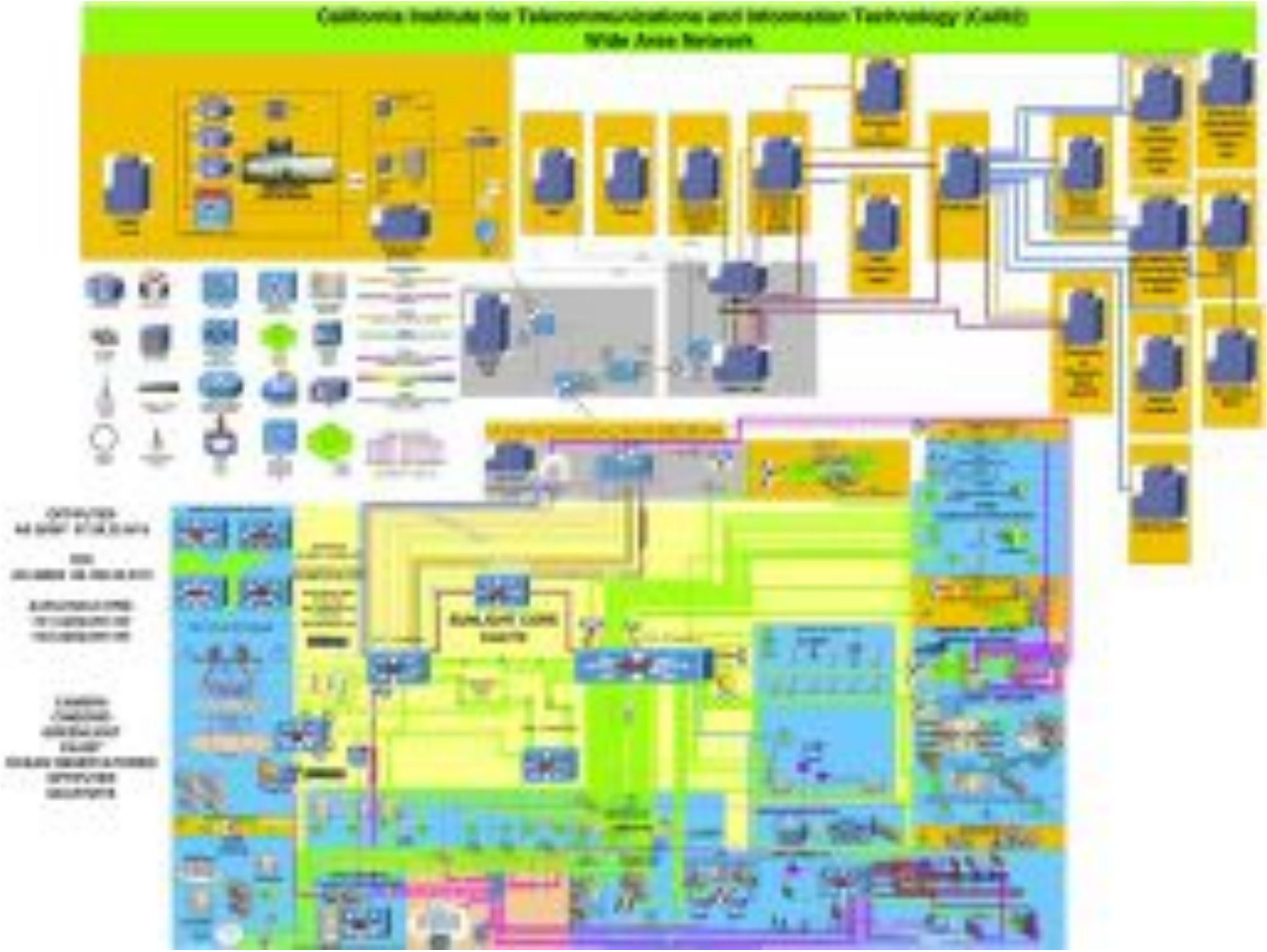





# C-Wave / CineGrid US Backbone



# OptiPuter Main Network


California Institute for Telecommunications and Information Technology (Calit2)
Wide Area Network

# 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
  - RTT is reaction and recovery time
  - TCP is friendly protocol, can adapt but performance usually takes severe hit (more RTT -> bigger hit)


# 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 solid 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

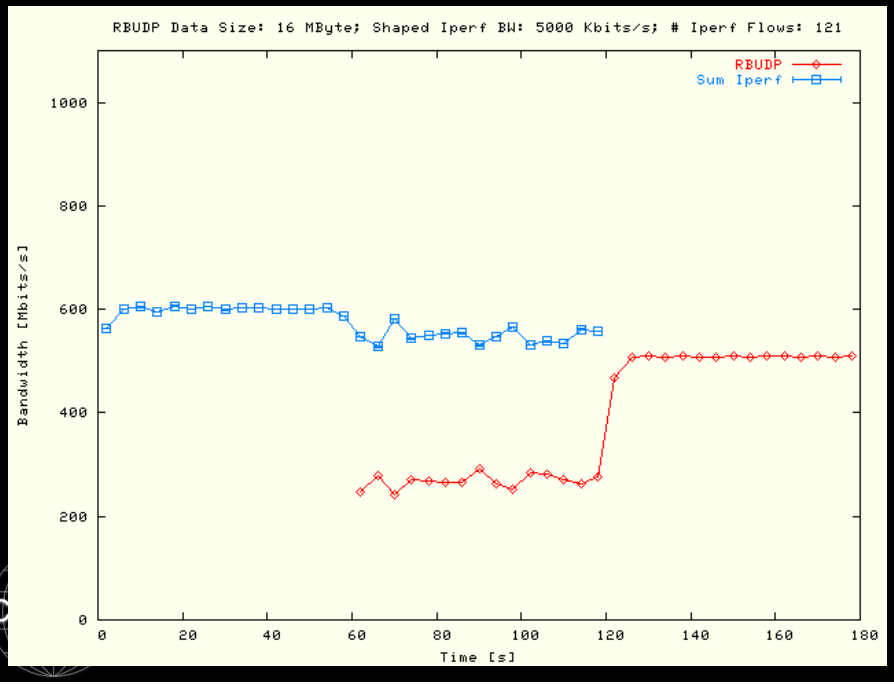
| RTT | 100 Mbit/s | 1 Gbit/s | 10 Gbit/s | 40 Gbit/s |
|---|---|---|---|---|
| 1 | 12.5 kB | 125 kB | 1.25 MB | 5 MB |
| 2 | 25 kB | 250 kB | 2.5 MB | 10 MB |
| 5 | 62.5 kB | 625 kB | 6.25 MB | 25 MB |
| 10 | 125 kB | 1.25 MB | 12.5 MB | 50 MB |
| 20 | 250 kB | 2.5 MB | 25 MB | 100 MB |
| 50 | 625 kB | 6.25 MB | 62.5 MB | 250 MB |
| 100 | 1.25 MB | 12.5 MB | 125 MB | 500 MB |
| 200 | 2.5 MB | 25 MB | 250 MB | 1 GB |
| 500 | 6.25 MB | 62.5 MB | 625 MB | 2.5 GB |
| 1000 | 12.5 MB | 125 MB | 1250 MB | 5 GB |

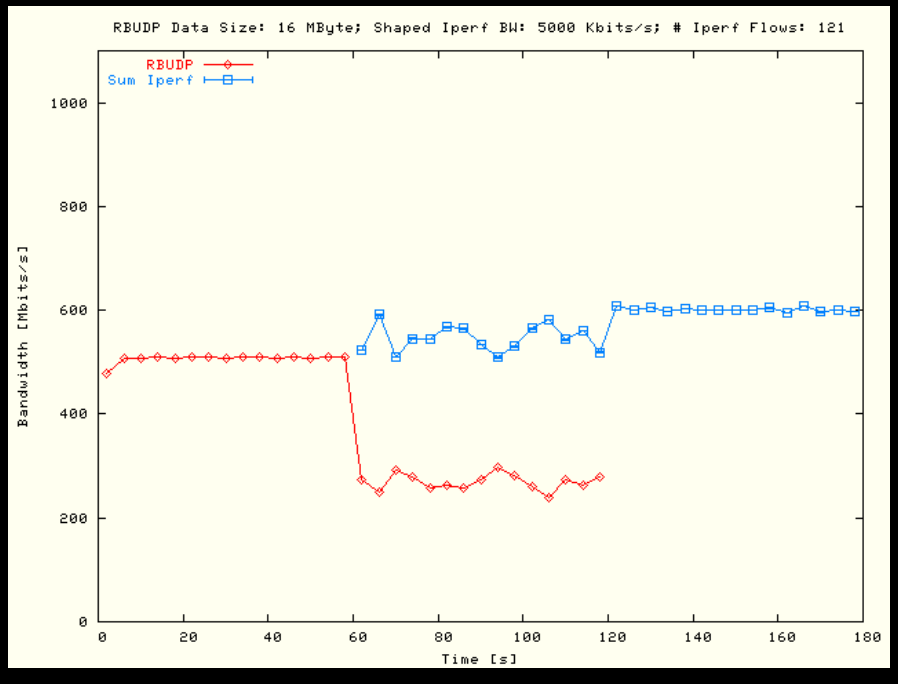

# TCP Tuning (if not auto-tuning)

- 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!
- Ethernet switches forget destination!
- About 10 other non standard protocols
- FAST TCP → Modified receiver algorithm
- RBUDP → simple back-off and retransmission scheme



# Contents

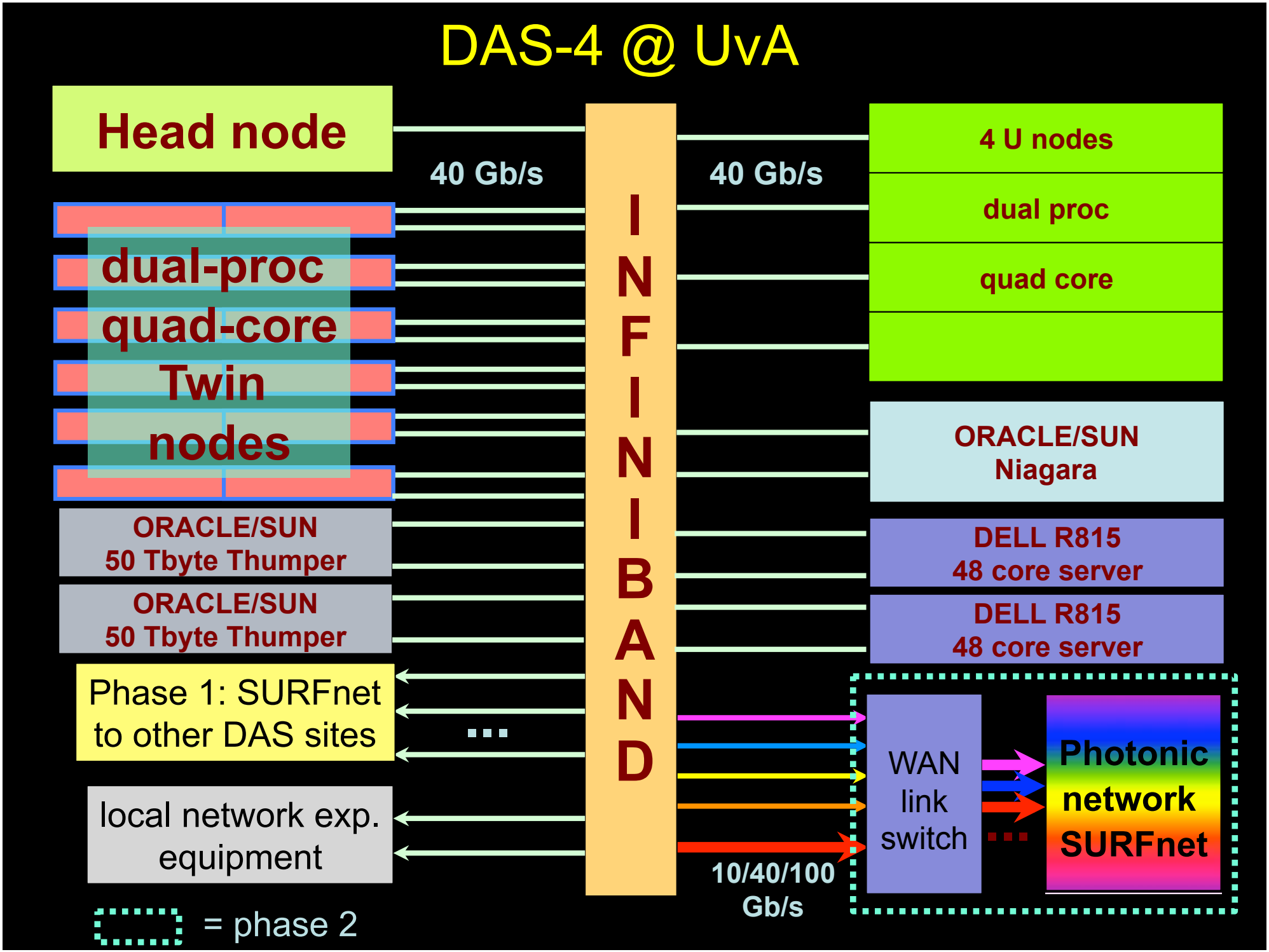
1. Use cases CineGrid & Networks
2. Formats - Numbers - Bits
3. Global Lambda Integrated Facility
4. A LightPath
5. Transport Protocol issues
6. End System Issues
7. Network Storage
8. Q/A

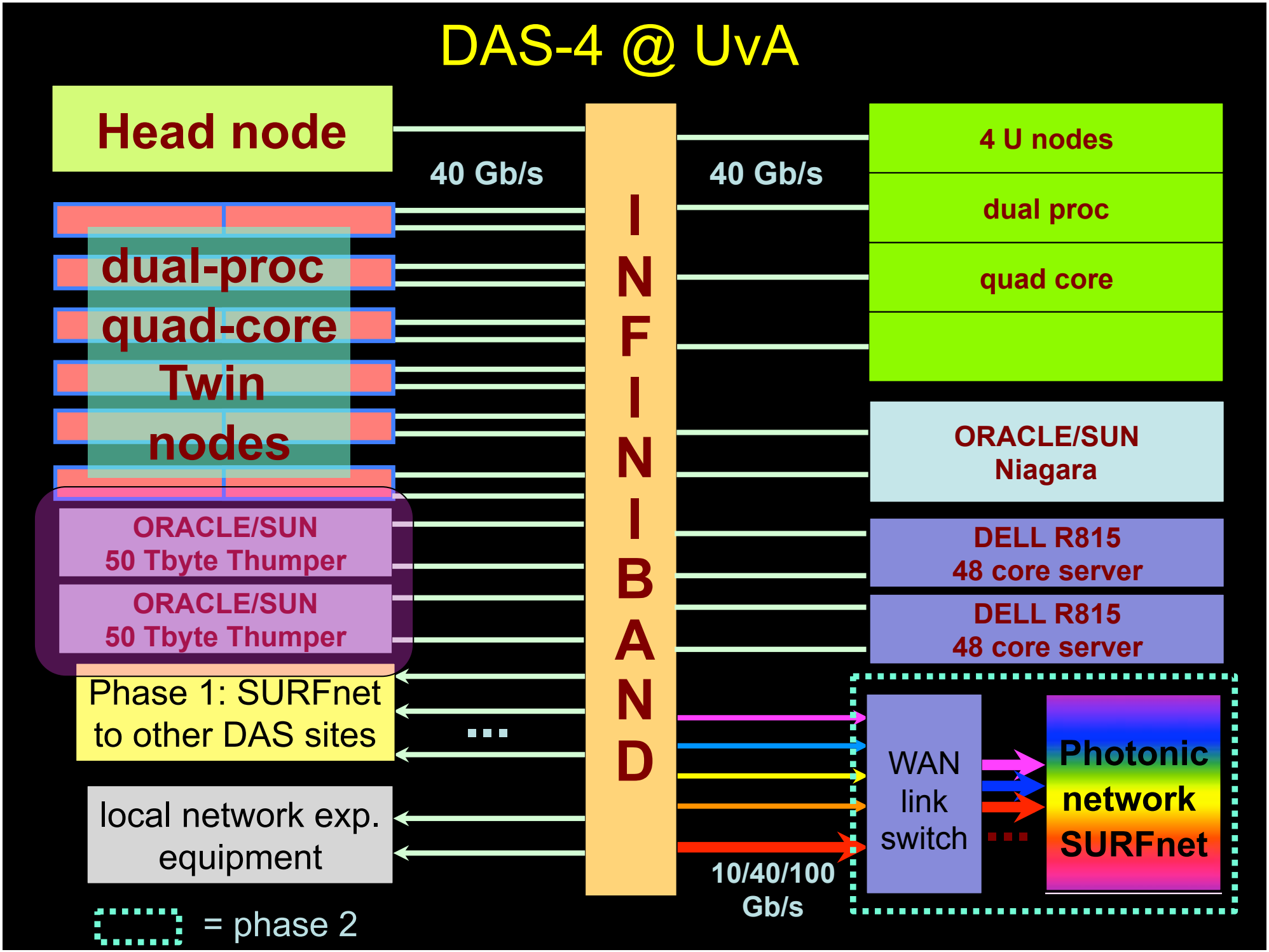
# End System Issues

- 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 (2.0)
    - 5 Gbit/s per lane: 4, 8, 16 lanes
    - 4/5 of that because of coding, 15-20% protocol overhead
- Memory organization, or avoid memory
- 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 “old” 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

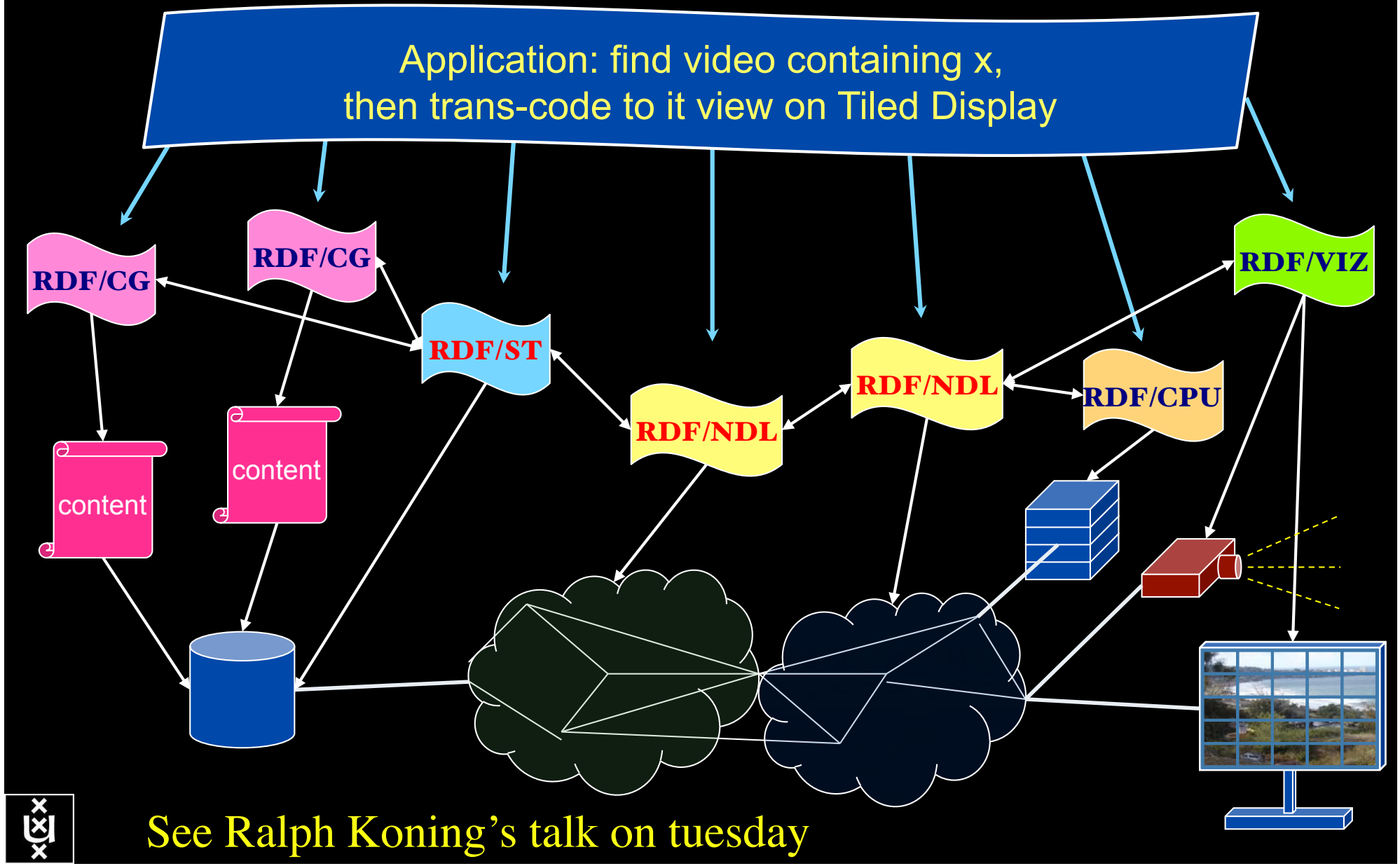


DAS-4 @ UvA
Head node
40 Gb/s
dual-proc
quad-core
Twin
nodes
ORACLE/SUN
50 Tbyte Thumper
ORACLE/SUN
50 Tbyte Thumper
Phase 1: SURFnet
to other DAS sites
local network exp.
equipment
= phase 2
INFINIBAND
40 Gb/s
4 U nodes
dual proc
quad core
ORACLE/SUN
Niagara
DELL R815
48 core server
DELL R815
48 core server
WAN
link
switch
Photonic
network
SURFnet
10/40/100
Gb/s


DAS-4 @ UvA
Head node
40 Gb/s
dual-proc
quad-core
Twin
nodes
ORACLE/SUN
50 Tbyte Thumper
ORACLE/SUN
50 Tbyte Thumper
Phase 1: SURFnet
to other DAS sites
local network exp.
equipment
= phase 2
INFINIBAND
40 Gb/s
4 U nodes
dual proc
quad core
ORACLE/SUN
Niagara
DELL R815
48 core server
DELL R815
48 core server
WAN
link
switch
Photonic
network
SURFnet
10/40/100
Gb/s

# CineGrid portal

100 Tbyte
Cache & Store & Forward

# RDF describing Infrastructure




See Ralph Koning's talk on tuesday

# 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


www.cinegrid.org www.cinegrid.nl

ext.delaat.net

# Questions?