Optical Networking

Charlie Catlett
Senior Fellow, Computation Institute
catlett@mcs.anl.gov

- “High Performance Nets”
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures

February 2003
The Internet: 1969 through Today

Sources: SIGCOMM Review, CAIDA/UCSD, AT&T Labs
NSFNET 56 Kb/s Site Architecture

Fuzzball

Vax

1024 MB 4 MB/s 1024 s (17 min) 150,000 s (41 hrs)

256 s (4 min) 1 MB/s .007 MB/s

128 MB (16 MW)

Charlie Catlett (catlett@mcs.anl.gov)
NSFNET T1 Backbone

Topology data from NLANR MOAT (www.nlanr.net)
T1 Site Architecture

- 8,192 MB
- 100 MB/s
- 12 MB/s
- .2 MB/s

80 s (1.5min)
7k s (11 min)
41,000 s (11 hrs)

Charlie Catlett (catlett@mcs.anl.gov)
NSFNET T3 Backbone

Core nodes
Exterior nodes

1992 – 1995

Topology data from NLANR MOAT (www.nlanr.net)
T3 Site Architecture

Router

T3 NSS

RS/6000

16,384 MB   12 MB/s   6 MB/s

1400 s (23 min)  27k s (45 min)

1024 MB

Charlie Catlett (catlett@mcs.anl.gov)
vBNS Logical Map

The National Science Foundation Very-High-Speed Backbone Network Service
Logical Network Map

Source: MCI vBNS project (www.vbns.net)
vBNS Site Configuration

16,384 MB  100 MB/s  18 MB/s

1024MB

Charlie Catlett (catlett@mcs.anl.gov)

vBNS schematic: MCI vBNS project (www.vbns.net)
Today’s Architecture

Charlie Catlett (catlett@mcs.anl.gov)

OC-48 Cloud

| 64 GB |
| 1024 MB |

GbE

OC-12

2000 s (33 min) 13k s (3.6h)

1 TB 0.5 GB/s 78 MB/s
To Build a Distributed Terascale Cluster...

Big Fast Interconnect

5 GB/s = 200 x 25 MB/s

25 MB/s = 200 Mb/s or 20% of GbE

2000 s (33 min)

100 TB  5 GB/s  100 TB

4096 GB

64 GB

Charlie Catlett (catlett@mcs.anl.gov)
Optical Networks

- “High Performance Nets”
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures
Recap: Layered Architecture

Source: Ian Foster
Recap: The Internet Routes Packets

Source: Ian Foster
Layered Network View

IP Routers

ATM Switches

SONET MUXES, DWDM Terminals

Wavelengths

Source: Cisco
IP Transport Alternatives

B-ISDN

IP over ATM

IP over SONET/SDH

IP over Optical

Multiplexing, Protection, and Management at Every Layer

Lower Cost, Complexity, and Overhead

Source: Cisco
Optical Networking Building Blocks

- Optical Fiber
  - Types of fiber
  - Fiber characteristics
- Attenuation and Dispersion
- Amplification
- Filters, Multiplexors, Demultiplexors
- Transmission Systems

- Vocabulary…
  - Wavelength = channel = lambda (\( \lambda \))

Charlie Catlett (catlett@mcs.anl.gov)
Total Internal Reflection & Numerical Aperture

- Total internal reflection allows light to propagate down fiber
- NA defines envelope within which light will enter the core and propagate
- Light must be “launched” into the core within the NA
- Light detectors must take into account NA as well

Source: Corning
Fiber Types

Multimode Fiber

• Different components of the light signal entering at different launch angles → “modes”
• Each mode travels a different path through the fiber, thus a different path length
• **Disadvantage**: modes arrive at different times – “modal dispersion” (more on dispersion later)
• **Advantage**: easier coupling to launch signal into fiber → cheaper components

Single mode Fiber

• Only allows one mode
• **Disadvantage**: tighter tolerances for coupling
• **Advantage**: no modal dispersion!

Source: Corning
Attenuation: Loss in Optical Fiber

Source: Corning
Amplification

- Laser signal at X db
- Receiver threshold at Y db (Y << X)
- Maximum distance before amplification determined by fiber characteristics (db loss per km at signal wavelength)
- More on this later!

- Amplification requires no conversion to electronics
- Amplifiers tuned to a range of wavelengths – on amplifier per strand

Source: www.lightreading.com
Dispersion

Chromatic Dispersion

Units: ps / nm * km

Polarization Mode Dispersion

<table>
<thead>
<tr>
<th>Bitrate</th>
<th>Allowable Dispersion (DL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5 Gb/s</td>
<td>12-16,000 ps/nm</td>
</tr>
<tr>
<td>10 Gb/s</td>
<td>800-1,000 ps/nm</td>
</tr>
<tr>
<td>40 Gb/s</td>
<td>60-100 ps/nm</td>
</tr>
</tbody>
</table>

Source: Corning
Dispersion at 40 Gb/s

40 Gb/s signal at transmitter (top) and after 80 km of NZ-DSF fiber without dispersion compensation.

Source: Lasercom, Inc
Dispersion-Shifted Fiber

Source: Lasercom, Inc
Filtering, Muxing, Demuxing

Fiber Bragg Grating

UV radiation
Optical Fiber
cladding

reflective index changes

Light at \textit{Bragg wavelength} is reflected. \textit{Bragg wavelength} is $f (\text{spacing, } \Delta \text{index})$

Array Waveguide Grating (Optical Phased Array)

- Multiple waveguides at fixed length intervals
- All wavelengths pass through each waveguide, arrive slightly out of phase
- Interference “focuses” individual wavelengths at different points on receiver surface, effectively splitting out individual wavelengths
- Same principle as phased array antenna

Source: Corning
Dispersion Compensation

With Bragg Grating

- Use a “loopback” section of fiber with series of Bragg filters, sending “fast” wavelengths on a longer loop than “slow” wavelengths.

Insert section of Negative NZ-DSF Fiber

Source: Corning
Optical Networks

- "High Performance Nets"
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures
Recap: Multiplexing

- Time-Division Multiplexing (TDM)
- Frequency-Division Multiplexing (FDM)

Source: Ian Foster
Recap: Statistical Multiplexing

- On-demand time-division
- Schedule link on a per-packet basis
- Packets from different sources interleaved on link
- Buffer packets that are *contending* for the link
- Buffer (queue) overflow is called *congestion*
Traditional Transmission Systems

Data transmitted over dedicated cables or fiber

Data transmitted in discrete time slots over a single wavelength
Two Changes: WDM and Optical Amplifiers

Then

72 Regenerators, 8 Fiber Pairs

• One channel per pair of fiber

• Electrical regeneration every ~60km

Now

4 Amplifiers, 1 Fiber Pair

• Wave Division Multiplexing allows all 8 channels to share one fiber pair

• Optical amplifiers allow for greater distances, less equipment in line

Source: Cisco
Wave Division Multiplexing (WDM)

Data transmitted via multiple “colors” (wavelengths, or “lambdas”)…

…on a single strand of fiber.

Charlie Catlett (catlett@mcs.anl.gov)
Attenuation: Loss in Optical Fiber

Region of operation of EDFA’s

Source: Corning
## Dense Wave Division Multiplexing: ITU Grid

### L-Band
(1565-1620nm)

<table>
<thead>
<tr>
<th></th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>186.00</td>
<td>1611.79</td>
<td>186.05</td>
<td>1611.35</td>
<td>186.00</td>
<td>1611.79</td>
</tr>
<tr>
<td>2</td>
<td>186.10</td>
<td>1610.92</td>
<td>186.15</td>
<td>1610.49</td>
<td>186.10</td>
<td>1610.92</td>
</tr>
<tr>
<td>3</td>
<td>186.20</td>
<td>1610.06</td>
<td>186.25</td>
<td>1609.62</td>
<td>186.20</td>
<td>1610.06</td>
</tr>
<tr>
<td>4</td>
<td>186.30</td>
<td>1609.19</td>
<td>186.35</td>
<td>1608.76</td>
<td>186.30</td>
<td>1609.19</td>
</tr>
<tr>
<td>5</td>
<td>186.40</td>
<td>1608.33</td>
<td>186.45</td>
<td>1607.90</td>
<td>186.40</td>
<td>1608.33</td>
</tr>
<tr>
<td>44</td>
<td>190.30</td>
<td>1575.37</td>
<td>190.35</td>
<td>1574.95</td>
<td>190.30</td>
<td>1575.37</td>
</tr>
<tr>
<td>45</td>
<td>190.40</td>
<td>1574.54</td>
<td>190.45</td>
<td>1574.13</td>
<td>190.40</td>
<td>1574.54</td>
</tr>
<tr>
<td>46</td>
<td>190.50</td>
<td>1573.71</td>
<td>190.55</td>
<td>1573.30</td>
<td>190.50</td>
<td>1573.71</td>
</tr>
<tr>
<td>47</td>
<td>190.60</td>
<td>1572.89</td>
<td>190.65</td>
<td>1572.48</td>
<td>190.60</td>
<td>1572.89</td>
</tr>
<tr>
<td>48</td>
<td>190.70</td>
<td>1572.06</td>
<td>190.75</td>
<td>1571.65</td>
<td>190.70</td>
<td>1572.06</td>
</tr>
<tr>
<td>49</td>
<td>190.80</td>
<td>1571.24</td>
<td>190.85</td>
<td>1570.83</td>
<td>190.80</td>
<td>1571.24</td>
</tr>
<tr>
<td>50</td>
<td>190.90</td>
<td>1570.42</td>
<td>190.95</td>
<td>1570.01</td>
<td>190.90</td>
<td>1570.42</td>
</tr>
</tbody>
</table>

### C-Band
(1530-1565nm)

<table>
<thead>
<tr>
<th></th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>191.00</td>
<td>1569.59</td>
<td>191.05</td>
<td>1569.18</td>
<td>191.00</td>
<td>1569.59</td>
</tr>
<tr>
<td>2</td>
<td>191.10</td>
<td>1568.77</td>
<td>191.15</td>
<td>1568.36</td>
<td>191.10</td>
<td>1568.77</td>
</tr>
<tr>
<td>3</td>
<td>191.20</td>
<td>1567.95</td>
<td>191.25</td>
<td>1567.54</td>
<td>191.20</td>
<td>1567.95</td>
</tr>
<tr>
<td>4</td>
<td>191.30</td>
<td>1567.13</td>
<td>191.35</td>
<td>1566.72</td>
<td>191.30</td>
<td>1567.13</td>
</tr>
<tr>
<td>5</td>
<td>191.40</td>
<td>1566.31</td>
<td>191.45</td>
<td>1565.90</td>
<td>191.40</td>
<td>1566.31</td>
</tr>
<tr>
<td>44</td>
<td>195.30</td>
<td>1535.04</td>
<td>195.35</td>
<td>1534.64</td>
<td>195.30</td>
<td>1535.04</td>
</tr>
<tr>
<td>45</td>
<td>195.40</td>
<td>1534.25</td>
<td>195.45</td>
<td>1533.86</td>
<td>195.40</td>
<td>1534.25</td>
</tr>
<tr>
<td>46</td>
<td>195.50</td>
<td>1533.47</td>
<td>195.55</td>
<td>1533.07</td>
<td>195.50</td>
<td>1533.47</td>
</tr>
<tr>
<td>47</td>
<td>195.60</td>
<td>1532.68</td>
<td>195.65</td>
<td>1532.29</td>
<td>195.60</td>
<td>1532.68</td>
</tr>
<tr>
<td>48</td>
<td>195.70</td>
<td>1531.90</td>
<td>195.75</td>
<td>1531.51</td>
<td>195.70</td>
<td>1531.90</td>
</tr>
<tr>
<td>49</td>
<td>195.80</td>
<td>1531.12</td>
<td>195.85</td>
<td>1530.72</td>
<td>195.80</td>
<td>1531.12</td>
</tr>
<tr>
<td>50</td>
<td>195.90</td>
<td>1530.33</td>
<td>195.95</td>
<td>1529.94</td>
<td>195.90</td>
<td>1530.33</td>
</tr>
</tbody>
</table>

### S-Band
(1490 – 1530nm)

<table>
<thead>
<tr>
<th></th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
<th>THz</th>
<th>nm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>196.00</td>
<td>1529.55</td>
<td>196.05</td>
<td>1529.16</td>
<td>196.00</td>
<td>1529.55</td>
</tr>
<tr>
<td>2</td>
<td>196.10</td>
<td>1528.77</td>
<td>196.15</td>
<td>1528.38</td>
<td>196.10</td>
<td>1528.77</td>
</tr>
<tr>
<td>3</td>
<td>196.20</td>
<td>1527.99</td>
<td>196.25</td>
<td>1527.60</td>
<td>196.20</td>
<td>1527.99</td>
</tr>
<tr>
<td>4</td>
<td>196.30</td>
<td>1527.22</td>
<td>196.35</td>
<td>1526.83</td>
<td>196.30</td>
<td>1527.22</td>
</tr>
<tr>
<td>5</td>
<td>196.40</td>
<td>1526.44</td>
<td>196.45</td>
<td>1526.05</td>
<td>196.40</td>
<td>1526.44</td>
</tr>
<tr>
<td>44</td>
<td>200.30</td>
<td>1496.72</td>
<td>200.35</td>
<td>1496.34</td>
<td>200.30</td>
<td>1496.72</td>
</tr>
<tr>
<td>45</td>
<td>200.40</td>
<td>1495.97</td>
<td>200.45</td>
<td>1495.60</td>
<td>200.40</td>
<td>1495.97</td>
</tr>
<tr>
<td>46</td>
<td>200.50</td>
<td>1495.22</td>
<td>200.55</td>
<td>1494.85</td>
<td>200.50</td>
<td>1495.22</td>
</tr>
<tr>
<td>47</td>
<td>200.60</td>
<td>1494.48</td>
<td>200.65</td>
<td>1494.11</td>
<td>200.60</td>
<td>1494.48</td>
</tr>
<tr>
<td>48</td>
<td>200.70</td>
<td>1493.73</td>
<td>200.75</td>
<td>1493.36</td>
<td>200.70</td>
<td>1493.73</td>
</tr>
<tr>
<td>49</td>
<td>200.80</td>
<td>1492.99</td>
<td>200.85</td>
<td>1492.62</td>
<td>200.80</td>
<td>1492.99</td>
</tr>
<tr>
<td>50</td>
<td>200.90</td>
<td>1492.25</td>
<td>200.95</td>
<td>1491.88</td>
<td>200.90</td>
<td>1492.25</td>
</tr>
</tbody>
</table>

*0.8 nm spacing*  *0.4 nm spacing*

Charlie Catlett (catlett@mcs.anl.gov)
Long Haul Transmission with WDM

- **Course Wave Division Multiplexing** (~8 channels per fiber)
- **Dense Wave Division Multiplexing** (>>16 channels per fiber)

- Optical regeneration every ~60-100 km (20-25 dB actually).

- Full electrical “3R” (Re-amplification, Re-shaping, Re-timing) every 4-5 spans or ~3-400 km. Requires O-E-O.

Source: Cisco
Optical Network Cost Components

- **Packets**: IP Routers
- **Wavelengths**: Optronics
  - {amplifiers, regeneration, DWDM terminals}
- **Medium**: Fiber and facilities

- Often vendors will do back-to-back terminals in major hubs to increase port capacity. This increases cost, as transponders are required at each point.

- Largest cost component: interfaces and transponders
Ultra-Long Haul DWDM

Conventional DWDM

400 Km Engineering Rules

4 Regens & 2 End Terminals

Ultra-Long Haul DWDM

2000+ Km Engineering Rules

ULH DWDM reduces or removes regenerators saving ~ 30% of the overall equipment costs

Source: Qwest
ULH DWDM Technical Challenges

Ultra-Long Haul DWDM + Fiber = System performance

Fiber Dispersion

Dispersion Compensation Custom to the Fiber Plant

Other Noise Sources:
- Four Wave Mixing
- Polarization Mode Dispersion
- Splice reflectances
- Amplified Spontaneous Emission

ULH DWDM and Fiber Integration

- Not Plug and Play; Analog link not Digital
- Each fiber optic link is a custom engineered
- In an optically transparent network each $\lambda (W + P)$ path is custom
- Early deployments at half the distances had unexpected results

Source: Qwest
Optical Networks

- “High Performance Nets”
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures
Case Study: TeraGrid

For more information: www.teragrid.org

Source: NCSA
TeraGrid Clusters

- OC192 / 10GbE IP Router
- Juniper T640 Backplane Border Router
- Force10 Cluster Aggregation Switch
- Common Storage
- Fibre Channel: Storage Network
- Myrinet: Messaging Interconnect Network
- Interactive Nodes
- Globus Services Nodes
- Cluster Management Nodes
- TeraGrid National Backplane Network
864-node Terascale Architecture

(a) Terascale Architecture Overview
- Clos mesh Interconnect
- Each line = 8 x 2Gb/s links

(b) Example 320-node Clos Network
- Spine Switches
- 128-port Clos Switches
- 64 hosts

(c) I/O - Storage
- 64 inter-switch links
- 64 TB RAID

(d) Visualization
- Local Display
- Networks for Remote Display

(e) Compute
- 64 inter-switch links
- 100Mbits Switched Ethernet Management Network

Charlie Catlett (catlett@mcs.anl.gov)
Example: 320-node Clos Network

All 8 units at layer n+1

8 hosts, or...
8 units at layer n-1

l = layer
u = unit
p = port

64 hosts  64 hosts  64 hosts  64 hosts  64 hosts

128-port Clos Switches

Charlie Catlett (catlett@mcs.anl.gov)
Initial Distributed Terascale Facility Design

- GbE interfaces in Myrinet root switches
  - Myrinet scaling, GbE interface feasibility

- Layer 2 topology
  - It’s just a big bunch of GbE nodes…
  - Requires spanning tree: sub-optimal routing here

- Lambda switching
  - Immature technology
  - Makes less sense with small number of λ’s

- Fixed Lambda Mesh
  - Very inefficient, inflexible bandwidth allocation

Charlie Catlett (catlett@mcs.anl.gov)
(1) Chicago and LA switches for dynamic topology changes.

(2) Border router or switch/router with 6x10GbE and multiple GbE.

(3) Cluster GbE switch fabric consists of multiple GbE switch-routers. Large clusters will use 2 layers of switches, small may have only one switch. 3x10GbE at the top, nx10GbE at the bottom going out to n bottom layer switch/routers. Bottom layer switch/routers have 10GbE at the top and nxGbE at the bottom, connecting to individual cluster servers.
Physical to Logical Topology for TeraGrid

Phase 0

Physical

# denotes \( \lambda \) count

Pasadena

LA

La Jolla

San Diego

Argonne

Chicago

Urbana

Phase 1

Light Paths

(Logical)

Caltech/JPL

ANL

SDSC/UCSD

NCSA/UIUC

Charlie Catlett (catlett@mcs.anl.gov)
Example Costs: 2000-mile Optical Network

**Optical Transport System**

**Capital Costs**
- One Pair Fiber (2000 mi x 700/mi x 2) = $2.80M
- Optical Amplifiers (every 50mi, 2000 mi, $100k/pair) = $4.00M
- DWDM Base System (10 regen at $100k) = $1.00M
- Four 10Gb/s Lambda ($75k/transponder x 18 x 4) = $5.40M
  Total for 4-lambda system = $13.2M

**Recurring Costs**
- Fiber Maintenance (@$200/mi) = $0.20M
- Equipment space & power ($10k, every 50 mi) = $0.40M
- Amplifier and DWDM Maintenance (15%) = $1.15M
  Total for 4-lambda system = $1.75M

**Integration with LANs**

**Capital Costs**
- IP Routers (2, $400k) = $0.80M
- WAN interfaces (8, $200k) = $1.60M
- LAN interfaces (8, $200k) = $1.60M
  Total for 4-lambda system = $4.00M

**Recurring Costs**
- 15% of capital = $0.60M

**Total 4-Lambda Network**

$17.2M, $2.35M/yr
DTF Network Architecture Options

**One Wilshire** (Carrier Fiber Collocation Facility)

- Los Angeles
- 2200mi from Chicago

**Qwest POP at JPL**

- 15mi from Los Angeles

**Qwest San Diego POP**

- 115mi from Los Angeles

**455 N. Cityfront Plaza** (Qwest Fiber Collocation Facility)

- Chicago
- 140mi from Starlight

**Starlight**

- 25mi from Los Angeles

**Additional Sites and Networks**

- 2mi
- 140mi
- 20mi
- 115mi
- 2mi

**Routers / Switch-Routers**

- DTF Local Site Resources and External Network Connections

**Vendor TBD**

- DWDM (operated by site)

**Vendor TBD**

- Long-Haul DWDM (Operated by Qwest)

- Metro DWDM (Operated by Qwest)

**Vendor TBD**

- CoreStream™ (operated by Qwest)

**DTF Backbone Core Switch**

**Site Border Switch**

**Cluster Aggregation Switch**

- **Caltech Cluster**
- **SDSC Cluster**
- **NCSA Cluster**
- **ANL Cluster**

Charlie Catlett (catlett@mcs.anl.gov)
Optical Networks

• “High Performance Nets”
• Optical Network Building Blocks
• Optical Transport
• Case Studies:
  • TeraGrid
  • I-WIRE
  • National Light Rail
• Architectures and Futures
Case Study: I-WIRE

• State Funded Dark Fiber Optical Infrastructure
  • $7.5M Total Funding 1998-2003
  • Application Driven
    • Access Grid: Telepresence & Media
    • TeraGrid: Computational and Data Grids
  • New Technologies Proving Ground
    • Optical Network Architecture
    • Dense Wave Division Multiplexing
    • Optical network control and management
    • Advanced middleware infrastructure

(Illinois Wired/Wireless Infrastructure for Research and Education)
I-WIRE Geography

- Investigating extensions to Northwestern Evanston, Fermilab, Northern Illinois University (DeKalb), Chicago State University, Milikin University (Decatur), Bradley University (Peoria)

Charlie Catlett (catlett@mcs.anl.gov)
Initial I-WIRE DWDM Systems

I-WIRE commercial hubs also connect AT&T, Norlight, Bell Canada, etc.

Argonne National Laboratory

Production
(Abilene, ESnet, MREN, StarTap…)

Experimental
(TeraGrid, CA*Net4, OmniNet, Europe, Japan…)

Gleacher Center
(University of Chicago)

NCSA

QWest

Level(3)

Starlight
(Northwestern University)

University of Chicago

University of Illinois-Chicago

Illinois Century Network (K-20)

University of Technology

McLeodUSA

5 DWDM Systems
(660 Gb/s capacity each)
I-WIRE Economics

• Initial Costs
  • $4M for fiber
    • 20-year IRU Fiber (existing fiber) ($700 to $5,000/strand-mile)
    • New construction (for ‘last mile’) ($30 to $100/foot ($160-530k/mile))
  • $2.5M for equipment
• Annual costs
  • $100k for fiber maintenance, equipment space/power
  • $250k for equipment maintenance, engineering staff
• Example of Potential Savings to the State of Illinois
  • NCSA: 622 Mb/s Urbana to Chicago, $50,000 per month
    • Replace with 2.5 Gb/s channel (4x capacity) using I-WIRE
      • $35k equipment, ~$k/mo maintenance
• Benefits to the State of Illinois (so far)
  • Chicago area institutions saving $3-400,000/year in Internet costs
  • Leveraged to win $88M TeraGrid project (joint with NCSA and 3 other labs).
  • Creates a unique environment for attracting research funding, world-class researchers, and companies to the Chicago area and to Urbana.
Proposed 2003 I-WIRE Expansion

- McLeodUSA fiber
- Level(3) fiber
- Existing I-WIRE fiber
- AT&T fiber
- Peoria Next Metro network (not funded via I-WIRE)

I-WIRE metro ring (may require add'l site(s))

Charlie Catlett (catlett@mcs.anl.gov)
Optical Networks

- “High Performance Nets”
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures
Case Study: National Light Rail (in progress)

- 7000 mile national fiber footprint
- Four initial lambdas
- Three “networks”
  - IP routed network (1 lambda)
  - 10 Gb/s experimental wavelengths (2 lambdas)
  - 1 Gb/s experimental GbE service (1 lambda)
- Total Cost $800-100M depending on final topology
  - 5 year cost, includes $2-3M/yr operations
- Status
  - Timeframe: 2003-4
  - Funding: Final negotiations with partners, enough momentum (i.e. partners with money committed) to build at least 1/2 of proposed system
NLR PoP Architecture

Black = Included
Gray = Extra

Source: NLR, Cisco

Possible addition of 15530s for 1GE (Layer1) waves
NLR Footprint and Layer 1 (optical) Topology

Source: NLR, Cisco
Evaluation Strategies and Factors

• Cost (Optical portion only)
  • Metric: Cost of lambda per mile per year
  • Examples:
    • *TeraGrid Qwest Partnership: Under $500*
    • *Recent Vendor Quotes: $6-800*
    • *NLR: $6-800 for initial waves, lower for additional waves*

• Notes
  • Lambdas are incremental relative to base infrastructure
  • More lambdas, lower per-lambda cost

• Options
  • Lease from Vendors (see recent vendor quotes above)
  • Build Dedicated
    • *Low cost for many (>16) lambdas, high initial cost*
  • Build with Consortium
    • *Share high initial cost with others*
  • Build with Vendor
    • *Leverage vendor infrastructure to lower initial cost*
Selecting Fiber Sources

• Footprint Routes and Coverage
  • AT&T goes North, others cut across middle (skipping Northern Plains). Inter-city route distances vary.

• Fiber types
  • Level(3) is homogeneous LEAF fiber. AT&T is multiple types. Any impact? (maybe for ULH)

• Repeater hut spacing
  • AT&T spacing is closer. More amplifiers. But at higher speeds will closer spacing be critical?

• Space and Power costs, access policies
  • 250+ locations on a 16,000 mile footprint

• Metro space locations
  • relative to regional/local fiber, other vendors (carrier neutral colo or single vendor controlled?)

• Fiber Costs and Terms
  • IRU, $/strand-mile, willing to “strand”? $/mile maintenance (strand-mile or route mile?)
  • Legal issues- indemnification, terms for “forced relocation,” bankruptcy contingency!

Context: Fiber build costs range from $30-100/ft (up to $500k/mile). If you only want 2 strands… cost ranges from $1000-5000 per mile.
Optical Networks

- “High Performance Nets”
- Optical Network Building Blocks
- Optical Transport
- Case Studies:
  - TeraGrid
  - I-WIRE
  - National Light Rail
- Architectures and Futures
End-to-End Architecture Issues

- **Today’s High Performance Production Networks**
  - Packet Switched: IP Routers
  - Fixed topology between routers
    - *Optical networks used for fixed point-to-point transport*
    - *No ultra-long haul in production, anywhere*
    - *DWDM systems use fixed-wavelength optics*
    - *Provisioning takes hours to days*
  - Maximum 10 Gb/s Individual Channel speeds
    - *Typically 2.5 Gb/s backbones*
    - *3-5 10 Gb/s backbones in service as of late 2002*
    - *1 multi-10 Gb/s backbone, TeraGrid at 40 Gb/s*

- **Testbeds and Proven in the Laboratory**
  - Lambda Switching/Routing
  - Tunable lasers
  - Optical Burst Switching
Optical Bandwidth on Demand

- **Digital Theater**
  - Each movie theater in a large area (SF, New York, Houston) requests 1 hour of bandwidth a week (OC192)
  - All movies transferred during this time
  - Efficient use of expensive but necessary fat pipe

- **Sporting Events**
  - E.g. Football Stadiums - need high-bandwidth 8-10 times/year
  - Today’s answer: drive a truck to the event with a satellite dish
    - *OK for today’s model*
  - Experiments with multiple HDTV streams
    - *Too much for satellite*
    - *Experiments in stadiums with user-selectable camera view*

Source: George Porter, Tal Lavian UC-Berkeley
Many-to-Many Interactive Collaboration

- **Access Grid**
  - Enable collaborative work at dozens of sites worldwide, with strong sense of shared presence
  - Combination of commodity audio/video tech + Grid technologies for security, discovery, etc.
  - 50+ sites worldwide, number rising rapidly

- **Exploration of Game Consoles**
  - PS2 or Xbox as Access Grid node

- **Network Flow Engine**
  - Heavy Dependence on Multicast, moving toward SSM
  - Need to break large streams into multiple formats (variable quality), bind some streams, present participants (sites) with information about streams, etc.

Source: Rick Stevens, ANL
"Cause Computing"

**Mathematical Research**

Push the limits of our theoretical understanding of the abstract mathematics tools that civilization is built upon and explore these new frontiers with your computer.

**Fighting Diseases**

Striving to improve quality of life and eliminate suffering, Entropia members can support disease research projects.

**Economics Research**

Long term stability of the world economies has become crucial to growing prosperity. Entropia members can help illuminate the meanings of global economic behavior through rigorous research models.

**Environmental Research**

Responsible management and preservation of Earth's environment requires a deep understanding of the complex effects of many factors. Entropia members can help researchers determine the most important aspects of planet stewardship for future generations.

**Entertainment**

Entropia is building new technologies to help accelerate wonderful new digital entertainment productions. Your computer can be a key part of bringing this exciting new technology to life!

**Scientific Research**

Science is a foundation on which many of the greatest human achievements rest. The Entropia community can participate in some of the most intense scientific research underway today.

**Product Design**

Researching safe product designs quickly and effectively requires an ever increasing amount of computing power to test and refine them before manufacturing even begins. Your computer can play a crucial role in making safer medicines, transportation, appliances, clothing, toys and more!

Source: Entropia
Medical Simulation: Remote Visual Steering

Example Combustion C-SAFE SCIRun Network

Source: Chris Johnson, Univ. of Utah
The Brain Data Grid

Objective: Form a National Scale Testbed for Federating Large Databases Using NIH High Field NMR Centers

Cyber Infrastructure Linking Tele-instrumentation, Data Intensive Computing, and Multi-scale Brain Databases.

Source: Mark Ellisman and Larry Smarr, UCSD
Provisioning and Guaranteed Bandwidth

- **MPLS**
  - Multiprotocol Label Switching (MPLS)
  - Routers set up Label Switch Paths through the network
    - *Collections of routes are fast-tracked through these paths without the router having to do a route lookup.*
  - Assumes large shared pipes, MPLS paths of some smaller amount of bandwidth

- **Example - TeraGrid**
  - Application running on 100 cluster nodes at SDSC and 200 cluster nodes at NCSA
  - Set up MPLS path for all traffic between these nodes, e.g. at 25 Gb/s. Leave remaining bandwidth for general traffic

- **Future: G-MPLS**
  - Generalized MPLS (aka MP\$S)
  - Routers set up optical paths through the network
  - Assumes large number of available optical paths
Evolution of IP Routers

Source: P. Kaiser
Canarie’s approach

- OBGP (Optical BGP)
- Routers spawn “virtual BGP” processes that peers can connect to
- By modifying BGP messages, lightpath information can be traded between ASes
Canarie Virtual Routers and OBGP

1) BGP OPEN message sent to router with information about optical capabilities
   • A virtual BGP process is spawned
   • A BGP session is initiated independently with new BGP process
   • The virtual process (running on the router) configures the OXC to switch the proper optical wavelengths

2)
Optical BGP Networks

Figure 12.0
Source: CANARIE
Burst Switching

• Separate control and data channels
• Burst Header Cell sent to set up next-hop path. BHC contains offset (time from first bit of BHC to first bit of Burst) and length (of Burst)
• Without optical buffering, critical design point is to avoid blocking.

Source: Jonathan Turner, Washington University St. Louis
At Issue for *Optical* Burst Switching: Blocking

Source: Jonathan Turner, Washington University St. Louis
Futures- Optical Chip Interconnects

Example Optoelectronic Device for Chip Interconnects
- Quantum Well Electroabsorption Modulator

essentially no internal speed limitations
no “threshold”
successfully fabricated and bonded in large arrays (e.g., 4000)
use in transmission or, with internal mirror, reflection
same mechanism as high-performance modulators on externally modulated lasers (EMLs)

Source: David Miller, Stanford
Optical Module of the Future

Hypothetical module with
- silicon CMOS chips with optoelectronic devices and lenslet arrays
- electrical wiring and waveguide layer
- massively parallel free space optics within and between chips
- flexible fiber connections off the module

T. J. Drabik and D. A. B. Miller, Stanford
2/15/01

Source: David Miller, Stanford
WDM to the Chip

Short (e.g., 100 fs) pulse is broad band (e.g., 10 nm wavelength range) source
- spread wavelengths over reflective modulator array
- send reflected signals over single fiber to receiver array

Multiple channel interconnect with single fiber and single laser


2/15/01

David A. B. Miller, Stanford

Source: David Miller, Stanford
More Information (from SURA OpCook)

- **SURA Optical Network Cookbook** – [www.opcook.sc.edu](http://www.opcook.sc.edu)

- **Books**
  - "Understanding Fiber Optics" by Jeff Hecht, Prentice Hall; ISBN: 0130278289

- **Web**
  - Light Reading Beginner’s Guides
    - [www.lightreading.com/](http://www.lightreading.com/)
  - Corning Cable Systems Basic Principles of Fiber Optics
  - Illustrated Fiber Optics Dictionary
    - [www.fiber-optics.info/glossary-a.htm](http://www.fiber-optics.info/glossary-a.htm)
  - How Fiber Optics Works
    - [www.howstuffworks.com/fiber-optic.htm](http://www.howstuffworks.com/fiber-optic.htm)
  - Introduction to Fiber Optics
    - [www.commspecial.com/fiberguide.htm](http://www.commspecial.com/fiberguide.htm)
  - Optics for kids
    - [www.opticalres.com/kidoptx.html](http://www.opticalres.com/kidoptx.html)