

# An Introduction to Fibre Channel SANs

Presenter:

**Mel Tsai**

[mtsai@eecs.berkeley.edu](mailto:mtsai@eecs.berkeley.edu)

11/20/2002

# Outline

1) Fibre Channel Basics

2) SAN Design

- From Brocade's "SAN Design Guide"

# The Goals of Fibre Channel

- A high-performance communications protocol & physical transport
  - 100 and 200 MBytes/sec today
- “Flexible” topologies that range up to 10-120 km
- Support **any** higher-level protocol you want
  - Currently **SCSI** and **IP** are the popular upper-layer protocols
  - You can use FC to completely **replace** SCSI’s physical transport
    - FC Disks, FC HBAs, etc.

# Fibre Channel Media

- Spelling changed to the French “**fibre**” instead of “fiber” because support for **copper links** were added
  - Today’s 100 Mbyte/sec FC networks can be **either** fiber or copper
  - 200 Mbyte/sec links are **always fiber**.
  - Many 200 Mbyte/sec FC switches are auto-sensing

# FC Connections

- **Point-to-Point**
  - No media sharing, highest performance
  - Least flexibility
- **Arbitrated Loop**
  - Can connect 127 devices (a 1-byte ID) without an FC switch
  - Not token passing... Devices arbitrate and gain control of the loop, then it becomes a point-to-point link
  - Bandwidth is shared among all devices
- **Fabric** (most interesting topology for SANs)
  - Very flexible, devices are addressed by a 3-byte ID ( $2^{24}$  nodes)
  - Requires **FC switches** to connect devices

# FC Initialization: Login

- When a node (i.e. a disk or host) connects to the fabric, it “logs in” to the known fabric address 0xFFFFFE
- Fabric responds by assigning the node a dynamic 3-byte ID
- Initializes flow control & class of service

# Flow Control

- Receiving nodes **cannot** always process data at the transmission rate...  
Need **flow control**.
- Simple flow control mechanism in FC:
  - Receiver tells sender how much buffer space it has (“**buffer credit**”), and vice-versa
  - When buffers **run out**, they must be **renegotiated**
  - Two types of flow control
    - **buffer-to-buffer, end-to-end**

# Class of Service

- Class of service (established during login)
  - Class 1:
    - Full bandwidth allocated, in-order delivery guaranteed
  - Class 2:
    - More like a LAN: connectionless transmission, dropped frames okay but get a notification, shared bandwidth among other traffic
  - Class 3:
    - Similar to class 2, only one flow control method allowed, used when upper-layer protocol guarantees transmission (SCSI)
  - Class 4:
    - Establishes a bandwidth-limited Virtual Circuit (VC) path across the fabric, used in switched topologies
  - Class 5:
    - Defunct?
  - Class 6:
    - Multicast capabilities

# FC's five-layer stack

- FC-0:
  - physical media specs
- FC-1:
  - 8B/10B character encoding
- FC-2:
  - Framing (up to 2148 bytes/frame), flow control, class of service
- FC-3:
  - Login, topologies, SAR
- FC-4:
  - Multiple-port services on one node
- FC-5:
  - Upper-Layer Protocol (ULP)
  - Can be SCSI, IP, HIPPI, ATM, IPI-3, SBCCS, etc.

# Port Terminology

- Node connections: **N\_Ports**
- AL-capable ports: **L\_Ports**
- Fabric ports: **F\_Ports**
- Switch-to-switch ports: **E\_ports**
- Dual-purpose fabric/switch-switch ports: **G\_ports**
- Variations
  - **NL\_Ports, FL\_ports, GL\_ports**

# FC-based SANs

- **Why FC?**
  - High-performance, relatively simple, connection-oriented
  - SCSI naturally fits on top of FC
    - Arbitrated-Loop FC can completely replace SCSI physical transport
- **Why SANs?**
  - Storage consolidation
  - LAN-free backup
  - Clustering
  - High Availability

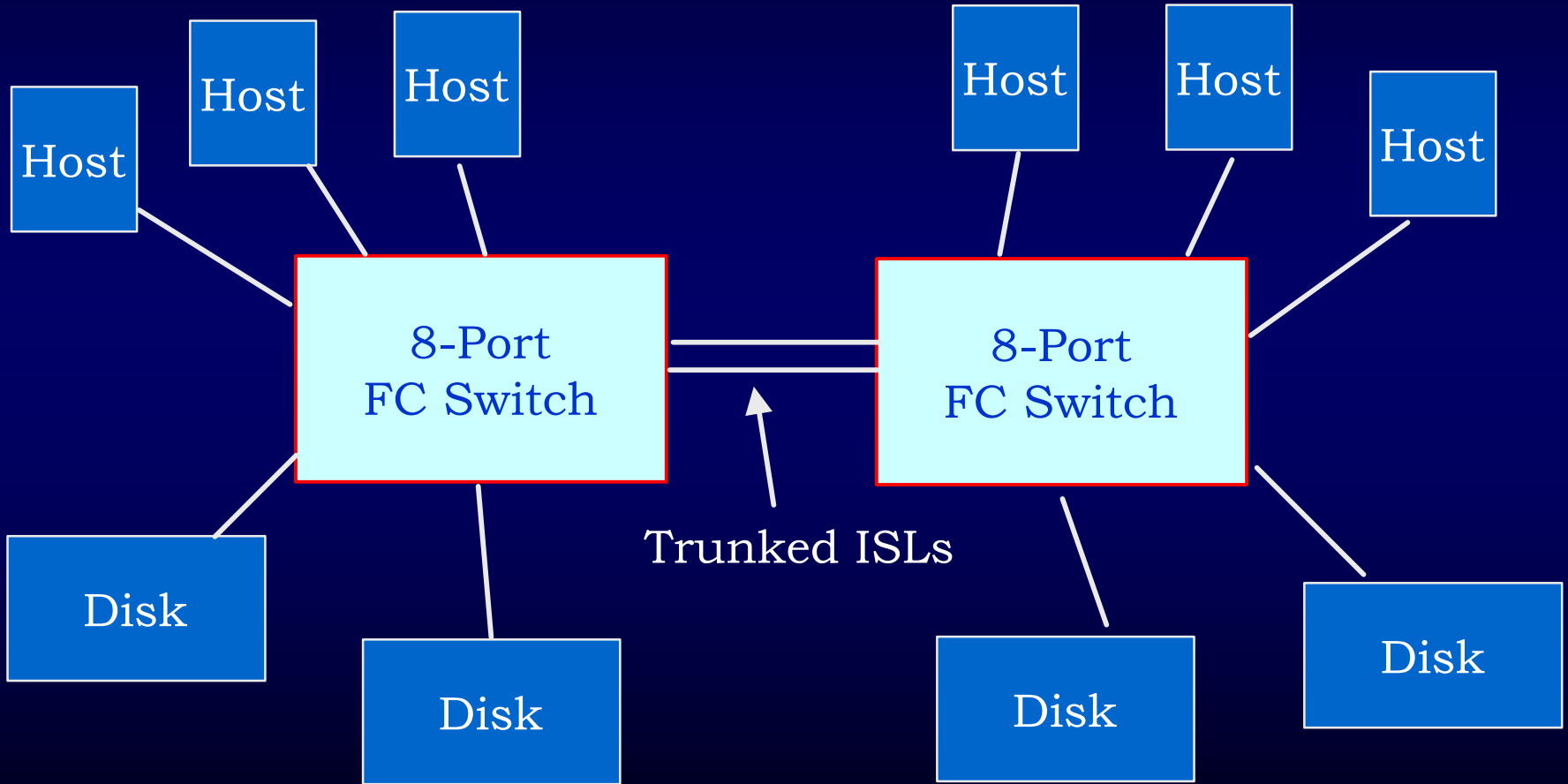
# SAN Fabric Design

- The SAN “Fabric” is everything that **connects** the hosts and disks
  
- The (basic) building blocks of an FC SAN
  - Copper/fiber **cables**
  - **Fabric Switches** (8, 16, 32, 64+ port switches)
  - **Redundant** power supplies & power feeds
  - Disk & tape drives with **FC interfaces**
  - Hosts & servers with **FC HBAs**

# Important Terms

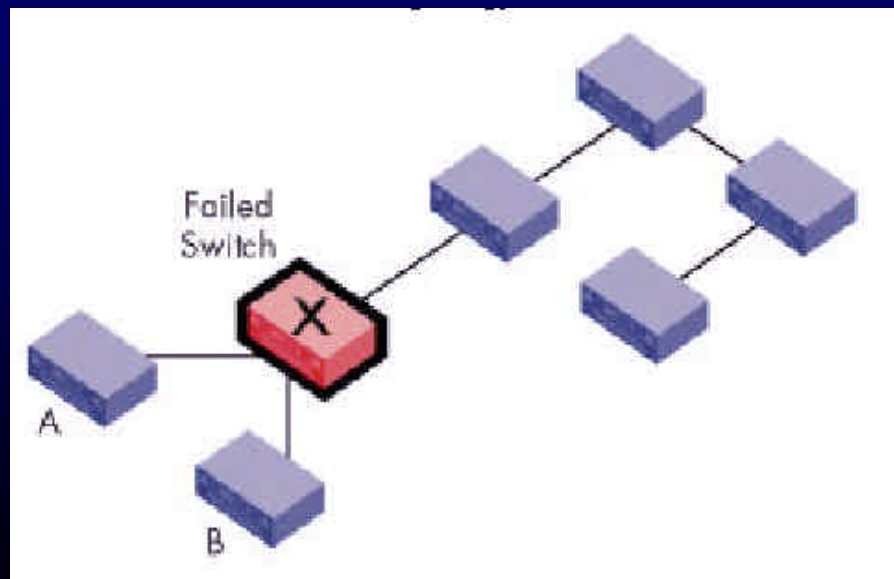
- **Fabric topology**: the layout of the SAN
- **Core/edge switch**: a “hop” in the SAN
- **Hop count**: # of switches a frame must traverse
- **ISL**: inter-switch link, connection between switches
- **Fan-in**: many storage devices to one host
- **Fan-out**: many hosts to one storage device
- **Blocking** vs. **Congestion**
- **Locality**: SAN traffic tends to cross ~1 hop
- **Tiered**: Devices of the same type are organized in the same location (i.e. on the same switch)
- **SAN Port Count**: Total # of available ports to connect nodes

# Basic 12-port SAN



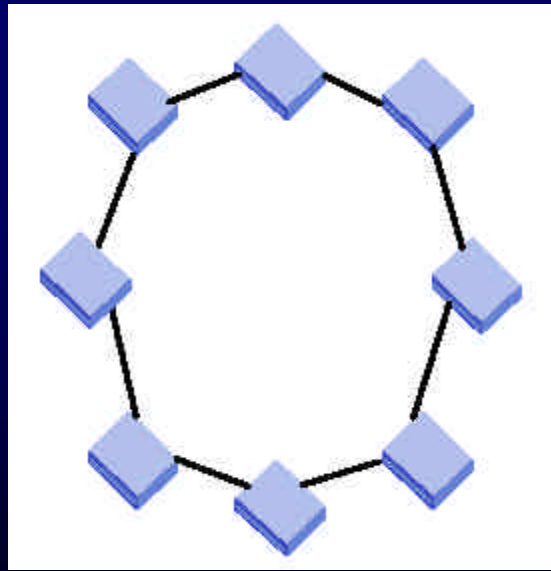
# Cascaded Topology

- **Inexpensive**, easy to expand
- Low reliability, low scalability
- Good for **localized traffic**
- **114** max ports for eight 16-port switches



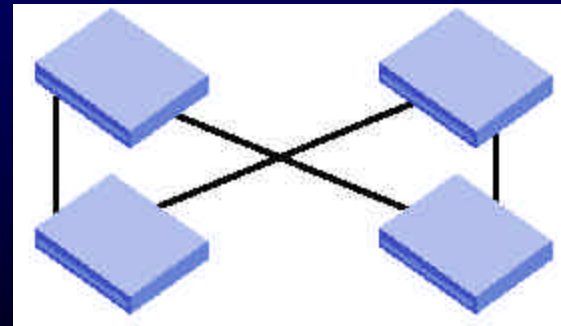
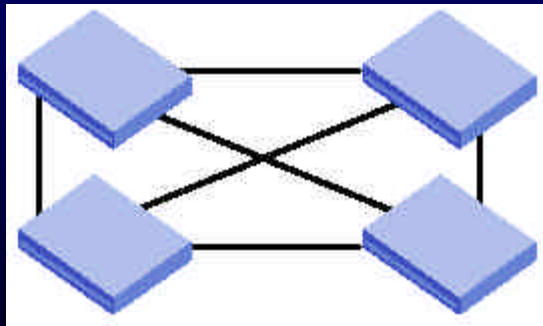
# Ring Topology

- Similar cost to cascaded, better reliability, good for small SANs
- 112 max ports for eight 16-port switches



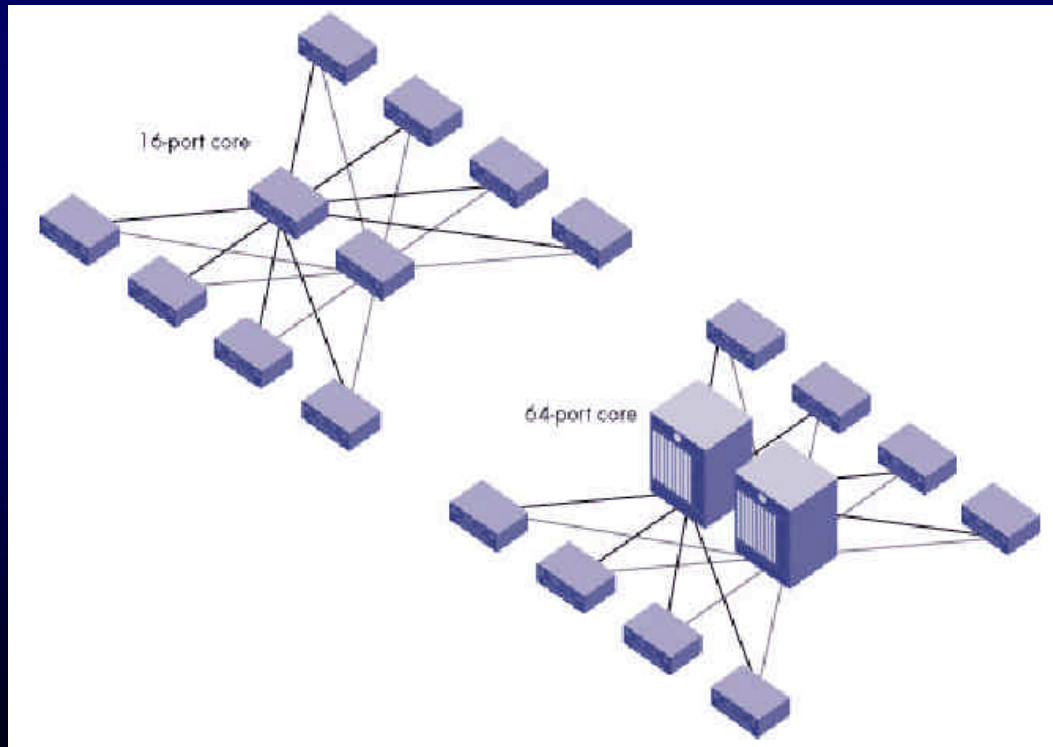
# Full/Partial Mesh

- Good for any-any traffic
- Not very scalable (**ISLs use up valuable ports**)
- 72 max ports for full mesh

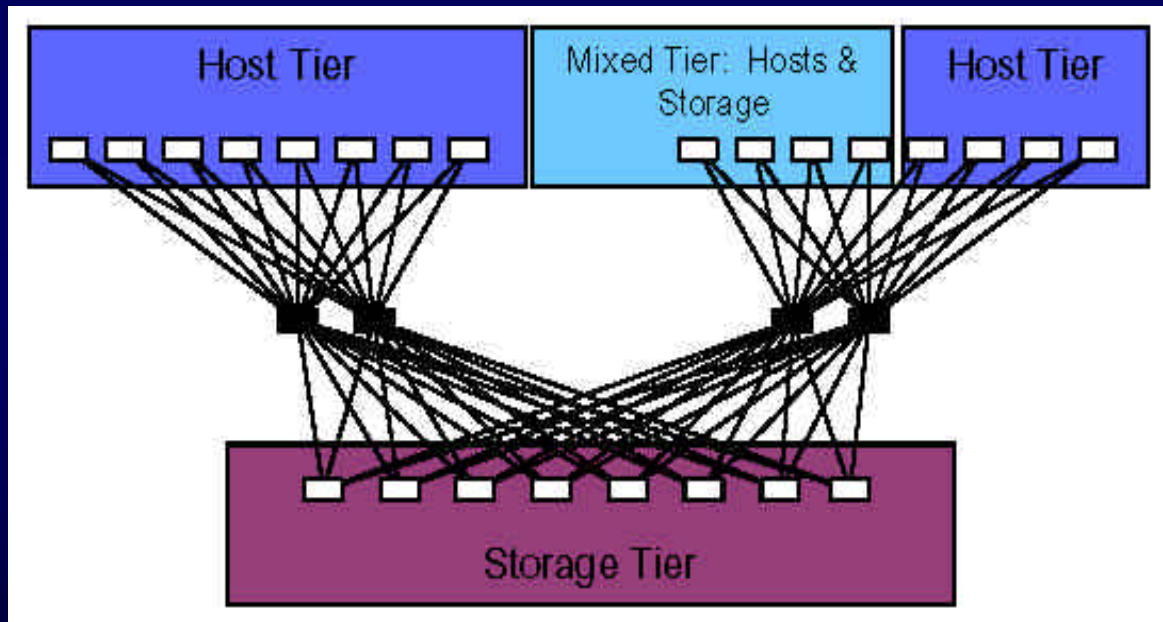


# Core/Edge Topology

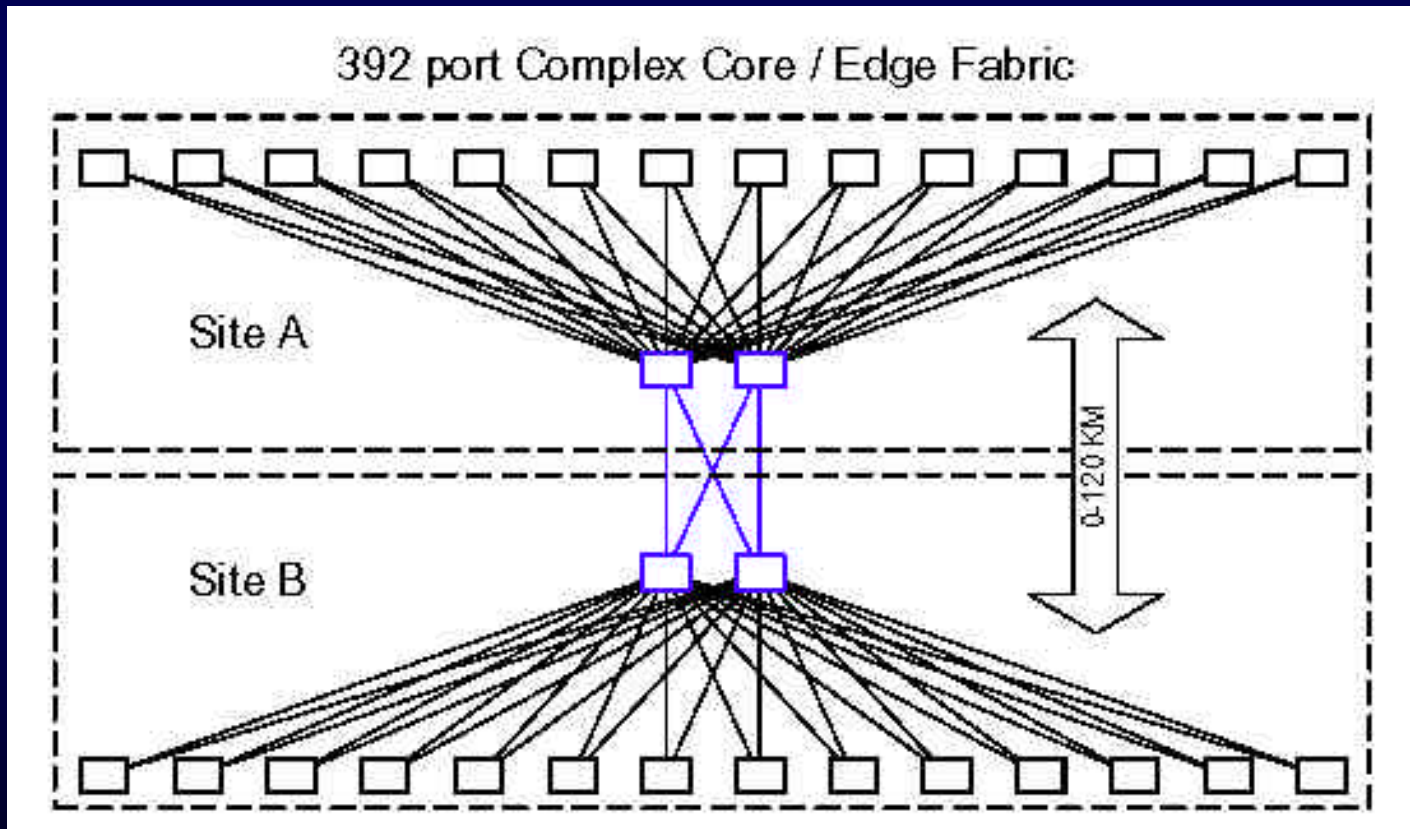
- Best scalability, reliability, flexibility



# Tiered Hybrid Core/Edge Topology



# Complex Core/Edge Topology



# SAN Goal 1: High Availability

- “Single” vs. “Multiple” Fabrics
  - One network vs. redundant/independent networks
  - You can put **two** FC HBAs on a PC...
- “Resilient” vs. “Non-Resilient” Fabrics
  - Non-resilient: a **single point of failure** can bring the fabric down, e.g. cascaded topologies
  - Resilient: **two** or more failures required, e.g. ring topology, mesh topology

# SAN Goal 2: Scalability

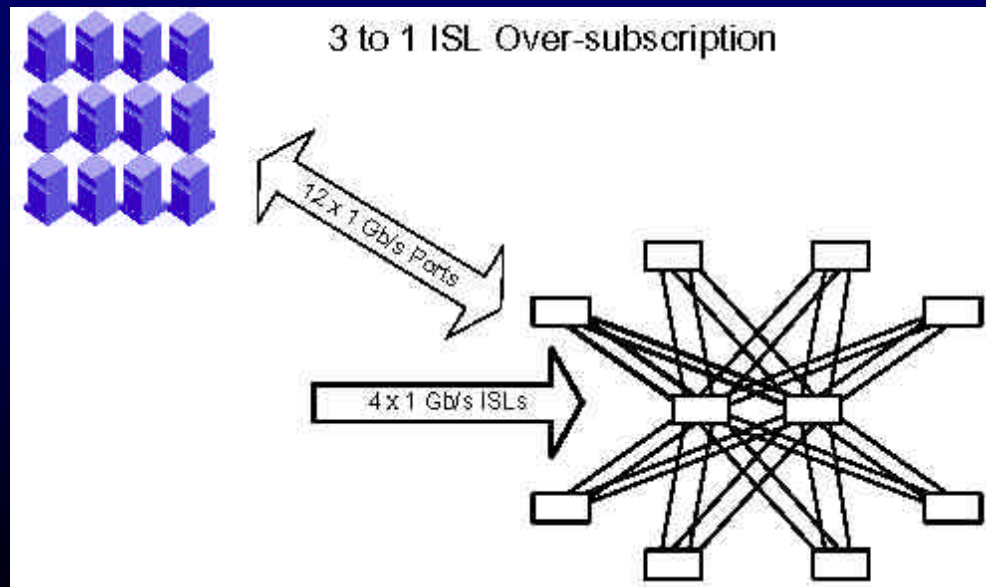
- Different topologies support different levels of scalability
  - Arbitrated loops
  - Cascaded topologies
  - Rings
  - Full meshes
  - Core/edge topologies
- Investment Protection
  - If you replace switches in the core with faster ones, can you move it to the edge?

# SAN Goal 3: High Performance

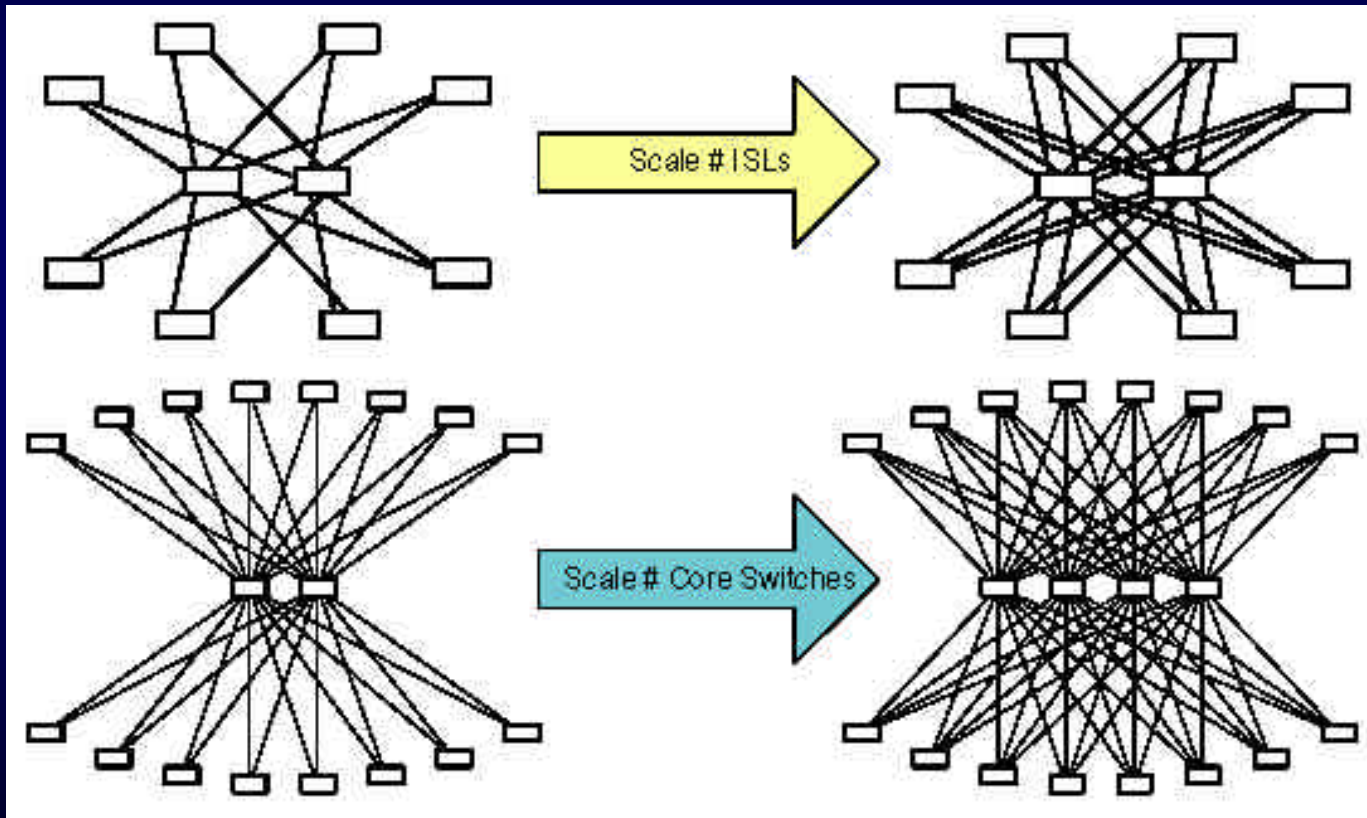
- **Locality** is usually good, but not always required
  - Misconception that hops introduce latency
    - **Latency is a non-issue for SANs!**
  - Tiered approaches have poor locality, but **simplify management** without much cost in performance
    - But this depends on the application
- Bigger concern: **Bandwidth**
  - Need to worry about **ISL Over Subscription**

# ISL Over-Subscription

- To lower costs, must **statistically multiplex** bandwidth requirements over the ISLs
- ISL Over-Subscription (and storage fan-out) is **usually 7:1**
- Worst-case is 15:1 for a 16-port switch



# Bandwidth Scaling of Core/Edge



# Latest Brocade Switches



**Silkworm 3900**  
1.5U 32-port  
(~\$50,000)



**Silkworm 3800**  
16-port  
(~\$20,000)



**Silkworm 3200**  
8-port  
(~\$8,000)



**Silkworm 12000**  
64 port  
(~\$150K)  
128 port  
(~\$250K)

# Coming soon....?

- FC over WAN techniques: ATM, iFCP, FCIP
- Storage virtualization
- Overview of SAN vendors & available products

# Image Sources & References

- Interoperability Lab's "**Fibre Channel Tutorial**,"  
[http://www.iol.unh.edu/training/fc/fc\\_tutorial.html](http://www.iol.unh.edu/training/fc/fc_tutorial.html)  
→ What: info on FC basics
- Brocade's excellent 69-page "**SAN Design Guide**"  
→ What: images & information for basically the entire 2nd half of this presentation