

A Topologically Optimal Internet

Alan Huang, Ph.d. '81

and

Scott Knauer, Ph.d. '80

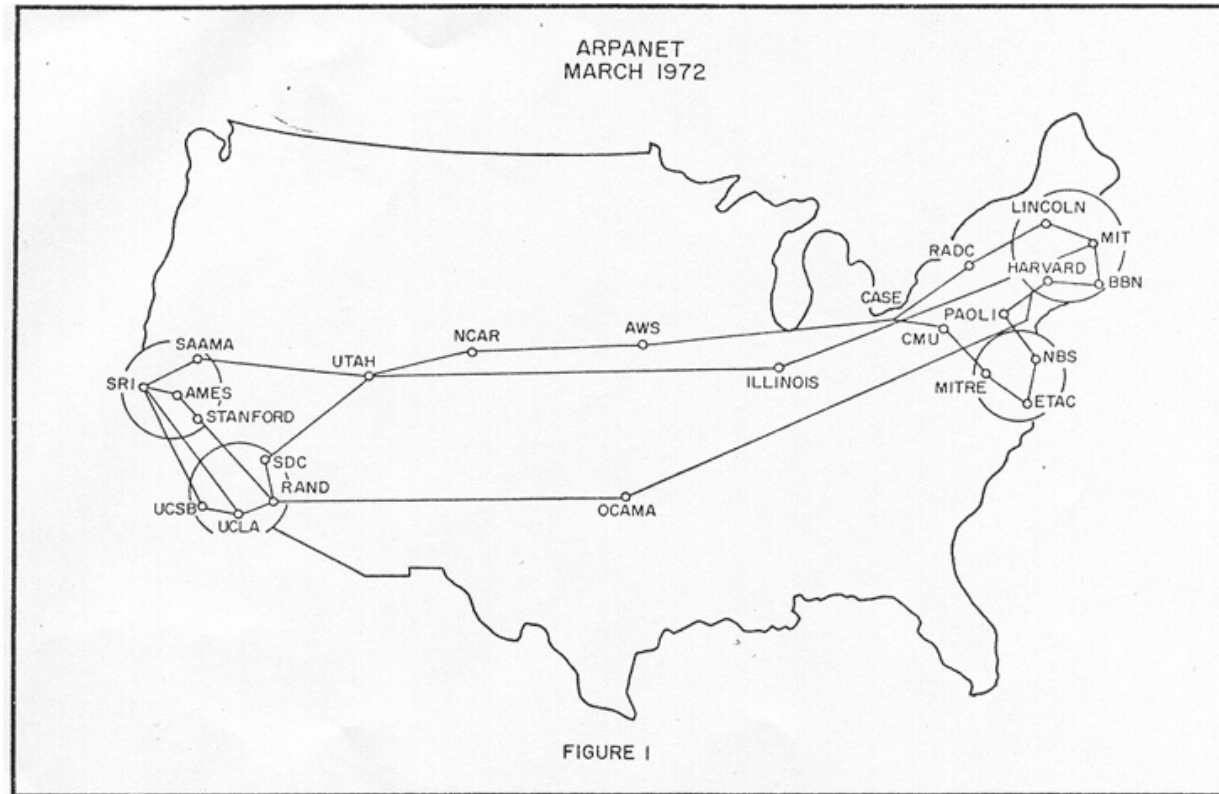
September 28, 2016

Circuit Switched Network



- Circuit Switched networks are optimized to minimize total link length

Packet Network



- Packet Networks should be optimized to minimize the number of Hops

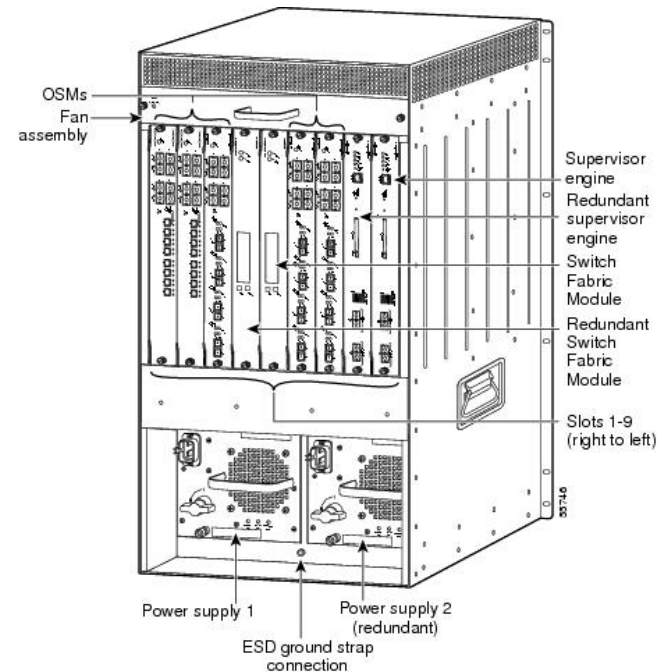
Unfortunately, the Packet Network was Superimposed on the Circuit Switched Network



- The Network should be designed to minimize the number of Hops and not minimize the total path length

Why are the number of Hops so important?

	<i>Latency</i>	<i>Power</i>	<i>Cost</i>
<i>Buffering</i>	X	X	X
<i>Processing</i>	X	X	X
<i>Routing</i>	X	X	X



How can the number of Hops be Reduced?



- By transforming the network into a toroid the number of hops can be reduced by 2X.

Where is all the additional bandwidth and connectivity going to come from?

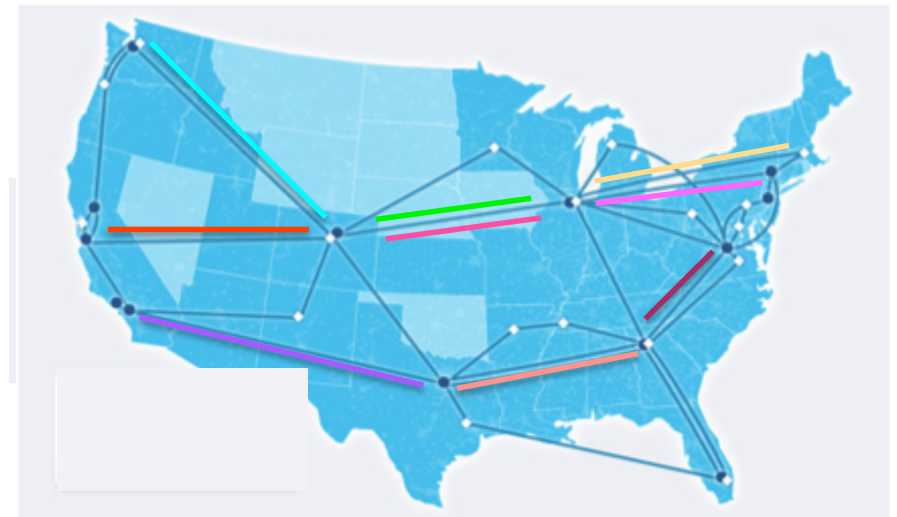


- The toroidal connections can be created by re-purposing existing links

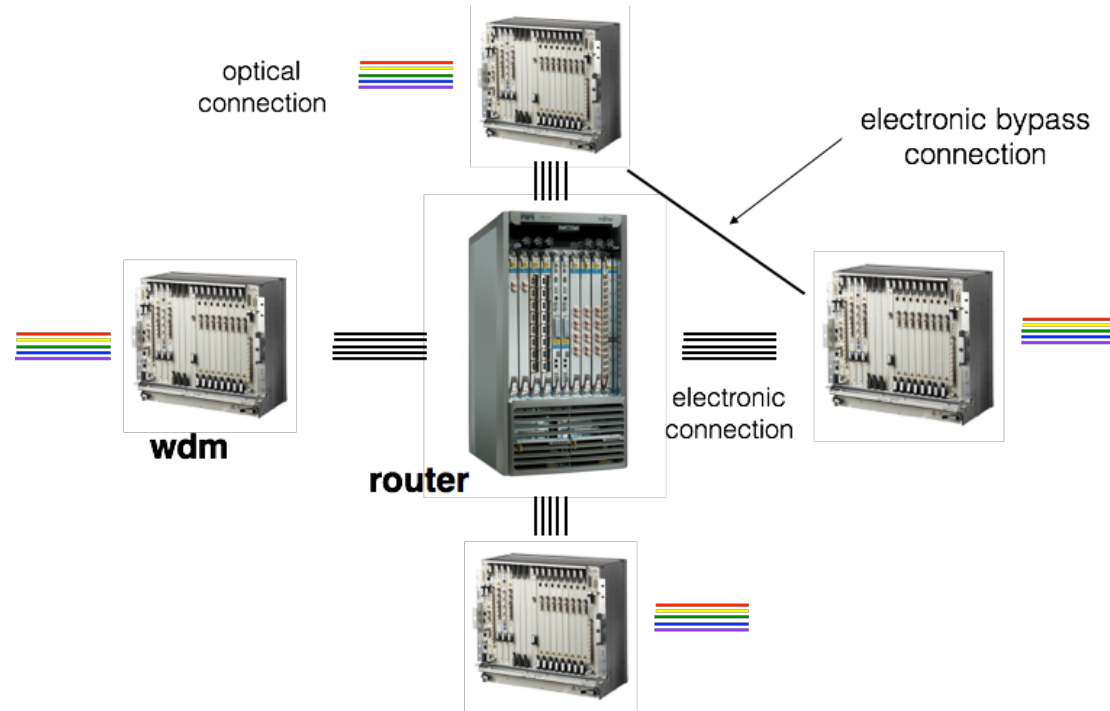
Re-purposing Existing Bandwidth and Connectivity

Steps:

- Add enough toroidal bandwidth to divert half of the current traffic
- Move the toroidal connections to the bandwidth previously used by the diverted traffic
- Eliminate the initially added toroidal bandwidth

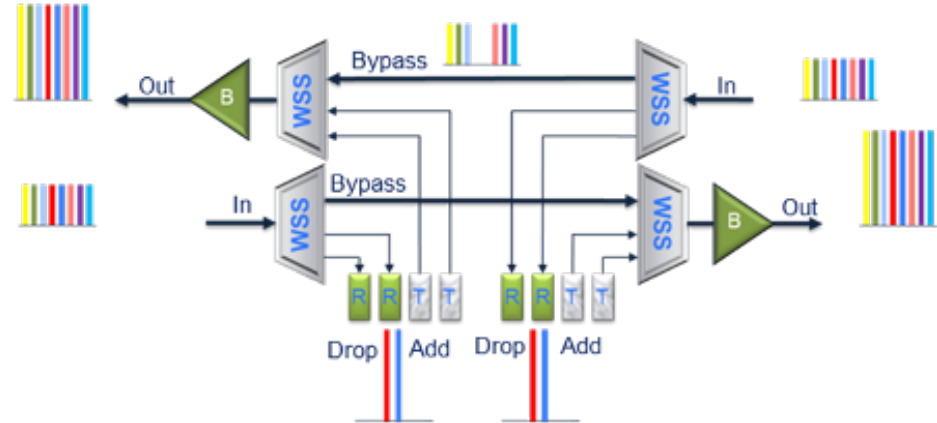


How are you going to create the toroidal connections?



- Each Bypassed Node eliminates the need for two Router Ports.
- This means that you can halve the number of hops while eliminating half the router ports. Less is more.

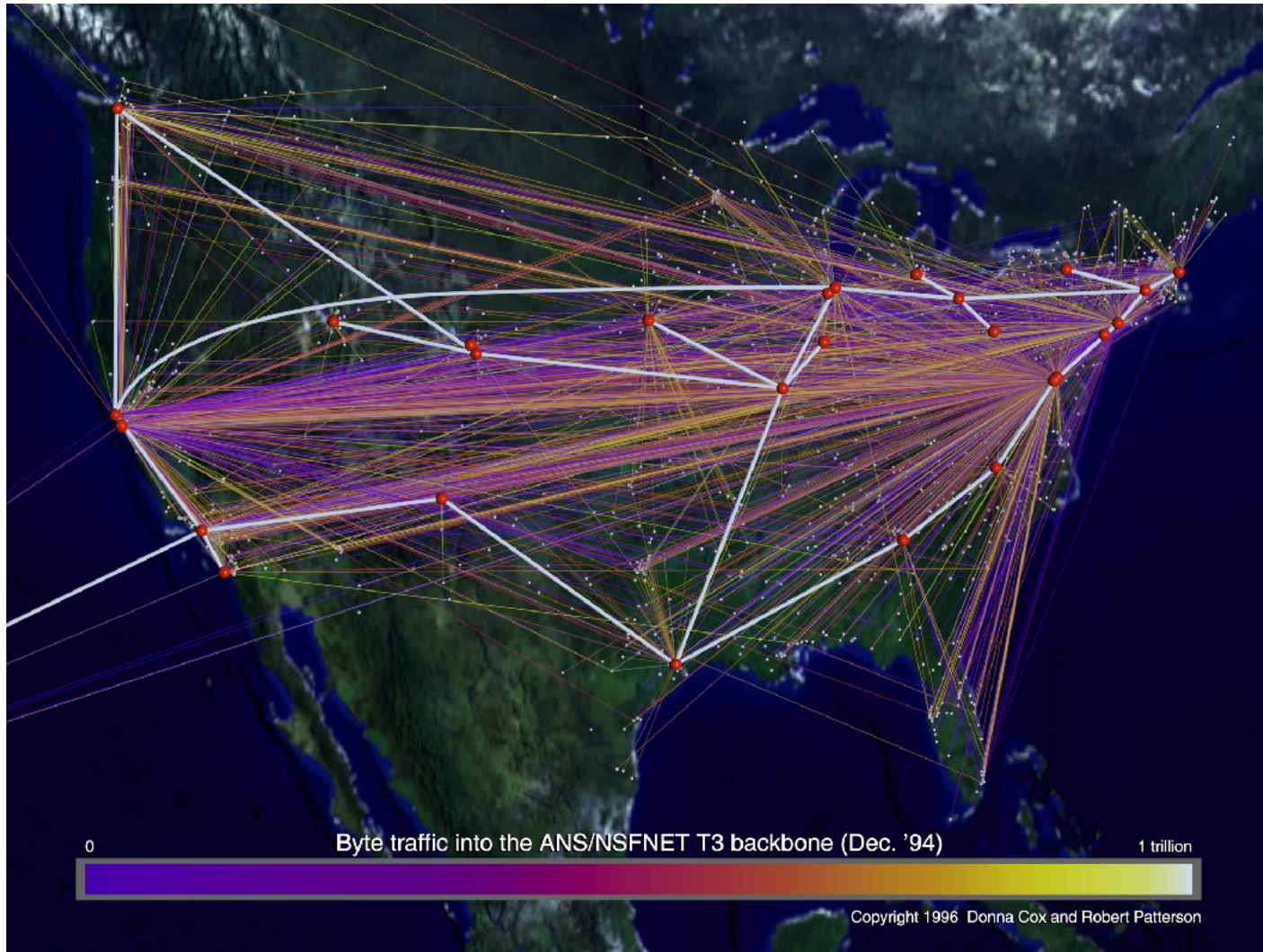
Continued: How are you going to create the toroidal bypass connections?



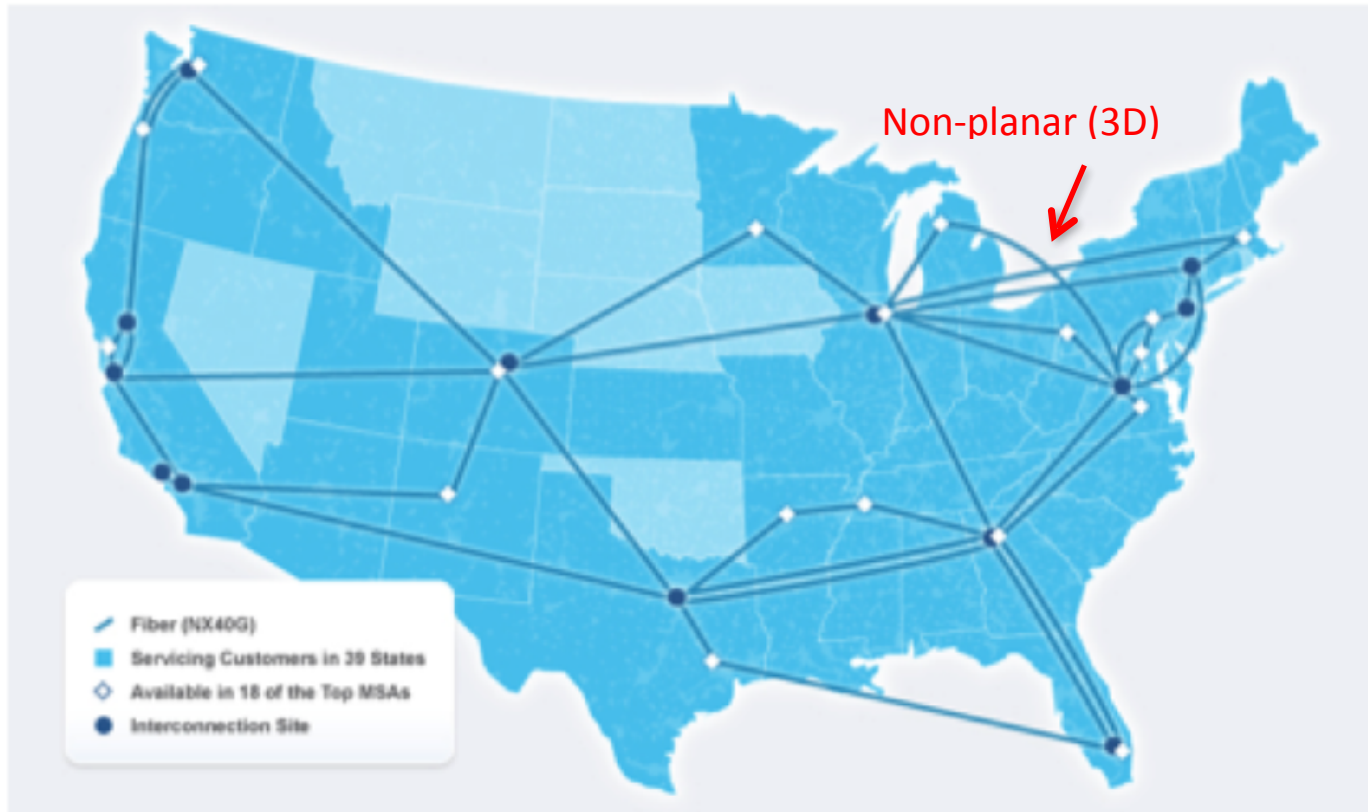
Some nodes are on Rings:

- This means you can create a bypass connection by reconfiguring a Reconfigurable Optical Add Drop Multiplexer (ROADM).
- This eliminates the need for both a transceiver port and two router ports.

Importance of Toroidal Connections

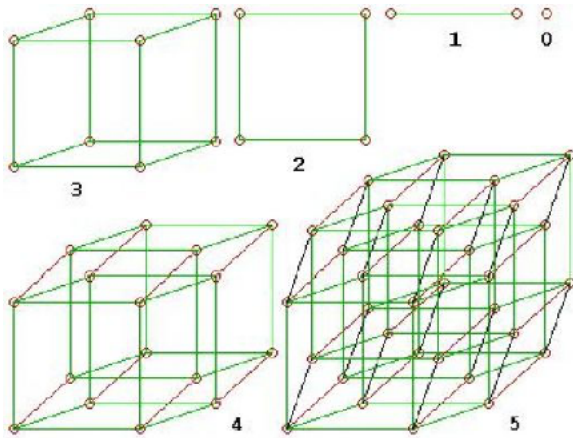


Can the number of Hops be reduced further?



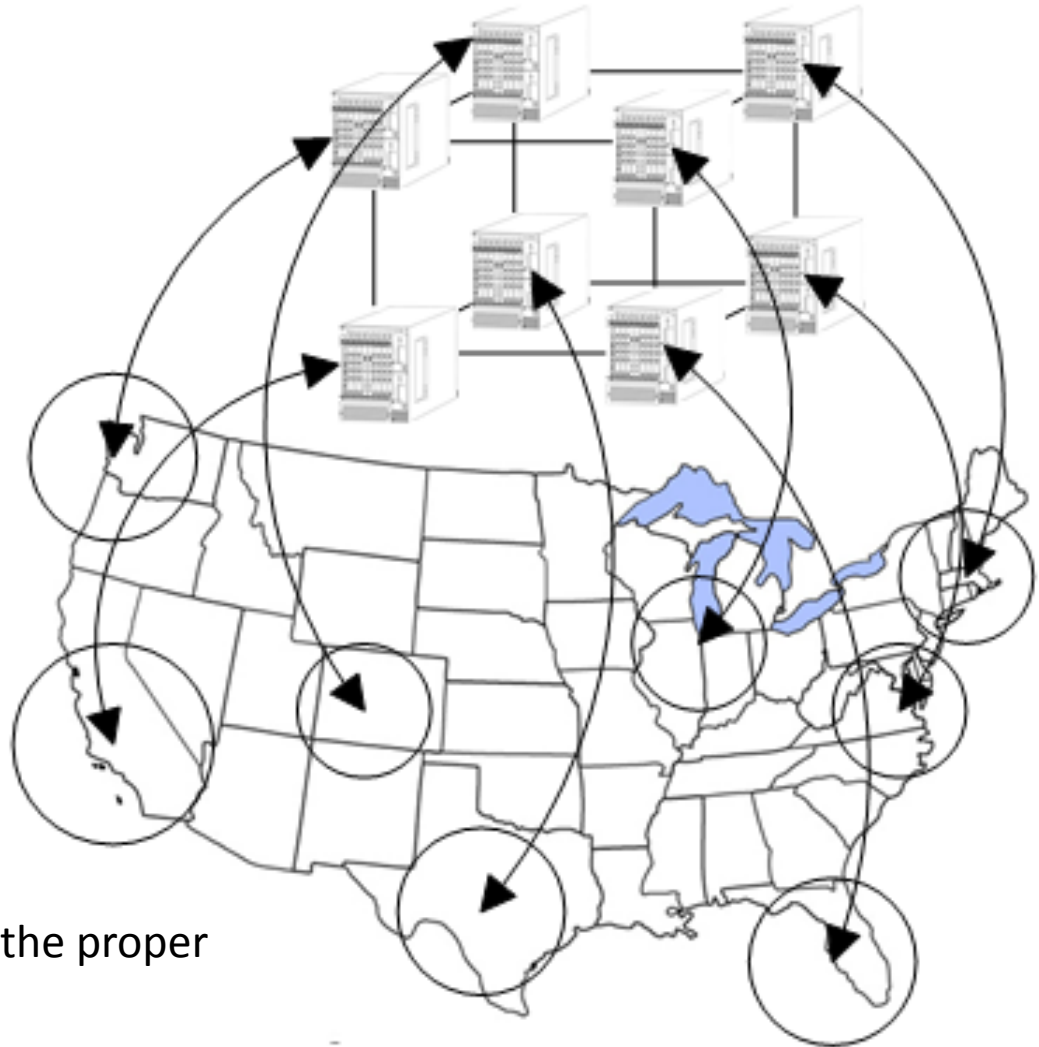
- The number of Hops can also be reduced by using a higher dimension topology such as a hypercube.

A Cube Based Network

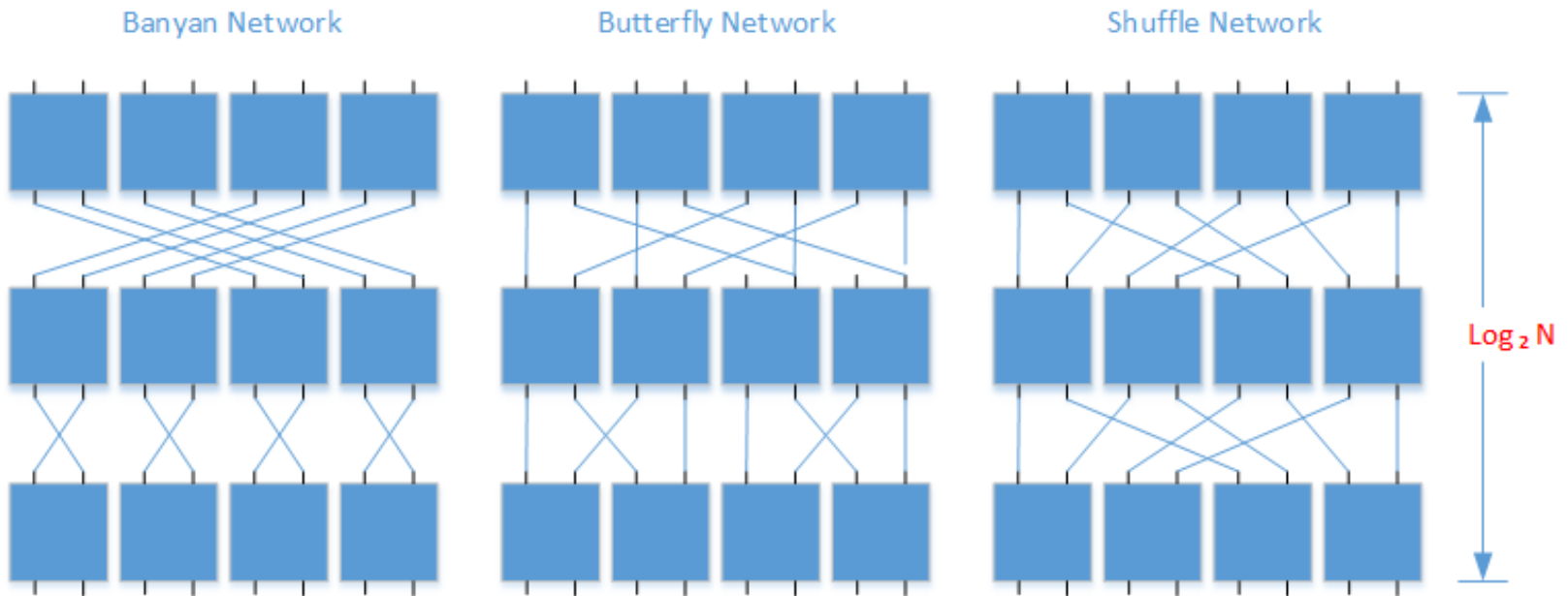


Steps:

- assign a city to each node
- “pancake” the cube
- “rubber band” each node to the proper geographic location

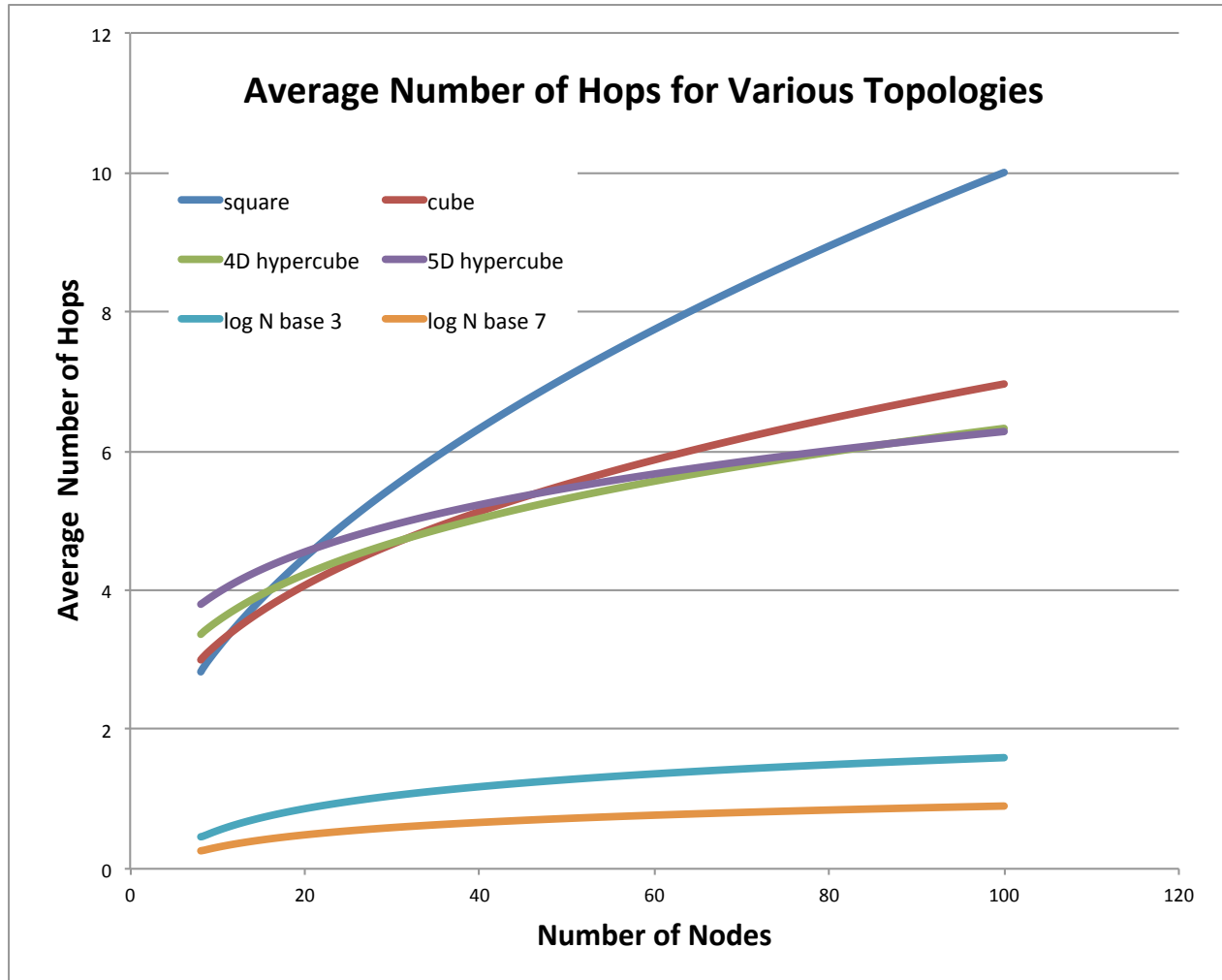


Multistage Switching Networks

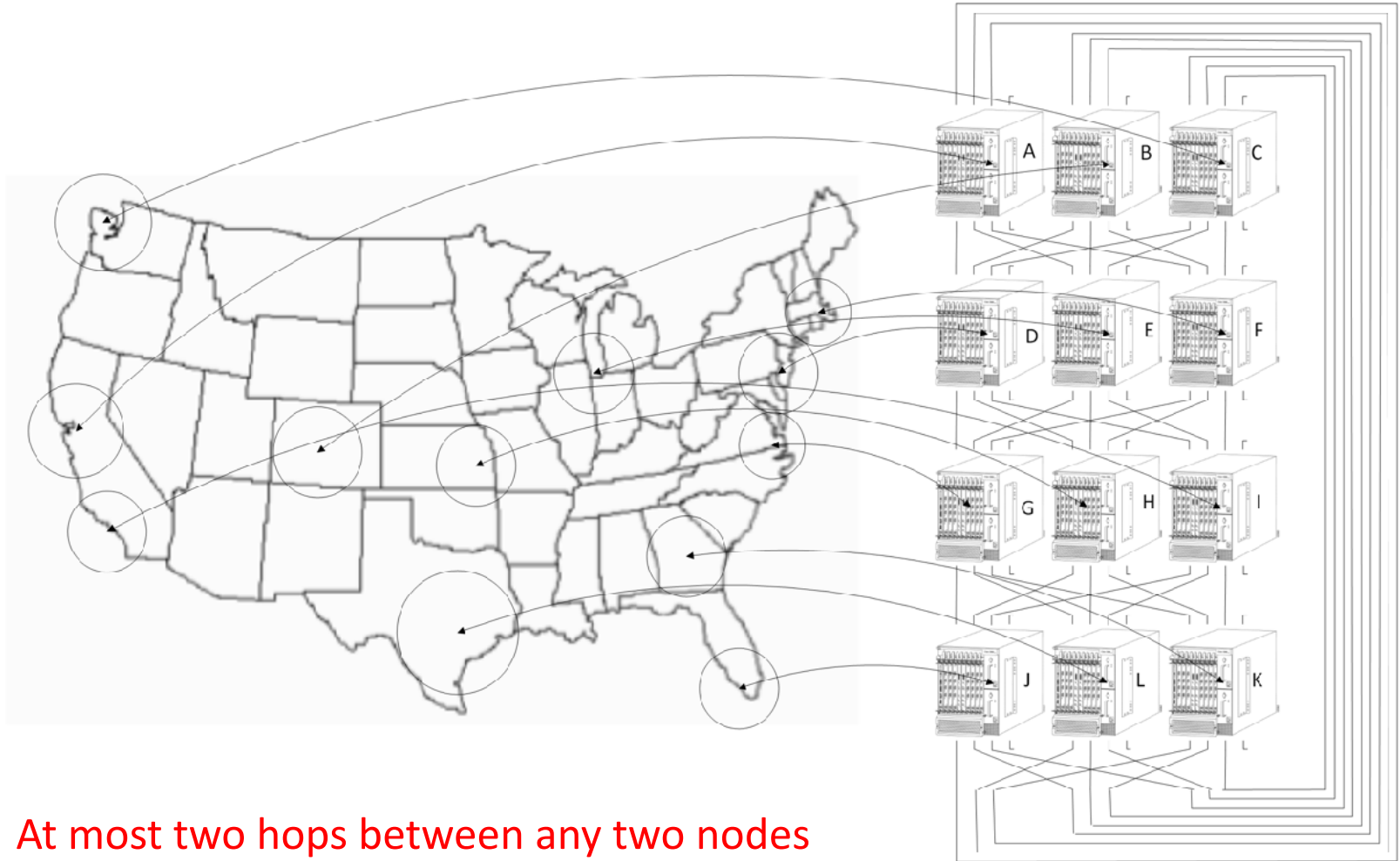


- The number of Hops can also be reduced by using multistage networks

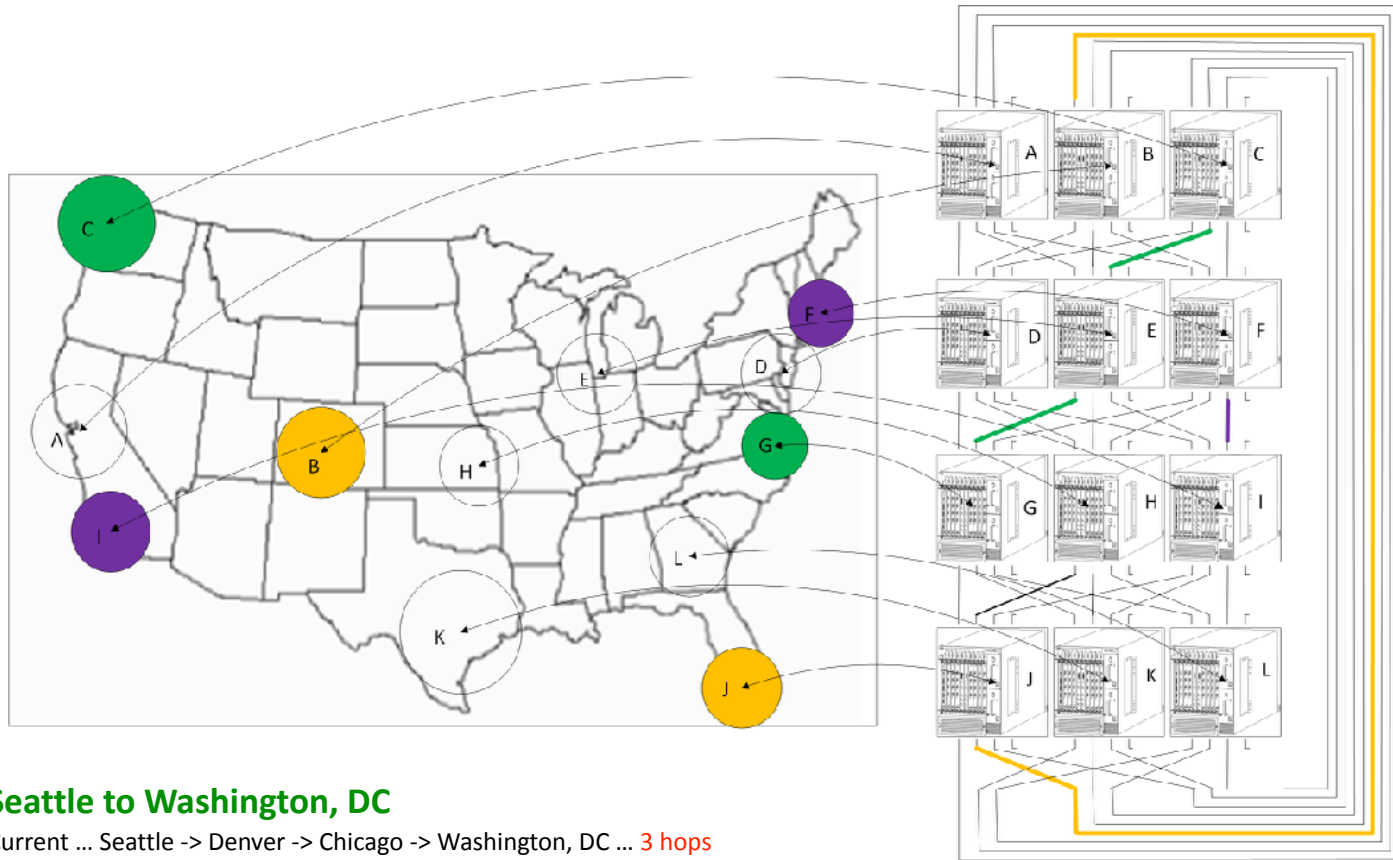
Average Number of Hops for Various Topologies



A Multistage Based Network



Examples of a multistage topology



Seattle to Washington, DC

Current ... Seattle -> Denver -> Chicago -> Washington, DC ... 3 hops

Multistage ... Seattle -> Chicago -> Washington, D.C. ... 2 hops

LA to Boston

Current ... LA -> Dallas -> Atlanta -> Washington, DC -> NY -> Boston ... 5 hops

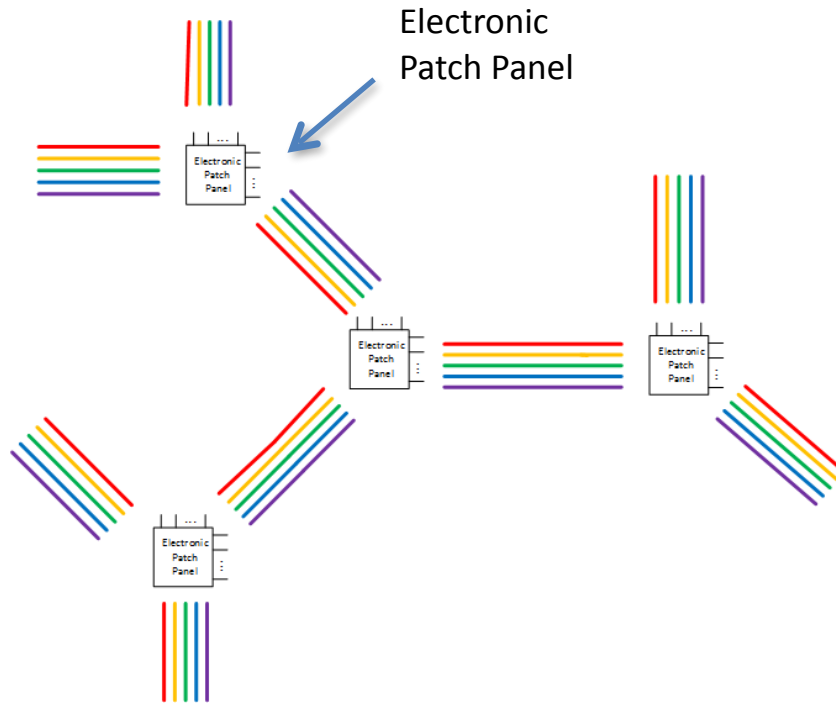
Multistage ... LA -> Boston ... 1 hop

Denver to Miami

Current ... Denver -> Dallas -> Atlanta -> Miami ... 3 hops

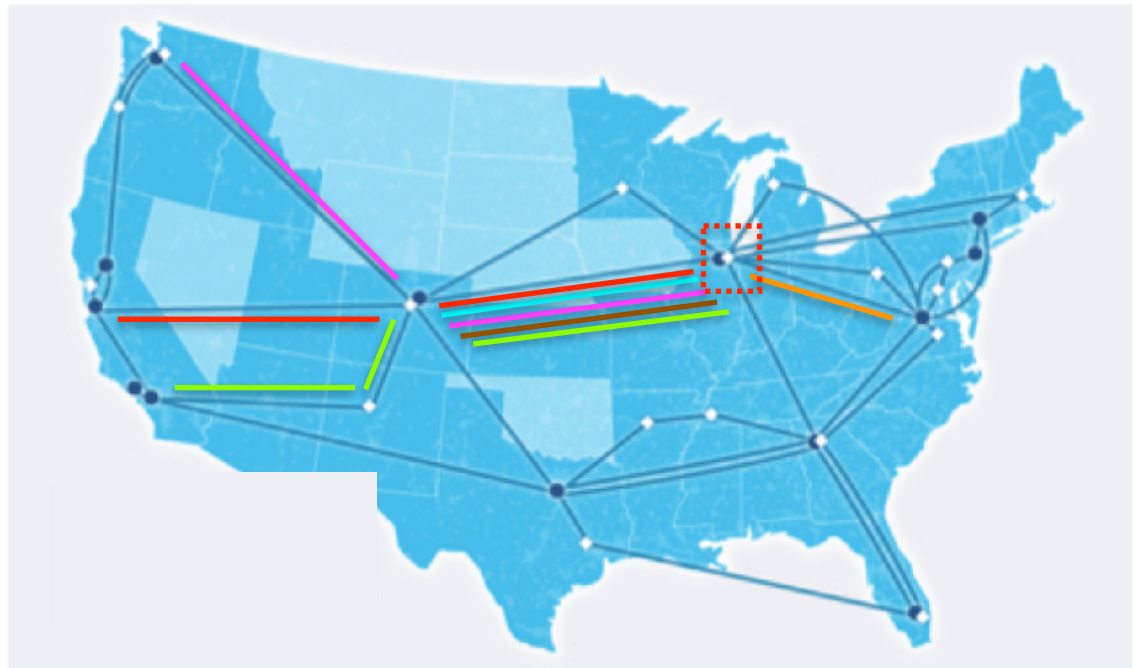
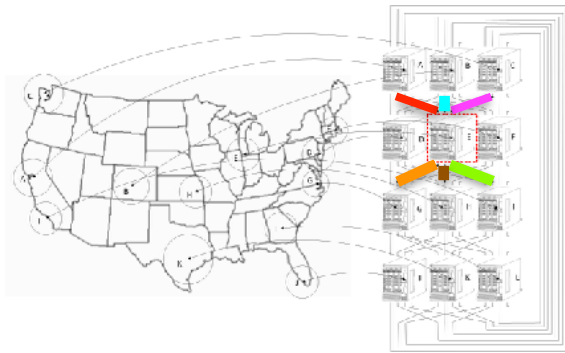
Multistage ... Denver -> Miami ... 1 hop

How can these advanced topologies be implemented?



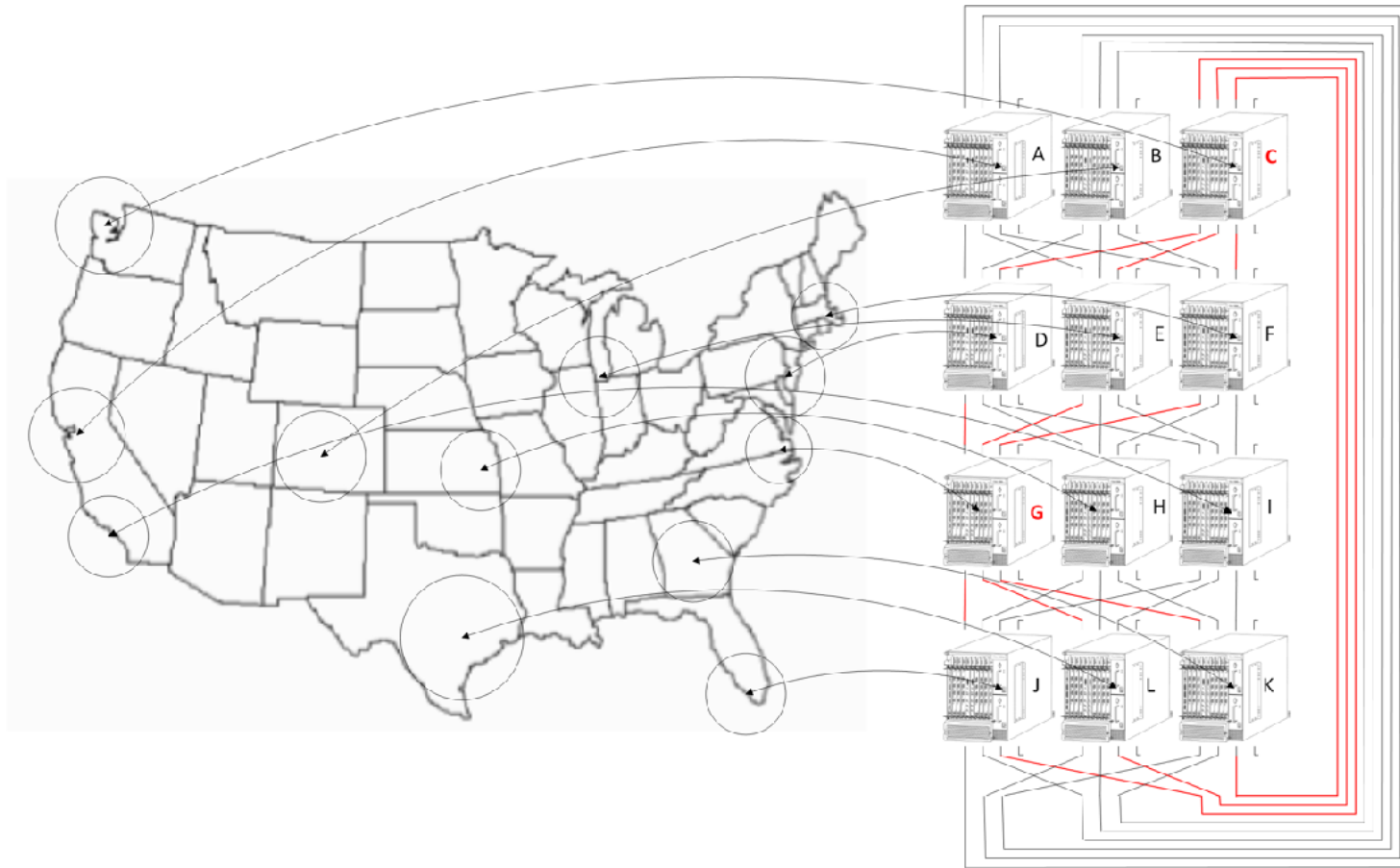
- WDM can be used to decouple the topology of the network from geographical constraints.

Some of the actual geographic connections



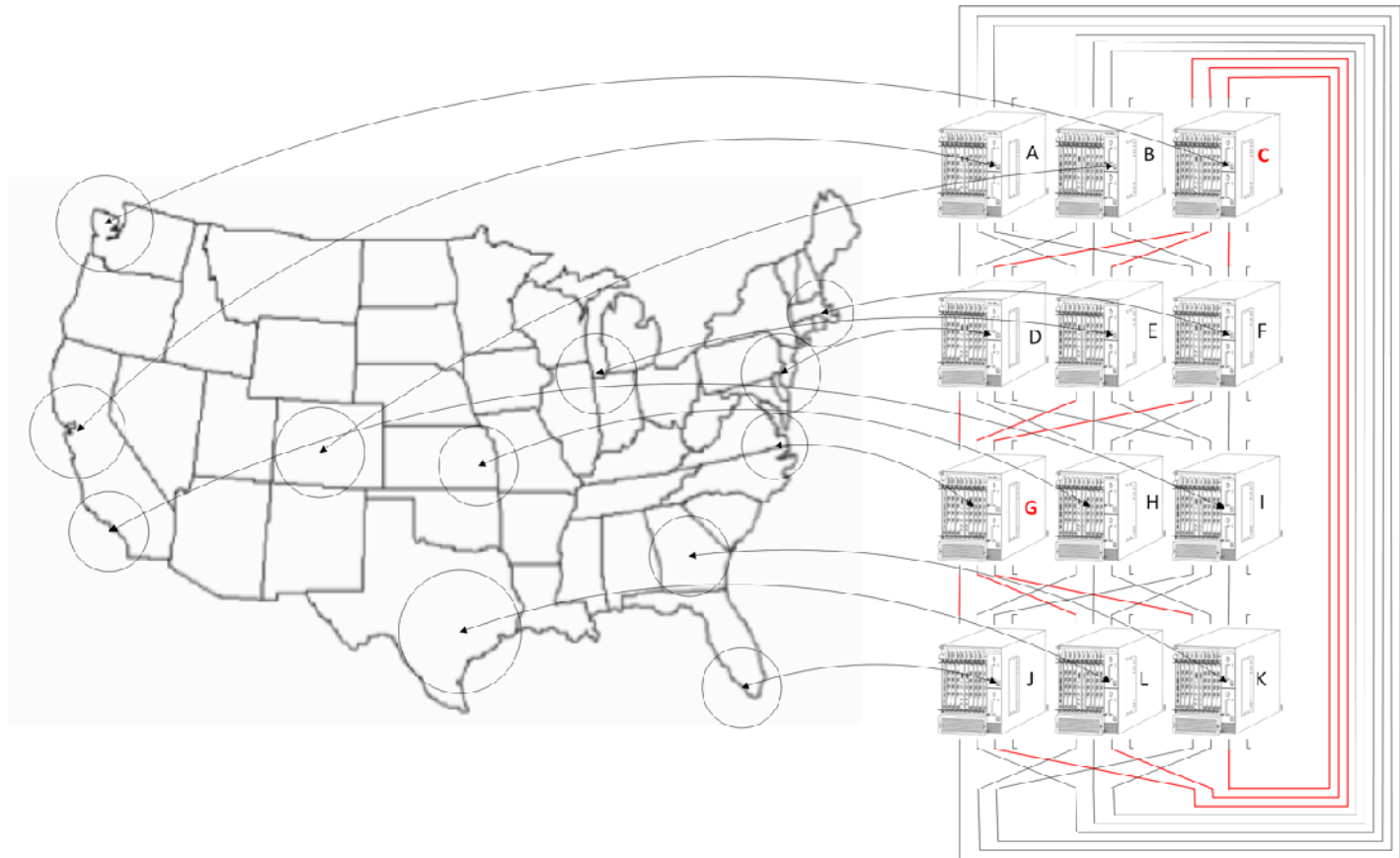
- Each bypassed node spares two router ports and a possible optical transceiver card

Fault Tolerance can be increased by (6X) by adding an extra stage to the Toroidal, $N \times \log N$ Network



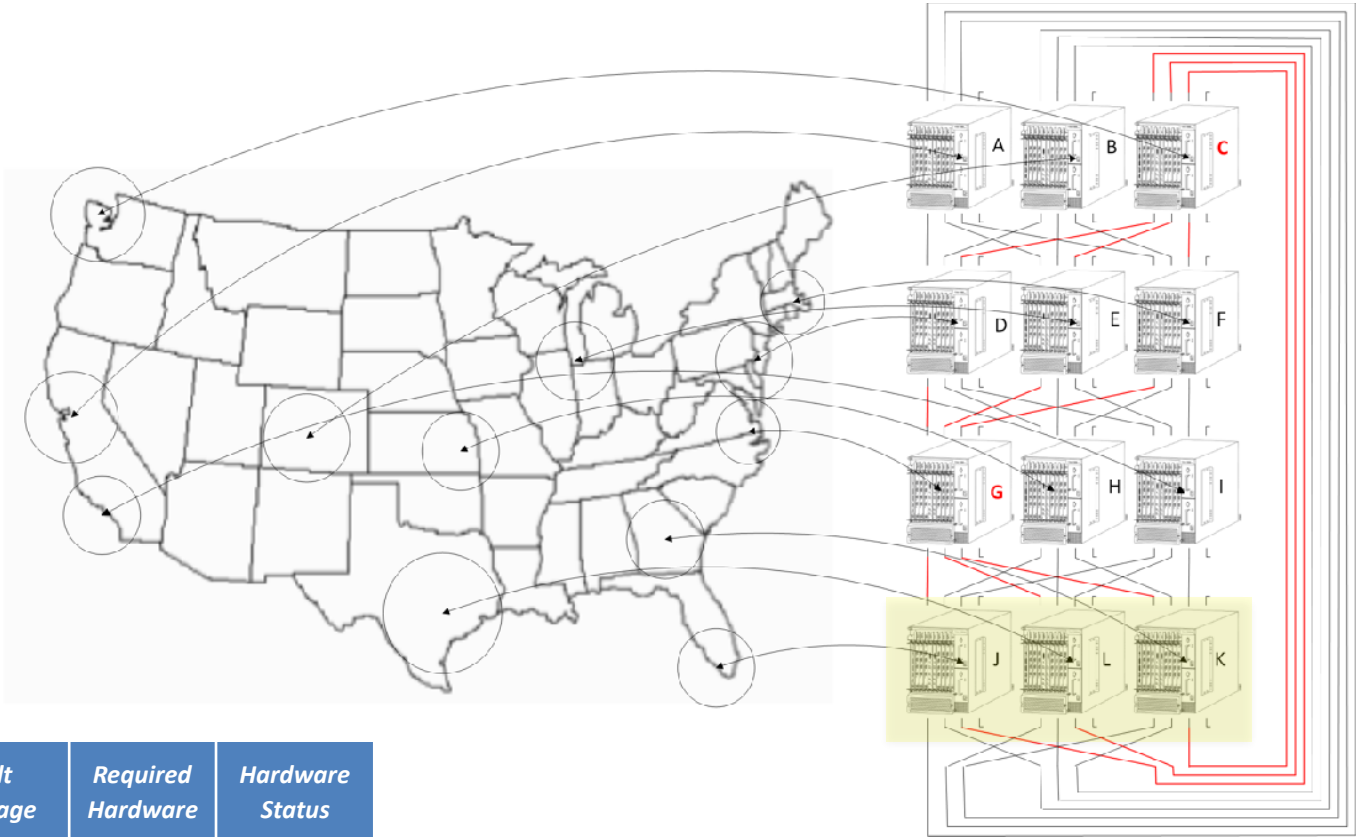
- There are 6 paths between any two nodes

Voice and Video throughput is also increased by the extra stage, since it decreases fabric blocking (6X)



- As Internet traffic becomes more isochronous, fabric blocking becomes more of an issue.

How much does this Path Redundancy cost?



	<i>Fault Coverage</i>	<i>Required Hardware</i>	<i>Hardware Status</i>
<i>Duplex</i>	<i>1X</i>	<i>2X</i>	<i>Standby</i>
<i>Majority Voting</i>	<i>1X</i>	<i>3X</i>	<i>Standby</i>
<i>Redundant Routing</i>	<i>6X</i>	<i>0</i>	<i>Active</i>

Average Packet Path Increase

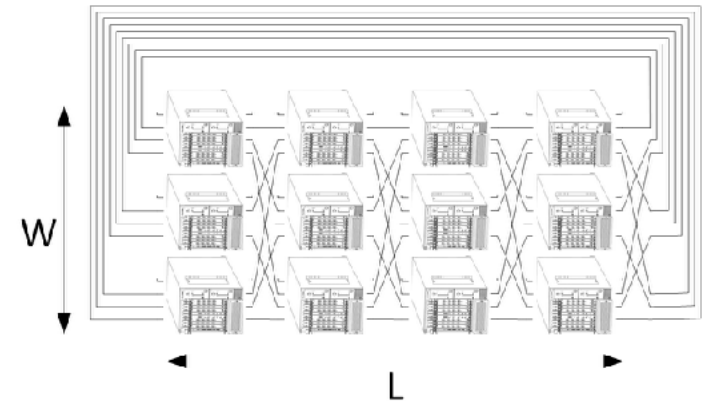
Old average lengthwise path length = $L/2$

New average lengthwise path length = $L/2$

Old average widthwise path length = $W/2$

New average widthwise path length
= (average number of hops)
* (average length of perfect shuffle)
= $(3/2) * (W/2) = 3W/4$

Difference of New - Old paths =
= $(L/2 - L/2) + (3W/4 - W/2) = W/4$



Cost versus Benefits

Cost

If $L=3000$ miles, $W=1200$ miles, and $C=186,000$ miles / sec

then Difference = $1200/4 = 300$ miles

= $300 / (3000/2 + 1200/2) = 300/2100 = 14\%$ increase in path length

= $300 / (2/3 * 186,000) = 2.4$ ms in travel time

Benefits

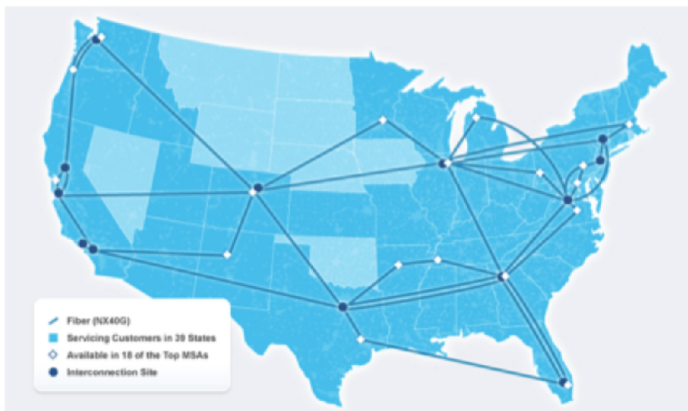
$(\text{old_average_hops}) / (\text{new_average_hops}) = ((\text{old_max_hops})/2) / ((\text{new_max_hops})/2)$

$(6/2) / (2/2) = 3\text{X}$ less Hops

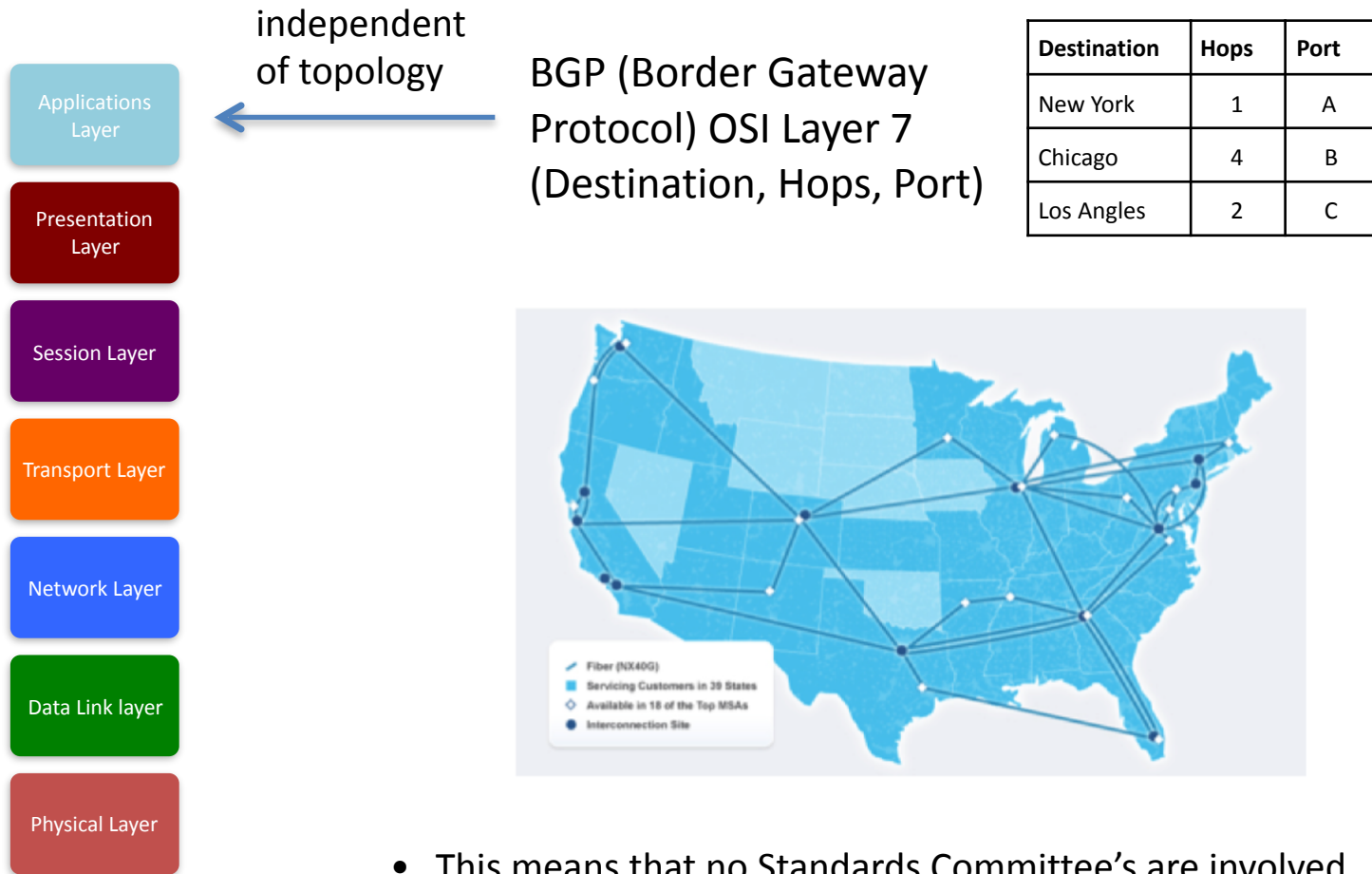
=> 3X less latency, power, and cost

$(9/2) / (2/2) = 4.5\text{X}$ less Hops

=> 4.5X less latency, power and cost



The Topology is independent of and thus compatible with all OSI Layers



OSI Layers

Border Gateway Protocol (BGP) Routing Table Construction

Destination	Hops	Port
New York	2	A
Chicago	5	B
Los Angeles	4	C

Port A



Destination	Hops	Port
New York	1+1	C
New York	2+1	A
New York	3+1	B
Chicago	4+1	C
Chicago	4+1	B
Chicago	5+1	A
Los Angeles	2+1	C
Los Angeles	4+1	A
Los Angeles	5+1	B

Port C



Destination	Hops	Port
New York	1	A
Chicago	4	B
Los Angeles	2	C

Port B



Destination	Hops	Port
New York	3	A
Chicago	4	B
Los Angeles	5	C

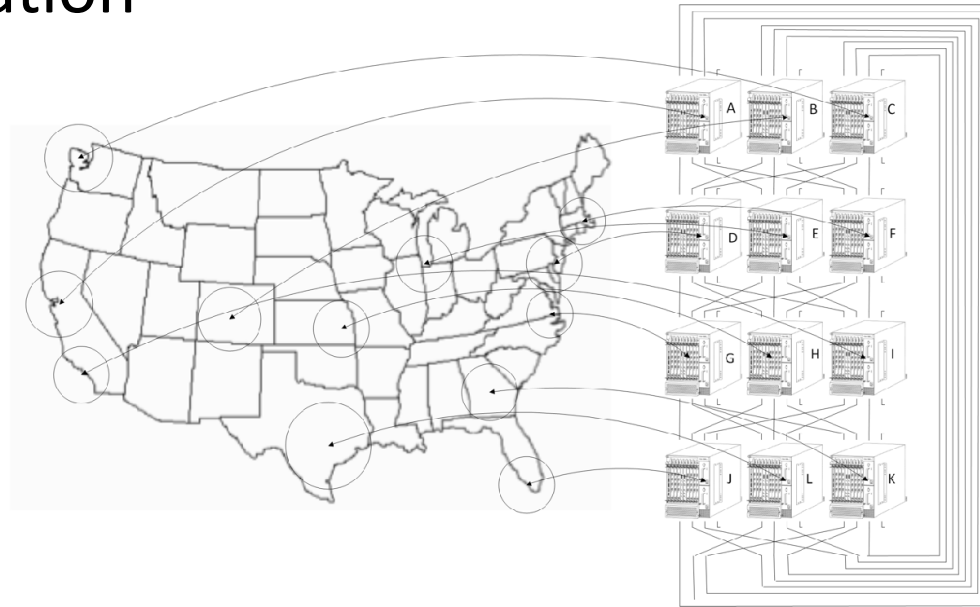
BGP:

- routes packets,
- load balances
- fault Tolerance

selected redundant path # = mod (source_addr[m:n], num_redundant_paths)

- BGP is essentially a distributed gradient decent algorithm

Implementation



- If the network was “seeded” with an optimal, $N \times \log N$, solution then BGP will automatically discover this solution and start directing traffic onto the optimal network.
- The non-“seeded” links are thus sub-optimal and can be cannibalized to strengthen “seeded” network.
- Network Zen ... A topological optimal network exists; however, the journey towards this goal is more important

Rant

- One carrier got fined \$100 M by the FCC for “throttling” their unlimited bandwidth customers.
- Another carrier wanted to abolish “Network Neutrality” because Netflix was causing their network to “melt down.”
- How many lawyers does it take to fix a network ... ?
- Math
- Optics
- Enough of this *Flatland* crap



Summary

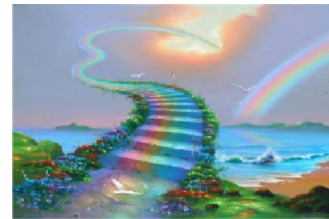
- Number of Hops matters



- Higher Dimensions and Groups will get you there faster



- WDM decouple network topology from geographical constraints

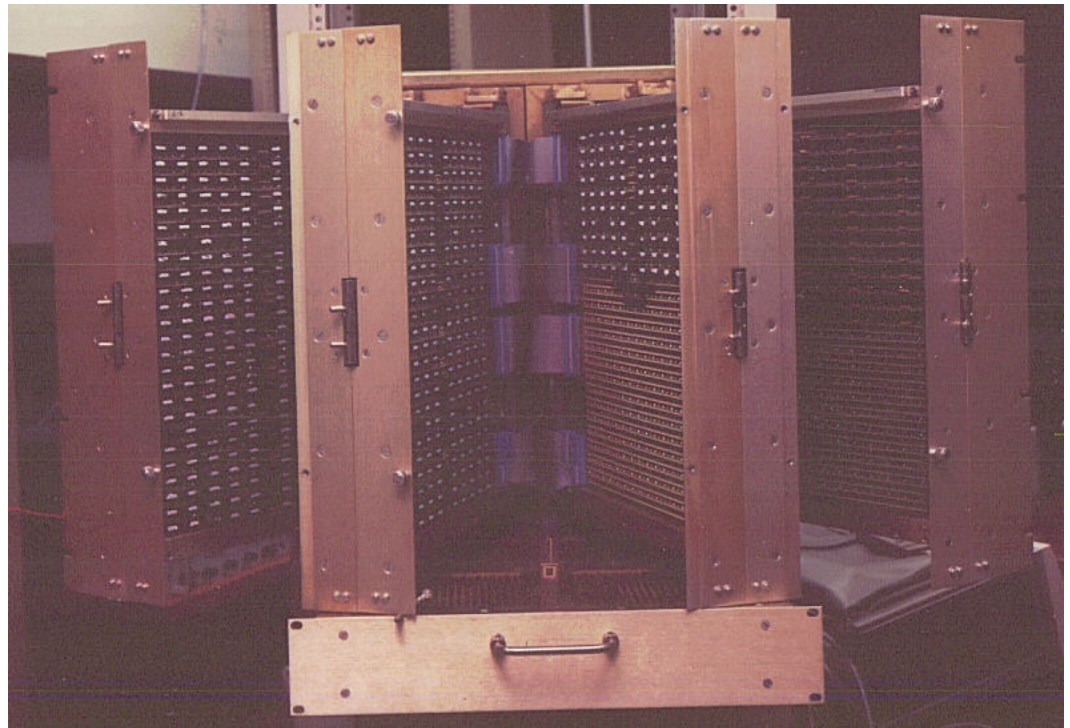
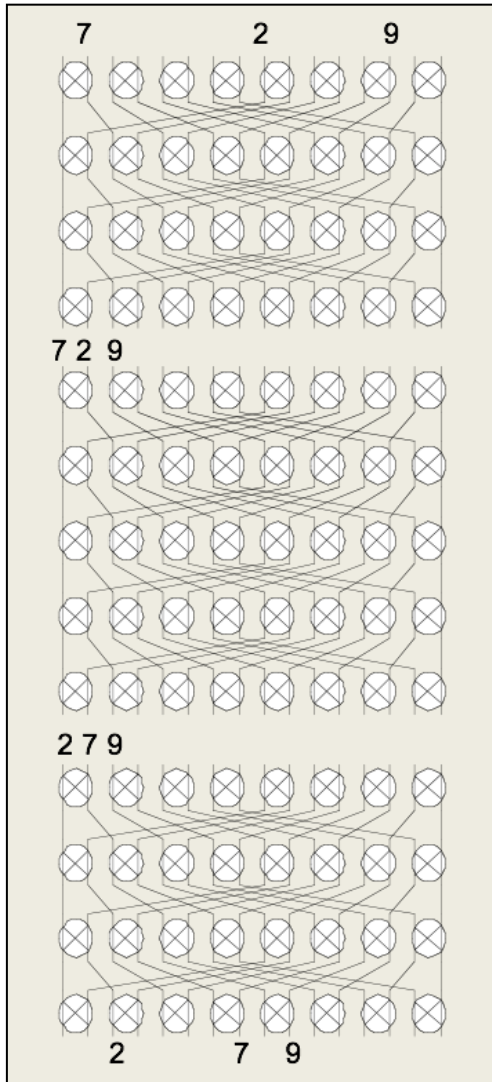


- Network Zen



Note: toroid

History ... How this all began



- “Starlite” Packet Switch (Batcher / Banyan)
 - 32 inputs each at 100 Mb/s (1982)
- “Starbrite” Multi-stage National Network

The Stanford Connection

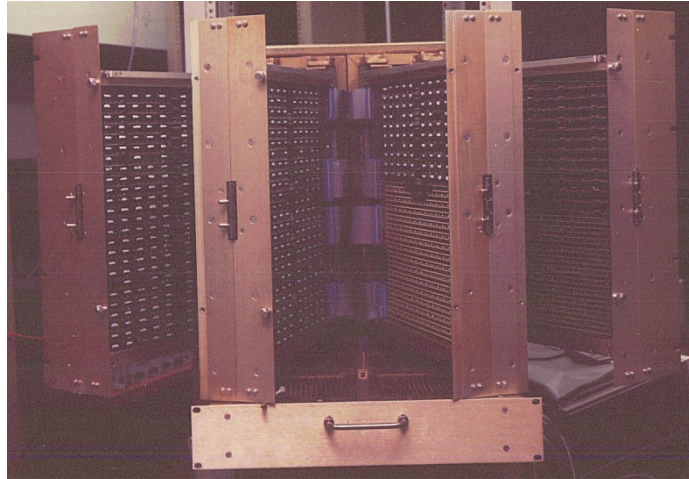
Prof. Harold Stone
(Perfect Shuffle)

Prof. Donald Knuth
(Sorting and Searching
and Discrete Math)

Prof. Mike Flynn
(Computer
Architecture)

Dr. Lynn Conway,
PARC - VLSI)

Prof. Vint Cerf
(Internet)



Prof. Joe Goodman
(Optics)

The Bell Labs Connection

Chuck Rutledge

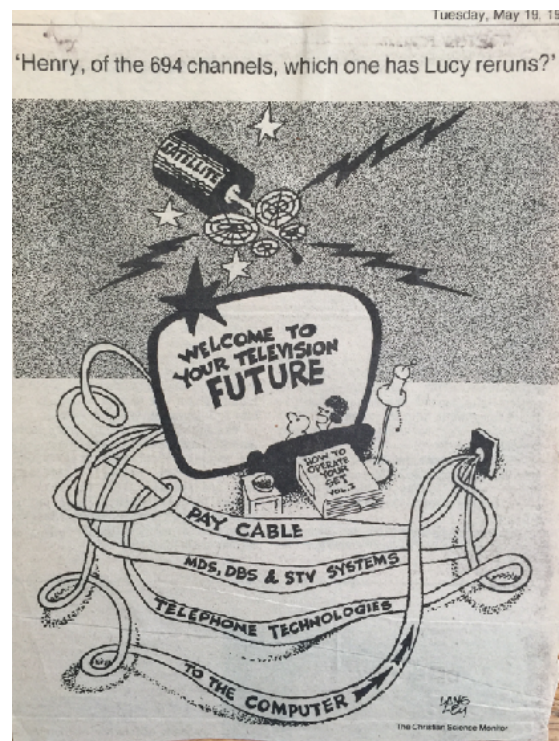
Jay O'Neil

Arun Netravali

Bob Lucky

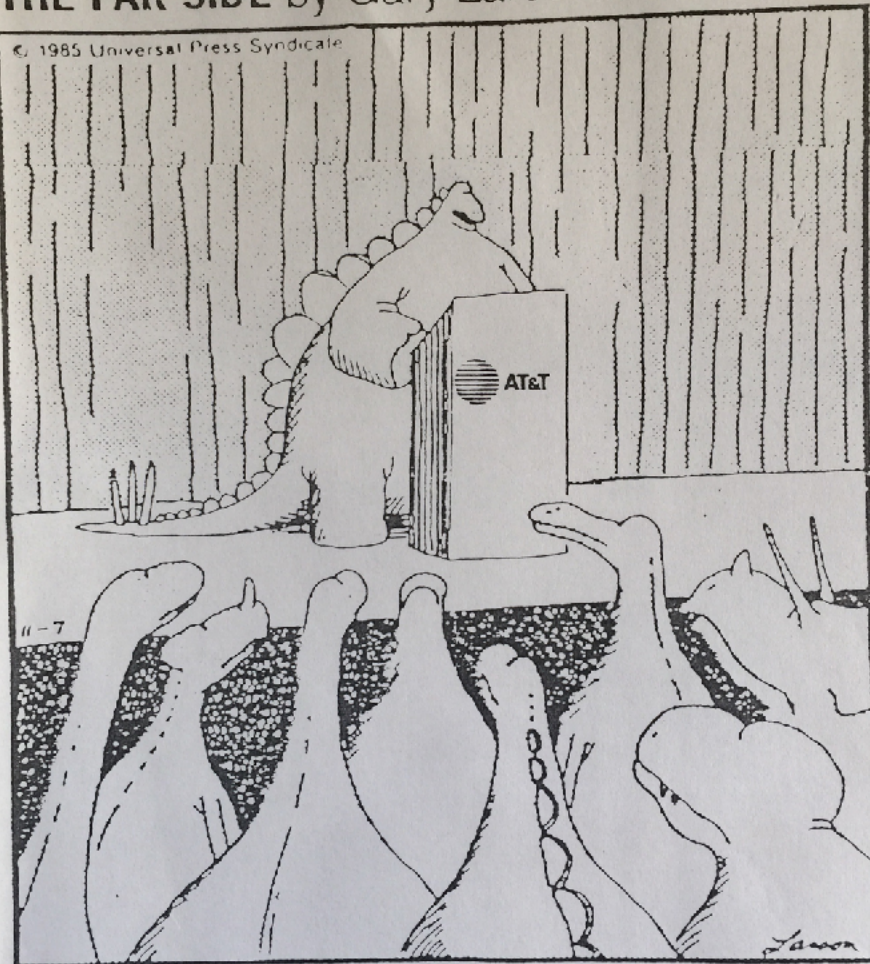
Haw-minn Lu

Some History



THE FAR SIDE by Gary Larson

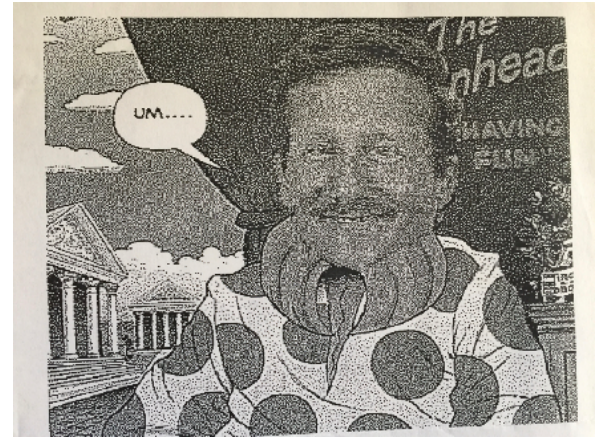
© 1985 Universal Press Syndicate

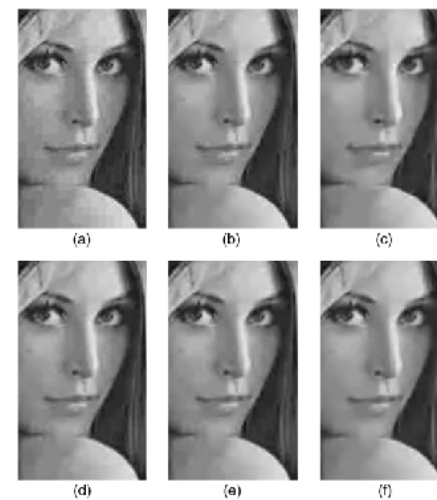
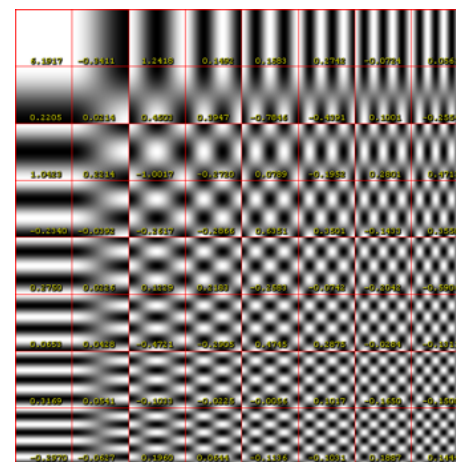


"The picture's pretty bleak, gentlemen. ... The world's climates are changing, the mammals are taking over, and we all have a brain about the size of a walnut."



Scott C. Knauer May 17, 1946 - March 16, 1993





A Topologically Optimal Internet

