# High-End Computing Systems

*EE380 State-of-the-Art Lecture*

**Hank Dietz**
Professor & Hardymon Chair in Networking
Electrical & Computer Engineering Dept.
University of Kentucky
Lexington, KY 40506-0046
`http://aggregate.org/hankd/`

UNIVERSITY OF KENTUCKY

# What Is A Supercomputer?

- One of the most expensive computers?
- A very fast computer?
- Really two key characteristics:
  - Computer that solves big problems...
    stuff that wouldn't fit on a PC
    stuff that would take too long to run
  - Performance can scale...
    more money buys a faster machine
- A supercomputer can be cheap!

# The Key Is
# Parallel Processing

- Process *N* “pieces” simultaneously, get up to factor of *N* speedup
- Modular hardware designs:
  - Relatively easy to scale – add modules
  - Higher availability (if not reliability)

# The Evolution Of Supercomputers

- Most fit survives, even if it's ugly
- Rodents outlast dinosaurs…
  and bugs will outlast us all!

# When Does Supercomputing Make Sense?

- When you need results **NOW!**
- *Top500* speeds up 1.4X every 6 months! Just waiting might work...
- Optimizing your code helps a lot; do that first!
- When your application takes enough time per run to justify the effort and expense
- Our technologies don't change the basics... they mostly improve price/performance

# What Is A
# Cluster Supercomputer?

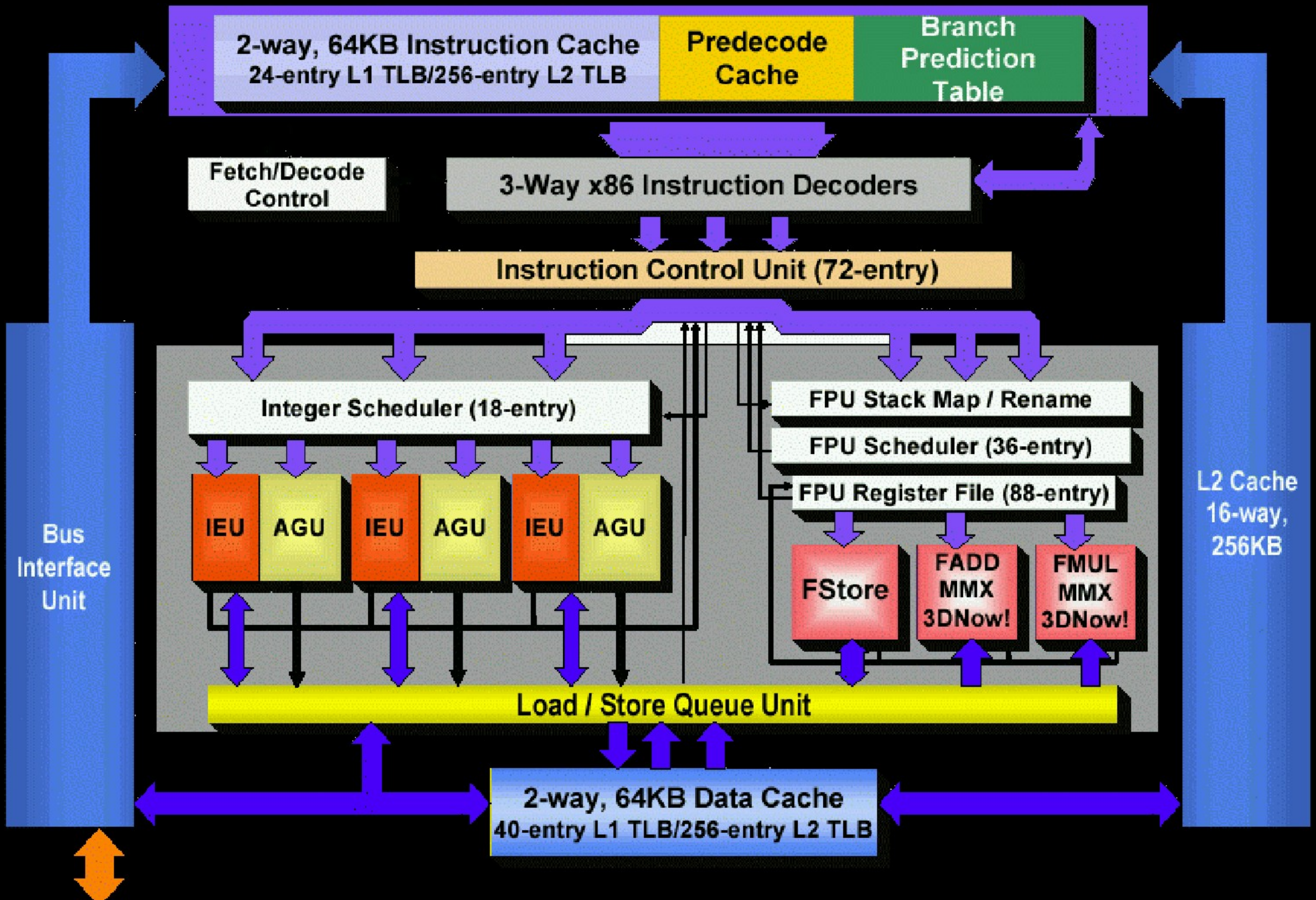
- Not a “traditional” supercomputer?
- Is The Grid a cluster?
- Is a Farm a cluster?
- A Beowulf?
- A supercomputer made from ***Interchangeable Parts*** (mostly from PCs)
- Some PC parts you don't need or want
- Often, Linux PC “nodes”

UNIVERSITY OF KENTUCKY

# Parts... Vs. In A Traditional Supercomputer

- Processors: AMD Athlon, Opteron; Intel Pentium 4, Itanium; Apple G5... within 2X of best @ very low cost
- Motherboards, Memory, Disks, Network, Video, Audio, Physical Packaging...
- Lots of choices, but parts tuned for PC use, not for cluster supercomputing

# AMD Athlon XP




UNIVERSITY OF KENTUCKY

# Types Of Hardware Parallelism

- Pipeline
- Superscalar, VLIW, EPIC
- SWAR (SIMD Within A Register)
- SMP (Symmetric MultiProcessor)
- Cluster
- Farm
- Grid

# Engineer To Meet Application Needs

- Know your application(s)
- Tune your application(s)
- Know your budget:
  Money, Power, Cooling, Space
- Hardware configuration options
- Software configuration options

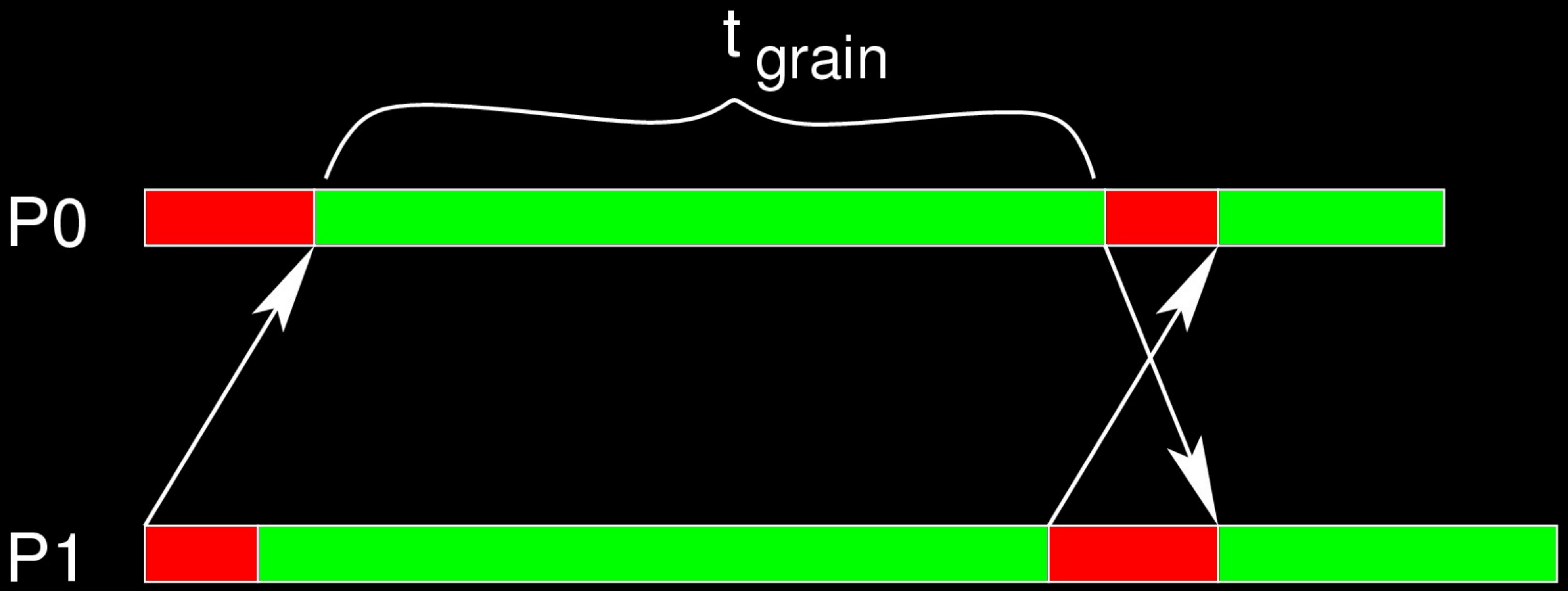
# Engineering A Cluster

- This is a ***systems*** problem
- Optimize ***integrated effects*** of:
  - Computer architecture
  - Compiler optimization/parallelization
  - Operating system
  - Application program
- Payoff for good engineering ***can be*** **HUGE!**
  (penalty for bad engineering ***is*** **HUGE!**)

# One Aspect: Interconnection Network

- Parallel supercomputer **nodes** interact
- **Bandwidth**
  - Bits transmitted per second
  - **Bisection Bandwidth** most important
- **Latency**
  - Time to send something here to there
  - Harder to improve than bandwidth....

# Latency Determines Smallest Useful Parallel Grain Size

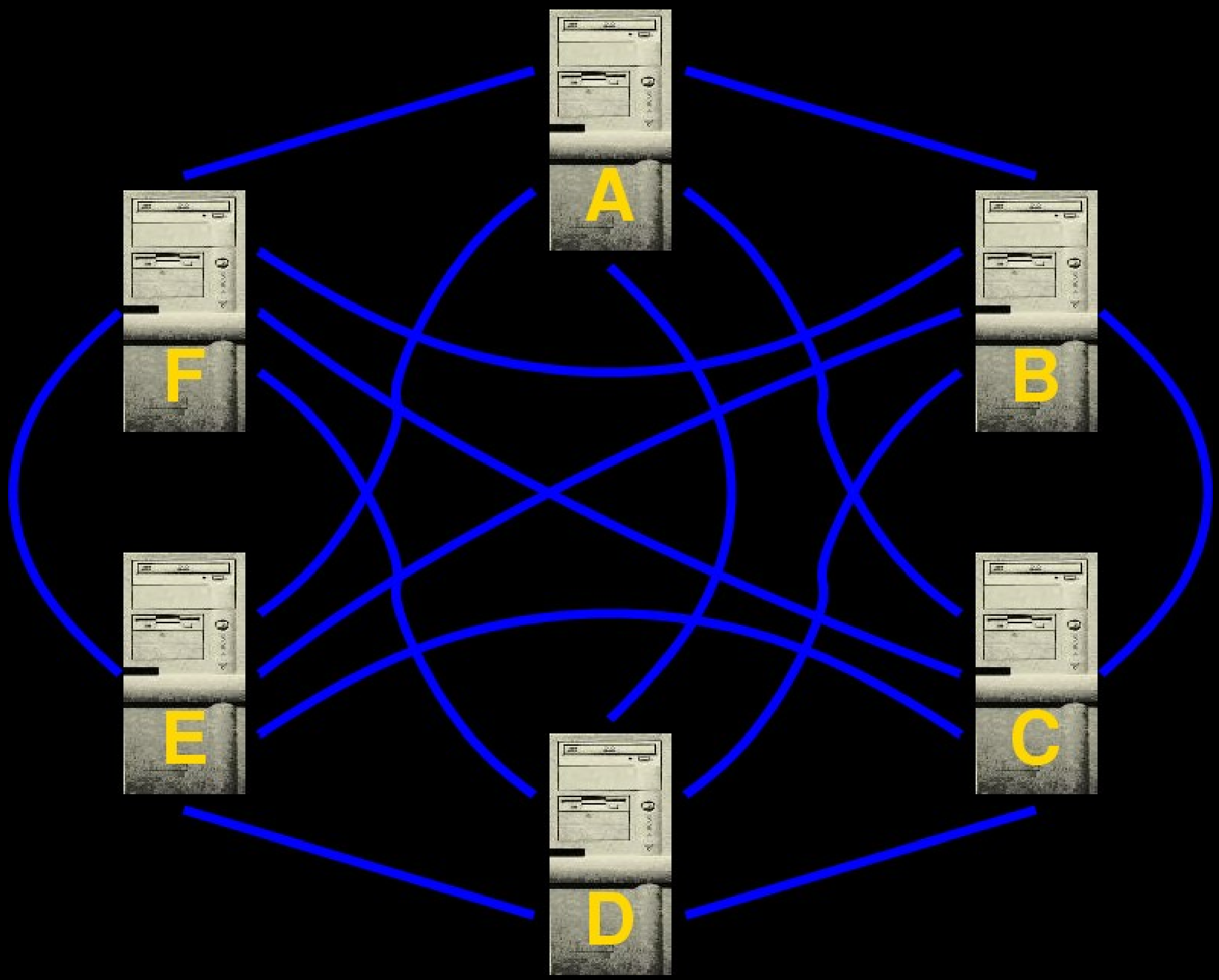


# Network Design

- Assumptions
  - Links are bidirectional
  - Bounded # of network interfaces/node
  - Point-to-point communications
- **Topology**
- Hardware
- Software

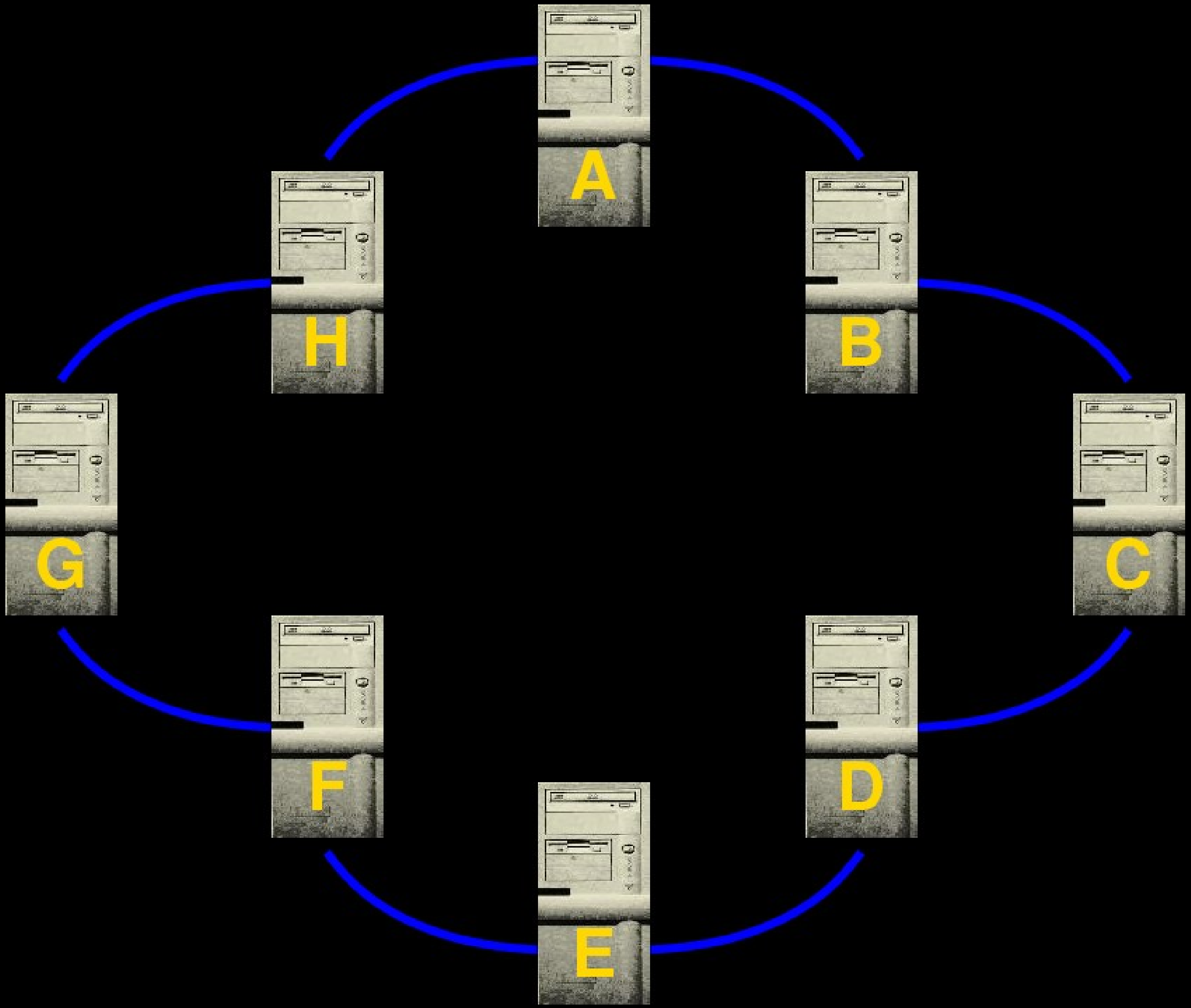
# No Network

# Direct Fully Connected



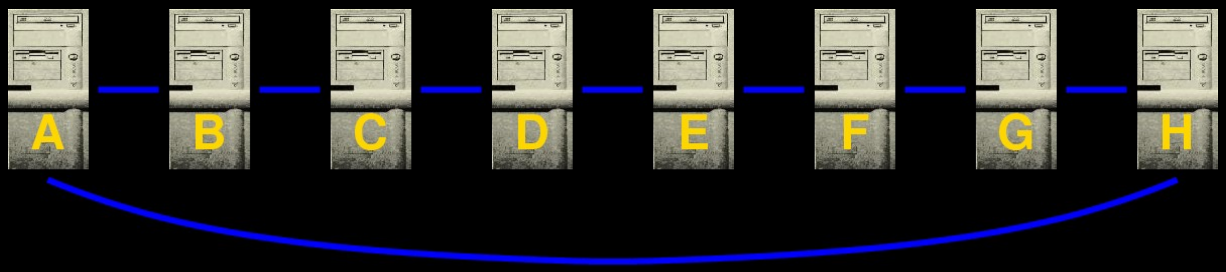

UNIVERSITY OF KENTUCKY

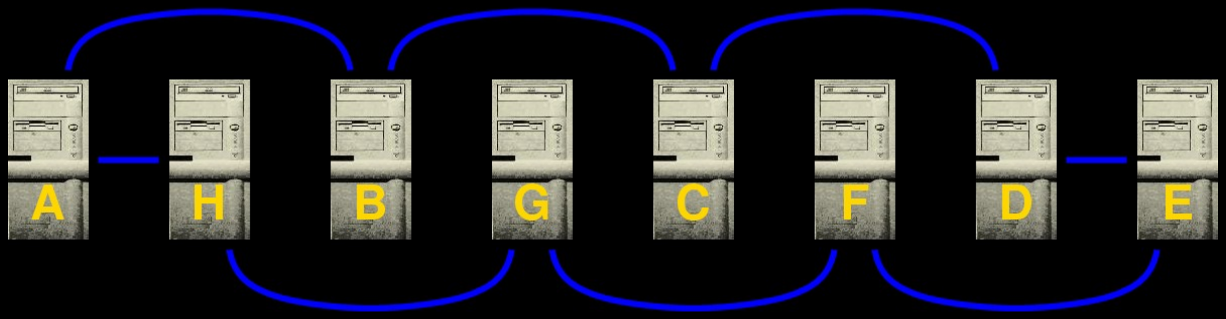
# Toroidal 1D Mesh (Ring)



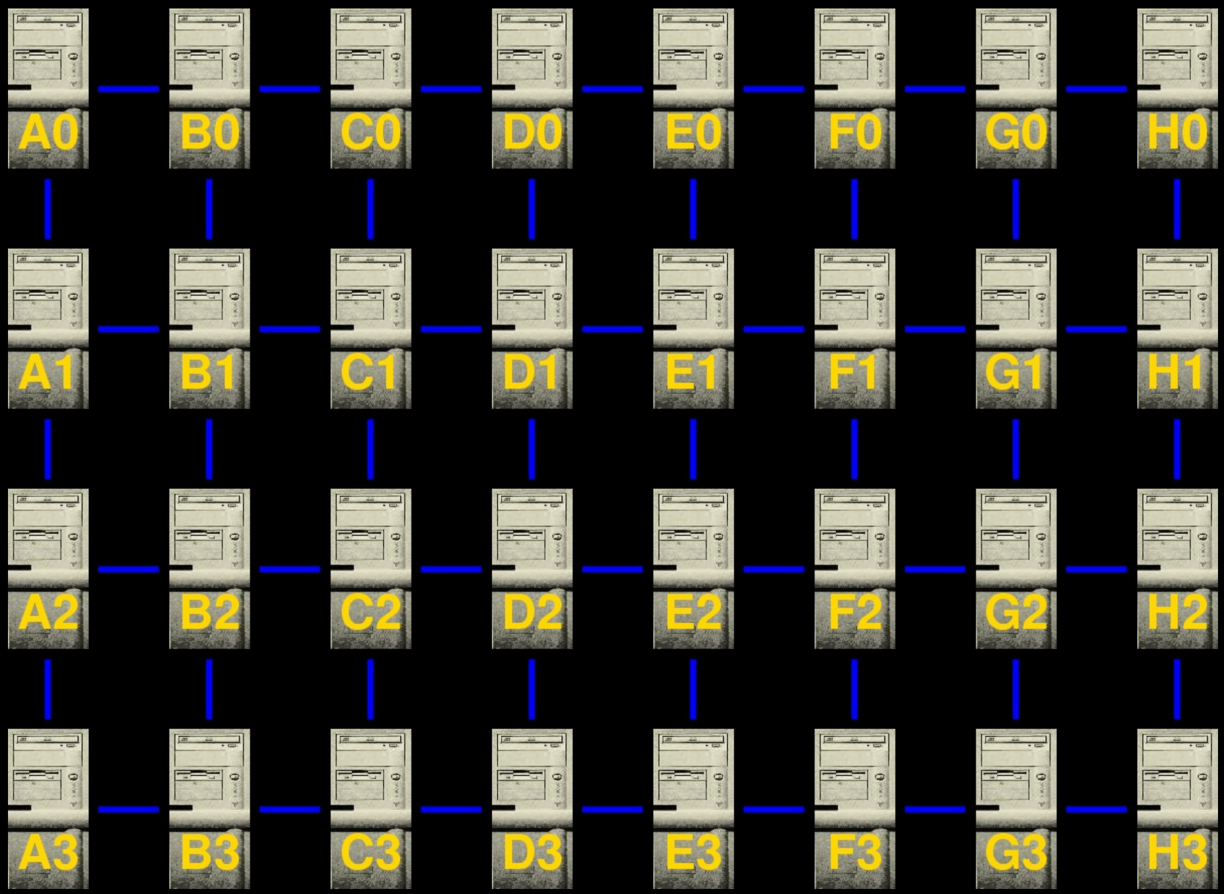

UNIVERSITY OF KENTUCKY

# Physical Layout Of Ring

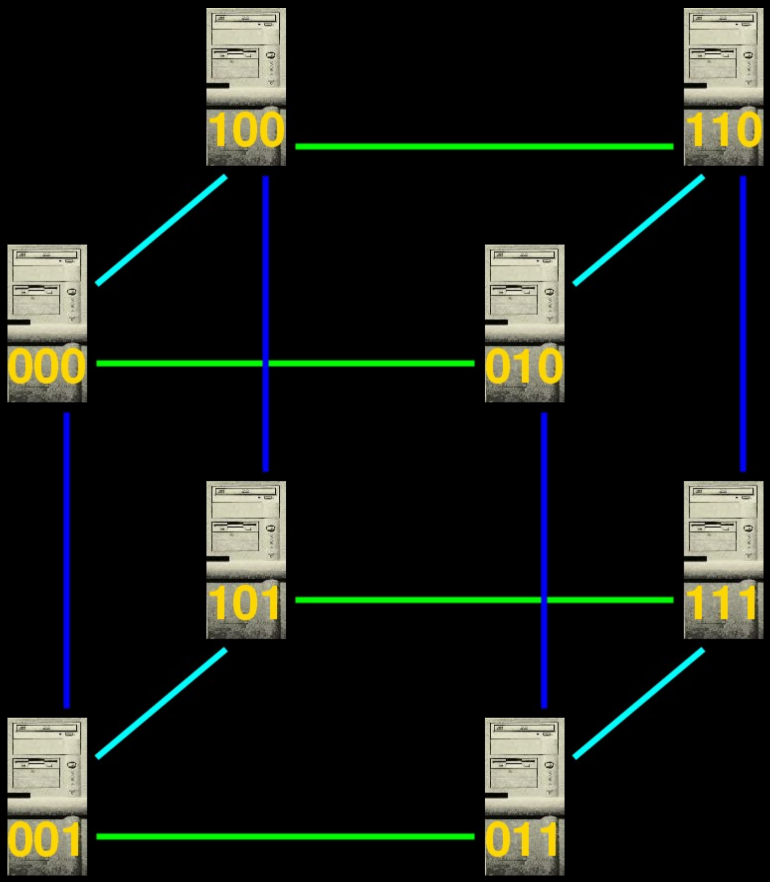


# Non-Toroidal 2D Mesh



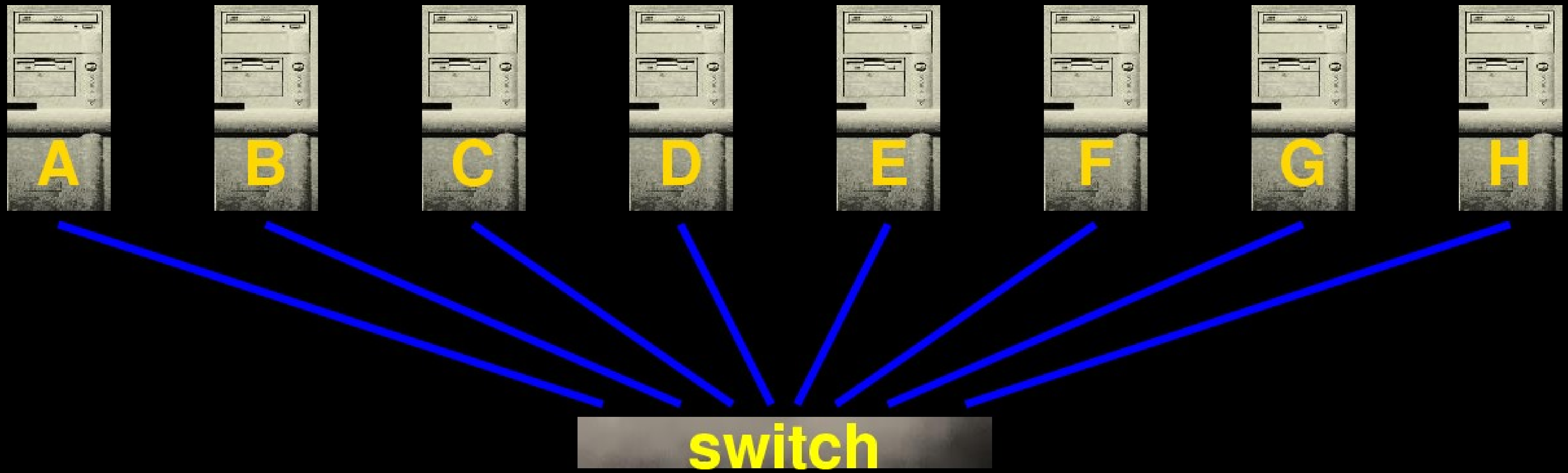
# 3-Cube (AKA 3D Mesh)



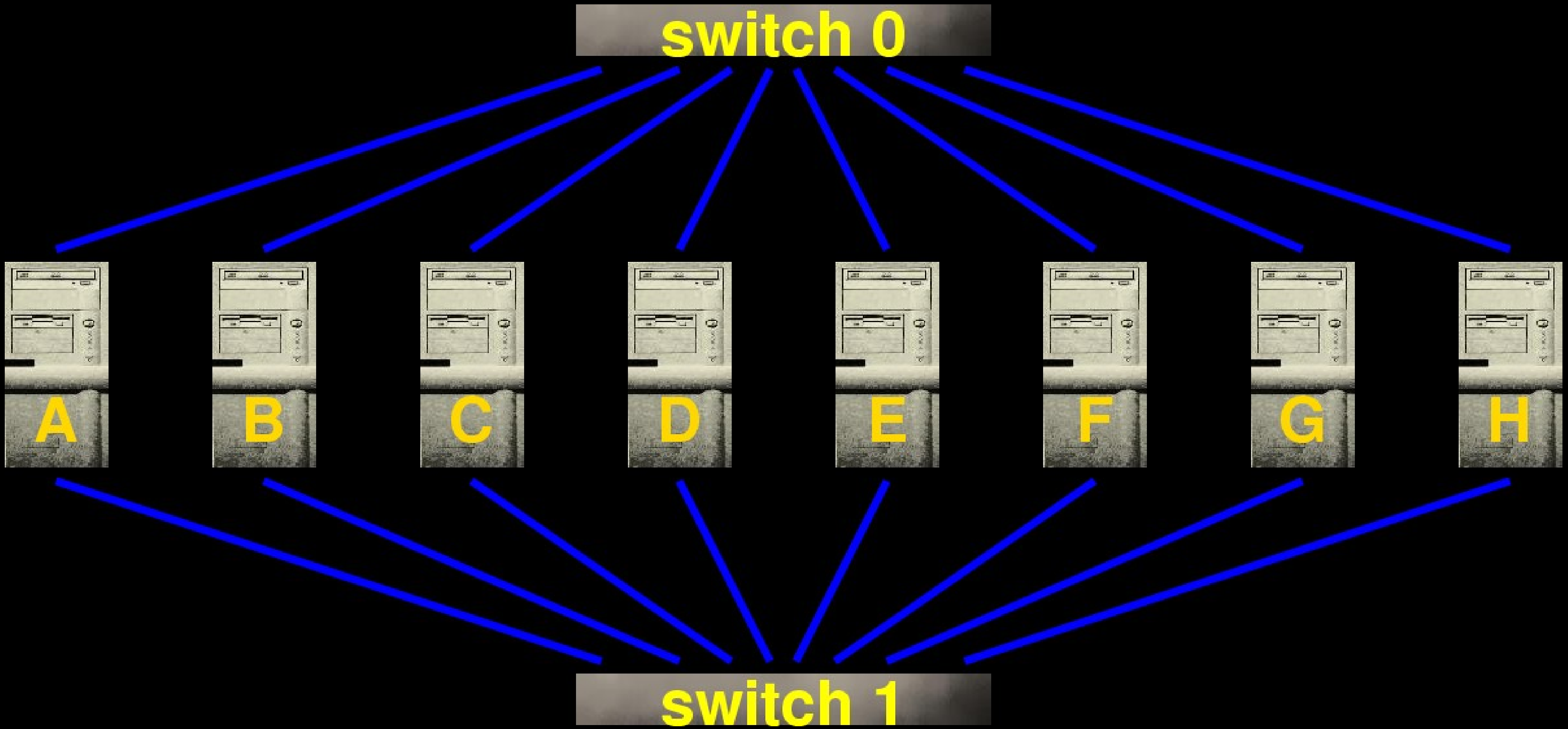
# Switch Networks

- Ideal switch connects *N* things such that:
  - Bisection bandwidth = # ports
  - Latency is low (~30us for Ethernet)
- Other switch-like units:
  - Hubs, FDRs (Full Duplex Repeaters)
  - Managed Switches, Routers
- Not enough ports, build a Switch Fabric

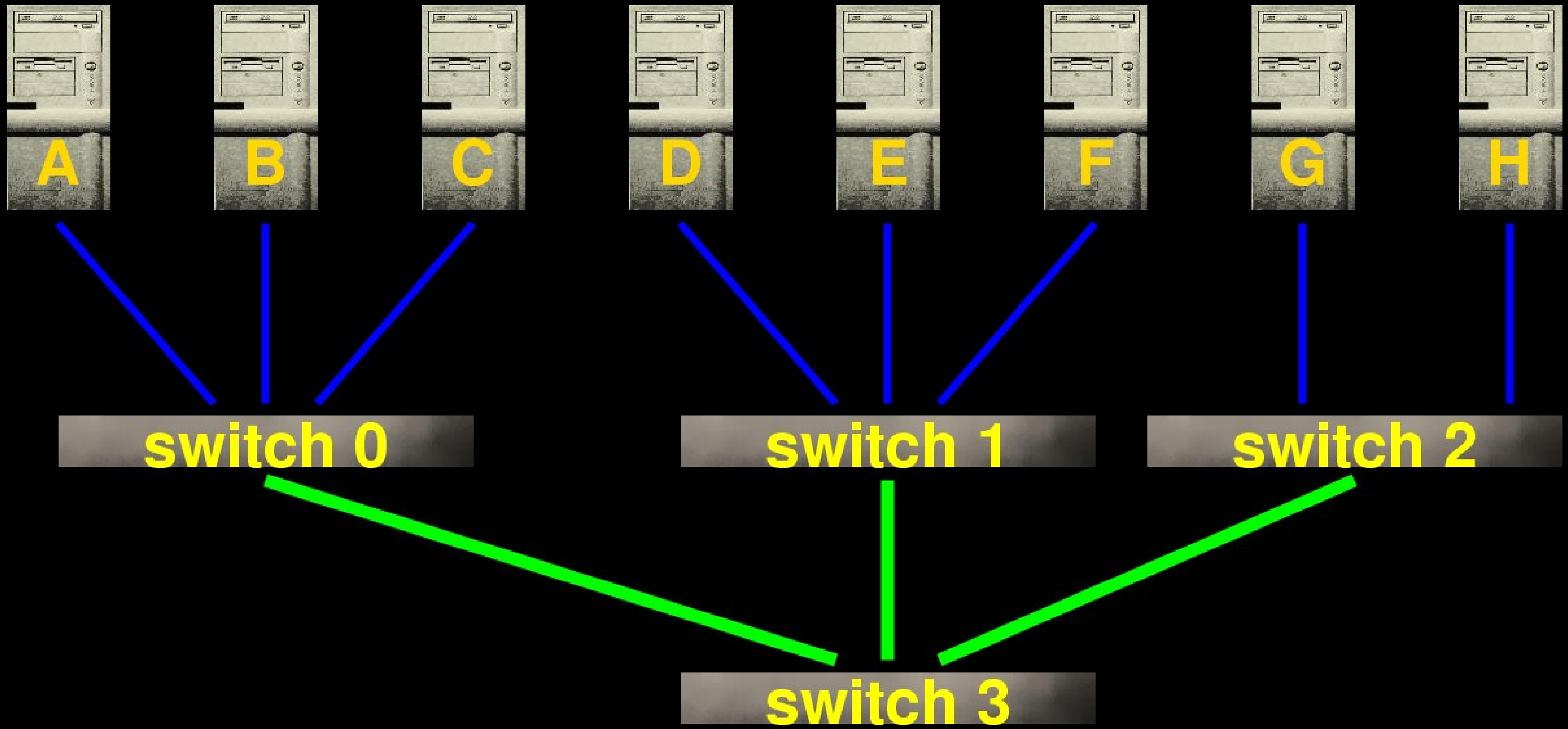
# Simple Switch (8-Port)



# Channel Bonding (2-Way)

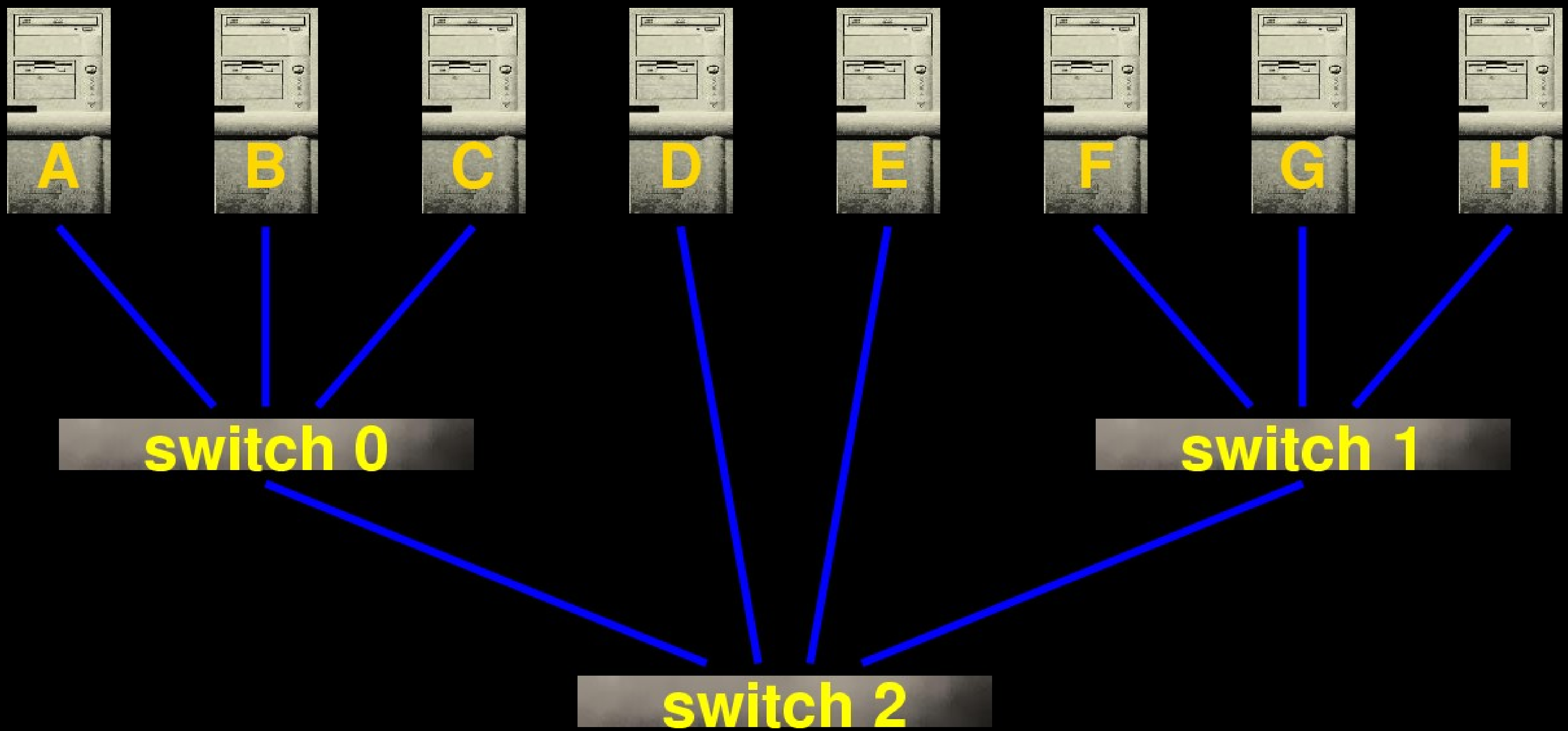


# Tree (4-Port Switches)



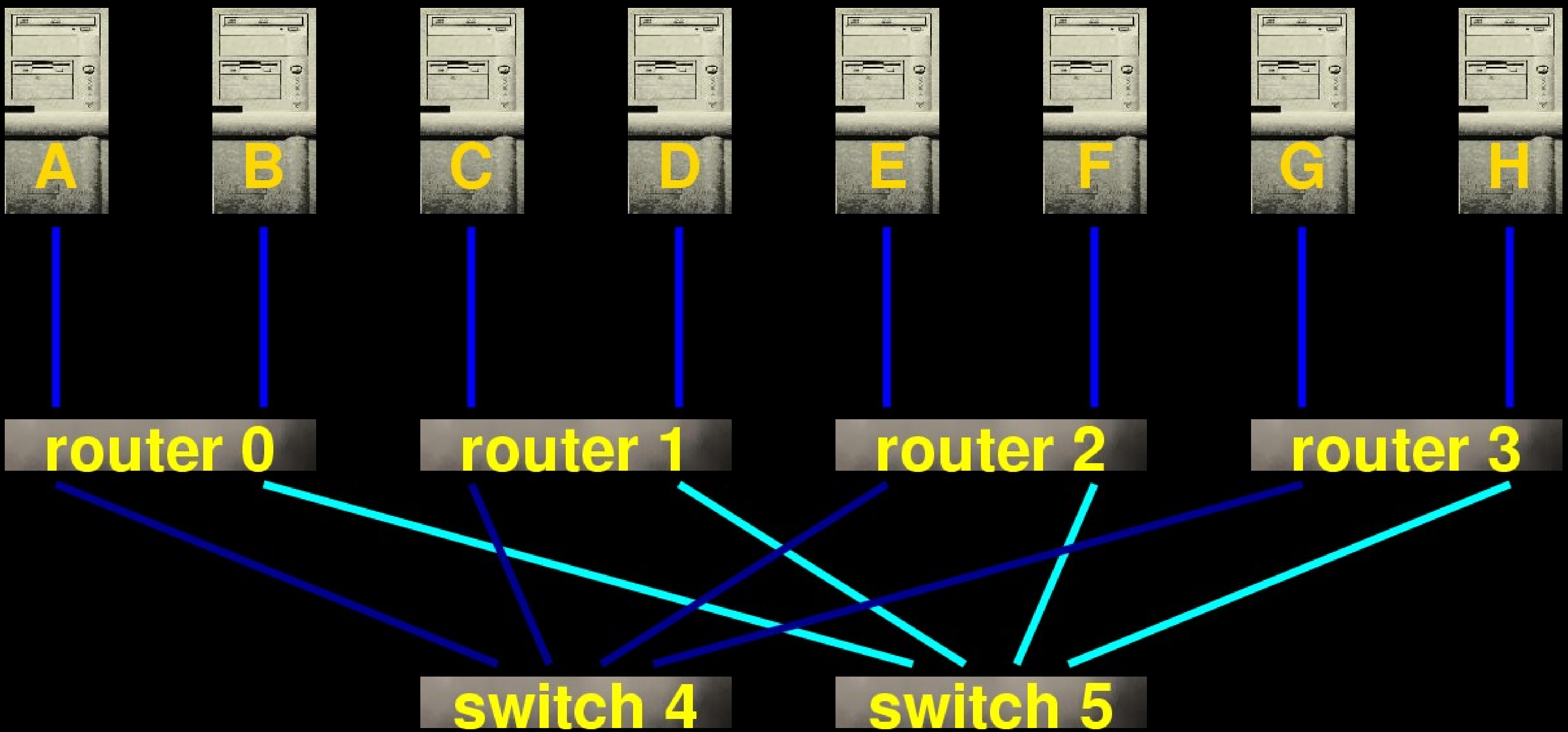

UNIVERSITY OF KENTUCKY

# A Better Tree




UNIVERSITY OF KENTUCKY

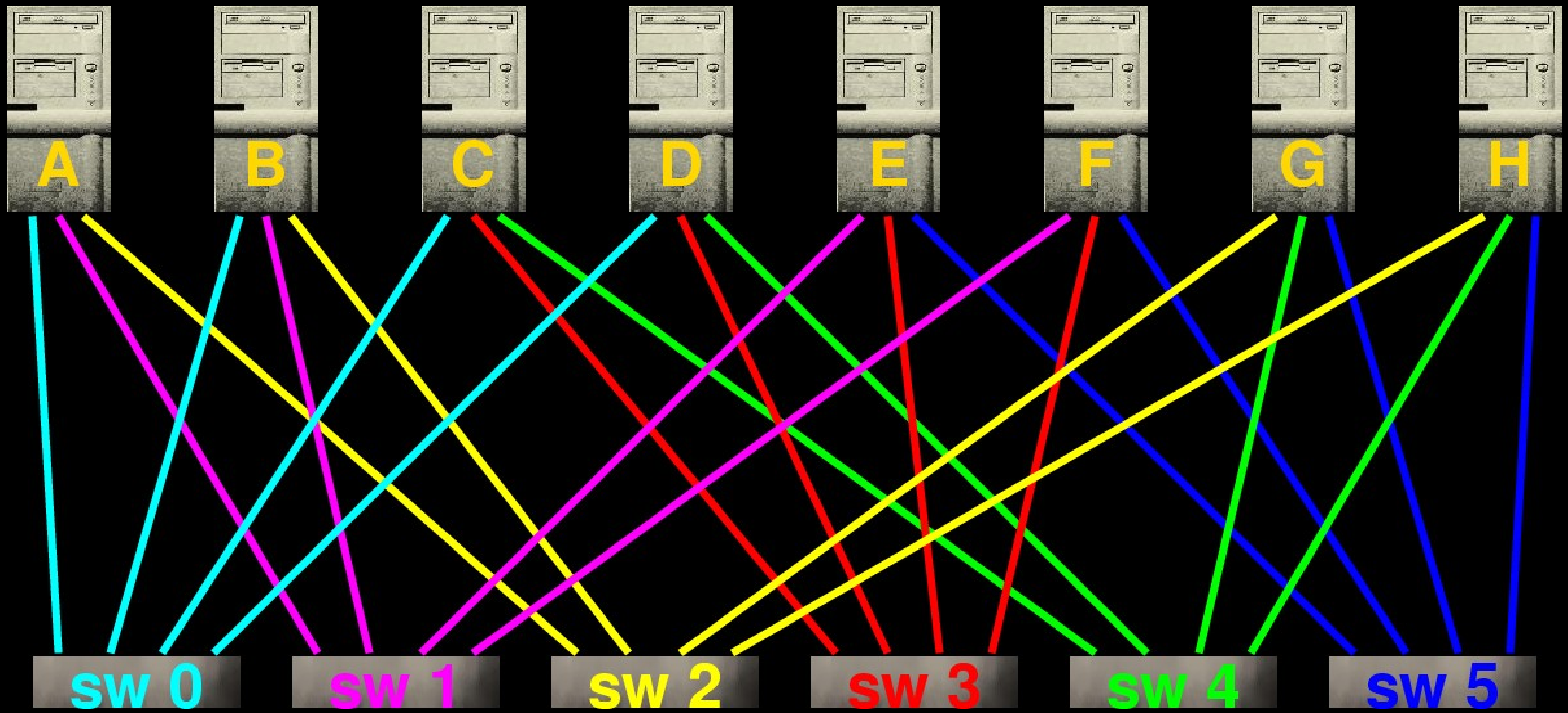
# Fat Tree


A
B
C
D
E
F
G
H
router 0
router 1
router 2
router 3
switch 4
switch 5

# Our Insights

- Want a “flat” single-level network
  - Top level determines bisection b'width
  - Multiple levels multiply latency
- Connect each node to multiple switches, talk with nodes ĩ“ n the same neighborhood”
- Use a wiring pattern such that each  node pair has at least one switch in common
  - Design is a problem in graph theory
  - Genetic Algorithm evolves a solution!

# Flat Neighborhood Network



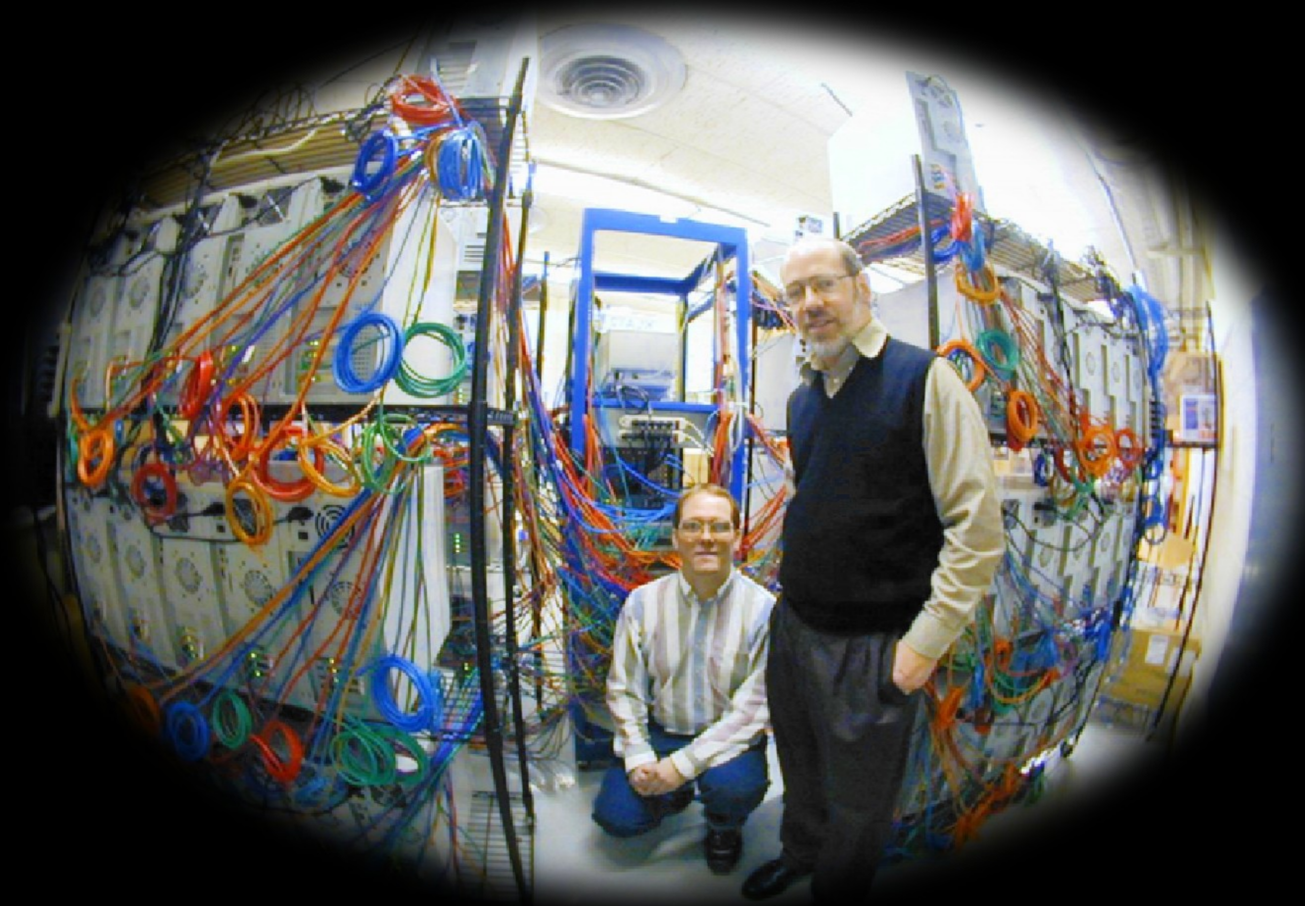
# Flat Vs. Fat

- Latency:
  - 8 node, 4 port: 1.0 vs. 2.7 switch delays
  - 64 node, 32 port: 1.0 vs. 2.5
- Pairwise bisection bandwidth:
  - 8 node, 4port: 1.29 vs. 1.0 units
  - 64 node, 32 port: 1.48 vs. 1.0
- Cost: more interfaces vs. smart routers
- Summary: **Flat Neighborhood wins!**

# KLAT2, Gort, & Klaatu


UK UNIVERSITY OF KENTUCKY

# Behind KLAT2



UNIVERSITY OF KENTUCKY

# KLAT2 Changed Everything

- KLAT2 (Kentucky Linux Athlon Testbed 2):
  - 1st network designed by computer
  - 1st network deliberately asymmetric
  - 1st supercomputer under $1K/GFLOPS
- 160+ news stories about KLAT2
- Various awards:
  - 2000 Gordon Bell (price/performance)
  - 2001 Computerworld Smithsonian, among 6 Its most advancing science

UK UNIVERSITY OF KENTUCKY

# Cool, But What Have You Done Recently?

- LOTS!
  - Nanocontrollers
  - GPUs for supercomputing
  - Warewulf & cAos systems software
  - etc., see:

**Aggregate.Org**

# Did I Mention SFNNs?

- Real parallel applications don't actually have every node talk to every other node
- Design the network to be "**Sparse**": FNN properties only for the node pairs that actually will talk to each other
- Network complexity apparently grows as **O(N*N)**, but this makes it **O(N*LogN)**!

# June 2003, KASY0


UK UNIVERSITY OF KENTUCKY

# KASY0

- 128-node system using 24-port switches!
- KASY0 (Kentucky ASYmmetric zero):
  - 1st Sparse FNN
  - 1st physical layout optimized by GA
  - 1st TFLOPS-capable computer in KY
  - 1st under $100/GFLOPS
  - World record fastest POVRay 3.5

# POVRay 3.5 Benchmark


UK UNIVERSITY OF KENTUCKY

# Supercomputers R Us

- We make supercomputing cheap!
- You can help...
  - Build parties
  - Weekly research group meetings
  - Projects
- Everything's at:

**Aggregate.Org**

UK UNIVERSITY OF KENTUCKY

Aggregate.Org

UNBRIDLED COMPUTING