

Genentech's IT Infrastructure: Evolving to support the Revolution

John "Scooter" Morris, Ph.D.
Distinguished Systems Architect
Computing Technologies

Genentech, Inc.

“Genentech is a leading biotechnology company that discovers, develops, manufactures and commercializes biotherapeutics for significant unmet medical needs.”

Statistics

- ~5,000 Employees
- ~\$2.2B in Revenue (2001)
- 10 products
 - Protropin[®], Nutropin[®], NutropinAQ[®], NutropinDepot[®], Cathflo[™] Activase[®], Activase[®], TNKase[™], Pulmozyme[®], Herceptin[®], Rituxan[®]
- 1 product awaiting FDA approval
 - Xolair[™]
- Three major sites
 - South San Francisco, California
 - Vacaville, California
 - Porriño, Spain
 - Several U.S. Sales offices

Genentech, Inc.

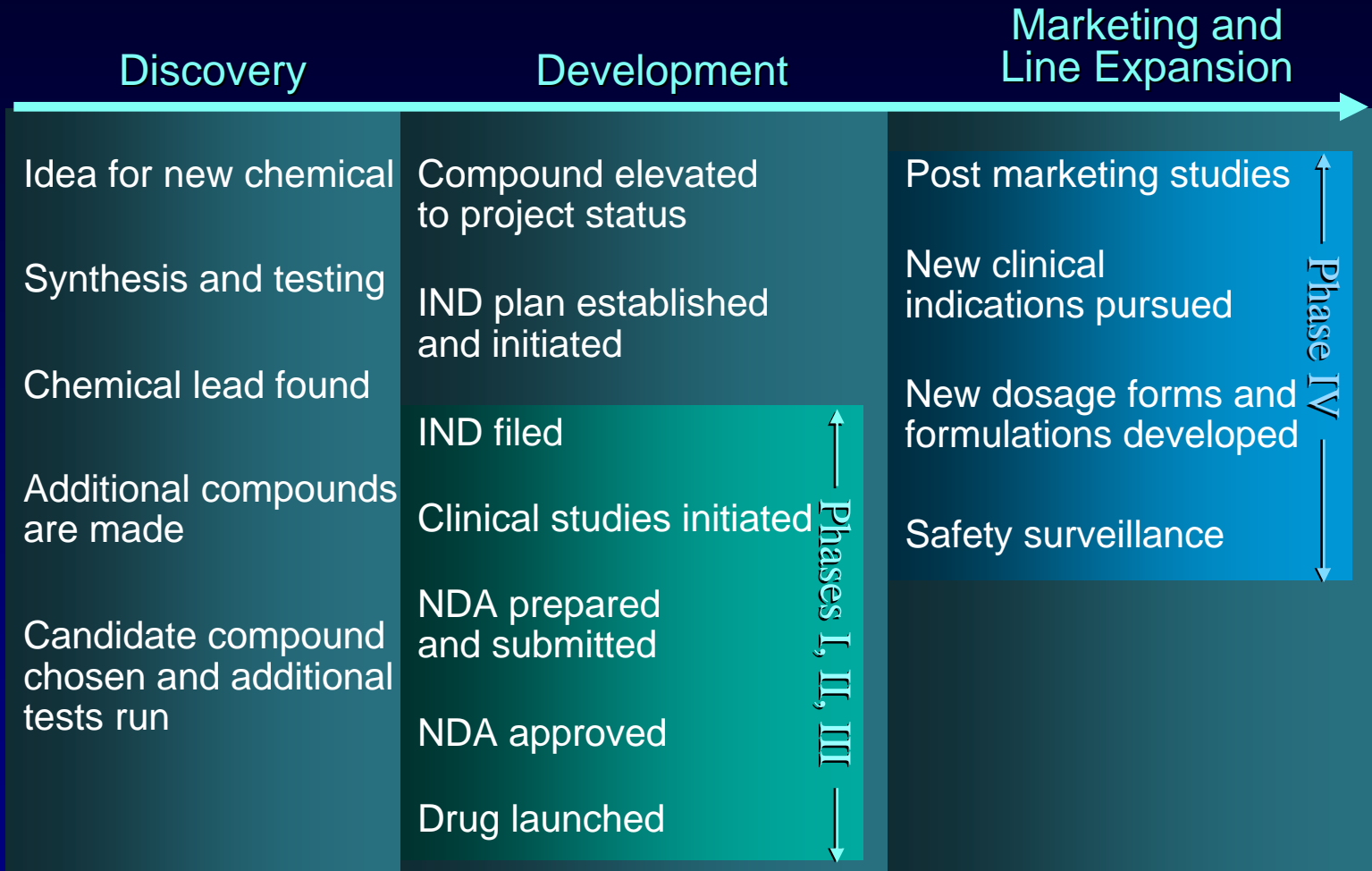
Founded in 1976

- Herb Boyer – UCSF Professor
- Bob Swanson – Entrepreneur, Venture Capitalist

Focus areas:

- Oncology
 - 7 drugs or new indications in pipeline
 - 4 in phase III
- Immunology
 - 5 drugs or new indications in pipeline
 - 2 in phase III, 1 awaiting approval
- Opportunistic
 - 3 drugs or new indications in pipeline
 - 1 in phase III

Clinical Development of Drugs



Outline

Current Infrastructure

- A bunch of details which I will skip

Research Computing

Revolution vs. Evolution

Evolution of Computing for Research

- Evolutionary tree

Future Directions

Current IT Infrastructure

Highly heterogeneous

- Servers: SGI, HP Alpha, HP PA-RISC, HP Intel, Sun
- Desktops: Mac, PC, some SGI

Primarily IP-based network

- AppleTalk also supported
- Routers and switches: Cisco

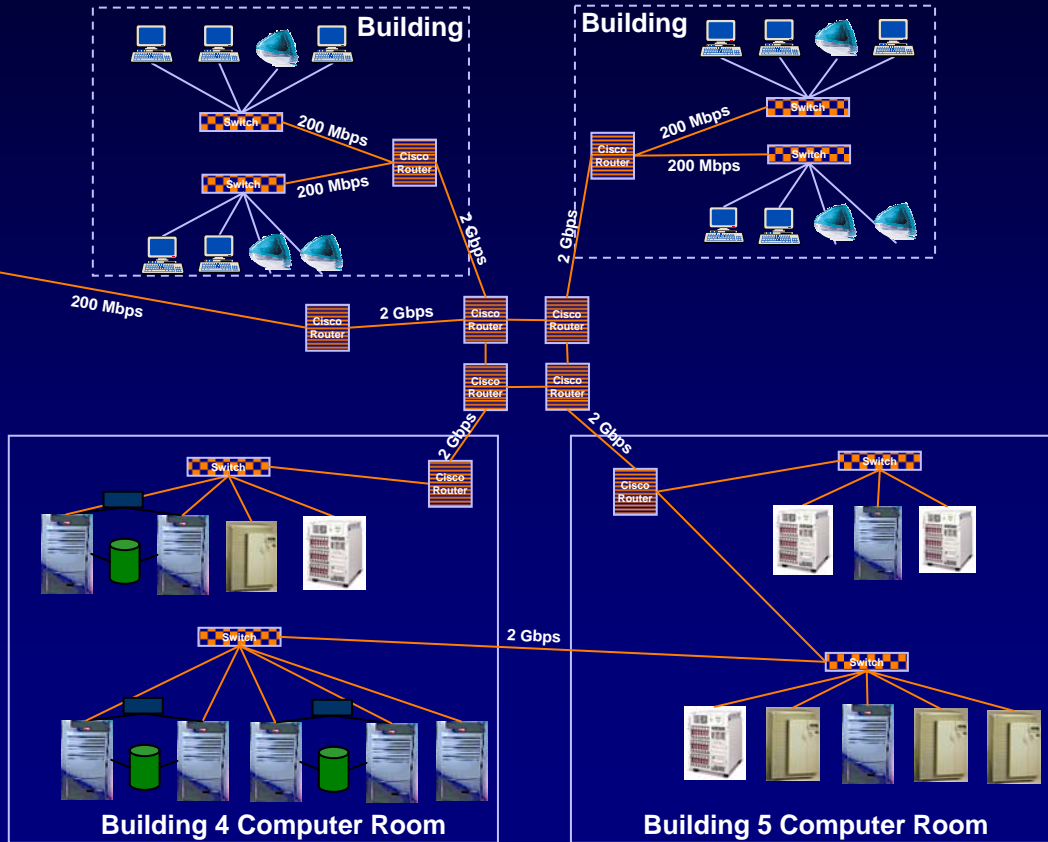
Security based on M&M principle

- Hard outer shell, soft inside
- Some “softness” appearing to support collaboration
- Important to maintain open environment

Current IT Infrastructure



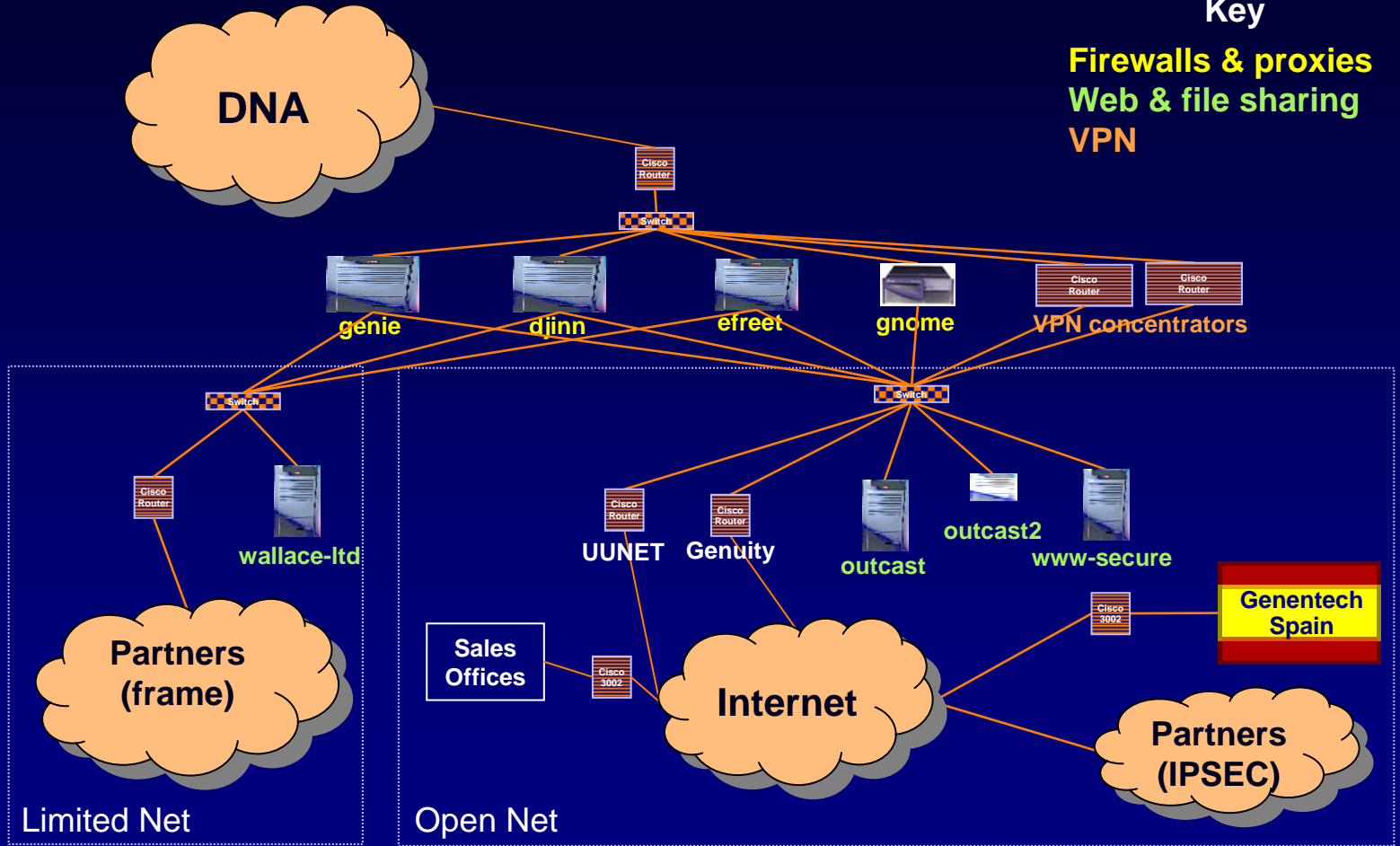
Vacaville



Current IT Infrastructure

Key

- Firewalls & proxies
- Web & file sharing
- VPN



Details

This starts the part I'm going to leave out....

Details - Network

Backbone: Switched Gigabit Ethernet

Vacaville link: 200 Mbps SONET Ring

Desktop: Switched 10/100 Ethernet

Routers: CISCO

Addressing: DHCP preferred

Naming: DNS (Bind 9.2.1), WINS, Active Directory, LDAP

Firewall: SOCKS5 (Aventail)

Monitoring: Big Brother, HP Open View

VPN: Cisco 3000, IPSEC

Details - Desktop

Compaq (now HP)

- Migrating to Windows 2000

Apple

- Migrating to Mac OS X

Primary Applications

- Microsoft Office 2000/2001
- Netscape Communicator (Browser, Mail)
 - Migrating to Mozilla/Netscape 7
- Steltor CorporateTime
- Norton Antivirus
- FileMaker Pro

Details - Server

HP Tru64 Unix

- Web, E-Mail (IMAP), Bioinformatics, Infrastructure (DNS, Firewall, DHCP, backup/restore, LDAP), General computing, Oracle
 - 5.1A (TruCluster Server 5.1A)

HP/UX

- Manufacturing, Commercial Computing (Lawson, PeopleSoft)
- 10.20, 11.0

Solaris

- Medical Affairs, Infrastructure (Calendar, Remedy, Web Proxy), Research

SGI

- Molecular Modeling, Computational Chemistry

Linux

- Computational Chemistry (pilot)

NT

- Workgroup Computing, Specific Applications

Details - Database

Oracle 8.1.7

- Exploring Oracle 9iRAC

n-tier approach

- Web Browser for presentation
- Web server for static pages
- Application servers for business logic
- Database server for data store

Details - Web

Server: Netscape Enterprise Server 4.6

- Migrating to Apache

Programming: Perl/CGI, JSP, Javascript

Application Servers:

- WebLogic

Distributed Computing:

- Direction is towards Enterprise Java Beans
 - WebLogic
 - Tuxedo in use for Manufacturing applications

Development Tools:

- Dreamweaver, JBuilder, Visual Age

Details - Security

Security is based on Kerberos V5

Provides authentication for Unix and Windows 2000

Oracle accounts often use Unix username, but also lots of application-specific accounts

- Exploring use of Kerberos for Oracle accounts

VPN and dial-in access through SecurID tokens

LDAP is used for Directory services

- Netscape Directory Server

Serious regulatory restrictions (21CFR Part 11)

Details - Internet/Extranet

Firewall is based on SOCKS5 (RFC1928)

Totally Proxy-based (*very secure*)

Firewall has three parts:

- Internal
- Internet
- Limited net

Internet link is redundant

- 42 Mbps link with Genuity
- T1 (1.54 Mbps) link with UUNET
 - Migrating to a redundant 42 Mbps link
- Uses OSPF for dynamic fail-over
- Using IPSEC tunnels over the Internet for secure communications with partners

Cisco 3000 VPN concentrators

- Employee access only

Reverse-web Proxy

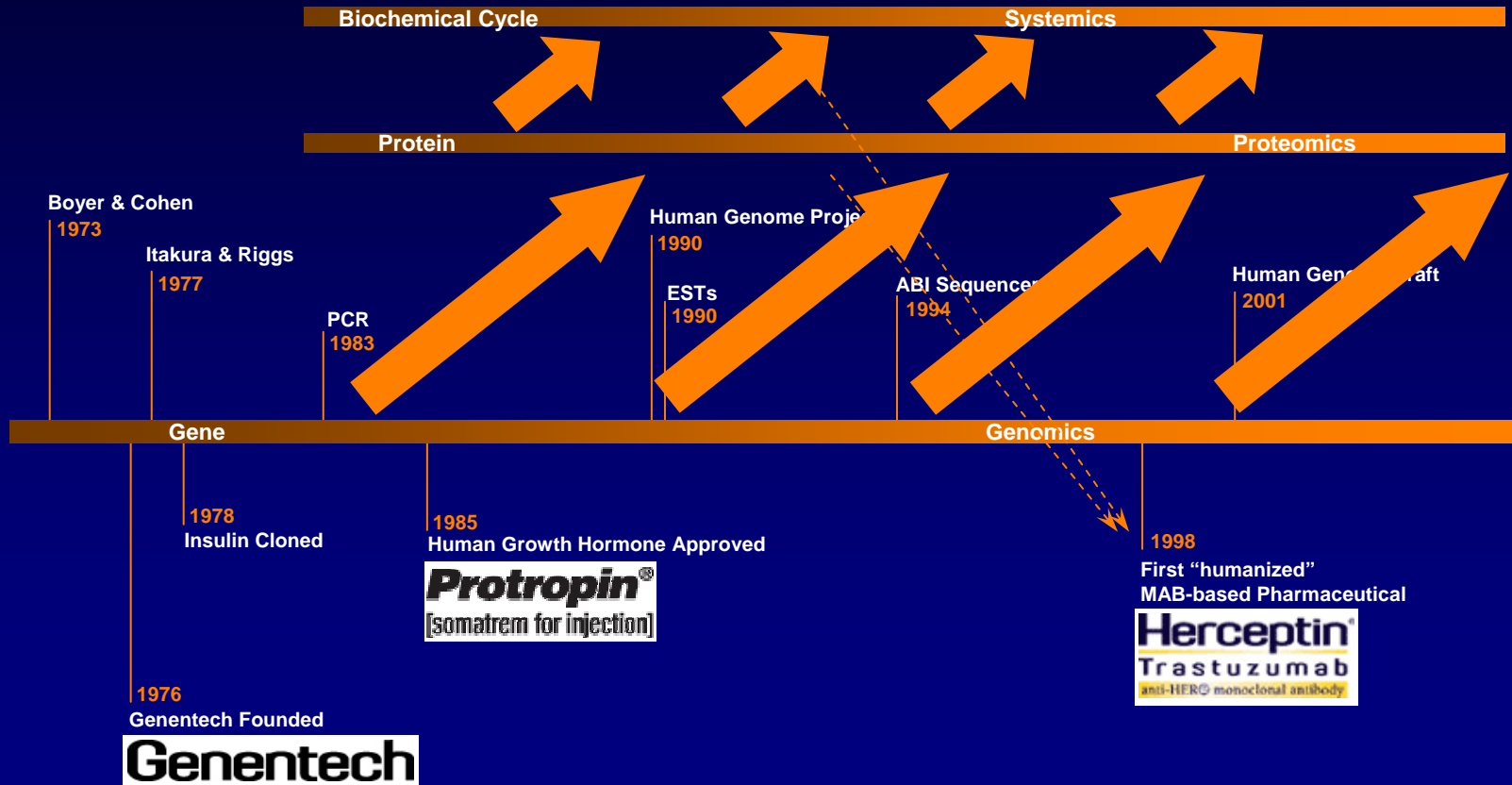
- Allows external partners access to selected internal web sites

Details

Any questions on the details?

- I didn't think so....

The Revolution (a parochial view)



The Revolution

More than changes in scale

Fundamental changes in *scope* and *complexity*

- Gene→Genome→Proteome→Systems

Dependent on computation

- Dramatic increases:
 - Performance
 - Storage capabilities
 - Communication capabilities

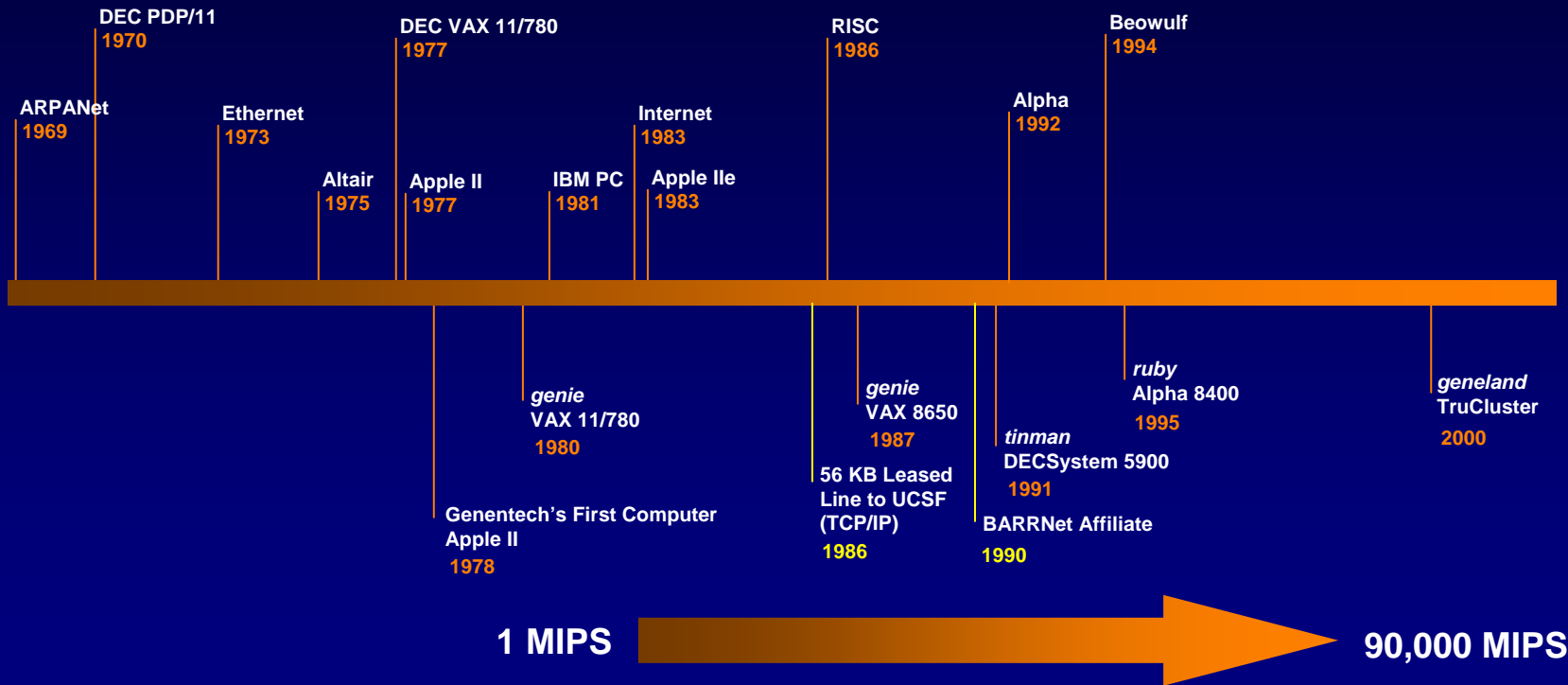
Evolution of computing at DNA

In the beginning was the.....



Apple II

Computing “Revolution” (a parochial view)



MIPS = Meaningless Indicator of Processor Speed

Computing “Revolution”

Changes in scale

- System performance
 - 1 MIPS → 90,000 MIPS (clustered)
- Memory capacity/system
 - 4MB → 88GB (~4 orders of magnitude)
- Global connectivity
 - 9.6 Kbps → 42 Mbps

Changes in complexity

- Multiprocessing (threaded code)
- Clustering (distributed code)

Strategies for IT Change

Change is required to:

- support changing research needs
- take advantage of changing technology

Strategies

- Revolutionary
 - Rip and replace
 - Takes maximum advantage of changing technology
- Evolutionary
 - Incremental changes
 - Minimizes impact to users
 - Might result in slower adoption of technology
- Genentech has taken an evolutionary approach

Setting - Research

Goal

- Basic research
- Human pharmaceuticals

Organization (roughly)

- Discovery
 - e.g. Molecular Oncology, Immunology
- Technology
 - e.g. Bioinformatics, Protein Engineering, Bioorganic Chemistry

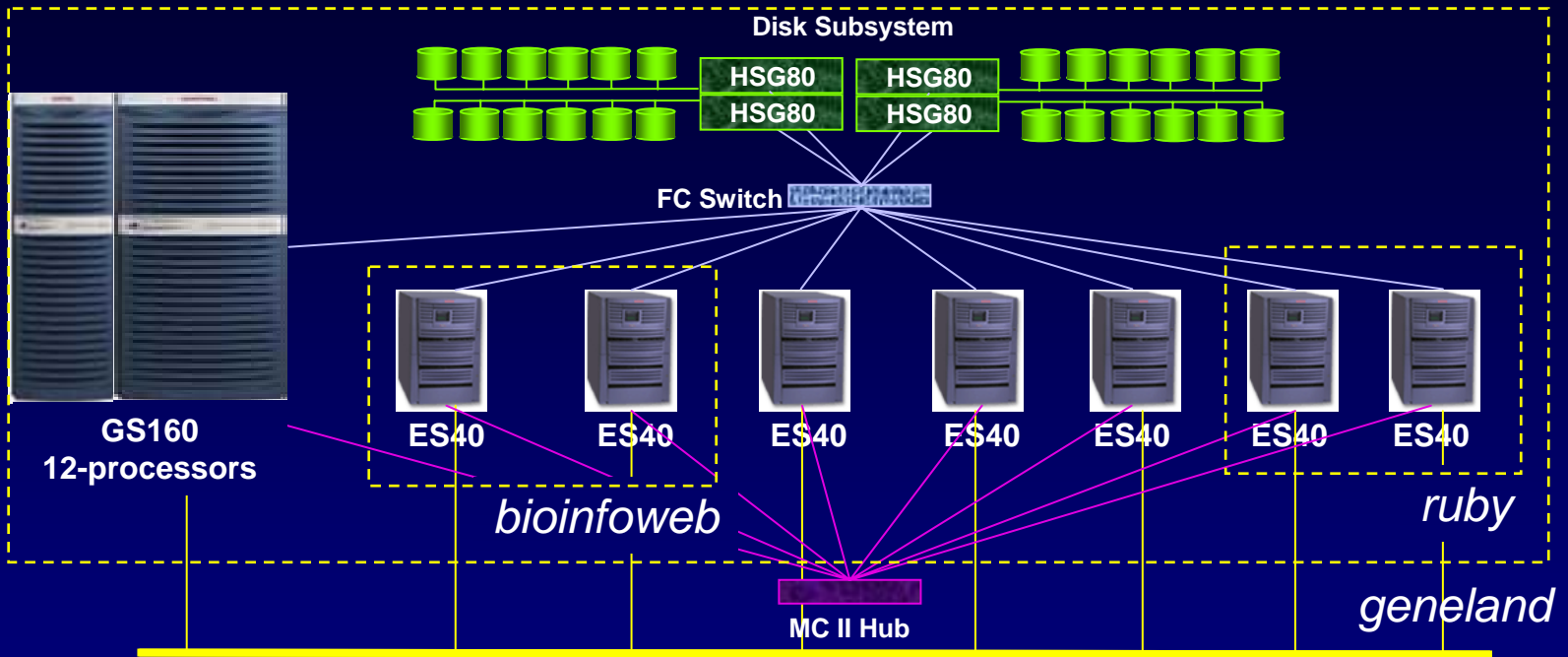
Academic culture

- Open, fast-paced environment
- Need to provide tools as much as solutions

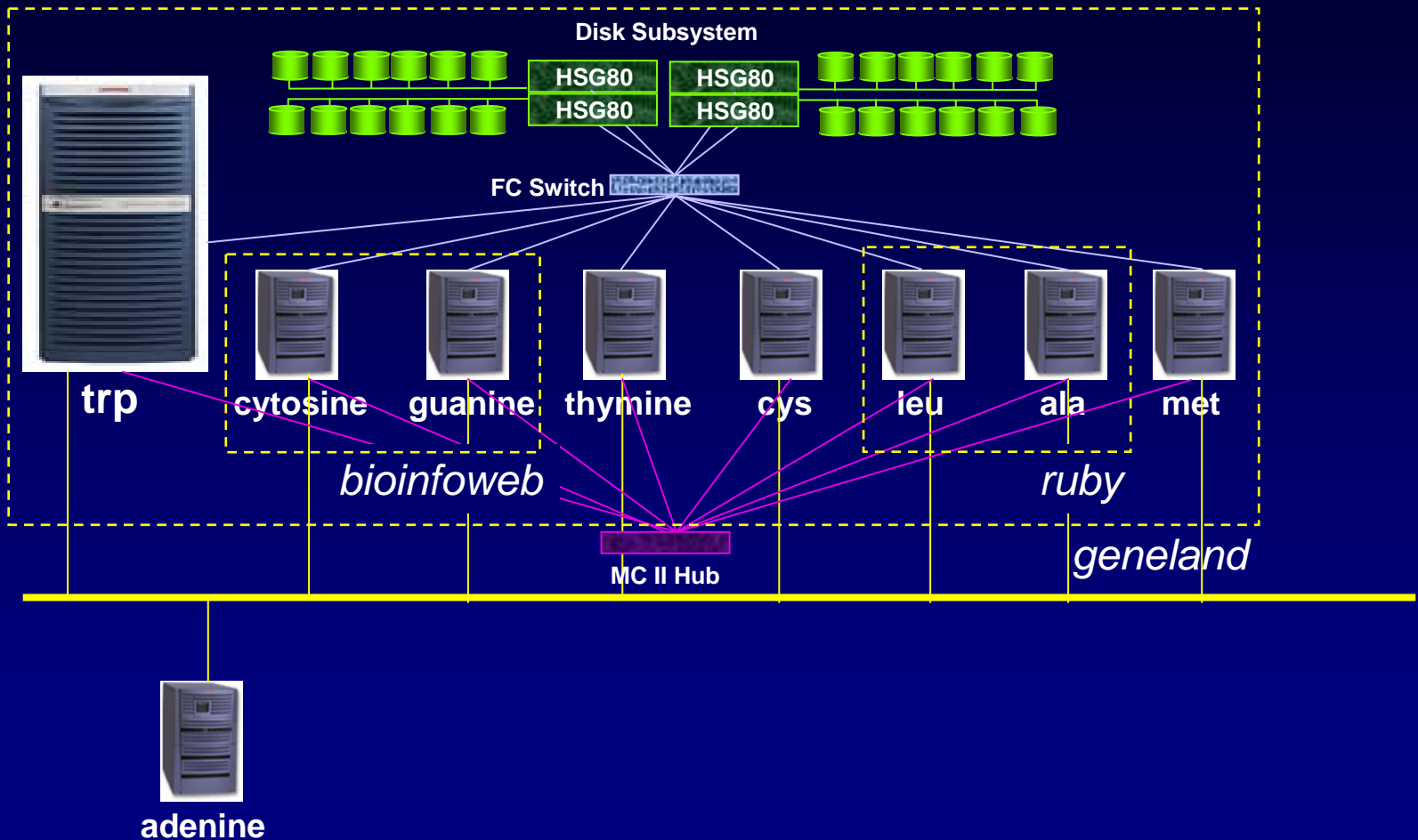
Computational needs are high

- Bioinformatics
- Molecular modeling
- Computational Chemistry

Bioinformatics Computing Evolution



Bioinformatics Cluster



Protein Engineering / Bioorganic

Linux Cluster



SGI Origin 2000
(12 processors)



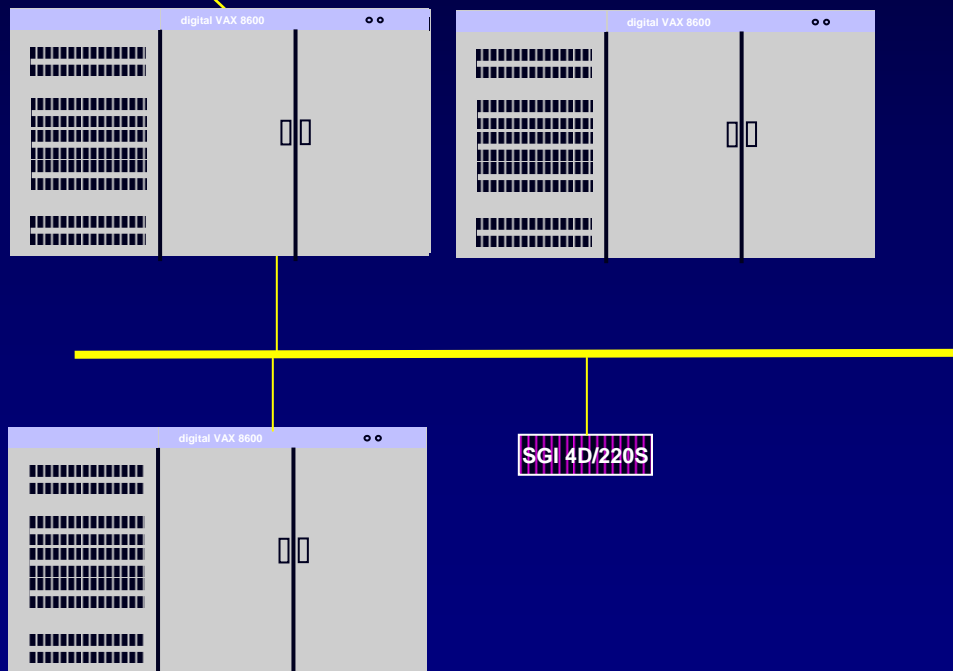
SGI Origin 2000
(16 processors)



SGI Origin 3000
(24 processors)



Evolution of computing at DNA



1980: VAX 11/780

- BSD Unix
- UUCP Dialup Connection to UCSF

1984: VAX 11/785

- BSD Unix
- “Corporate” network

1986: Connection upgrade

- 56Kb Leased line (TCP/IP)

1987: VAX 8600

- BSD Unix
- gene.com registered
- BARRNet affiliate (1990)

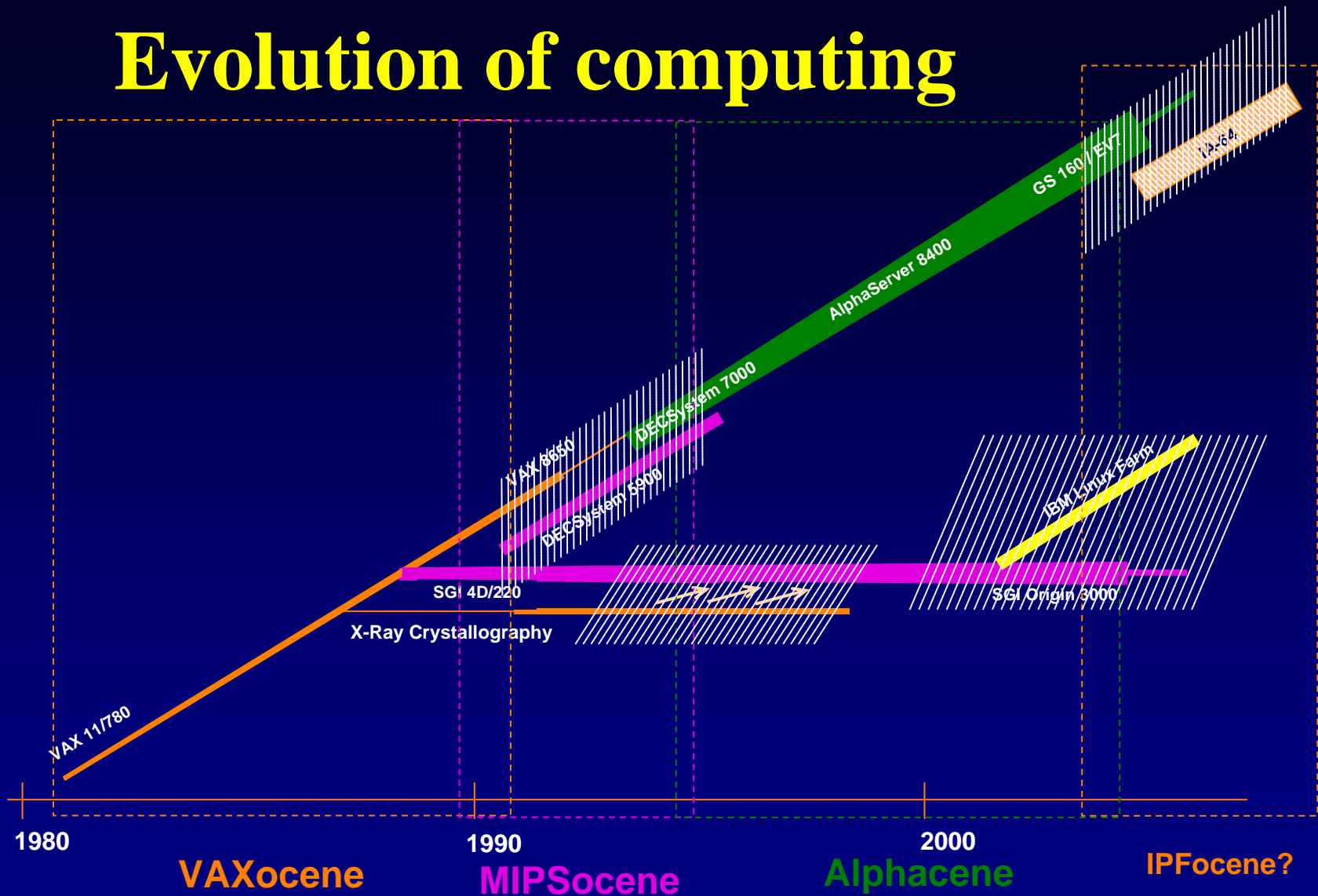
1989: VAX 8650

- BSD Unix
- Total of 3 shared systems

1989: SGI Server

- IRIX
- NFS Services
- Molecular modeling

Evolution of computing



Evolution of computing – Punctuated Equilibrium?

We have been in a period of relative stability

Episodes of change are “normal”

- Hardware changes
- OS changes
- Technological changes (Java, J2EE, etc.)
- Vendor changes

Our job:

- Making the change less intrusive
 - Attractiveness of Java
 - Vendor stability
 - Planning
- Avoid revolution

Futures

Linux Pilot

Purchased a 24-node Linux cluster

- 2 800MHz PIIIs in each node
- Myrinet
- Computational Chemistry Applications
 - Amber, Gaussian

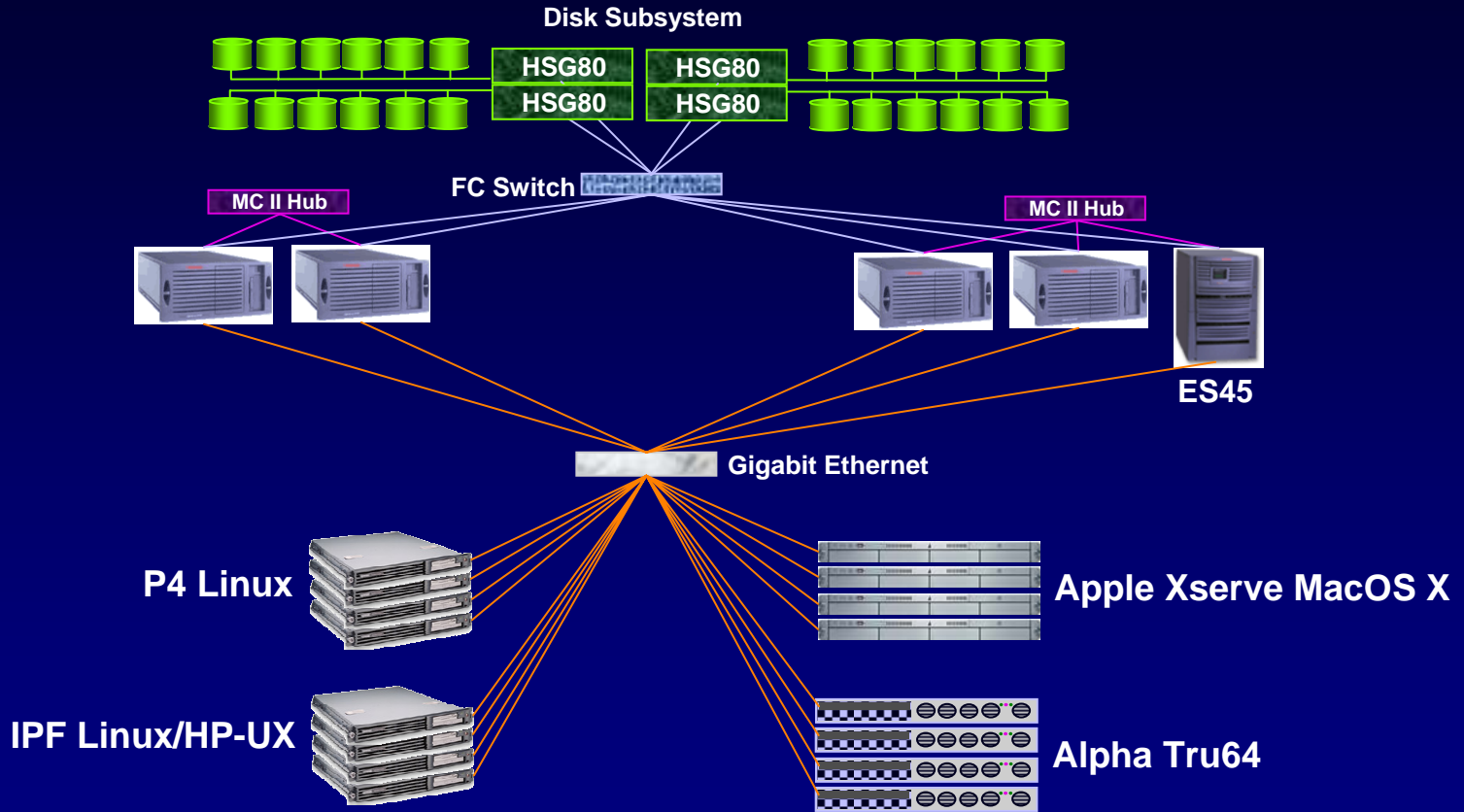
Results

- Memory bandwidth was a problem
- PIIIs already out-of-date
- Myrinet capabilities not heavily used

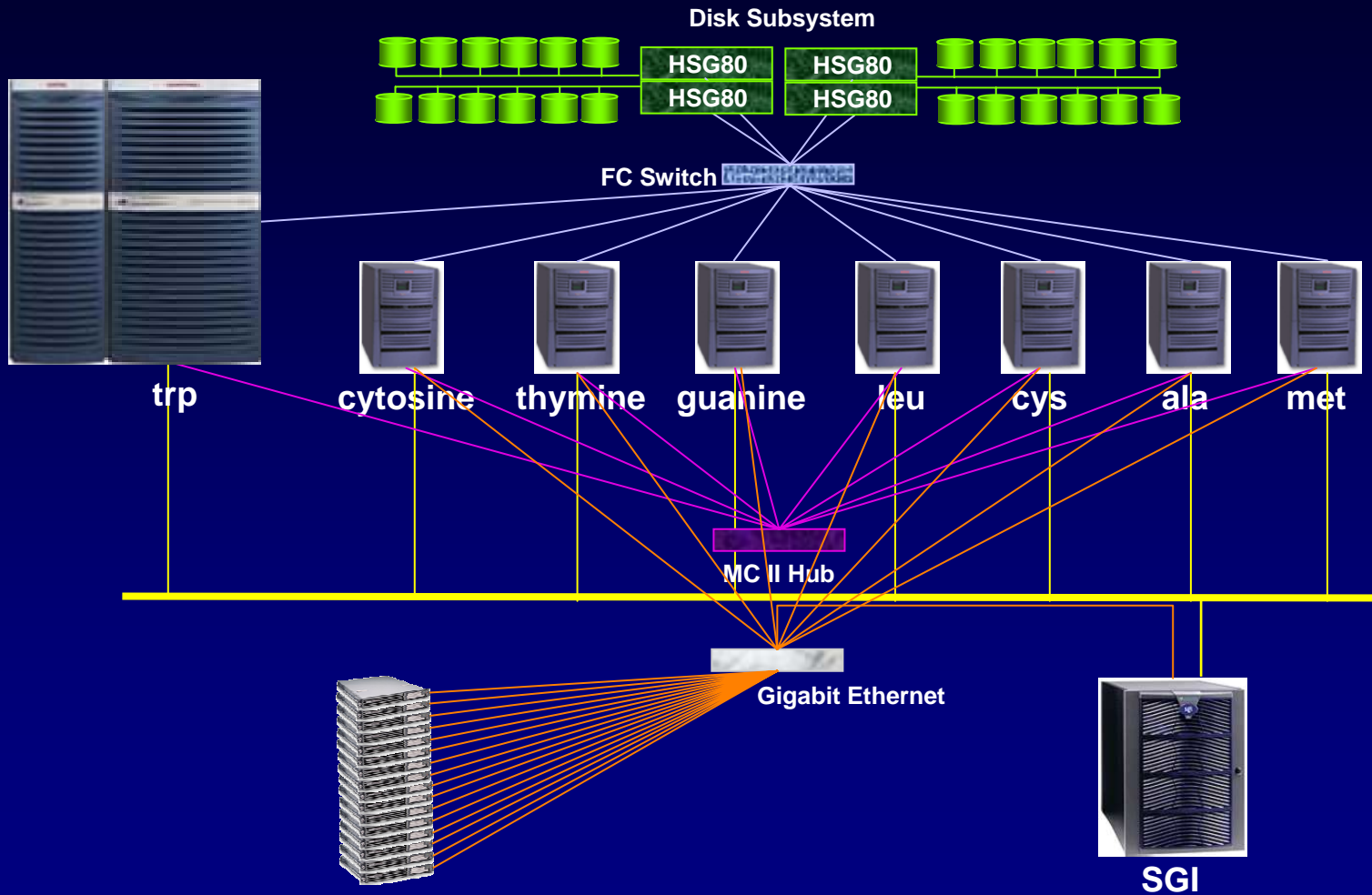
Hypothesis

- Could build a cluster with different types of nodes
- Submit jobs to appropriate node depending on computational needs
- May not need expensive cluster interconnect

Heterogeneous Compute Farm



Research Computing Environment – Future?



Questions?

Good, I've got some questions:

- **How many are Tru64 customers?**
- **How many are concerned about Itanium transition?**
 - **What are your concerns?**
- **How many are concerned about HP/UX transition?**
 - **What are your concerns?**

Thank you!

Acknowledgements:

Colin Watanabe

Carol Morita

Roy Takai

Pam Fitch