

# **Strategic Planning for NCAR's Computing Facility: A Status Report**

**Dr. Richard Loft  
Associate Director of R & D  
Scientific Computing Division**



**NCAR**

# Talk Outline

- **Driving Issues for a New Facility**
  - Geoscience: **Exponential demand for supercomputing**
  - Technology: **CMOS is getting hotter and cheaper**
  - Engineering: **Mesa Lab facility design is obsolescent**
- **Status of UCAR Response**
  - What are our peer supercomputing centers doing?
  - Short term (1-2 year):
    - Mesa Lab risk mitigation (RMH, Inc.)
  - Medium term (2-5 years):
    - lease options (Staubach Real Estate)
    - Colocation (Inflow, Inc.)
  - Long term (5+ years):
    - Refurbishment of existing structures (RNL Design, Inc.)
    - new construction (Uptime Institute, Inc.)

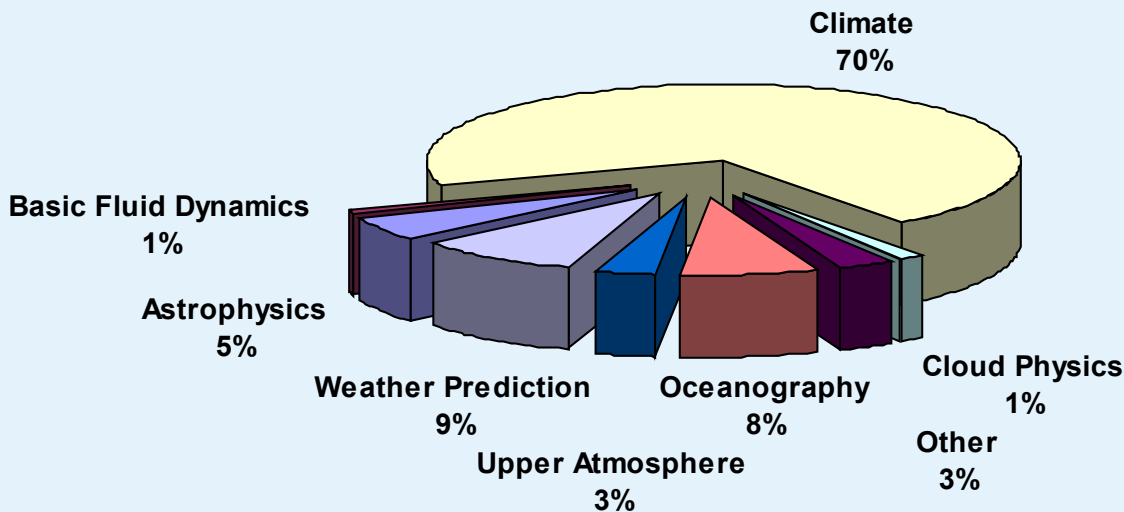


# The Science Drivers...



# Climate Simulation Dominates NCAR Resource Usage

- At the end of FY03, the combined supercomputing capacity at NCAR was approximately 8.5 TFLOPs
- Roughly 70% of that capacity was used for climate simulation and analysis

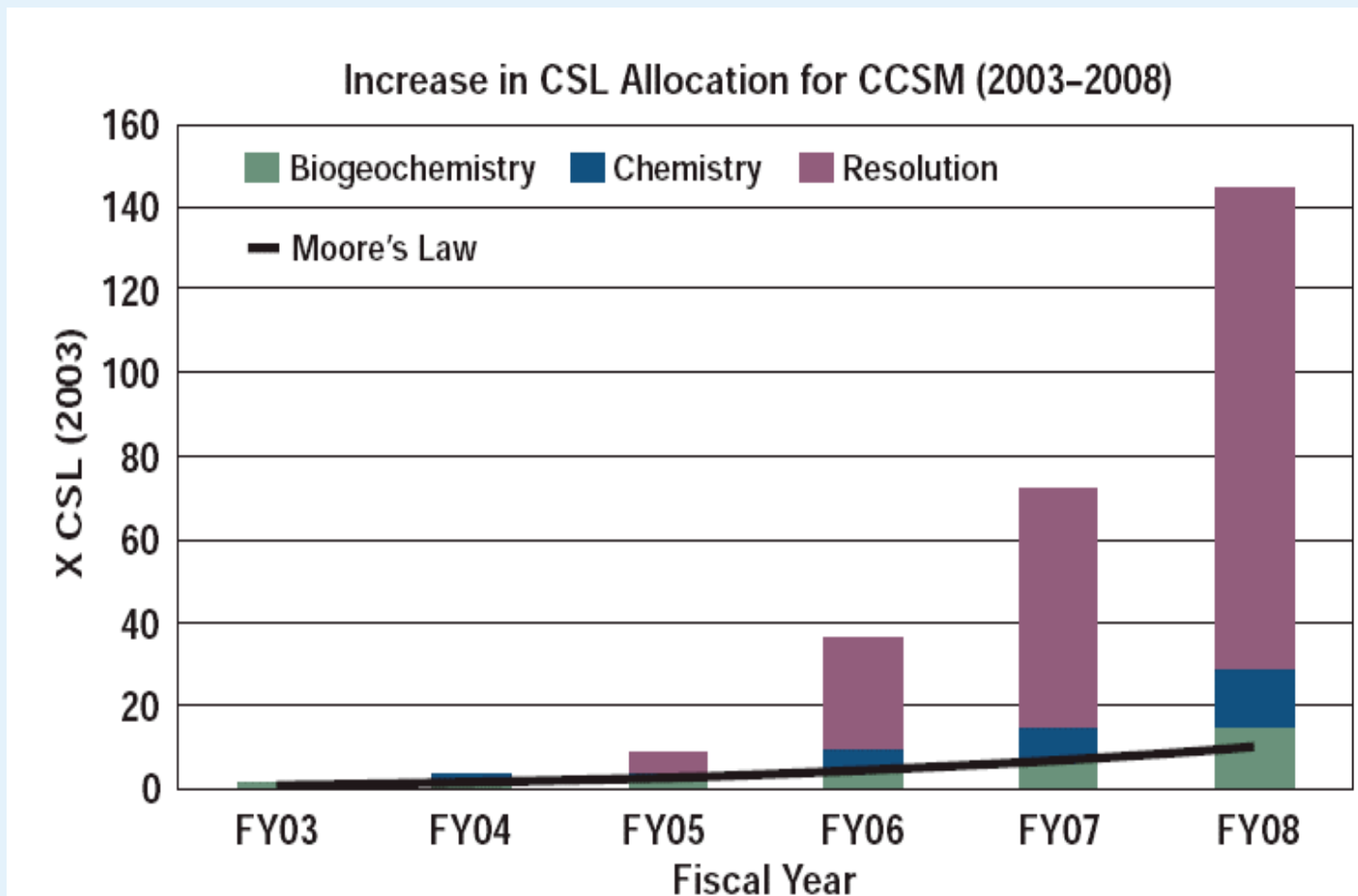


# General Trends

- Toward Higher Resolution
  - Eddy resolving oceans
  - Cloud resolving atmospheres
- Toward Higher Complexity
  - More components
  - More chemical species
  - More physical processes modeled



# Projected Climate Computing Requirements (CCSM)



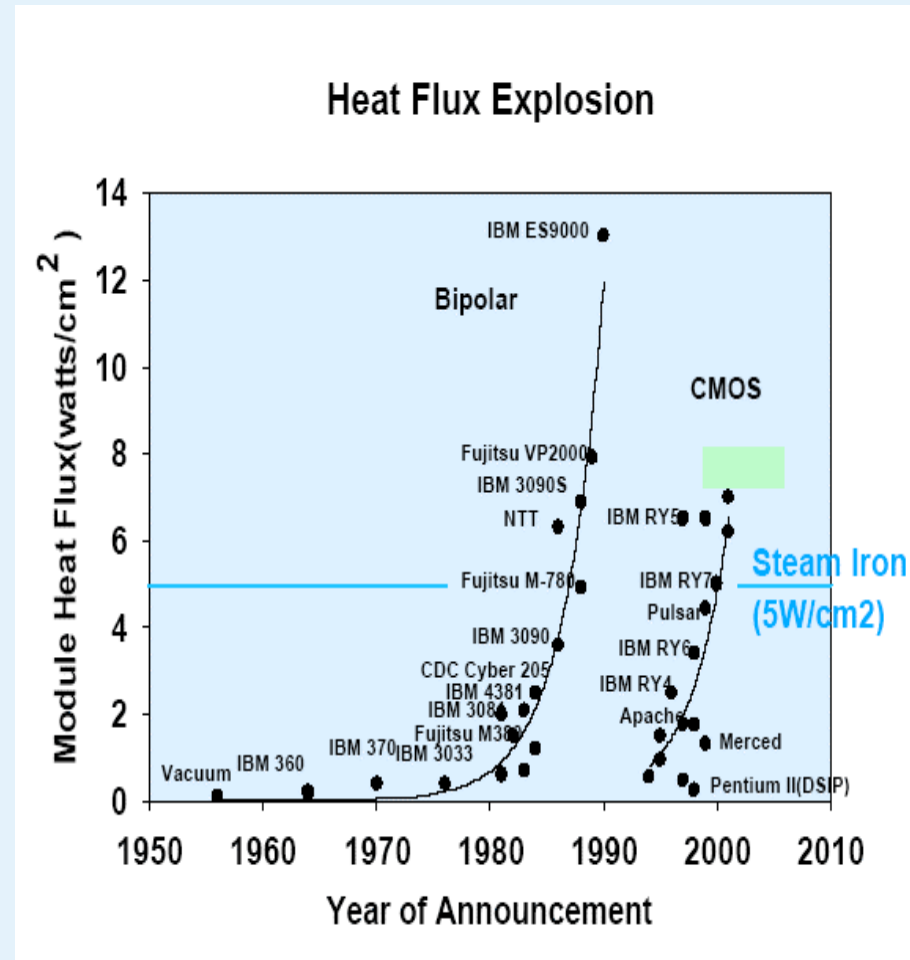
Thanks to Jeff Kiehl/Bill Collins

# Trends in Computer Technology...

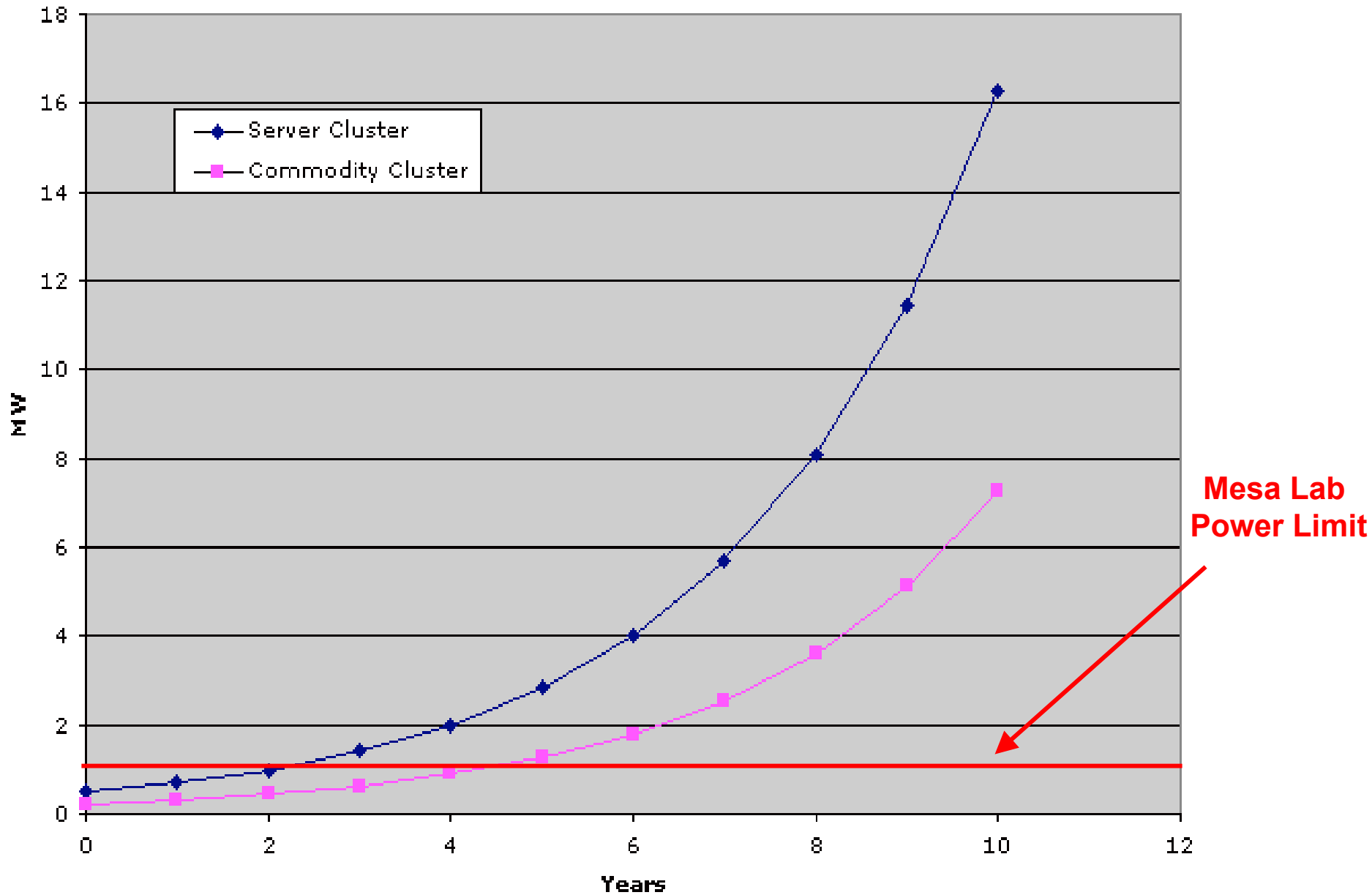


# Chips: Faster, Cheaper but **Hotter**

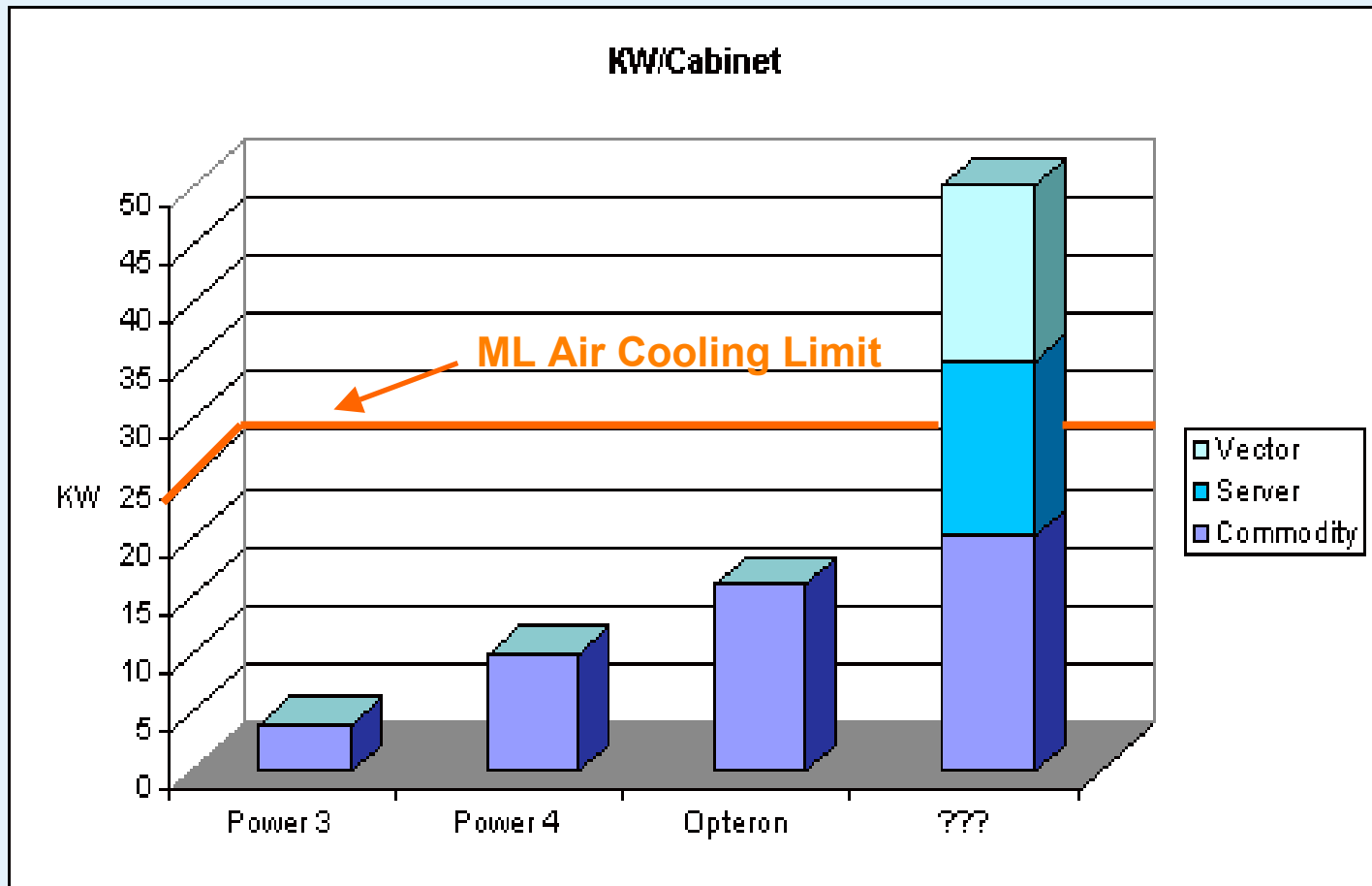
- Power problem:
- $P = C V^2 f$ 
  - Frequency (clock) increase driving Moore's Law
  - Decreasing voltage increases gate delay (hurts performance)
  - Capacitance increases as gate count increases.
- Chip power consumption will increase **30% over 5 years**.
- Processor costs decreasing at **29%/year**.



# 2004-2014 System Power Trends (10%/yr)



# These projections are Materializing at the Mesa Lab Facility



# CMOS Power Impact

- CMOS situation is similar to what happened to Bipolar in the early 1990s.
- Difference: **there is no ready replacement for CMOS on the horizon.**
- Additional pressures: as feature size decreases to about 50nm, static power (leakage) will become comparable to dynamic power (flipping bits).
- **Impacts on facility infrastructure of these developments have materialized and must be dealt with.**



# Engineering Realities...



# Computers as space heaters

- CPU's consume energy (electricity)
- This energy is primarily converted to heat
- To remove the heat more energy is consumed
- Net result for every kW into the system an equivalent kW is needed by the cooling system to remove the waste product (heat)
- Computers are really great space heaters with the side effect of calculating numbers!



# The Problem with the Mesa Lab Facility

- NCAR's machine room was built during the era of low power/high cost computing systems (1976).
  - **13000 sq ft**
  - **1.2 MW computer power limitation**
  - **80 W/sq ft cooling design**
- We live in an era in which high power/low cost computing systems are available.
- It is not cheap to modify the power infrastructure of Mesa Lab Computing Facility, nor does it make sense without expanding the machine room floor.
- Site situated on bedrock.
- **If we do nothing, the Mesa Lab power and design limitations will become a serious constraint on NCAR's scientific computing capacity. This is true for both future vector and microprocessor systems.**



# The problem is here: 2006 Procurement Scenario

- Assume \$16M 2006 acquisition.
- Estimated Power requirements: 970 kw
- Estimated used at ML facility: 650 kw
- Exceeds 1.2 MW for computing by 420kw.
- Thus, problems begin in as little as 12-18 months.



# **Status of UCAR Response...**



# What are other centers doing?



*Oak Ridge's new Computational Sciences Building (blue) and the Joint Institute for Computational Sciences (green).*

# Other Peer Centers

- **Several peer centers are augmenting their facilities.**
  - **NERSC – new facility in Oakland**
  - **CanadaMet – expansion of existing facility**
  - **SDSC – part of their five year plans**
  - **Oak Ridge – 40,000 sq. ft. expansion**
  - **UK Met – two 20,000 sq. ft. computer rooms.**
  - **ECMWF - new computer hall**
  - **Australian BoM has new facility in Melbourne.**
  - **DKRZ has added second floor.**



# Computer Facility Options

- Short term (1-2 years): stabilize Mesa Lab facility infrastructure
  - Chilled water expansion project
  - Risk assessment
- Medium term (2-5 years): obtain temporary space
  - Lease
  - Collocation
- Long term (5+ years): Build new facility
  - Retrofit existing space:
    - **Mesa Lab**
    - **Foothills Lab**
    - **Center Green (CG-3)**
  - Construct new space from “greenfield”.



# **Short Term: Mesa Lab Risk Mitigation**



# Chilled Water Expansion

- Goal:
  - Expand the cooling capacity to match maximum power to computing equipment.
  - Simplify design of chilled water system.
  - Improve cooling redundancy
- General contractor: **U.S. Engineering**
- Estimated completion date: March 2005



# Mesa Lab Risk Assessment

- Goal:
  - outside inspection to identify objective risks to Mesa Lab computing facility.
  - Assign investment priorities
- Scope of study
  - Electrical systems - e.g. utility feeders
  - Mechanical systems - water supply, pumps
  - Fire suppression systems
  - Maintenance records
- Competitive bid process held in September 2004
- Engineering firm selected: **RMH, Inc.**
- Preliminary Findings: Early October 2004
- Completion Date: November 1, 2004



# Mid-term Solutions...



# Lease Option

- Engaged **Staubach Real Estate**
- Toured Seven Existing Data Centers
  - 36 Corridor
  - Denver metro
- Pros
  - Expedient short term solution
- Cons
  - Most data centers on the lease market are also obsolete and would require significant modification
    - 1980's designs 30 – 50 Watts / sq. ft.
    - \$2 - \$5Million.
  - Good ones quickly exit the market
    - Best two of the seven visited are no longer available



# Collocation Option

- **This is truly renting vs. owning**
- **Vendors Contacted**
  - **Inflow, Inc.**
  - **Level 3**
- **Pros**
  - Expedient short term solution
- **Cons**
  - No residual value
- **Process**
  - Proposals received September 16, 2004
  - Inflow data center tour September 27, 2004



# Collocation Option (cont)

- **Recall 2006 Procurement Scenario**
  - 80 cabinets
  - 120 amps/cabinet
- **Level 3 Example**
  - \$80,000 setup charge
  - \$136,000 / month
- **Inflow Example**
  - \$60,000 - \$80,000 setup charge
  - \$160,000 - \$220,000 / month
- **This doesn't include network charges**



**Long term solutions...**



**NCAR**

# Retrofit Existing Facilities

- **Mesa Lab**
  - Uptime Institute, Inc. Executive Seminar (June 30, 2004)
  - Assessment of ML from seminar
    - Far from power substation
    - Very expensive location for new construction
    - Environmentally sensitive, visible site
    - Political issues with new construction
- **Center Green 3**
  - Study initiated June 2, 2004 by **RNL Design, Inc.**
  - Feasibility study completed September 22, 2004
  - Results
    - \$26M cost estimate for refurbishment
    - Resulting data center size is maxed out at 12000 sq. ft.
- **Foothills Lab 1**
  - Involves moving both ATD & SCD

# New Construction

- New “greenfield” construction
  - Possible locations
    - **CU Research Park (Preliminary Discussions)**
    - **Marshall Site (Federal Property)**
- Pros
  - Build to suite organizational needs
  - Designed for current low cost high power technologies
  - Build in flexibility
    - Preserve capital and add capacity as needed
- Cons
  - Significant upfront cost
  - Long process
    - Best case is two years
    - Realistic case is three years



# Technology Warehouse Concept

- Presented at Executive Seminar conducted by **Uptime Institute, Inc.**
- “Extra” land is cheap insurance against premature obsolescence and the subsequent cost of migration to a successor site
- 20, 10, 5 principle
  - Land for 20 years
  - Building shell for 10 years
  - Capacity build-out for 5 years



# Realistic Data Center Cost Model Components

- The dominant term of the cost of a data center is power and cooling, or “kW”.
- The contribution of square footage is secondary.
- Raised floor space is likewise a minor contribution to total cost.



# Future Steps

- Complete due diligence, including financial analysis of all options
- Expand dialogue with community
- Recommend a course of action
  - Short term
  - Mid-term
  - Long term



**Questions?**



**NCAR**



# **Impact of Computer Architecture on Computing Facility Requirements**

Dr. Richard D. Loft  
Scientific Computing Division  
National Center for Atmospheric  
Research



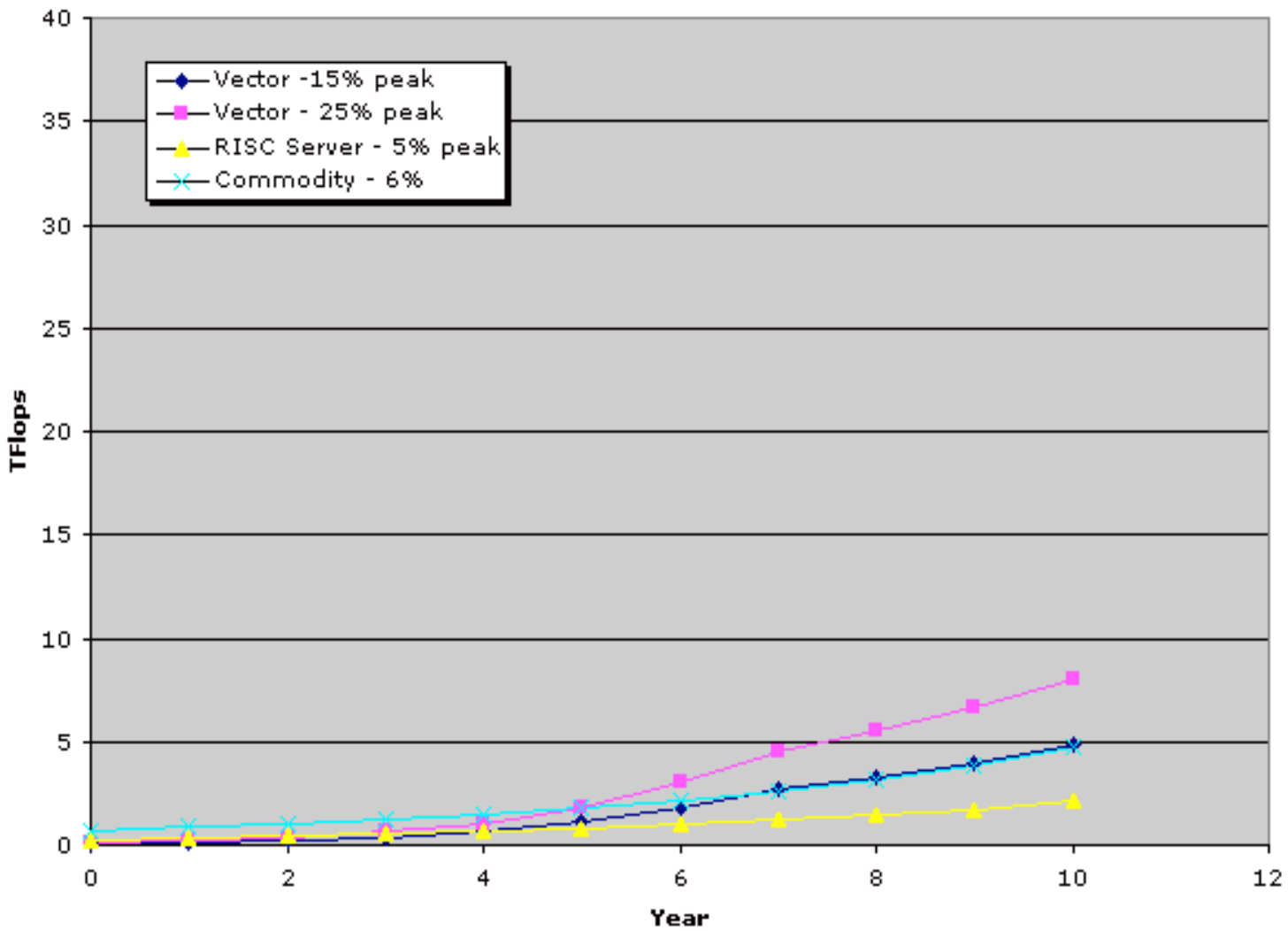
NCAR

# Assumptions governing Computing Infrastructure Model

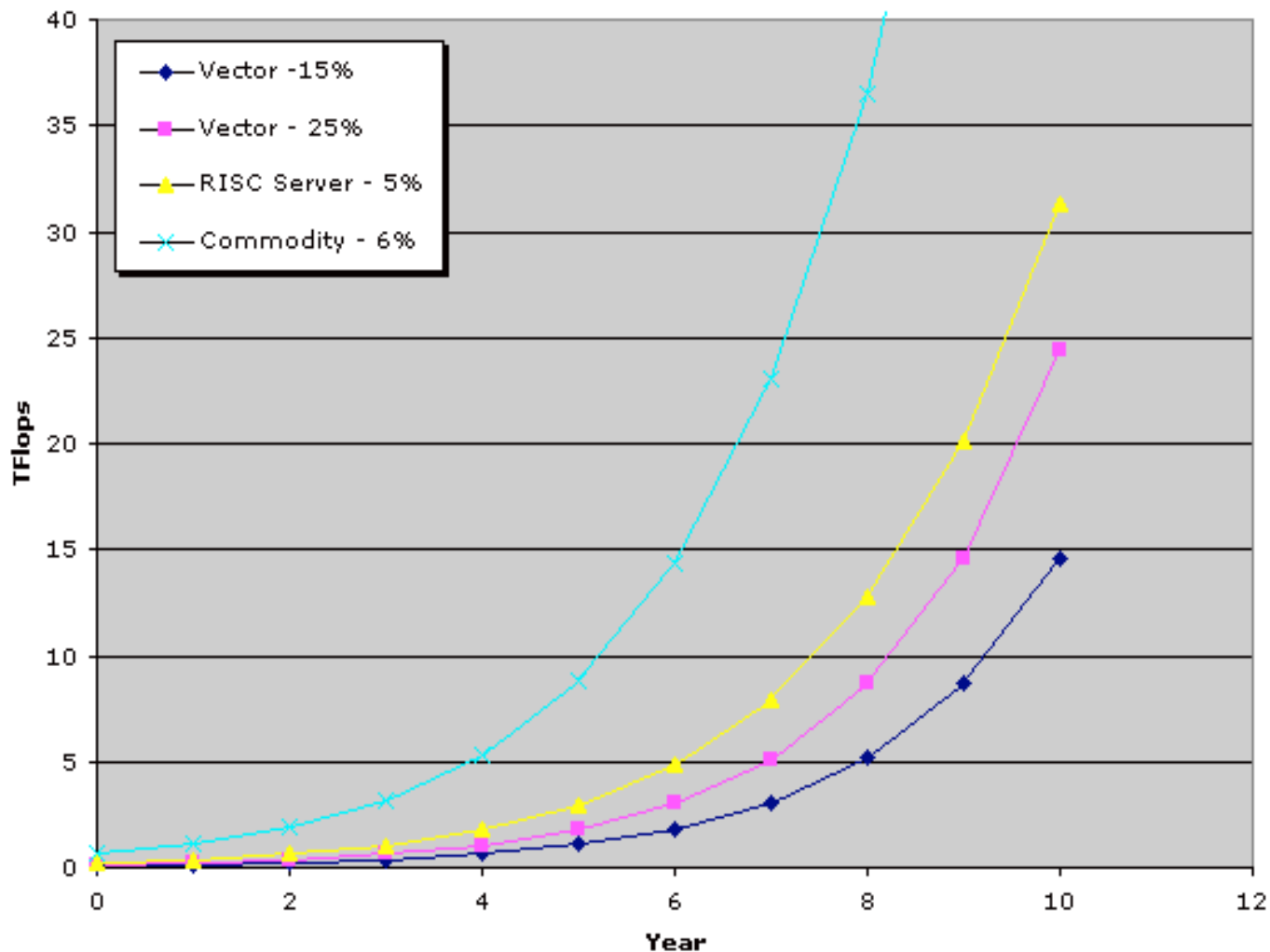
- All architectures are made from the same stuff: CMOS
- Chip power requirements continue to grow at **10%/year.**
- Per processor costs continue to decrease at **29%/year.**
- Mesa Lab computing system power supply limit remains **1.2 MW**, floor space remains **13,000 ft<sup>2</sup>.**
- machine room power density holds at **80 W/sq ft.**
- SCD computer acquisition budget and power costs grow at nominal inflationary rates.



## Power & Budget Constraint on SCD CCSM "Annual Buying Capacity"



# Budget only Constraint on SCD CCSM "Annual Buying Capacity"



# Computer Costs to Achieve 2008 CCSM Science Goals

| Architecture Type              | Sustained Price Perf. (Gflops/\$M) | Cost of CCSM allocation (\$M) | Cost of 144 x CCSM allocation (\$M) |
|--------------------------------|------------------------------------|-------------------------------|-------------------------------------|
| BG/L++                         | 897                                | 0.15                          | 22                                  |
| $\mu$ proc cluster (commodity) | 770                                | 0.17                          | 25                                  |
| Vector cluster [1]             | 281                                | 0.48                          | 69                                  |
| $\mu$ proc cluster (server)    | 250                                | 0.54                          | 78                                  |

[1] based on X1e

# Power and Space Estimates to Meet CCSM 2008 Science Goals

| Architecture Type         | System Power Requirement (MW) | Space Requirement (ksq. ft) [2] | Annual Power Costs (\$M) [1] | New Space Construction Costs (\$M) [3] |
|---------------------------|-------------------------------|---------------------------------|------------------------------|--|
| Vector cluster            | 1.4                           | 7.0                             | 0.49                         | 3.5                                    |
| BG/L++                    | 1.5                           | 7.5                             | 0.53                         | 3.8                                    |
| μproc cluster (commodity) | 5.6                           | 27.9                            | 1.96                         | 14                                     |
| μproc cluster (server)    | 11.5                          | 57.6                            | 4.0                          | 30                                     |

[1] \$350 K/MW-year

[2] 200 W/sq ft

[3] Tier 2+ (interest 3 years @ 5%)



# Total Costs of Ownership over 3 years

| Architecture Type              | Total Cost of 144 x CCSM allocation (\$M) |
|--------------------------------|---|
| BG/L++                         | 27  |
| $\mu$ proc cluster (commodity) | 45  |
| Vector cluster [1]             | 74  |
| $\mu$ proc cluster (server)    | 120                                       |

[1] based on X1e

# Conclusions

- CMOS trends affect all architectures.
- ML Facility will eventually constrain vector systems (6-7 years), others commodity architectures almost immediately.
- Cost model shows NCAR science constrained by factor of at least 3 in 10 years, more if commodity systems are considered.
- Total cost of ownership model shows vectors unlikely to be cost effective, even considering facility costs.
- Low power MPP commodity systems most cost effective, but experimental at this time.



# Semiconductor Details

Dr. Richard D. Loft  
Scientific Computing Division



# What's going on here?

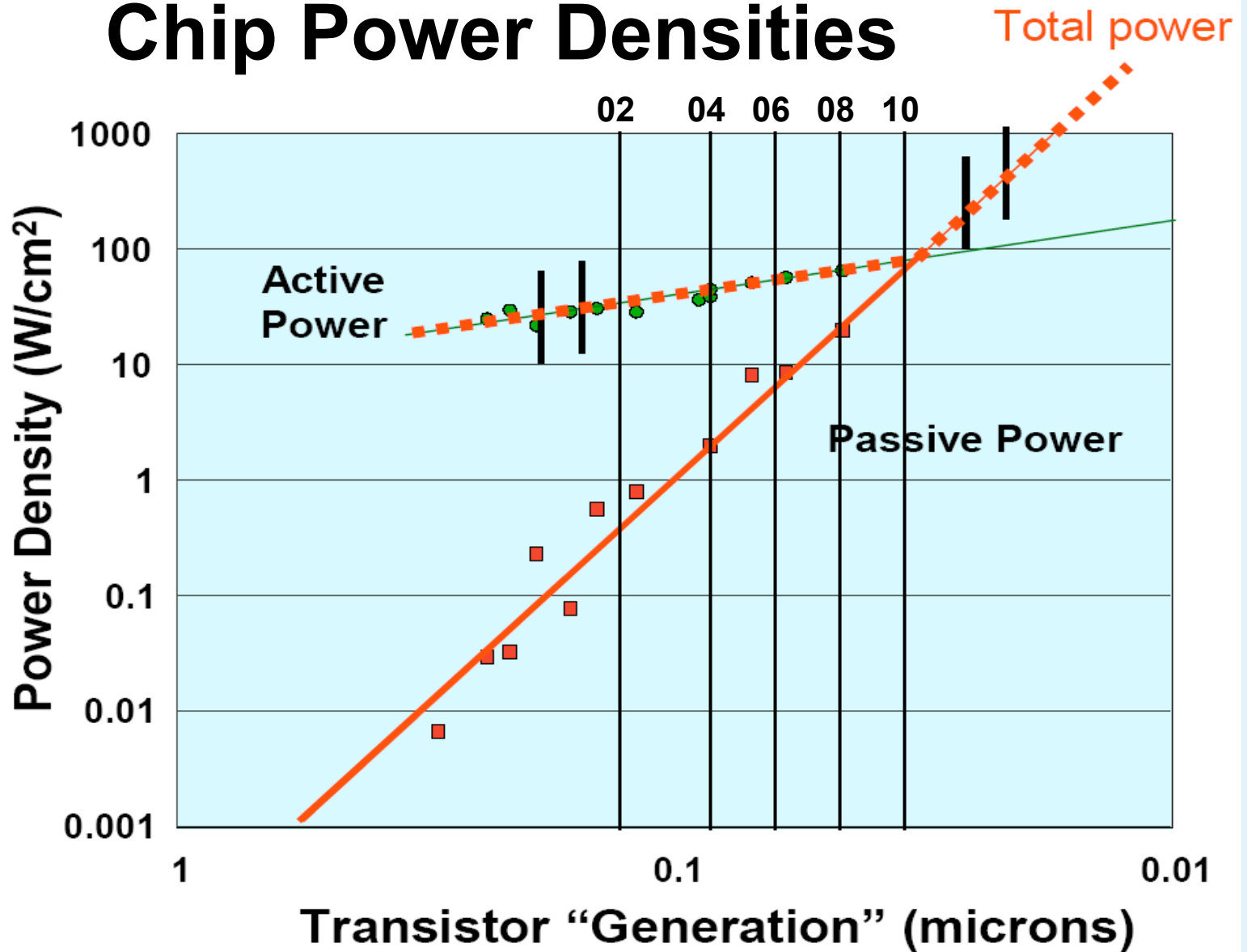
“Intel will reportedly announce that it will discontinue the development process of the next-generation NetBurst microprocessors code-named Tejas and Jayhawk. **The chips are apparently too hot to go the market.**”

According to reports over *Reuters*, *The Wall Street Journal* and some other sources, Intel will announce on Friday its future plans that do not include Tejas and Jayhawk processors, but that are formed around the architecture used in the Pentium M chips. The move is believed to represent a significant shift in the development plans of the world's largest chipmaker and stems from its desire to build chips that are **powerful without generating excessive amounts of heat**, Reuters reported.”

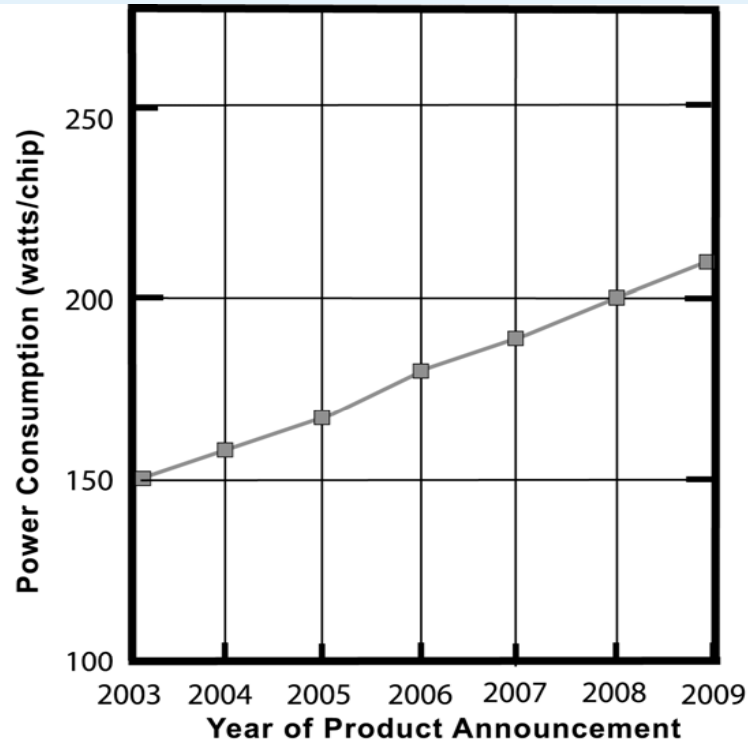
John Shilov, Xbit Laboratories, May 7, 2004



# Chip Power Densities



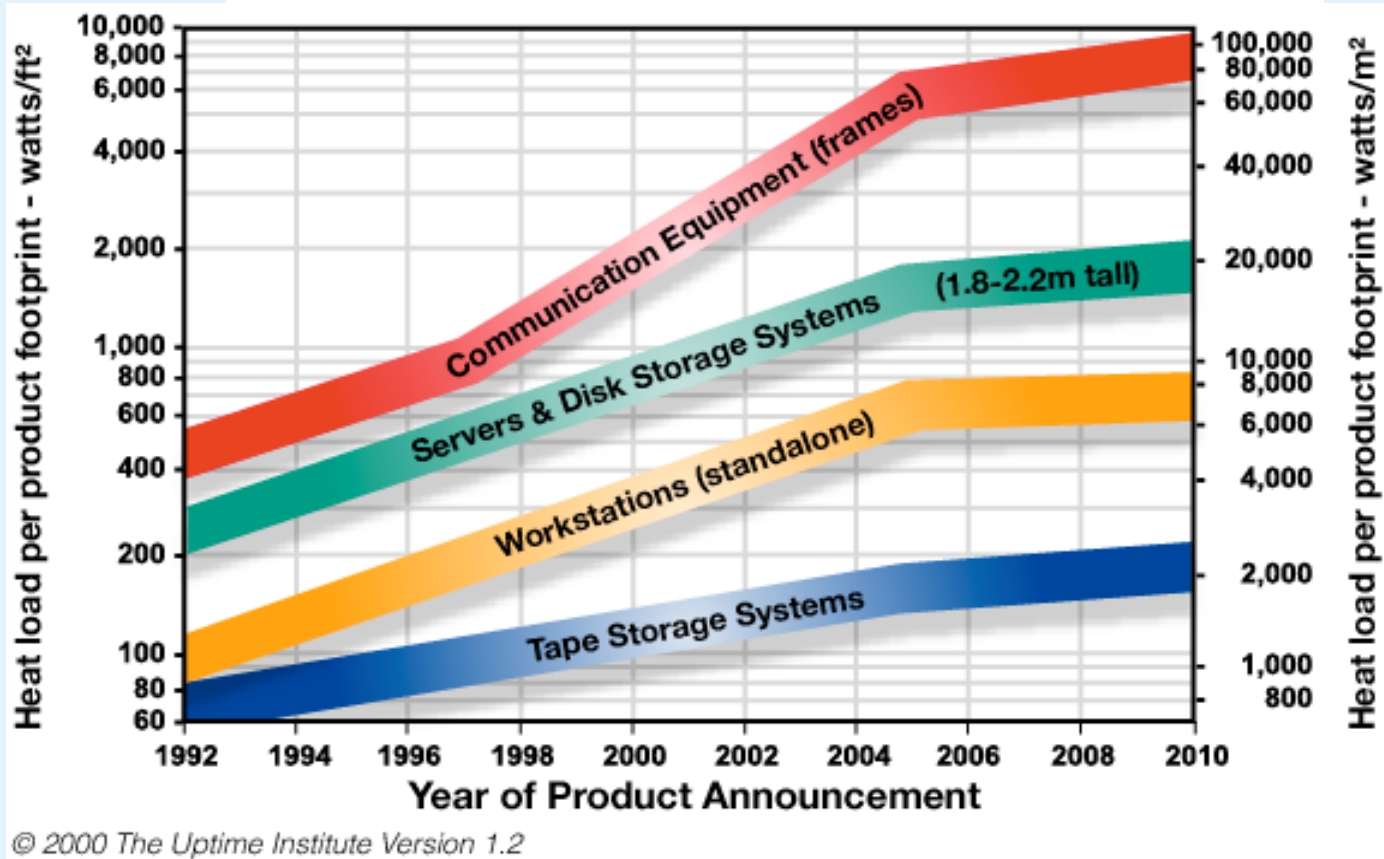
# High Performance Microprocessor Power Consumption (Maximum)



Source: Semiconductor Industry Association  
2003 International Technology Roadmap for Semiconductors

**Will increase 30% over next 5 years**

# Product Heat Density Trend Chart



© 2000 The Uptime Institute Version 1.2

Year of First Product Shipment

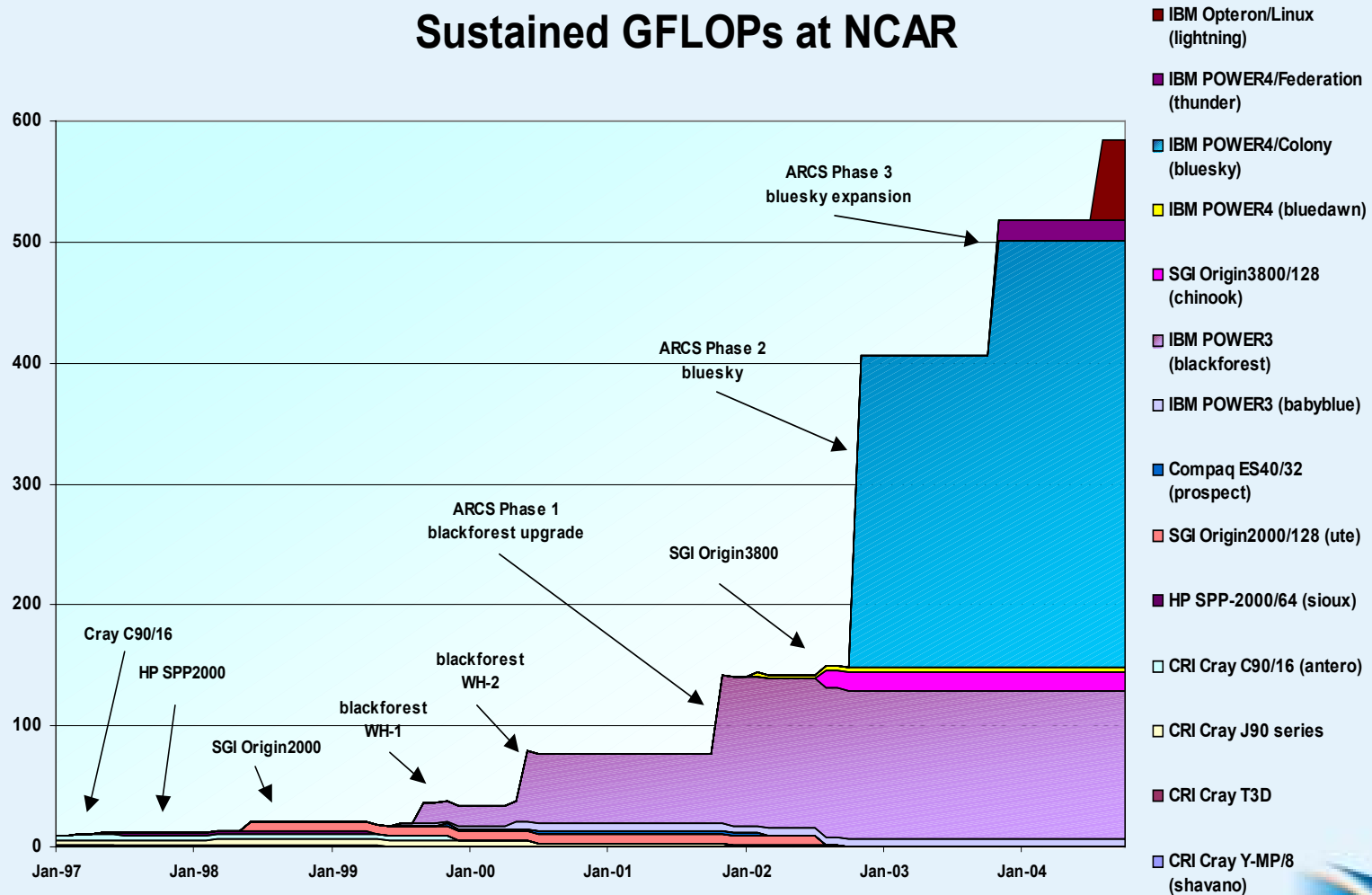
# Science Driver Details

Dr. Richard D. Loft  
Scientific Computing Division

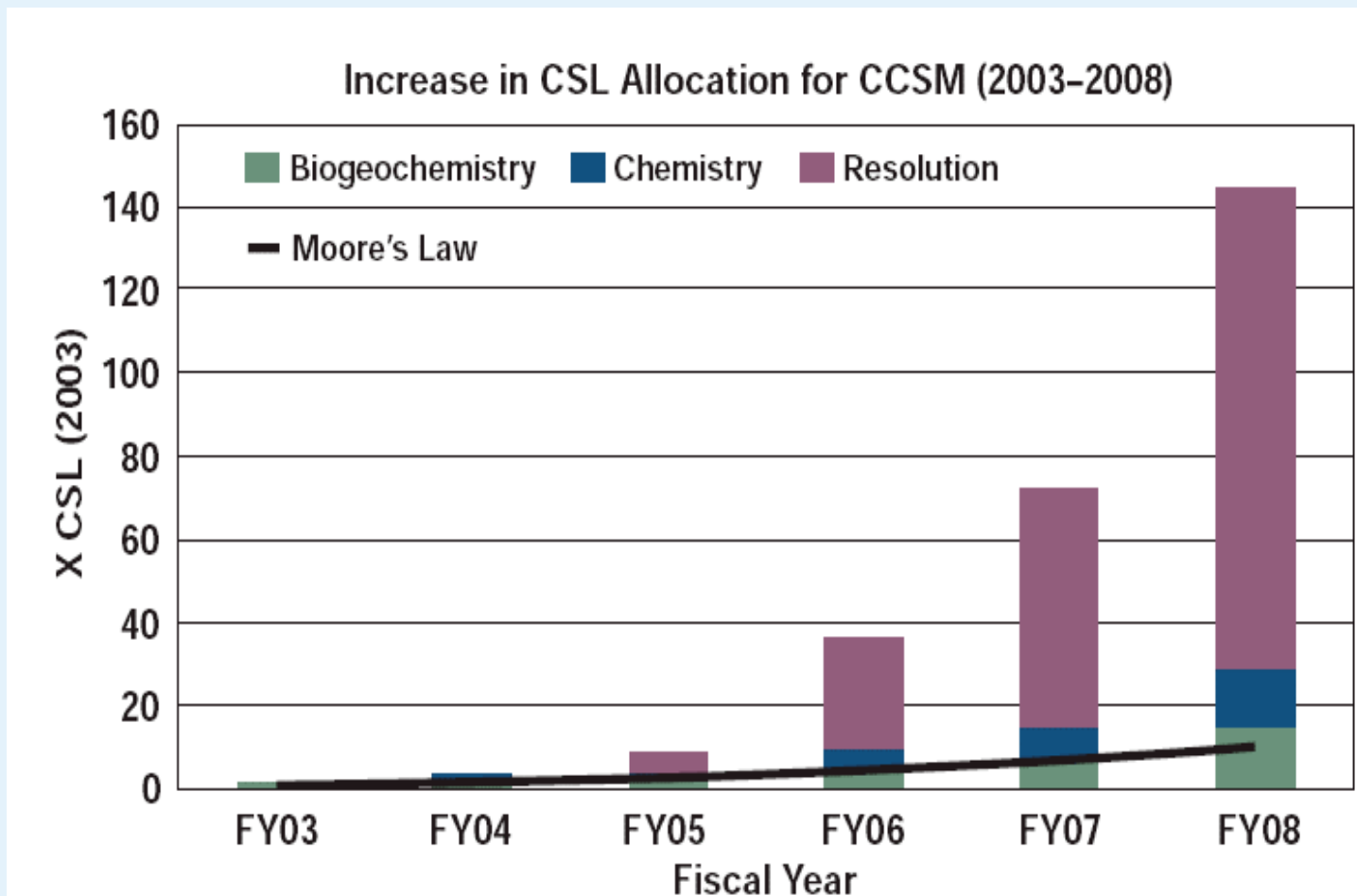


# SCD Computing Growth

## Sustained GFLOPs at NCAR



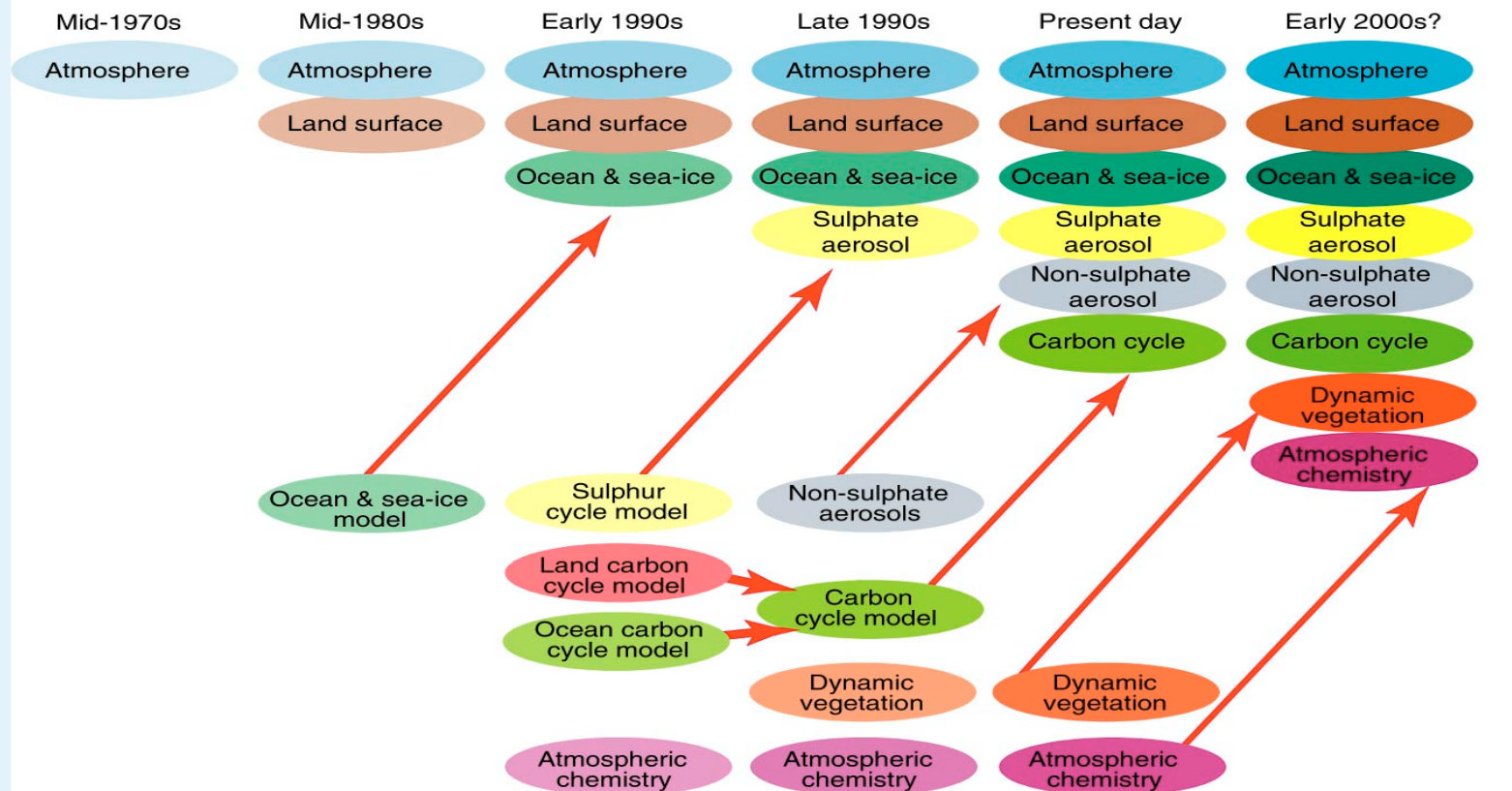
# Projected Climate Computing Requirements (CCSM)



Thanks to Jeff Kiehl/Bill Collins

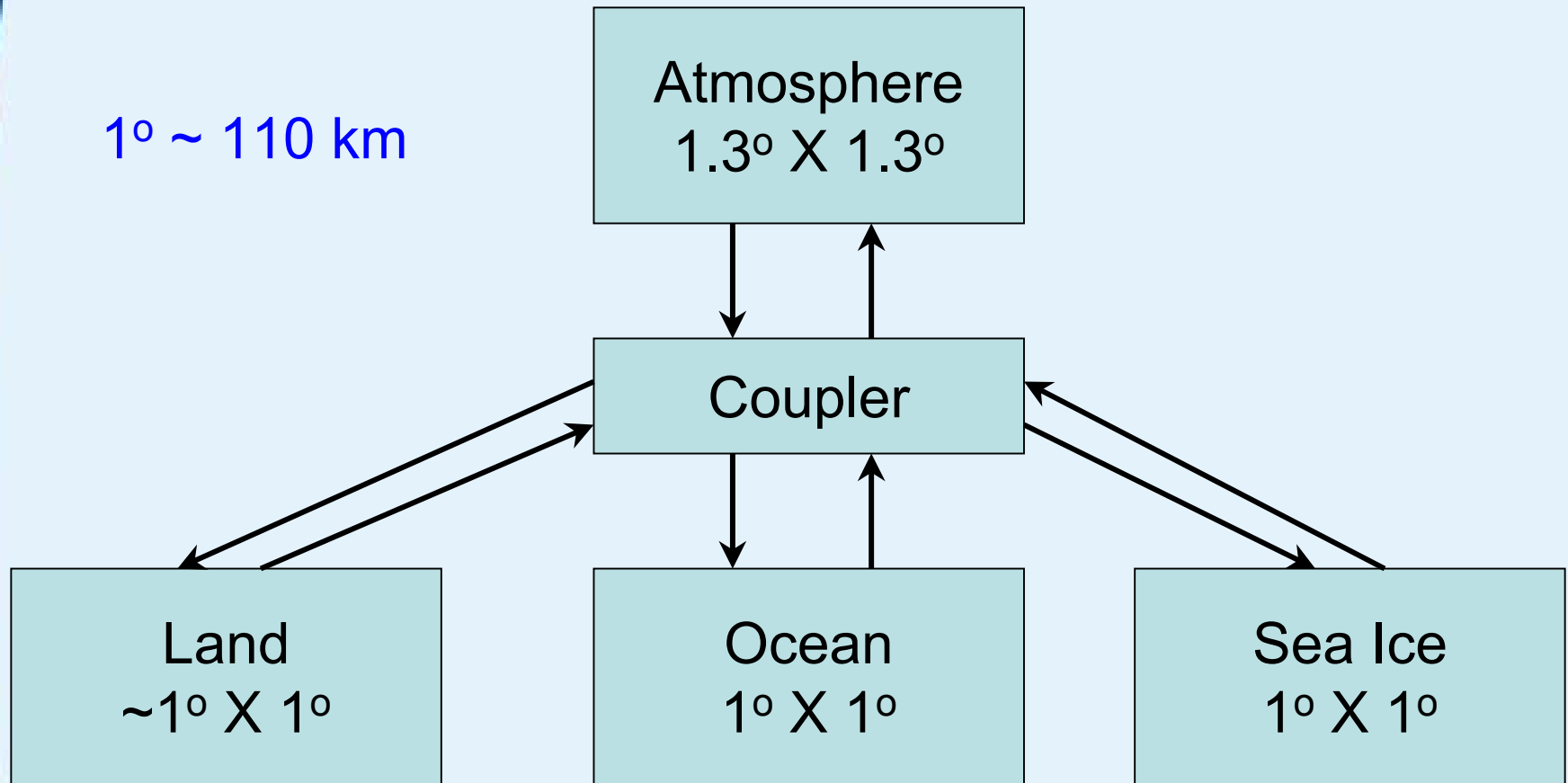
# Growing Model Complexity

## The Development of Climate models, Past, Present and Future



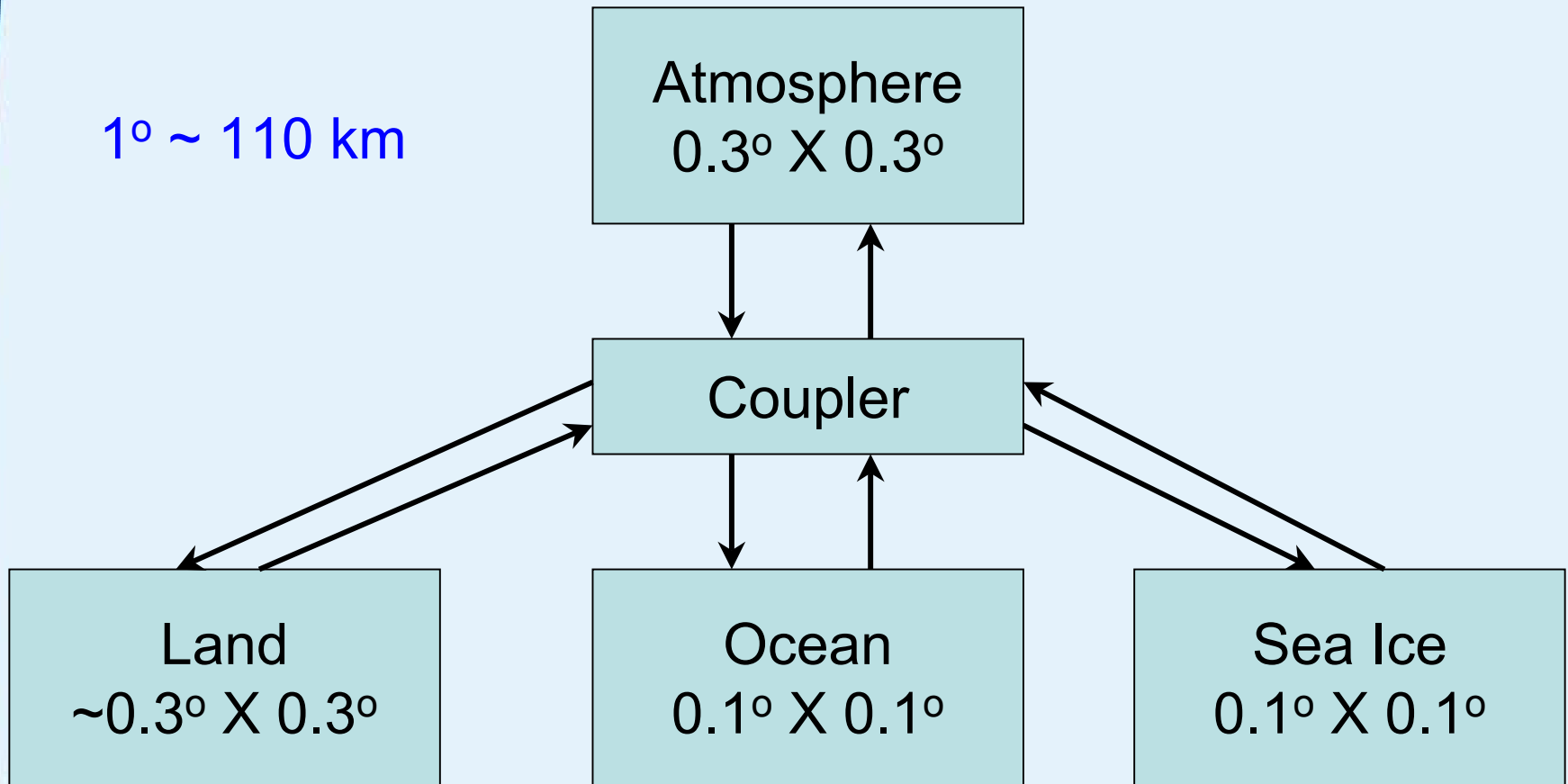
# Current CCSM3 Structure

1° ~ 110 km



# 2008 CCSM-X Resolutions

1° ~ 110 km

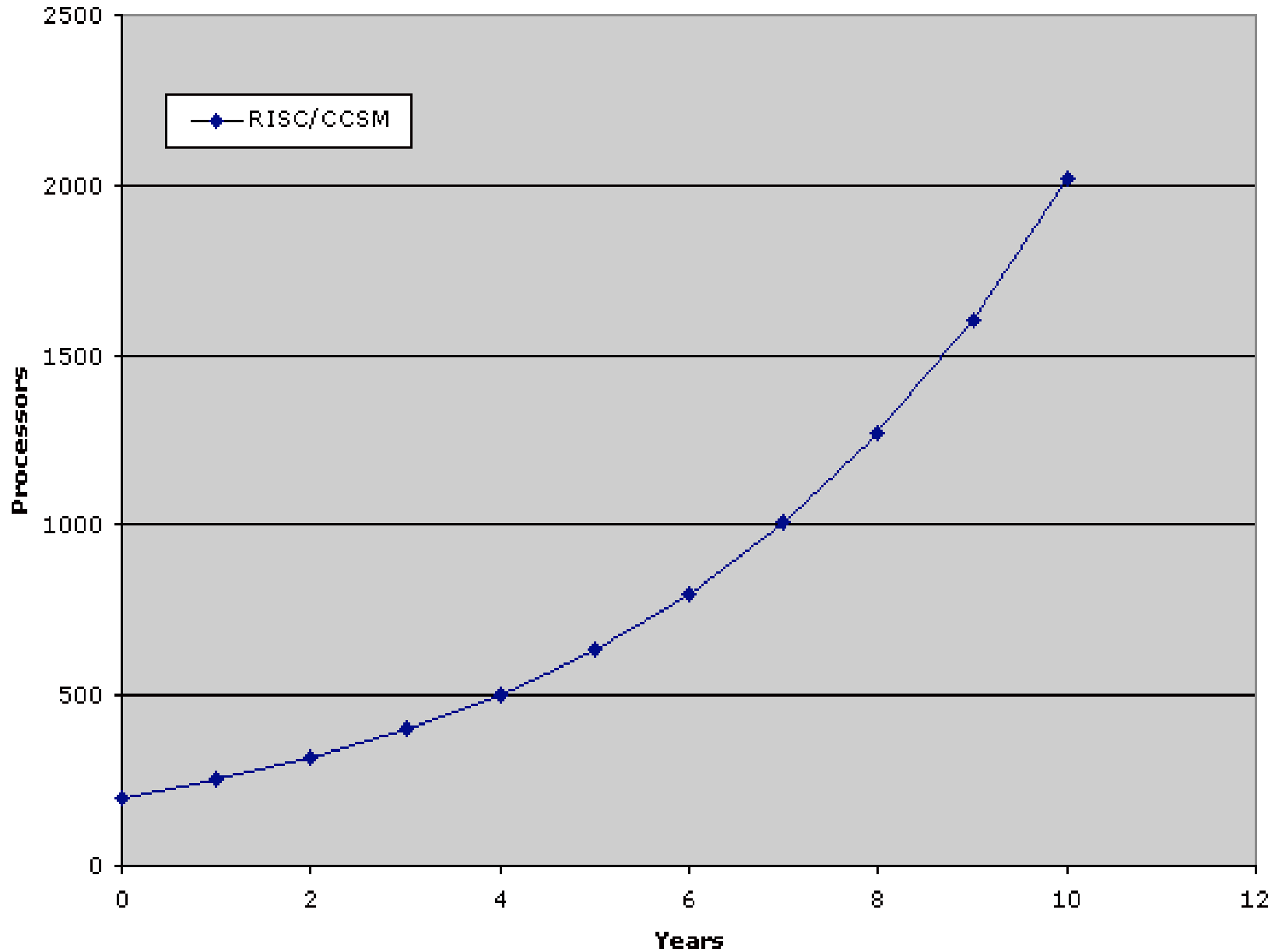


# Historical Application Trends

- Climate modeling is taken as representative.
  - 75% of allocated GAU's
- Climate model parallelism increased at **26%/year**.
- Sustained per-processor performance increased at **33%/year**.
- 2004 Starting point (CCSM example):
  - 1572 processor system image (bluesky)
  - 250 Mflops/proc ( CAM2, IBM 1.3 GHz Power-4)
  - 200 processes/CCSM instance
- If we extrapolate historical trends into the future...



## 2004-2014: Application Size



# Demand for Computing Resources

| Application                     | System Size $L$      | Outer Scale $l$    | Smallest Resolved Scale (current) | Inner Scale/ $\Delta_i$ | Degrees of Freedom | Computational Work                   |
|---------------------------------|----------------------|--------------------|-----------------------------------|-------------------------|--------------------|--------------------------------------|
| "1000-Cubed" Turbulence Problem | 10                   | 1                  | 0.01                              | 0.01                    | $10^9$             | <b>1</b>                             |
| Homogeneous Turbulence          | 10 $L$               | $L$                | $L/100 - L/1000$                  | $L/ R^{3/4}$            | $1000 R^{9/4}$     | <b><math>\sim(R/1000)^3</math></b>   |
| Magnetosphere                   | 600,000 km           | 60,000 km          | 3000 km                           | 200 km                  | $3 \times 10^{10}$ | <b>100</b>                           |
| Earth's Dynamo                  | 5000 km              | 2000 km            | 10 km                             | 2 km                    | $10^{10}$          | <b>100</b>                           |
| Solar Corona                    | 700,000 km           | 30,000 km          | 70,000 km                         | 10 m                    | $10^{15}$          | <b><math>10^8</math></b>             |
| Weather                         | 10,000 km            | 5000 km            | 20 km                             | <1 km                   | $10^{15}$          | <b><math>10^8</math></b>             |
| Climate                         | 30,000 km            | 5000 km            | 100 km                            | <1 km                   | $3 \times 10^{16}$ | <b><math>10^{10}</math></b>          |
| Solar Wind                      | $1.5 \times 10^8$ km | $2 \times 10^6$ km | $2 \times 10^6$ km                | 400 km                  | $10^{17}$          | <b><math>5 \times 10^{10}</math></b> |
| Global Heliosphere              | $2 \times 10^{10}$   | $2 \times 10^6$ km | $2 \times 10^8$ km                | 400 km                  | $10^{22}$          | <b><math>5 \times 10^{14}</math></b> |

From the CyRDAS committee report *Cyberinfrastructure for the Atmospheric Sciences in the 21<sup>st</sup> Century*, June 2004.

# Conclusions

- Climate Science demands described by CCSM business plan increase by a factor of **144 times through 2008**.
- Moore's Law will allow SCD to increase computing by factor of 15-20 during this same time period.
- Future demand for climate cycles exceeds SCD's projected budget by factor of **seven**.
- Just one science domain we serve.

