IEEE Cool Chips VII: An International Symposium on Low-Power and High-Speed Chips, Yokohama, Japan, April 2004.

The Origin and Evolution of Green Destiny

W. Feng and C. Hsu
{feng, chunghsu}@lanl.gov

Research & Development in Advanced Network Technology (RADIANT)
Computer & Computational Sciences Division
Los Alamos National Laboratory
Los Alamos, NM 87545

Supercomputers are making less and less efficient use of the space that they occupy, as evidenced by the fact that supercomputer performance has increased by approximately 4000-fold since the Cray C90 vector supercomputer (circa early 1990s) while the performance-per-square foot has only increased by a factor of 60. The main reason for this inefficiency is the exponentially increasing power requirements of compute nodes, i.e., Moore's Law for Power Consumption (Figure 1). When nodes consume and dissipate more power, they must be spaced out and aggressively cooled.

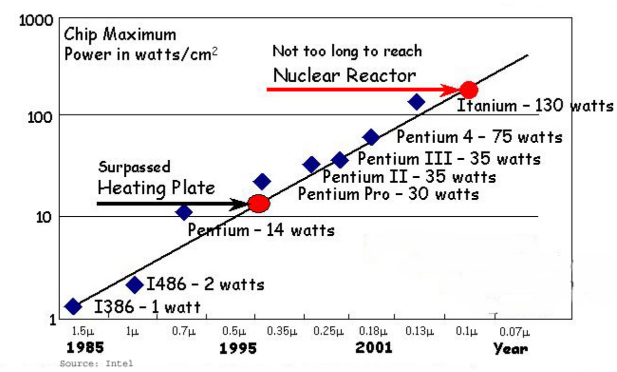


Figure 1. Moore's Law for Power Consumption

In addition, our own empirical data as well as unpublished empirical data from a leading vendor demonstrates that the failure rate of a compute node *doubles* with every 10˚C increase in temperature, and temperature is proportional to power density. Thus, traditional supercomputers require exotic cooling facilities; otherwise, they would be so unreliable (due to overheating) that they would be unavailable for use by the application scientist. For example, our 128-processor Linux cluster with dual 333-MHz Intel Pentium II processors failed on a weekly basis because it resided in a warehouse with no cooling facilities.

To address these problems, we identified low-power building blocks to construct our energy-efficient *Green Destiny* (see Figure 2), a 240-processor supercomputer in a telephone booth (five square feet) that sips less than 5200 watts at full load. The key component to Green Destiny was the 1-GHz Transmeta processor, which consumed only 6 watts of power. However, its Achilles' heel was its floating-point performance. Consequently, we modified Transmeta's code-morphing software to improve performance by 42%, thus matching the performance of a conventional mobile processor on a per-clock-cycle basis (i.e., 1.2-GHz Pentium III-M) but still lagging the performance of the fastest processors at the time by a factor of two.

Figure 2. Green Destiny

Thus, we propose a hybrid solution that uses widely available commodity CPUs from AMD to achieve better performance and its associated "Cool-n-Quiet" technology to reduce power consumption by as much as 40% while impacting peak performance by less than 5%. Like past solutions, we use a mechanism called dynamic voltage (and frequency) scaling. But instead of using a simple performance model that assumes that programs are CPU-bound, we formulate a new performance model that can be constructed on-the-fly and in real-time and applies to both CPU- and non-CPU-bound programs. This performance model then allows us to develop a fine-grained schedule of frequency-voltage settings that minimizes energy use without adversely affecting performance.

LA-UR-04-1924

The Origin and Evolution of Green Destiny

Wu-chun Feng and **Chung-hsing Hsu**

Research & Development in Advanced Network Technology (RADIANT)
Computer & Computational Sciences Division
Los Alamos National Laboratory

IEEE Cool Chips VII: A Symposium on Low-Power and High-Speed Chips,
Yokohama, Japan, April 2004.

LA-UR-04-1924

Outline

- Motivation
- The Origin of Green Destiny
- The Architecture of Green Destiny
- Performance Evaluation of Green Destiny (Relative to Performance)
- The Need for New Performance Metrics
- A Renaissance in Supercomputing
  - Supercomputing in Small Spaces → Bladed Beowulf
- Performance Evaluation of Green Destiny (Relative to Efficiency, Reliability, and Availability)
- The Evolution of Green Destiny
  - Real-time, Constraint-based Dynamic Voltage Scaling
- Conclusion

COMPUTER & COMPUTATIONAL SCIENCES

Motivation

- Operating Environment
  - 80-90°F (27-32°C) warehouse at 7,400 feet (2195 meters) above sea level.
  - No air conditioning, no air filtration, no raised floor, and no humidifier/dehumidifier.
- Computing Requirement
  - Parallel computer to enable high-performance network research in simulation and implementation.

Motivation

- Operating Environment
  - 80-90°F (27-32°C) warehouse at 7,400 feet (2195 meters) above sea level.
  - No air conditioning, no air filtration, no raised floor, and no humidifier/dehumidifier.
- Computing Requirement
  - Parallel computer to enable high-performance network research in simulation and implementation.
- Solution (circa 2001)
  - *Little Blue Penguin,* a 64-node dual-CPU Linux cluster.
    - Power Consumption: ~ 10 kilowatts.
    - Space Consumption: ~ 48 square feet.

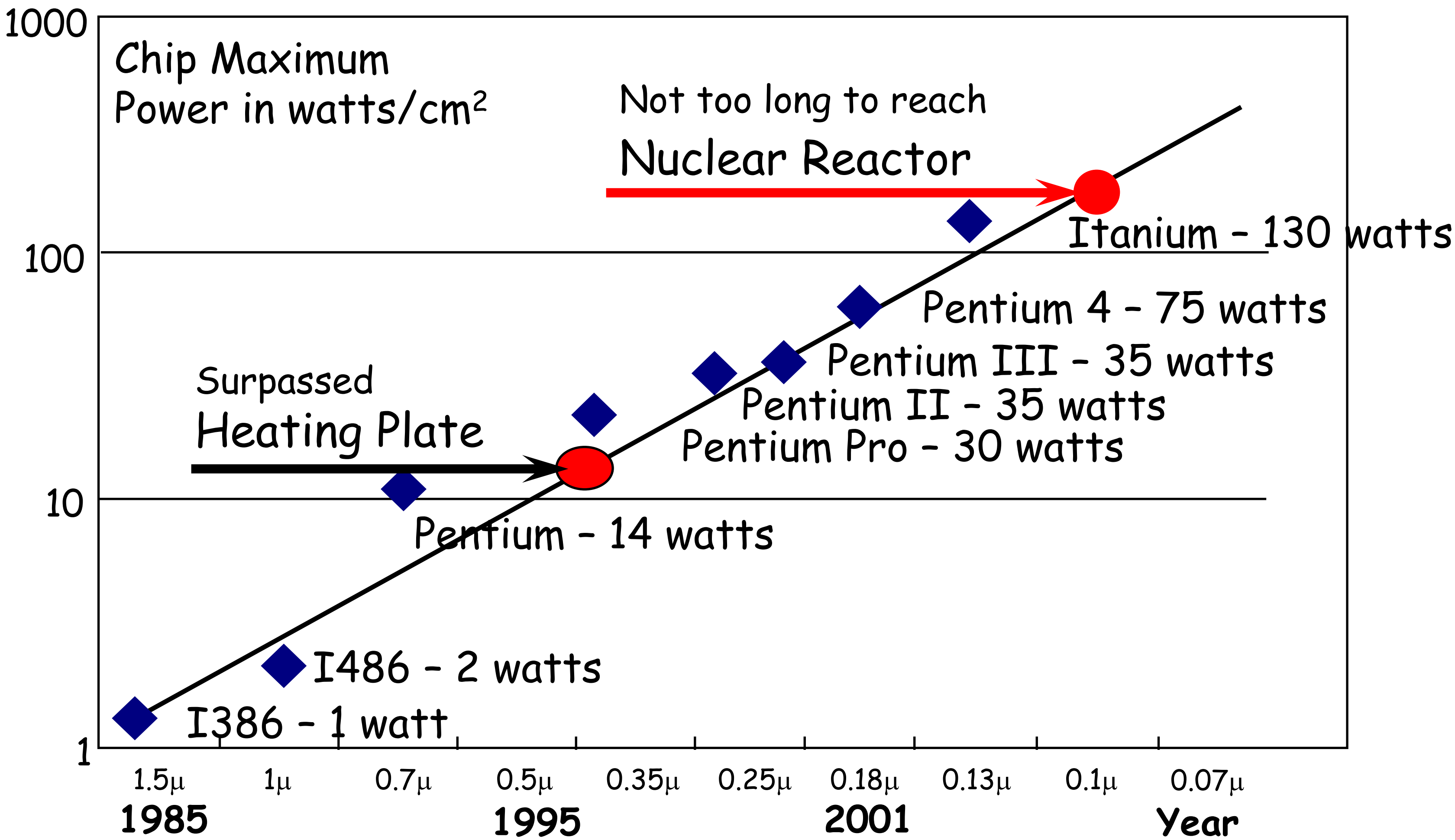
Motivation

- Operating Environment
  - 80-90°F (27-32°C) warehouse at 7,400 feet (2195 meters) above sea level.
  - No air conditioning, no air filtration, no raised floor, and no humidifier/dehumidifier.
- Computing Requirement
  - Parallel computer to enable high-performance network research in simulation and implementation.
- Solution
  - *Little Blue Penguin,* a 64-node dual-CPU Linux cluster.
- Problem
  - Our "Little Blue Penguin" cluster *failed* weekly. Why?

High Power → High Temp → Higher Unreliability

- Arrhenius' Equation
  (circa 1890s in chemistry → circa 1980s in computer & defense industries)
  - As temperature increases by 10° C …
    - The failure rate of a system *doubles*.
    - The reliability of a system is cut in *half*.
  - Twenty years of unpublished empirical data .
- Question
  - Can we build a low-power supercomputer that is still considered high performance? Yes.

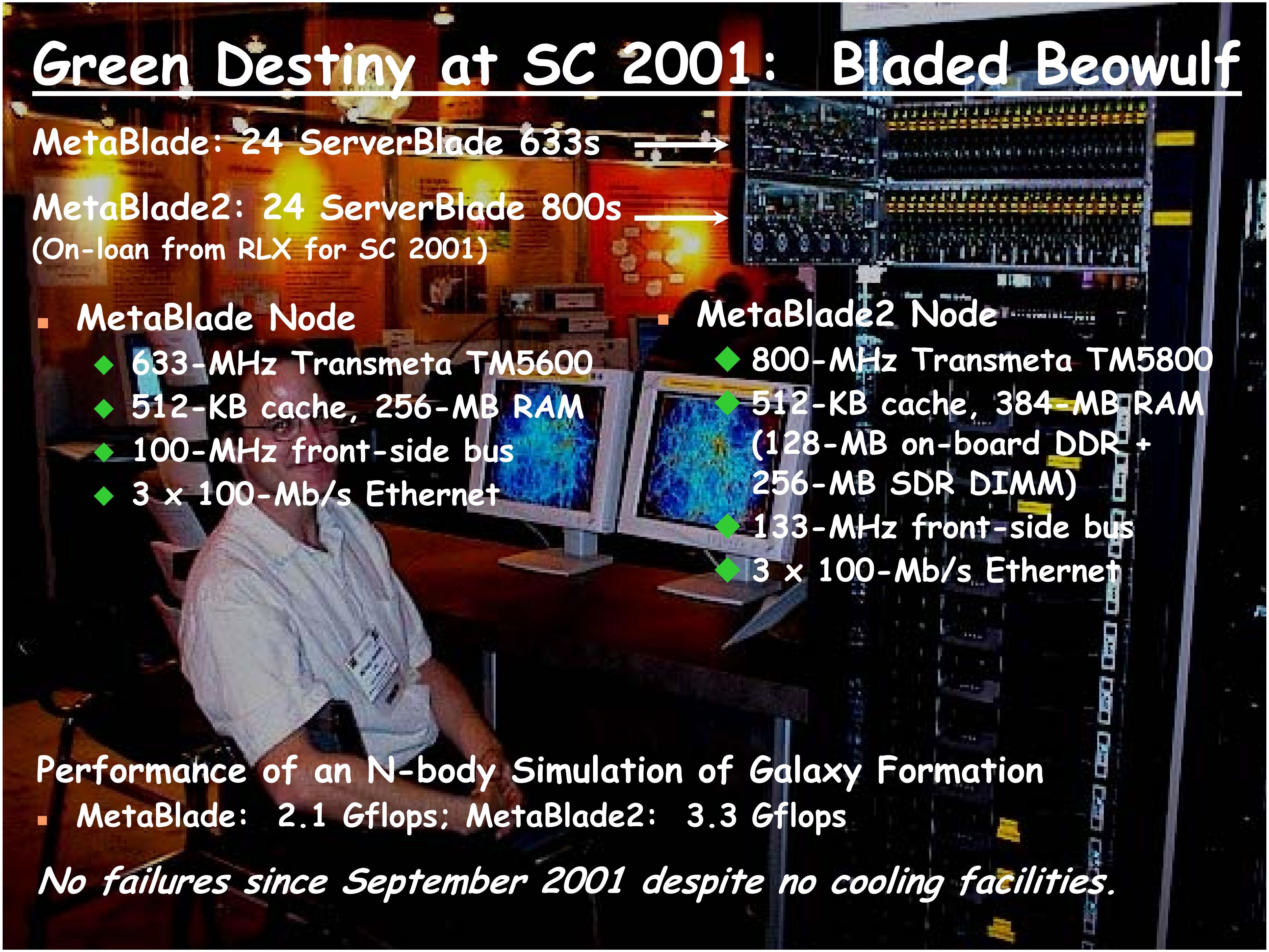
Moore's Law for Power



Source: Fred Pollack, Intel. New Microprocessor Challenges in the Coming Generations of CMOS Technologies, MICRO32 and Transmeta

The Origin of *Green Destiny*

- Project Conception: Sept. 28, 2001.
  - On a winding drive home through Los Alamos Canyon … the need for reliable compute cycles.
    - Leverage RLX web-hosting servers with Transmeta CPUs.
- Project Implementation: Oct. 9, 2001.
  - Received the "bare" hardware components.
  - Two man-hours later …
    - Completed construction of a 24-processor RLX System 324 and installation of system software.
  - One man-hour later …
    - Successfully executing a 10-million N-body simulation of a galaxy formation
- Public Demonstration: Nov. 12, 2001 at SC 2001.



Green Destiny at SC 2001: Bladed Beowulf

MetaBlade: 24 ServerBlade 633s →

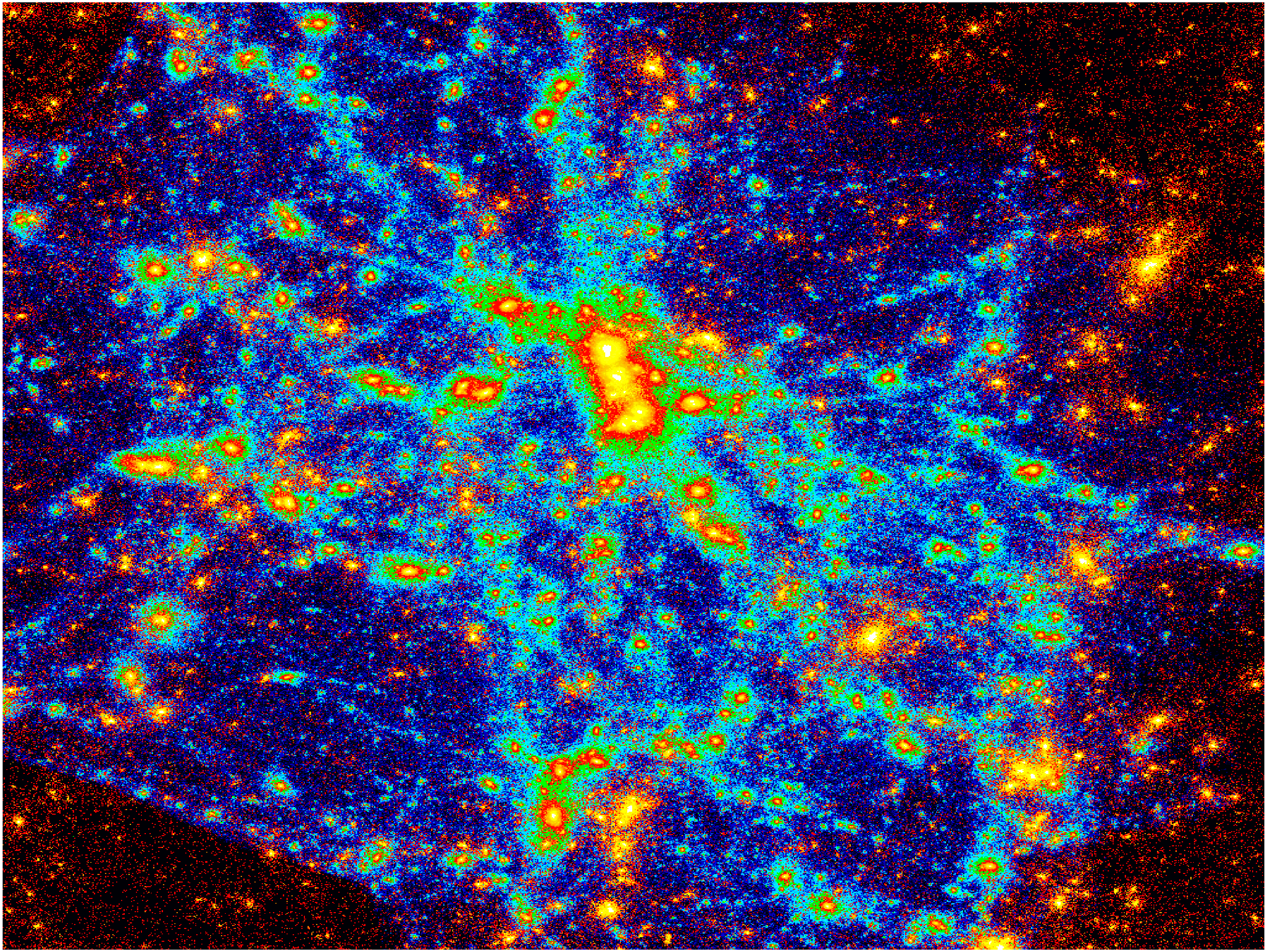
MetaBlade2: 24 ServerBlade 800s →
(On-loan from RLX for SC 2001)

- MetaBlade Node
  - 633-MHz Transmeta TM5600
  - 512-KB cache, 256-MB RAM
  - 100-MHz front-side bus
  - 3 x 100-Mb/s Ethernet
- MetaBlade2 Node
  - 800-MHz Transmeta TM5800
  - 512-KB cache, 384-MB RAM (128-MB on-board DDR + 256-MB SDR DIMM)
  - 133-MHz front-side bus
  - 3 x 100-Mb/s Ethernet

Performance of an N-body Simulation of Galaxy Formation

- MetaBlade: 2.1 Gflops; MetaBlade2: 3.3 Gflops

No failures since September 2001 despite no cooling facilities.



The Origin of *Green Destiny*

- Feedback on *MetaBlade* and *MetaBlade2* was huge!
  - Continual crowds over the three days of SC 2001.
- Analysis of Empirical Data (circa 2001)
  - Performance/Power: 4.12 Gflop/kW.
  - Performance/Space: 350 Gflop/sq. ft.

The Origin of *Green Destiny*

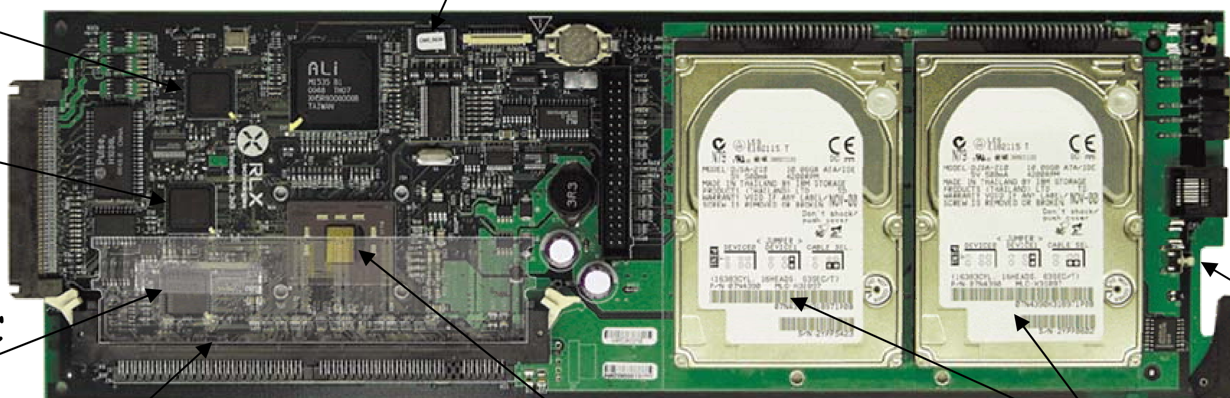
- Feedback on *MetaBlade* and *MetaBlade2* was huge!
  - Continual crowds over the three days of SC 2001.
- Analysis of Empirical Data (circa 2001)
  - Performance/Power: 4.12 Gflop/kW.
  - Performance/Space: 350 Gflop/sq. ft.
- Inspiration
  - Build a full rack of MetaBlade clusters.
    - Scales up performance/space to 3500 Gflop/sq. ft.
  - Problem: In 2001, the performance per node on MetaBlade was more than *three times worse* than the fastest processor at the time.
  - Can we improve performance while maintaining low power? Yes via Transmeta's code-morphing software.

The Origin of *Green Destiny*: RLX System™ 324

RLX System™ 300ex
- Interchangeable blades
  - Intel, Transmeta, or both.
- Switched-based management

- 3U vertical space
  - 5.25" x 17.25" x 25.2"
- Two hot-pluggable 450W power supplies
  - Load balancing
  - Auto-sensing fault tolerance
- System midplane
  - Integration of system power, management, and network signals.
  - Elimination of internal system cables.
  - Enabling efficient hot-pluggable blades.
- Network cards
  - Hub-based management.
  - Two 24-port interfaces.

The Origin of *Green Destiny:* RLX ServerBlade™ 633 (circa 2000)



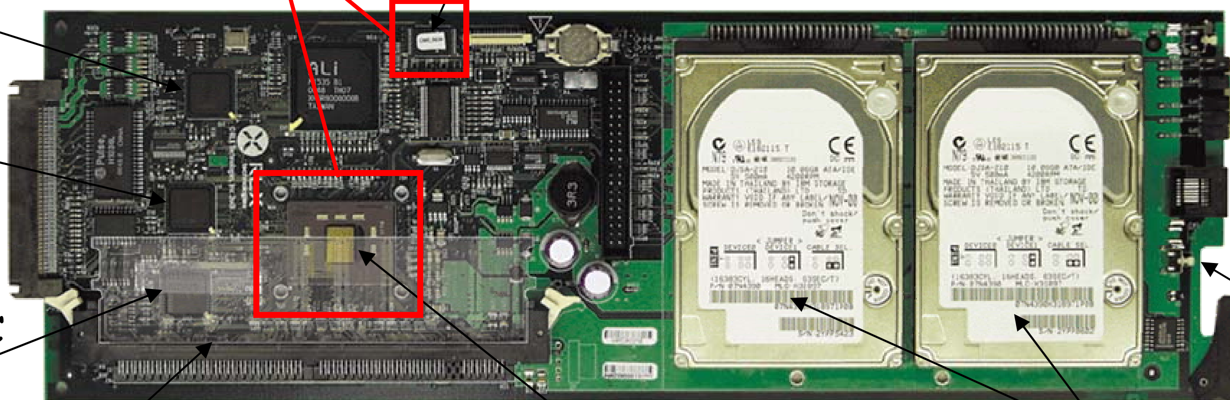
Crusoe™

128KB L1 cache, 512KB L2 cache
LongRun, Northbridge, x86 compatible

RLX ServerBlade™ 1000t
$999 (as of Dec. 2003)

The Origin of *Green Destiny:* RLX ServerBlade™ 633 (circa 2000)

Modify the Transmeta CPU software to improve performance.



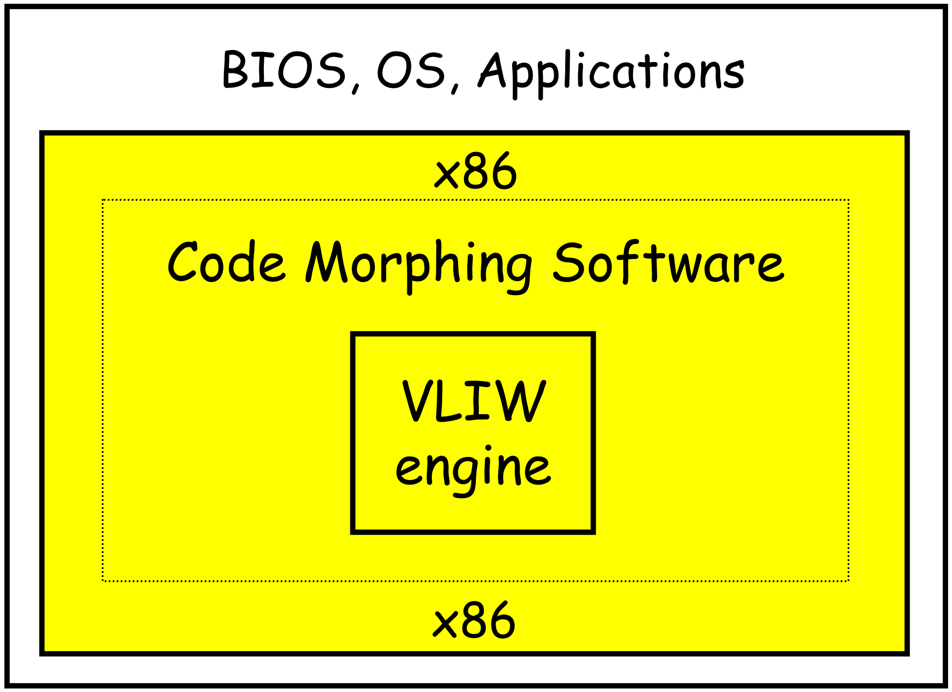
Crusoe™

RLX ServerBlade™ 1000t
$999 (as of Dec. 2003)

128KB L1 cache, 512KB L2 cache
LongRun, Northbridge, x86 compatible

Transmeta TM5600 CPU: VLIW + CMS

- VLIW Engine
  - Up to four-way issue
    - In-order execution only.
  - Two integer units
  - Floating-point unit
  - Memory unit
  - Branch unit



- VLIW Transistor Count ("Anti-Moore's Law")
  - $\sim\frac{1}{4}$ of Intel PIII → ~ 6x-7x less power dissipation
  - Less power → lower "on-die" temp. → better reliability & availability

The Origin of *Green Destiny*: Transmeta TM5x00 CMS

- Code-Morphing Software (CMS)
  - Provides compatibility by dynamically "morphing" x86 instructions into simple VLIW instructions.
  - Learns and improves with time, i.e., iterative execution.
- High-Performance Code-Morphing Software (HP-CMS)
  - *Optimized to improve floating-pt. performance by ~50%.*

COMPUTER & COMPUTATIONAL SCIENCES

The Origin of *Green Destiny:* Transmeta TM5x00 Comparison

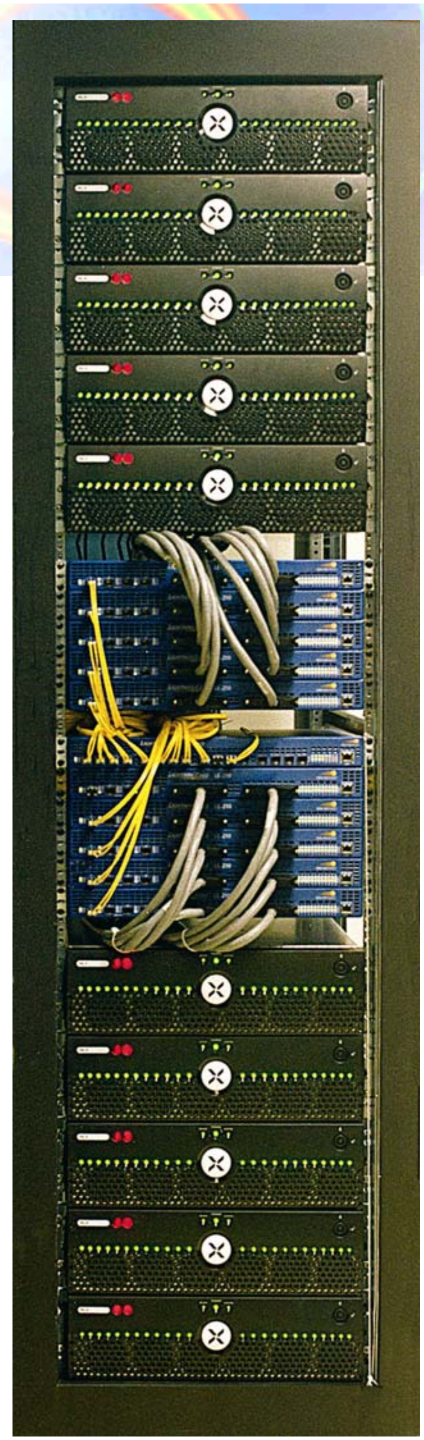
Intel P4	MEM	MEM	2xALU	2xALU	FPU	SSE	SSE	Br
Transmeta TM5x00	MEM		2xALU		FPU			Br

- Current-generation Transmeta TM5800 + HP-CMS
  - Performs comparably to an Intel PIII over iterative scientific codes on a clock-for-clock-cycle basis.
  - Performs only *twice* as slow as the fastest CPU (at the time) rather than three times as slow.
- Efficeon, the next-generation CPU from Transmeta, rectifies the above mismatch in functional units.

The Origin of *Green Destiny*: Low-Power Network Switches

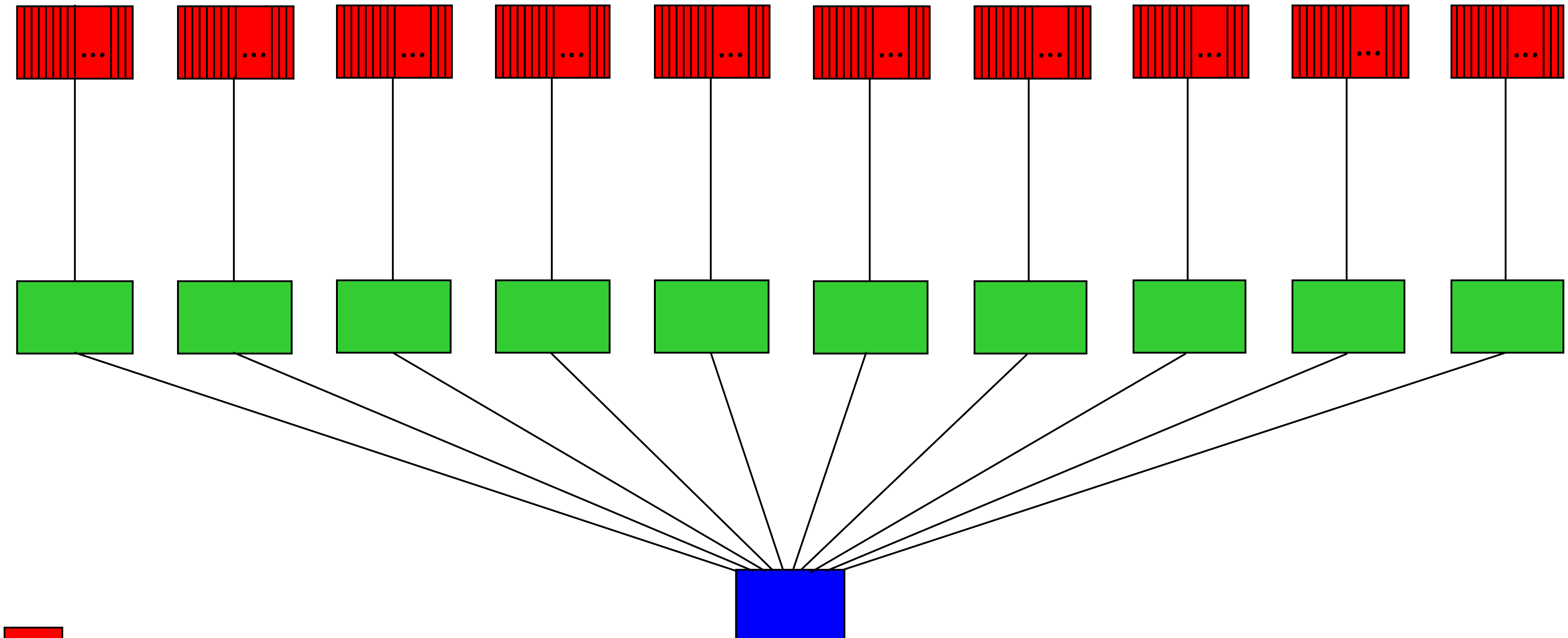
- WWP LE-410: 16 ports of Gigabit Ethernet
- WWP LE-210: 24 ports of Fast Ethernet via RJ-21s
- (Avg.) Power Dissipation / Port: A few watts.

"Green Destiny" Bladed Beowulf
(circa 2002)



- A 240-Node Beowulf in One Cubic Meter
- Each Node
  - 667-MHz Transmeta TM5600 CPU w/ Linux 2.4.x
    - Upgraded to 1-GHz Transmeta TM5800 CPUs
  - 640-MB RAM, 20-GB hard disk, 100-Mb/s Ethernet (up to 3 interfaces)
- Total
  - 160 Gflops peak (240 Gflops with upgrade)
  - 150 GB of RAM (expandable to 276 GB)
  - 4.8 TB of storage (expandable to 38.4 TB)
  - Power Consumption: Only 3.2 – 5.2 kW.
- Linpack: 101 Gflops (with upgrade)
- Reliability & Availability
  - *No unscheduled failures in 24 months.*

Architecture of *Green Destiny*: A Bladed Beowulf Cluster



- 24-node RLX 324 chassis
- 24-port Fast Ethernet switch
- 16-port Gigabit Ethernet switch
- 100-Mb/s Fast Ethernet link

Wu-chun Feng
feng@lanl.gov

Performance Evaluation of *Green Destiny*

- Gravitational Microkernel Benchmark (circa June 2002)

Processor	Math sqrt	Karp sqrt
500-MHz Intel PIII	87.6	137.5
533-MHz Compaq Alpha EV56	76.2	178.5
633-MHz Transmeta TM5600	115.0	144.6
800-MHz Transmeta TM5800	174.1	296.6
375-MHz IBM Power3	298.5	379.1
1200-MHz AMD Athlon MP	350.7	452.5

Units are in Mflops.

Bottom Line: CPU performance was competitive. Memory bandwidth was not (i.e., 300-350 MB/s with STREAMS).

COMPUTER & COMPUTATIONAL SCIENCES

Treecode Benchmark for n-Body

Site	Machine	CPUs	Gflops	Mflops/CPU
NERSC	IBM SP-3	256	57.70	225.0
LANL	SGI O2K	64	13.10	205.0
LANL	Green Destiny	212	38.90	183.5
SC'01	MetaBlade2	24	3.30	138.0
LANL	Avalon	128	16.16	126.0
LANL	Loki	16	1.28	80.0
NASA	IBM SP-2	128	9.52	74.4
SC'96	Loki+Hyglac	32	2.19	68.4
Sandia	ASCI Red	6800	464.90	68.4
CalTech	Naegling	96	5.67	59.1
NRL	TMC CM-5E	256	11.57	45.2

Treecode Benchmark for n-Body

Site	Machine	CPUs	Gflops	Mflops/CPU
NERSC	IBM SP-3	256	57.70	225.0
LANL	SGI O2K	64	13.10	205.0
LANL	Green Destiny	212	38.90	183.5
SC'01	MetaB[illegible]	24	3.30	138.0
LANL	[illegible]	[illegible]	[illegible]	126.0
LA[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
NAS[illegible]	[illegible]	[illegible]	[illegible]	[illegible].4
SC'96	Loki+Hyglac	[illegible]	2.19	68.4
Sandi[illegible]	[illegible]	[illegible]	[illegible]	68.4
CalTe[illegible]	[illegible]	[illegible]	[illegible]	59.1
NRL	TMC CM-5E	256	11.57	45.2

Upgraded "Green Destiny"
58 Gflops → 274 Mflops/CPU

Is it just about performance?

Wu-chun Feng
feng@lanl.gov

Metrics for Evaluating Supercomputers

- *Performance (i.e., Speed)*
  - Metric: Floating-Operations Per Second (FLOPS)
  - Examples: Japanese Earth Simulator and ASCI Q.
- *Price/Performance → Cost Efficiency*
  - Metric: Cost / FLOPS
  - Examples: SuperMike, Space Simulator, VT Apple G5.
- Performance & price/performance are important metrics, but …

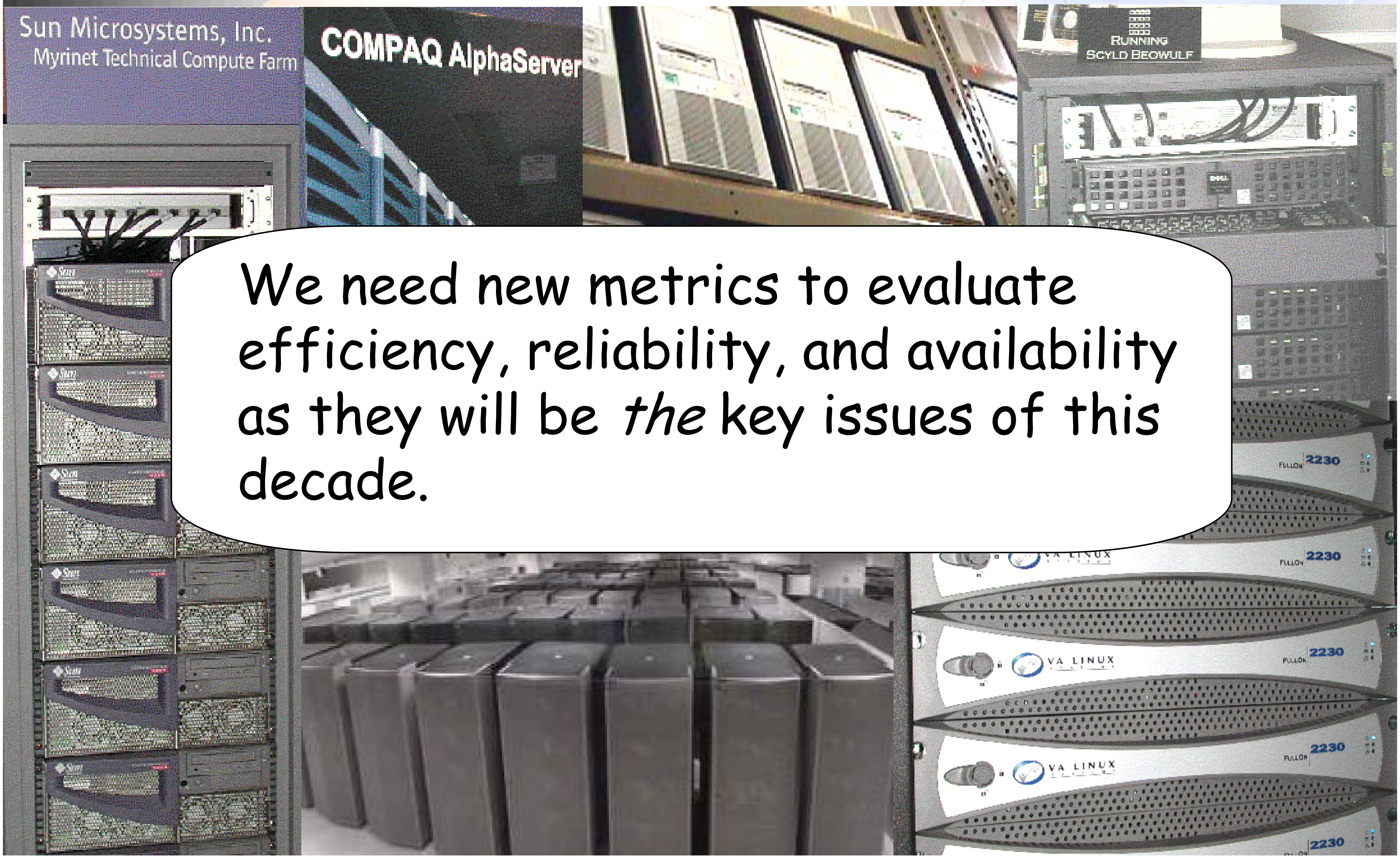
The Need for *New* Supercomputing Metrics

- Analogy: Buying a high-end car.
  Which metric to use?
  - *Raw Performance:* Ferrari 550.
  - *Price/Performance:* Ford Mustang GTO.
  - *Fuel Efficiency:* Honda Insight.
  - *Reliability:* Toyota Camry.
  - *Storage:* Honda Odyssey.
  - *Off-Road Worthiness:* Jeep Cherokee.
  - *All-Around:* Volvo XC90.
- So many metrics to evaluate a car …
  why not to evaluate a supercomputer?
- But which metrics?

COMPUTER & COMPUTATIONAL SCIENCES

Where is Supercomputing?

(Courtesy: Thomas Sterling, Caltech & NASA JPL)



We need new metrics to evaluate efficiency, reliability, and availability as they will be *the* key issues of this decade.

Why Efficiency, Reliability, and Availability (ERA)?

- Requirement: Near-100% availability with efficient and reliable resource usage.
  - E-commerce, enterprise apps, online services, ISPs.
- Problems (Source: David Patterson, UC-Berkeley)
  - Frequency of Service Outages
    - 65% of IT managers report that their websites were unavailable to customers over a 6-month period.
      - 25%: 3 or more outages
  - Cost of Service Outages
    - NYC stockbroker: **$ 6,500,000/hr**
    - Ebay (22 hours): **$ 225,000/hr**
    - Amazon.com: **$ 180,000/hr**
    - Social Effects: negative press, loss of customers who "click over" to competitor.

A Renaissance in Supercomputing: *Supercomputing in Small Spaces*

- Supercomputing in Small Spaces (http://sss.lanl.gov)
  - First instantiation: *Bladed Beowulf*
    - MetaBlade (24) and Green Destiny (240).
- Goal
  - Improve *efficiency*, *reliability*, and *availability* (ERA) in large-scale computing systems.
    - Sacrifice a little bit of raw performance.
    - Improve overall system throughput as the system will "always" be available, i.e., effectively no downtime, no hardware failures, etc.
  - Reduce the *total cost of ownership* (TCO).
- Analogy
  - Ferrari 550: Wins raw performance but reliability is poor so it spends its time in the shop. Throughput low.
  - Toyota Camry: Loses raw performance but high reliability results in high throughput (i.e., miles driven).

A Renaissance in Supercomputing: *Supercomputing in Small Spaces*

- Supercomputing in Small Spaces (http://sss.lanl.gov)
  - First instantiation: *Bladed Beowulf*
    - MetaBlade (24) and Green Destiny (240).
- Goal
  - Improve *efficiency*, *reliability*, and *availability* (ERA) in large-scale computing systems.
    - Sacrifice a little bit of raw performance.
    - Improve overall system throughput as the system will "always" be available, i.e., effectively no downtime, no hardware failures, etc.
  - Reduce the *total cost of ownership* (TCO).
- Analogy
  - Ferrari 550: Wins raw performance but reliability is poor so it spends its time in the shop. Throughput lo
  - Toyota Camry: Loses raw performance but h results in high throughput (i.e., miles driven).

Problem: How to quantify?

What is TCO?

- Cost of Acquisition ← Fixed, one-time cost
  - $$$ to buy the supercomputer.
- Cost of Operation ← Variable, recurring cost
  - Administration
    - $$$ to build, integrate, configure, maintain, and upgrade the supercomputer over its lifetime.
  - Power & Cooling
    - $$$ in electrical power and cooling that is needed to maintain the operation of the supercomputer.
  - Downtime → Reliability and Availability
    - $$$ lost due to the downtime (unreliability) of the system.
  - Space
    - $$$ spent to house the system.

Total Price-Performance Ratio

- Price-Performance Ratio
  - *Price = Cost of Acquisition*
  - Performance = Floating-Point Operations Per Second
- Total Price-Performance Ratio (ToPPeR)
  - *Total Price = Total Cost of Ownership (TCO)*
  - Performance = Floating-Point Operations Per Second
- Using "FLOPS" as a performance metric is problematic as well … another talk, another time …

Quantifying TCO?

- Why is TCO hard to quantify?
  - Components
    - Acquisition + Administration + Power + Downtime + Space

Quantifying TCO?

- Why is TCO hard to quantify?
  - Components
    - Acquisition + Administration + Power + Downtime + Space

Too Many Hidden Costs
Institution-Specific

Quantifying TCO?

- Why is TCO hard to quantify?
  - Components
    - Acquisition + Administration + Power + Downtime + Space

      Too Many Hidden Costs
      Institution-Specific
  - Traditional Focus: Acquisition (i.e., equipment cost)
    - Cost Efficiency: Price/Performance Ratio

Quantifying TCO?

- Why is TCO hard to quantify?
  - Components
    - Acquisition + Administration + Power + Downtime + Space

Institution-Specific

Too Many Hidden Costs

  - Traditional Focus: Acquisition (i.e., equipment cost)
    - Cost Efficiency: Price/Performance Ratio
  - New *Quantifiable* Efficiency Metrics
    - "Power" Efficiency: Performance/Power Ratio
    - "Space" Efficiency: Performance/Space Ratio

Related to efficiency, reliability, and availability.

Parallel Computing Platforms (An "Apples-to-Oranges" Comparison)

- Avalon (1996)
  - 140-CPU *Traditional Beowulf Cluster*
- ASCI Red (1996)
  - 9632-CPU *MPP*
- ASCI White (2000)
  - 512-Node (8192-CPU) *Cluster of SMPs*
- Green Destiny (2002)
  - 240-CPU *Bladed Beowulf Cluster*

Parallel Computing Platforms Running the N-body Code

Machine	Avalon Beowulf	ASCI Red	ASCI White	Green Destiny
Year	1996	1996	2000	2002
Performance (Gflops)	18	600	2500	39
Area (ft^2)	120	1600	9920	6
Power (kW)	18	1200	2000	5
DRAM (GB)	36	585	6200	150
Disk (TB)	0.4	2.0	160.0	4.8
DRAM density (MB/ft^2)	300	366	625	25000
Disk density (GB/ft^2)	3.3	1.3	16.1	800.0
Perf/Space ($Mflops/ft^2$)	150	375	252	6500
Perf/Power (Mflops/watt)	1.0	0.5	1.3	7.5

Computer & Computational Sciences

Parallel Computing Platforms Running the N-body Code

Machine	Avalon Beowulf	ASCI Red	ASCI White	Green Destiny
Year	1996	1996	2000	2002
Performance (Gflops)	18	600	2500	39
Area (ft^2)	120	1600	9920	6
Power (kW)	18	1200	2000	5
DRAM (GB)	36	585	6200	150
Disk (TB)	0.4	2.0	160.0	4.8
DRAM density (MB/ft^2)	300	366	625	25000
Disk density (GB/ft^2)	3.3	1.3	16.1	800.0
Perf/Space ($Mflops/ft^2$)	150	375	252	6500
Perf/Power (Mflops/watt)	1.0	0.5	1.3	7.5

Parallel Computing Platforms Running the N-body Code

Machine	Avalon Beowulf	ASCI Red	ASCI White	Green Destiny+
Year	1996	1996	2000	2002
Performance (Gflops)	18	600	2500	*58*
Area (ft^2)	120	1600	9920	6
Power (kW)	18	1200	2000	5
DRAM (GB)	36	585	6200	150
Disk (TB)	0.4	2.0	160.0	4.8
DRAM density (MB/ft^2)	300	366	625	25000
Disk density (GB/ft^2)	3.3	1.3	16.1	800.0
Perf/Space ($Mflops/ft^2$)	150	375	252	9667
Perf/Power (Mflops/watt)	1.0	0.5	1.3	11.6

Green Destiny vs. Earth Simulator: LINPACK

Machine	Green Destiny+	Earth Simulator
Year	2002	2002
LINPACK Performance (Gflops)	*101*	35,860
Area (ft^2)	6	17,222 * 2
Power (kW)	5	7,000
Cost efficiency ($/Mflop)	3.35	11.15
Space efficiency ($Mflops/ft^2$)	16,833	1,041
Power efficiency (Mflops/watt)	20.20	5.13

Disclaimer: This is not a fair comparison. Why?

(1) Use of area and power does *not* scale linearly.

(2) Goals of the two machines are different.

The Evolution of *Green Destiny*

- Problems with Green Destiny (even with HP-CMS)
  - An architectural approach that ties us to a specific vendor, i.e., RLX, who looks to be headed in a different direction.
  - Raw performance of a compute node.
    - Two times worse than the fastest CPU at the time of construction (2002). Now, upwards of four times worse (2004).

The Evolution of *Green Destiny*

- Problems with Green Destiny (even with HP-CMS)
  - An architectural approach that ties us to a specific vendor, i.e., RLX, who looks to be headed in a different direction.
  - Raw performance of a compute node.
    - Two times worse than the fastest CPU at the time of construction (2002). Now, upwards of four times worse (2004).
- Obvious Solution
  - Transform our architectural approach into a software-based one that works across a wide range of processors.
  - Start with higher-performing commodity components to achieve performance goals but use the above software-based technique to reduce power consumption dramatically.

The Evolution of *Green Destiny*

- Problems with Green Destiny (even with HP-CMS)
  - An architectural approach that ties us to a specific vendor, i.e., RLX, who looks to be headed in a different direction.
  - Raw performance of a compute node.
    - Two times worse than the fastest CPU at the time of construction (2002). Now, upwards of four times worse (2004).
- Obvious Solution
  - Transform our architectural approach into a software-based one that works across a wide range of processors.
  - Start with higher-performing commodity components to achieve performance goals but use the above software-based technique to reduce power consumption dramatically.
- But How?
  - Dynamic voltage scaling + efficient scheduling algorithm.

The Evolution of *Green Destiny*: Dynamic Voltage Scaling (DVS)

- DVS Technique
  - Trades CPU performance for power reduction by allowing the CPU supply voltage and/or frequency to be adjusted at run-time.
- Why is DVS important?
  - Recall: Moore's Law for Power.
  - CPU power consumption is directly proportional to the *square of the supply voltage* and to frequency.
- DVS Algorithm
  - Determines *when* to adjust the current frequency-voltage setting and *what* the new frequency-voltage setting should be.

The Evolution of *Green Destiny:* Real-Time Constraint-Based DVS

- Assume n frequency-voltage settings (f_i, V_i)
- P_i is the power consumption at setting i
- T_i is the total execution time running entirely at f_i
- D is the deadline

$$\min E = \sum_i P_i \cdot t_i$$

such that

$$\sum_i t_i \le D$$

$$\sum_i t_i / T_i = 1$$

$$t_i \ge 0$$

The Evolution of *Green Destiny*: Generalized Theorem

Theorem. If $f_1 < f_2 < \cdots < f_n$ and $T_1 > T_2 > \cdots > T_n$ and

$$0 \geq \frac{E_2 - E_1}{T_2 - T_1} \geq \frac{E_3 - E_2}{T_3 - T_2} \geq \cdots \geq \frac{E_n - E_{n-1}}{T_n - T_{n-1}}$$

then

$$t_i^* = \begin{cases} \frac{D - T_{j+1}}{T_j - T_{j+1}} \cdot T_j & i = j \\ D - t_j^* & i = j + 1 \\ 0 & \text{otherwise} \end{cases}$$

where

$$E_i = P_i \cdot T_i \quad \text{and} \quad T_{j+1} < D \leq T_j$$

(Note: the theorem generalizes the results developed by Ishihara and Yasuura at ISLPED-1998 which many DVS scheduling algorithms base on)

The Evolution of *Green Destiny*: Initial Experimental Results

- Tested on a mobile AMD Athlon XP system with 5 settings
- Measured through Yokogawa WT210 digital power meter
- $\beta \in [0, 1]$ indicates performance sensitivity to changes in CPU speed, with 1 being most sensitive.

program	β	T_{rel}/E_{rel}
swim	0.02	1.02/0.46
tomcatv	0.24	1.01/0.80
su2cor	0.27	1.02/0.81
compress	0.37	1.05/0.80
mgrid	0.51	1.04/0.84
vortex	0.65	1.06/0.85
turb3d	0.79	1.04/0.92
go	1.00	1.05/0.93

Conclusion

- Traditional Performance Metrics
  - Performance
  - Price/Performance
- New Performance Metrics
  - Overall Efficiency
    - ToPPeR: Total Price-Performance Ratio
  - Power Efficiency
    - Performance-Power Ratio
  - Space Efficiency
    - Performance-Space Ratio

Wu-chun Feng
feng@lanl.gov

Conclusion

- Performance Metrics for Green Destiny (circa 2002)
  - Performance
    - 2x to 2.5x worse than fastest Intel/AMD processor.
  - Price/Performance
    - 2x to 2.5x worse.
  - Overall Efficiency: ToPPeR
    - 1.5x to 2x better. (See related publications.)
  - Power Efficiency: Performance-Power Ratio
    - 10x to 20x better.
  - Space Efficiency: Performance-Space Ratio
    - 20x – 60x better.

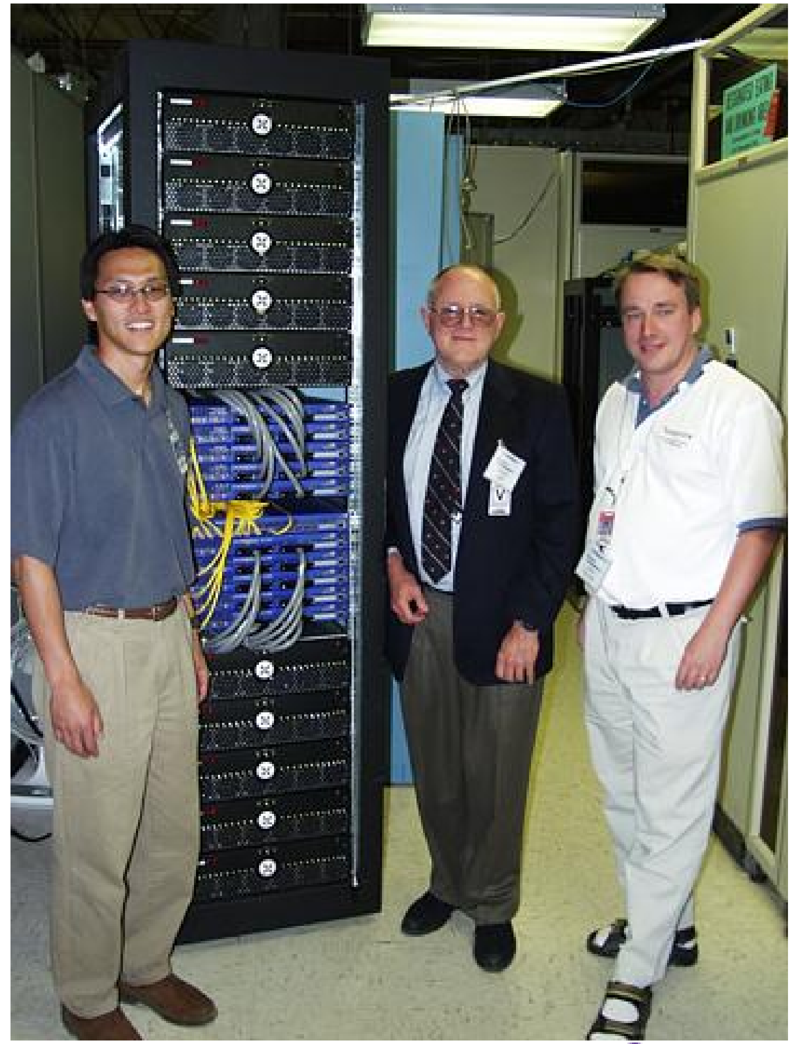
Conclusion

- Problem with Green Destiny
  - Architectural solution that sacrifices too much performance.
- Solution: Software-Based Solution
  - Real-time, constraint-based dynamic voltage scaling.
  - Performance on AMD XP-M
    - Power reduction of as much as 56% with only a 2% loss in performance.
- Future Directions
  - Refinement of the DVS scheduling algorithm.
  - Profiling on multiprocessor platforms and benchmarks.

COMPUTER & COMPUTATIONAL SCIENCES

Selected Recognition & Awards

- 2003 R&D 100 Award, 10/16/03.
- "Los Alamos Lends Open-Source Hand to Life Sciences," *The Register*, 6/29/03. http://www.theregister.com/content/61/31471.html.
- "LANL Researchers Outfit the 'Toyota Camry' of Supercomputing for Bioinformatics Tasks," *BioInform / GenomeWeb*, 2/3/03.
- "Developments to Watch: Innovations," *BusinessWeek*, 12/2/02.
- "Craig Venter Goes Shopping for Bioinformatics to Fill His New Sequencing Center," *GenomeWeb*, 10/16/02.
- "At Los Alamos, Two Visions of Supercomputing," *The New York Times*, 6/25/02.
- "Supercomputing Coming to a Closet Near You?" *PCworld.com*, 5/27/02.
- "Bell, Torvalds Usher Next Wave of Supercomputing," *CNN*, 5/21/02.



Acknowledgments

- Technical Co-Leads
  - Mike Warren and Eric Weigle
- Contributions
  - Mark Gardner, Adam Engelhart, Gus Hurwitz
- Encouragement & Support
  - Gordon Bell, Chris Hipp, and Linus Torvalds
- Funding Agencies
  - LACSI
  - IA-Linux

Wu-chun Feng
feng@lanl.gov

Los Alamos NATIONAL LABORATORY

SUPERCOMPUTING
in SMALL SPACES

http://sss.lanl.gov

Wu-chun (Wu) Feng
Research and Development in Advanced Network Technology
RADIANT

http://www.lanl.gov/radiant