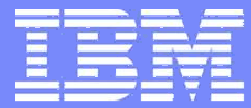


PowerScale™ Energy Management Technology For System p Servers

May 15, 2006

Michael Floyd, Senior Engineer
IBM Server & Technology Group

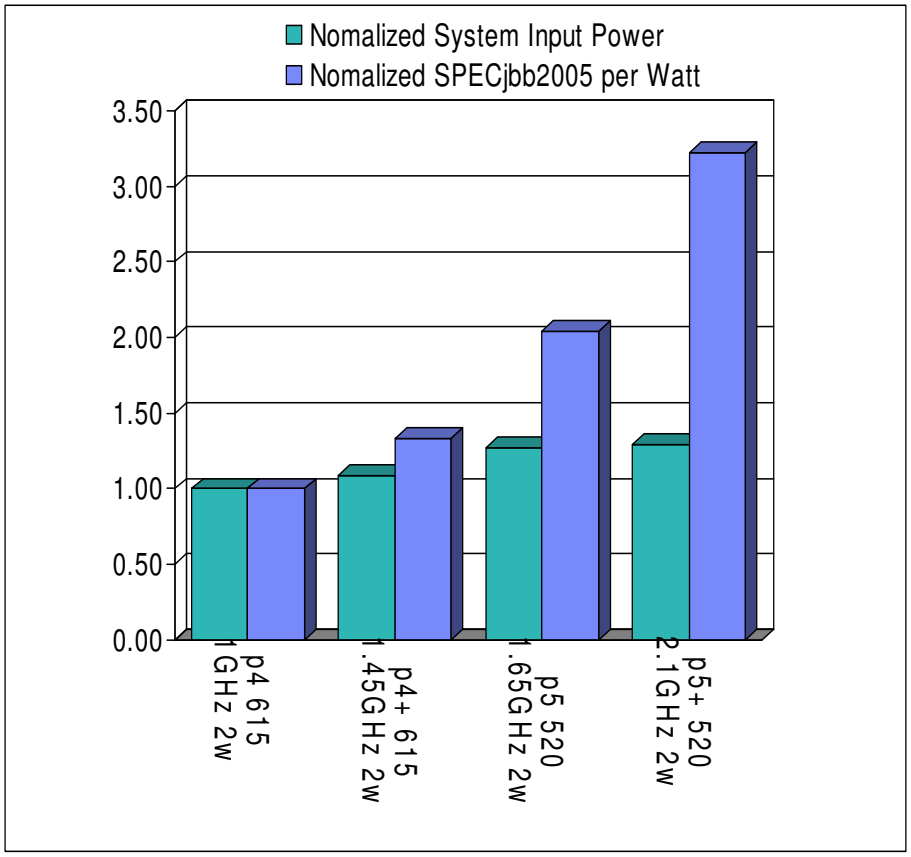
Overview | IBM's Big Green Innovation

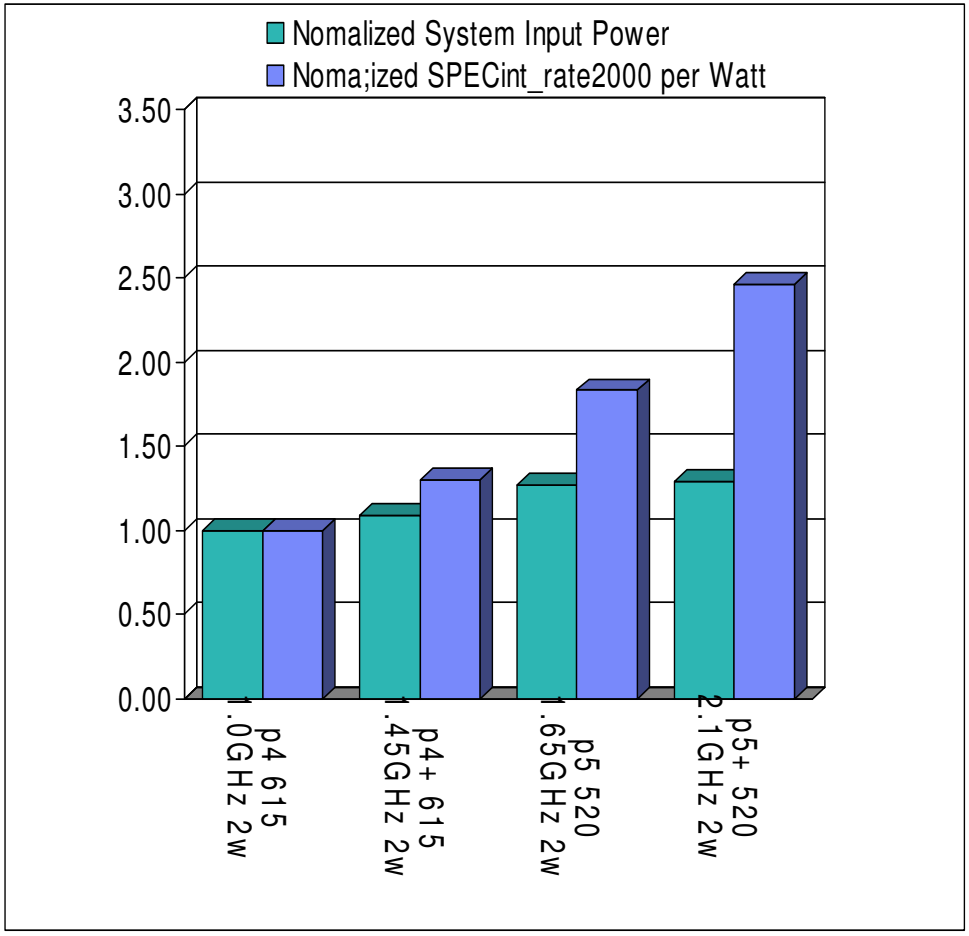
- IBM increasing push for *Datacenter Energy Efficiency & Power Management*
  - Founding member of the "The Green Grid" Initiative in early 2007
    - now a 39 company strong consortium of compute industry leaders
  - "Project Big Green," announced May 10, 2007 YouTube - Project Big Green - in Second Life
    - IBM is allocating $1 Billion per year towards the goal of making IT infrastructures more energy-efficient
    - IBM claims this effort will enable average 25,000-square-foot data center to cut energy bills by 42 percent.
    - IBM has also set a goal of doubling the computing capacity of its worldwide data centers by 2010 while keeping power consumption levels steady.

- Only IBM has the depth and breadth to create a comprehensive end-to-end solution for the datacenter
  - IBM engineers design energy-efficiency into all levels of computing:
    - From the layout of the datacenter floor space
    - Through the workload management and IT software stack
    - To the server hardware components and integration
    - To the transistors on the computer chips

System p | IBM Energy Efficient Hardware Design

IBM Server Performance/Power Efficiency Is Increasing





Hardware Configuration: CEC 2-way, 8 GB memory, 1 DASD

Overview | IBM's Big Green Innovation

- IBM's Cool Blue™ Technology, first outlined in 2005
  - Modified description from the IBM website

1. Start with a smartly designed server that takes less power to deliver the maximum function and performance
2. Pair the server with a well-planned rack and data center layout.
3. Manage IT infrastructure via industry leading portfolios, e.g.:
   - Tivoli Software Suite for workload management
   - IBM Systems Director for Partitioning and Virtualization
4. Finish with power management tools that measure power and heat while collecting and trending power consumption and temperature data server-by-server, enabled by IBM PowerExecutive™.
5. Now starting with POWER6, we complete the picture with IBM's EnergyScale™ technology that will allow the customer to manage the power/performance of their servers as per configurable policy.

Overview | IBM's Big Green Innovation

- IBM is leading with innovative hardware & software, products & services

Examples:

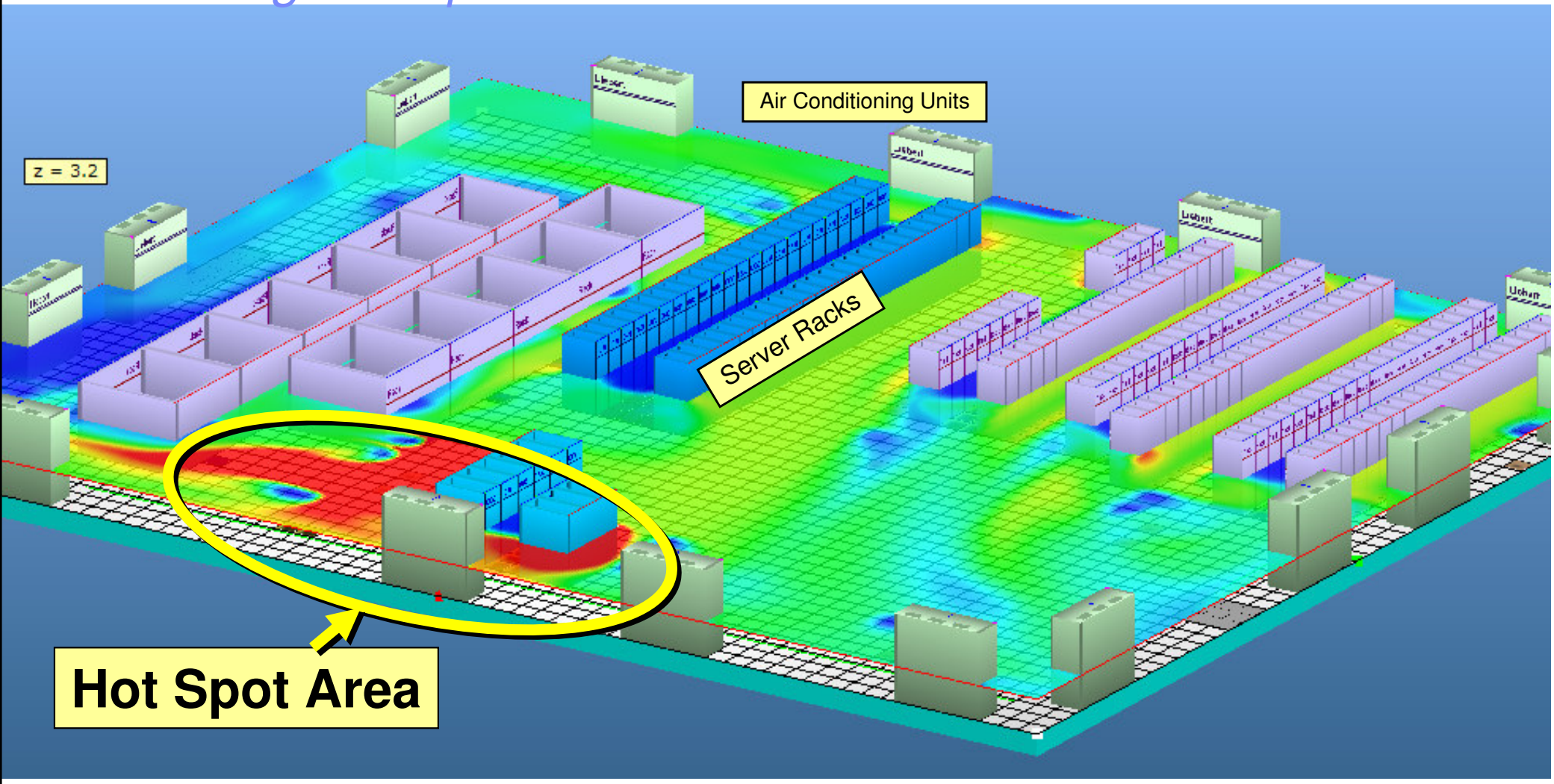
- Rear Door Heat eXchanger
  - Datacenter power and cooling management innovation
  - Introduced in July 2005
  - Use the existing chilled water supply already located in the majority of customer datacenters to reduce server heat emissions by up to 55 percent, and save up to 15 percent of energy costs

- PowerExecutive™
  - Allows the customer to monitor power consumption of a collection of servers
  - Started shipping in November 2005
  - Shipped on a majority of System X rack-mounted servers and blades in 2006
  - Recently added Power Capping capability
  - Begins shipping in June 2007 on System P servers, taking advantage IBM EnergyScale™ Technology on POWER6

- POWER Hypervisor Partitioning and Virtualization
  - Virtualization -- consolidates workloads onto a fewer number of processors, more efficient since has less unused server overhead
  - IBM Micro-partitioning™ lets workloads more efficently share the available processors
  - Dynamic LPAR and Capacity on Demand (CoD) enables only the processors needed, allowing unused processors to be turned off

IBM Technology

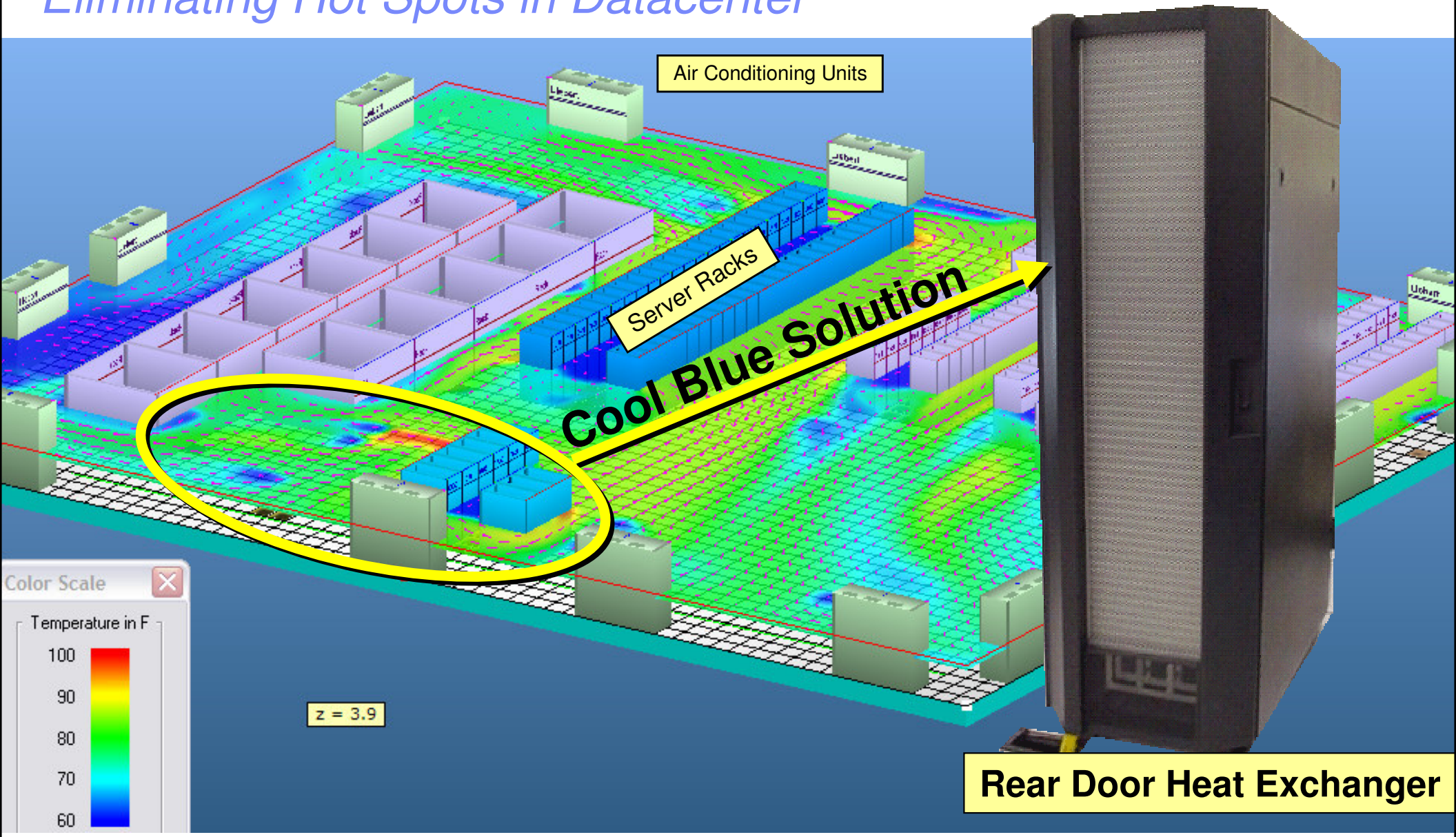
Eliminating Hot Spots in Datacenter



Coolblue

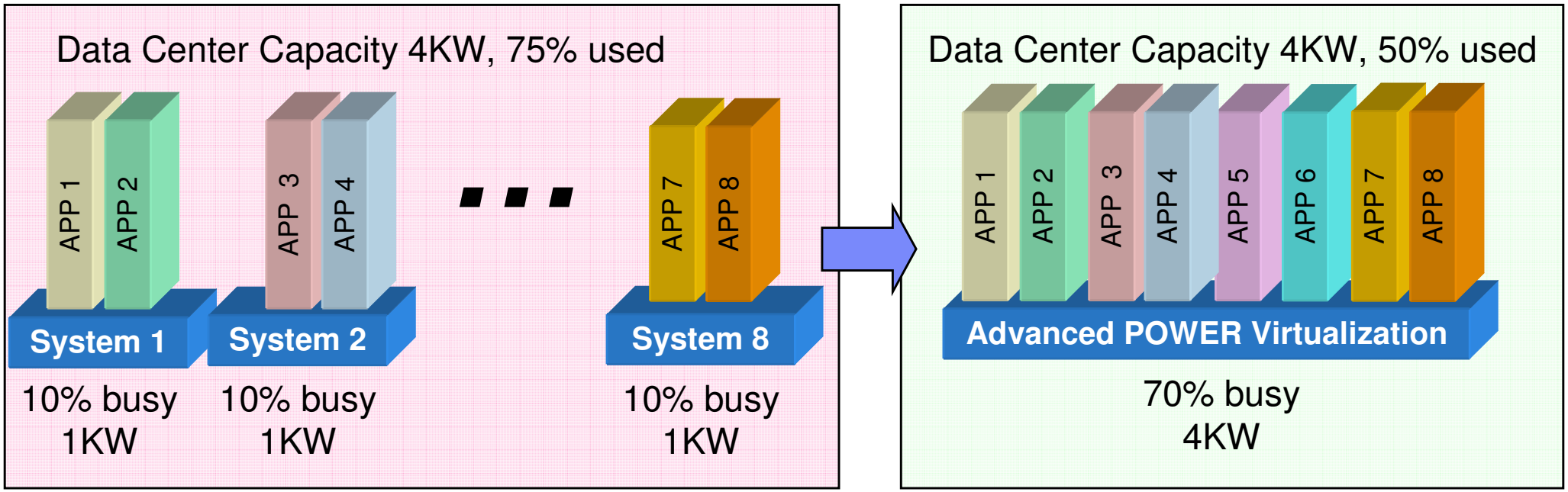
IBM Technology

Eliminating Hot Spots in Datacenter



Virtualization | System Level Energy Management

Physical Server Consolidation Conserves Energy



Total Power 8KW

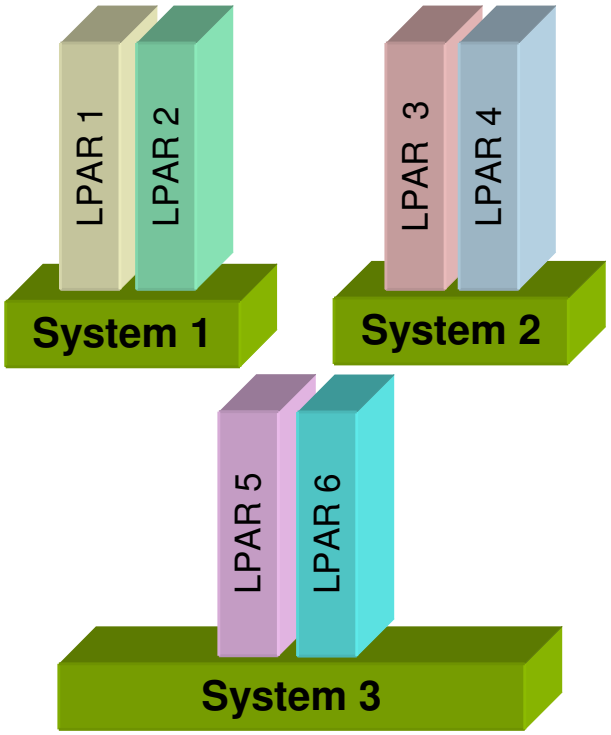
Total Power 4KW

Server consolidation exploiting virtualization is a very effective tool in reducing energy costs

Partition Management | Dynamic Energy Savings

Dynamically pack workloads to reduce overall power consumption

- *Use LPAR Migration to pack workload to single system*
- *Systems 1 and 2 enter low power state or go off*

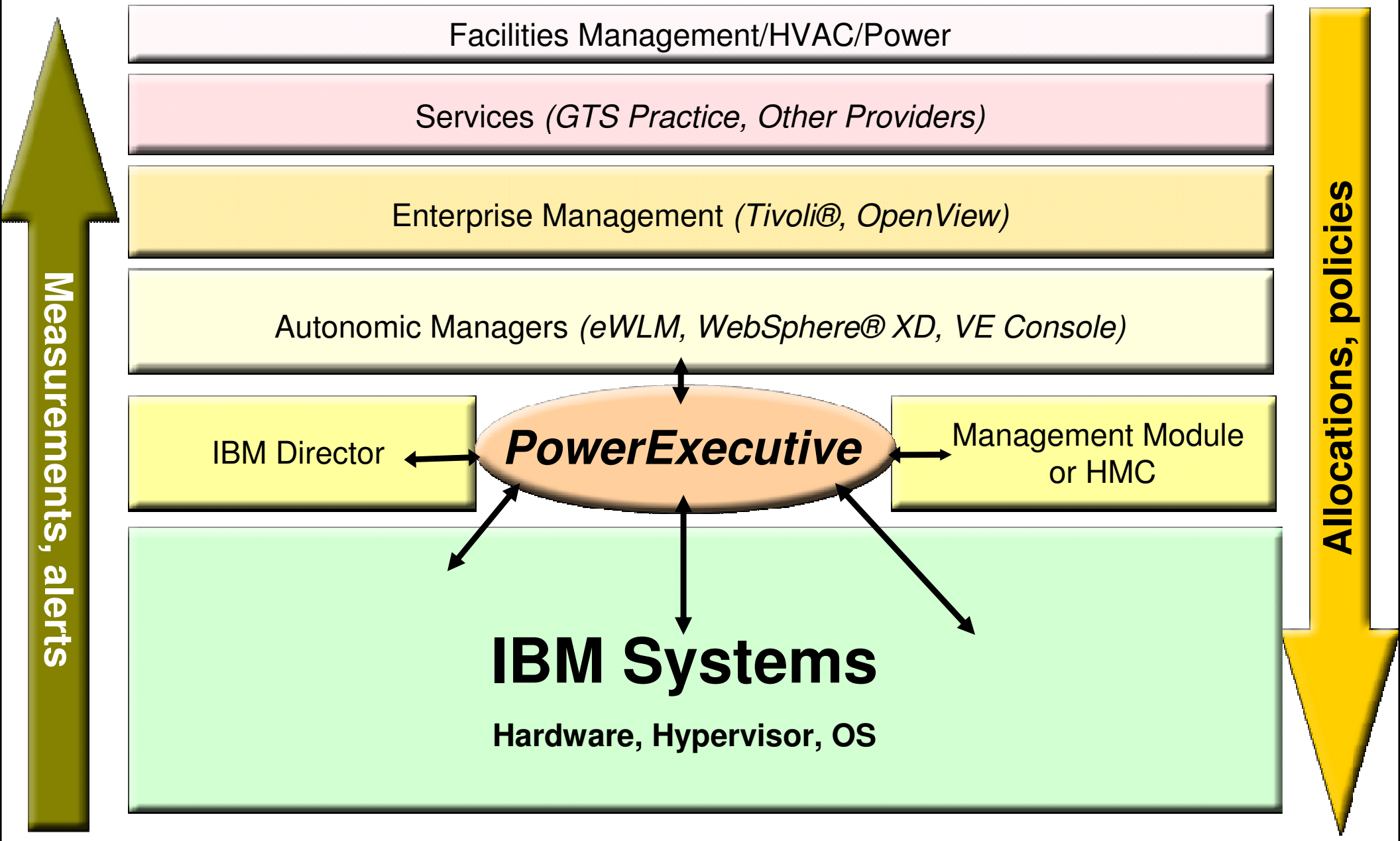


Use of hibernation, powering off servers, and other low power states in combination with other workload balancing and provisioning tools can provide a valuable tool in management of Power and Thermal issues.

Automate Power Control
Policy based automation

Control Power Consumption
Consolidate workloads to reduce

System Stack Enablement | PowerExecutive™



System Firmware | PowerExecutive™ Interface

- **IBM Strategic Interface**
- **Hardware Management Console may forward customer policy to PEx**
- **Select systems may communicate directly with PEx via the Common Information Model (CIM)**
- **P6 Blade conforms to the BladeCenter power and thermal management architecture**

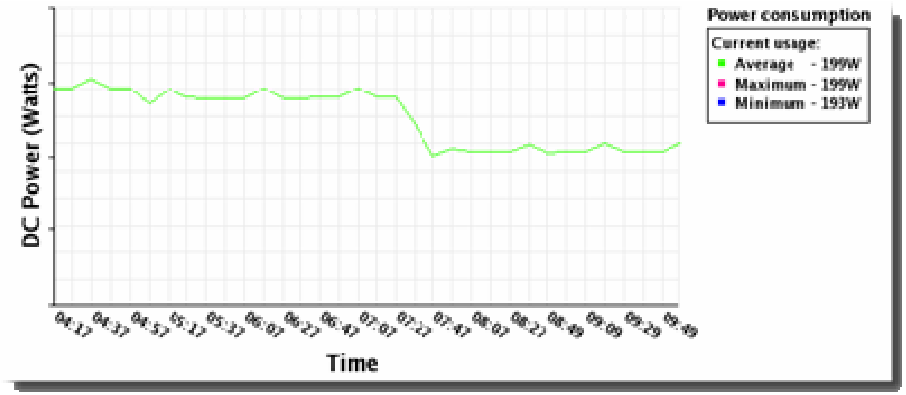


Figure 1: POWER6 BladeCenter Screenshot

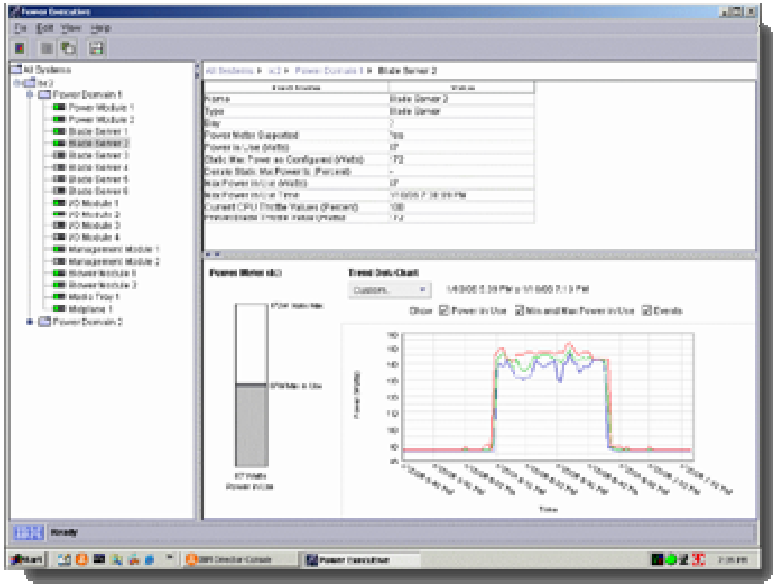


Figure 2: PowerExecutive Screenshot

View a demo online at http://www.ibm.com/systems/management/director/extensions/powerexec.html .

Demonstrations | IBM PowerExecutive™

PowerExecutive base features and function

IBM_Demo_IBM_Director_Power_Executive_Tool-1-Feb06.exe

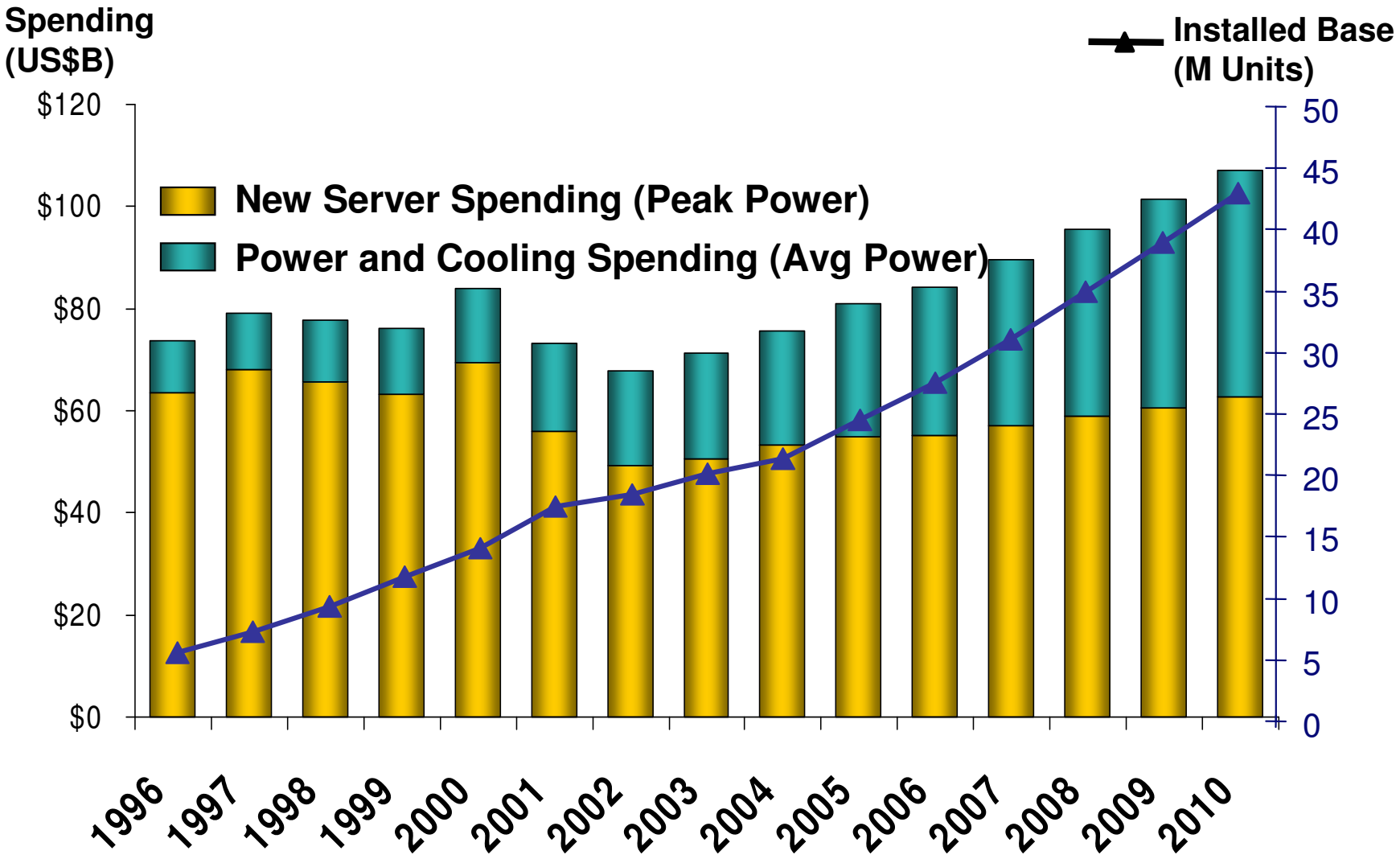
PowerExecutive -- Power Capping capability

IBM_Demo_IBM_PowerExecutive_Power_Capping-1-Mar07.exe

Overview | IBM's Big Green Innovation

- Datacenter Energy costs and concerns are increasing due to several converging factors:
  - Compute density and therefore power consumption is growing exponentially
    - Smaller chips with multiple processor cores per chip
    - Smaller transistors with higher energy leakage (tradeoff to maintain performance increases)
  - Rising energy costs
    - Including increased cost of electricity during hours peak consumption in some markets
  - Overstressed power grids in some locations/scenarios (rolling brownouts)
  - Increased sensitivity to Carbon Emissions and the Environment
  - Existing and Impending Government Regulations in Japan, EU, and the US

Datacenter Cost

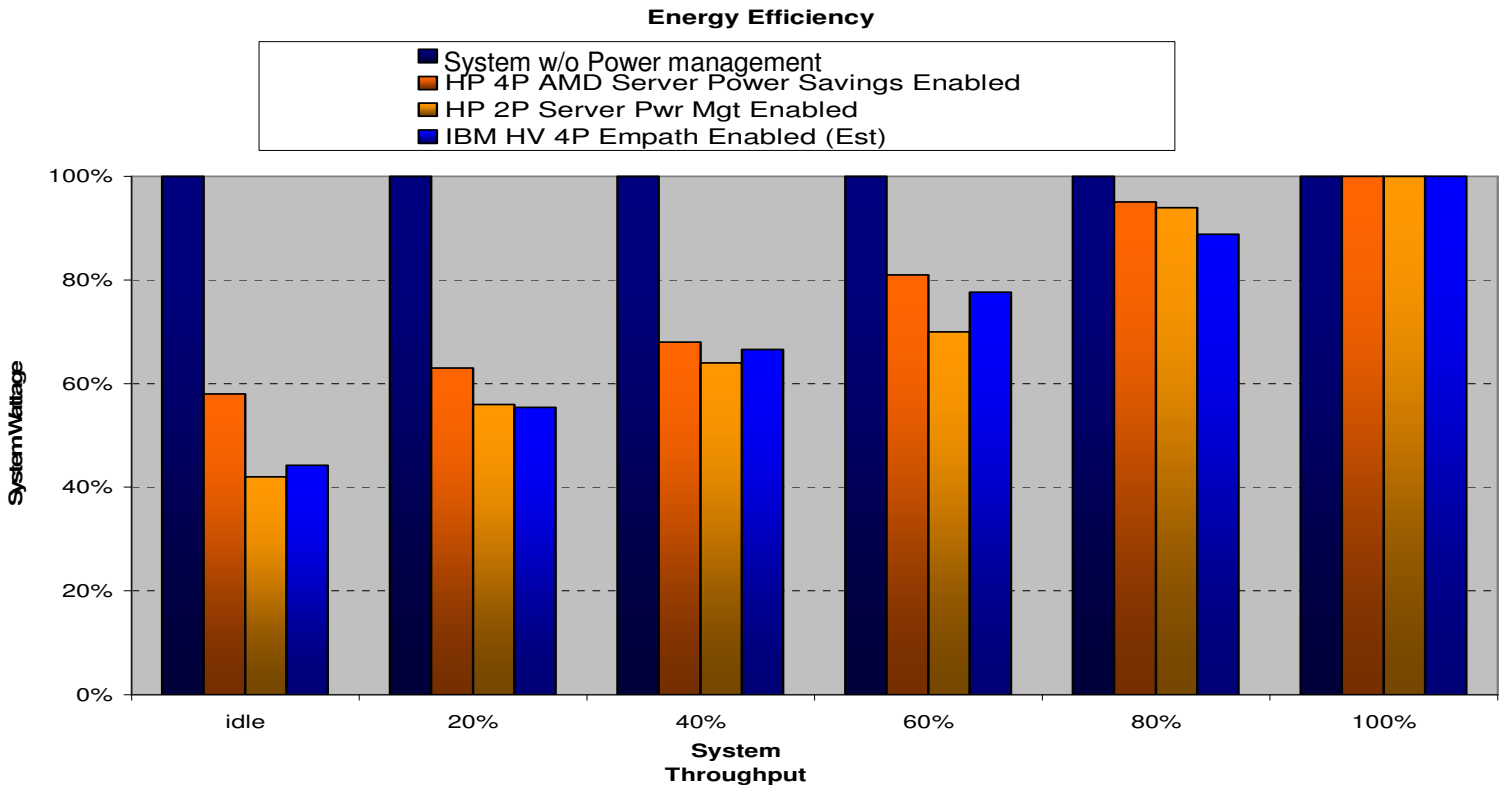


SOURCE: IDC, 'The Impact of Power and Cooling on Data Center Infrastructure,' Document #201722, May 2006"

Growing Issue | Average Power

- **Average power has had minimal impact on previous system designs**
  - Possible exception is RAS computations

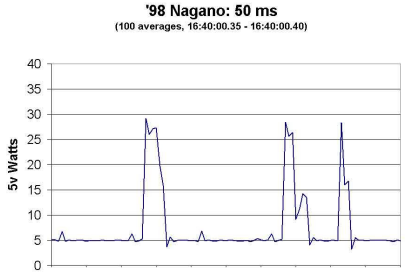
- **In the future average power will be as important as peak power is today**
  - SPEC Benchmarks will be published at a range of CPU utilizations
  - System power management strategy enables exploitation of variation in workloads or environment
  - POWER/PERF analysis at a single operating point is not sufficient
    - An optimal power performance curve is required



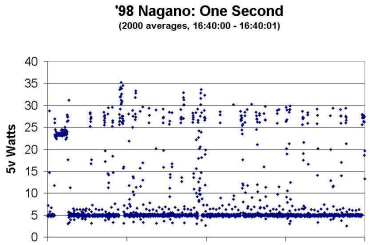
Average Power Minimization

- Fast Sense and Respond
- Capping Ability

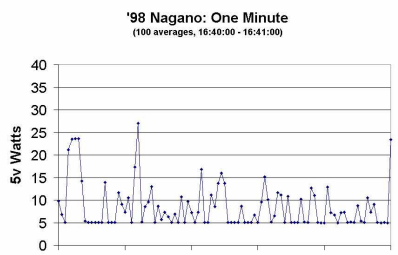
50 ms



1 Sec

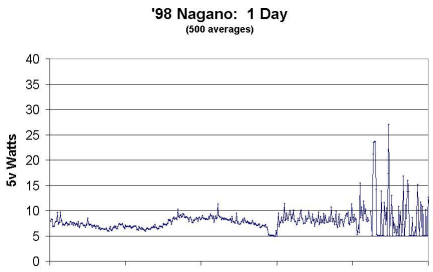


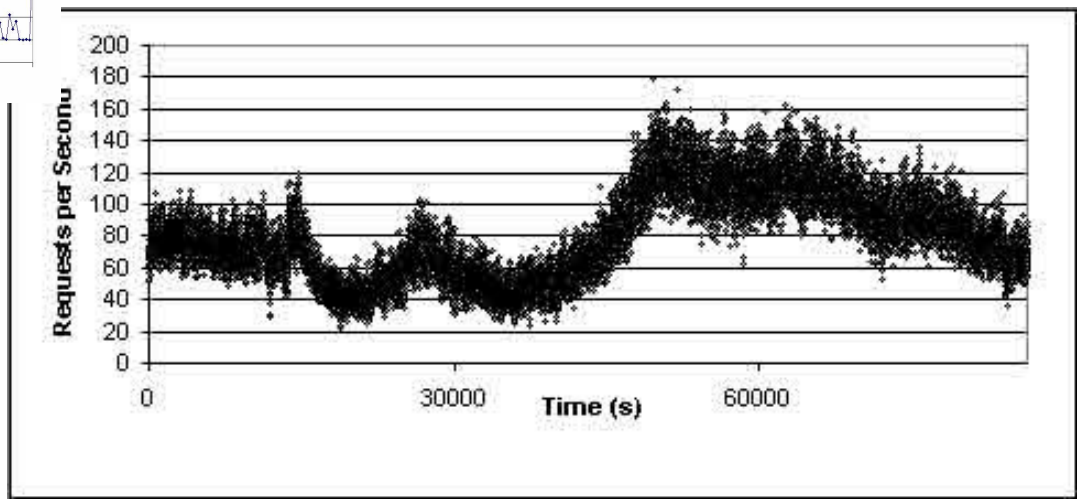
1 Min



Nagano Workload Run on Pentium CPU -- Wattage Measured Over Different Timescales

1 Day

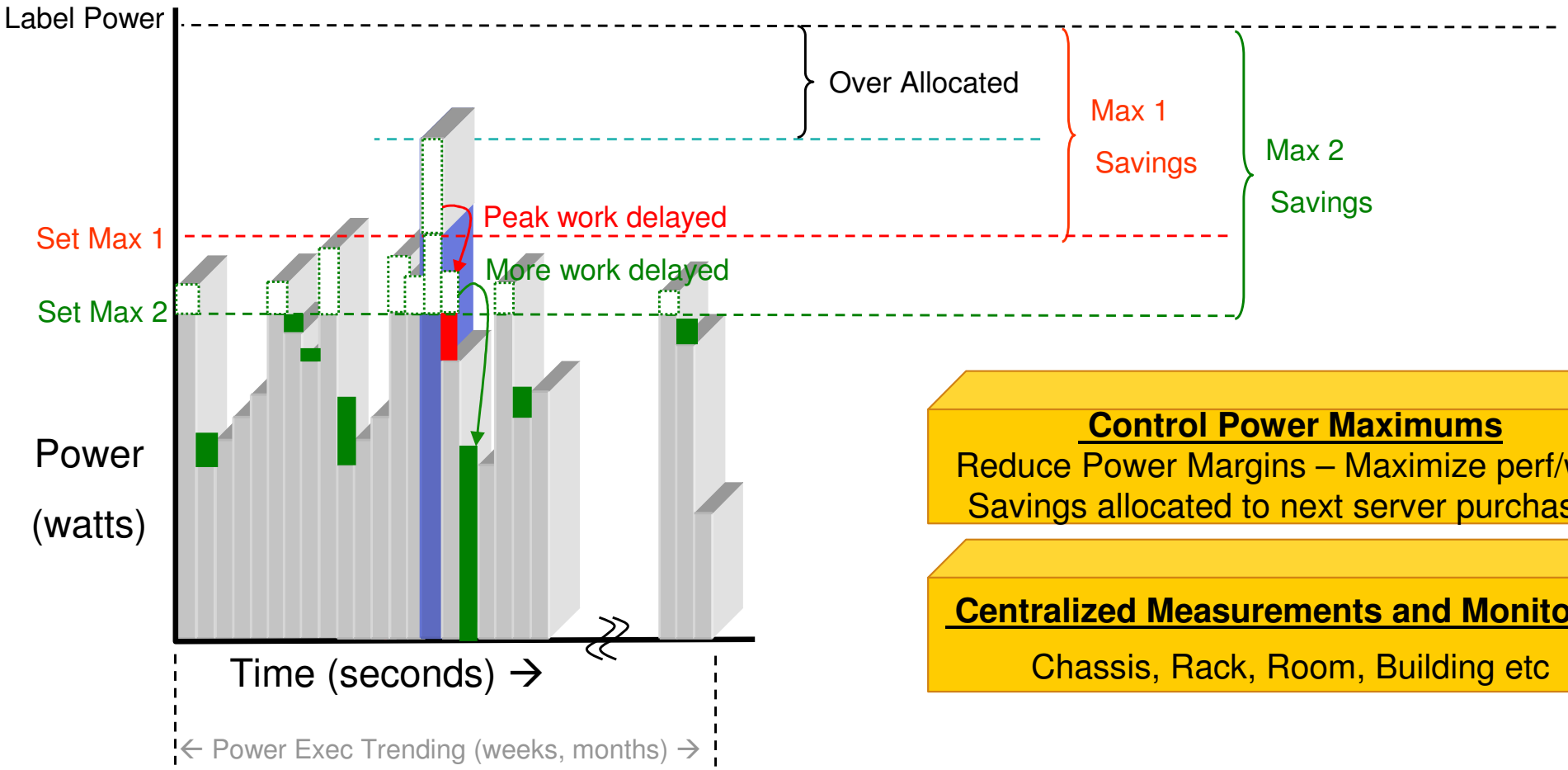




Managing Power vs. Performance

A graph of power consumed by the server over time.

- Today, label power is the only option within the server
- Set Max 1 Power Limit – Very conservative, little delay
- Set Max 2 Power Limit – Less conservative, more delay

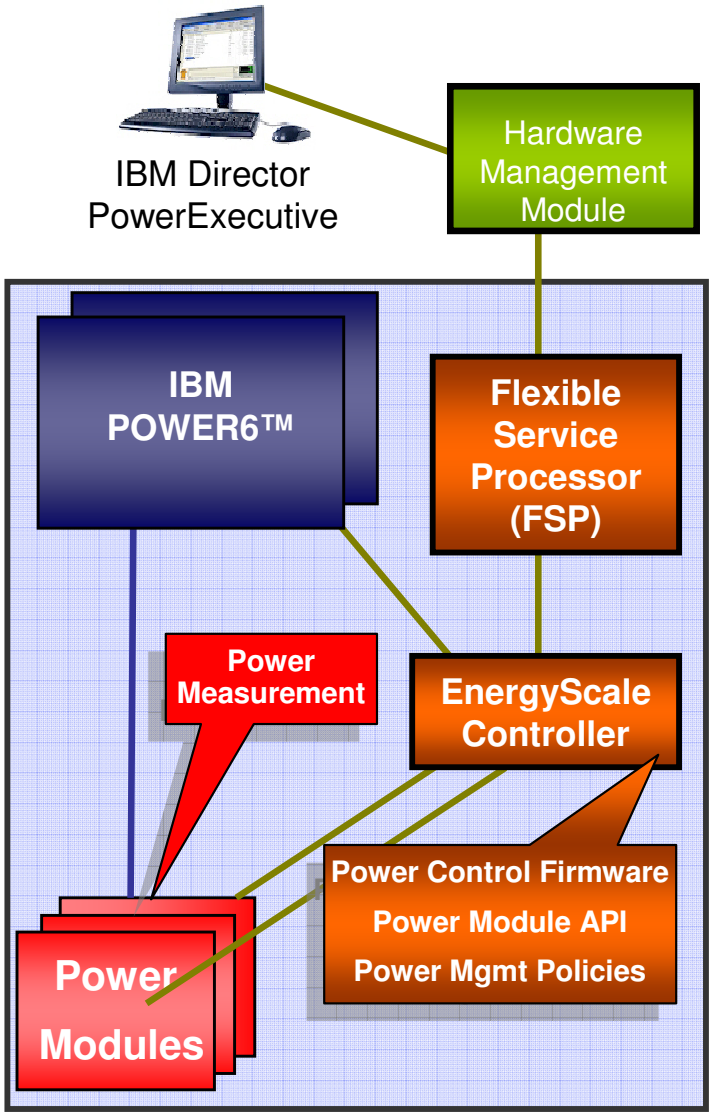


Control Power Maximums
Reduce Power Margins – Maximize perf/watt
Savings allocated to next server purchased

Centralized Measurements and Monitoring
Chassis, Rack, Room, Building etc

Overview | POWER6 EnergyScale™ Functions

Architecture Illustration



- **Thermal / Power Measurement**
  - Read thermal data from processor chip thermal sensors
  - Measure power data from system level sensors
  - Report data via PowerExecutive™
- **System Health Monitoring/maintenance**
  - Use of hardware sensors to ensure system is operating within predefined safety bounds
- **Power Capping**
  - Precision control to keep system power under a specified limit
- **Power Saving**
  - Operation at reduced power/energy when workload and/or policy allows
    - Can be a static policy (e.g. overnight reduction)
    - Can be dynamic (when absolute max performance is not always required).
- **Performance-aware Power Management**
  - Dedicated sensors to guide power and thermal management
  - Policy-guided to enhance performance under power/thermal constraints or enhanced power savings

Overview | POWER6 EnergyScale™ Customer Value

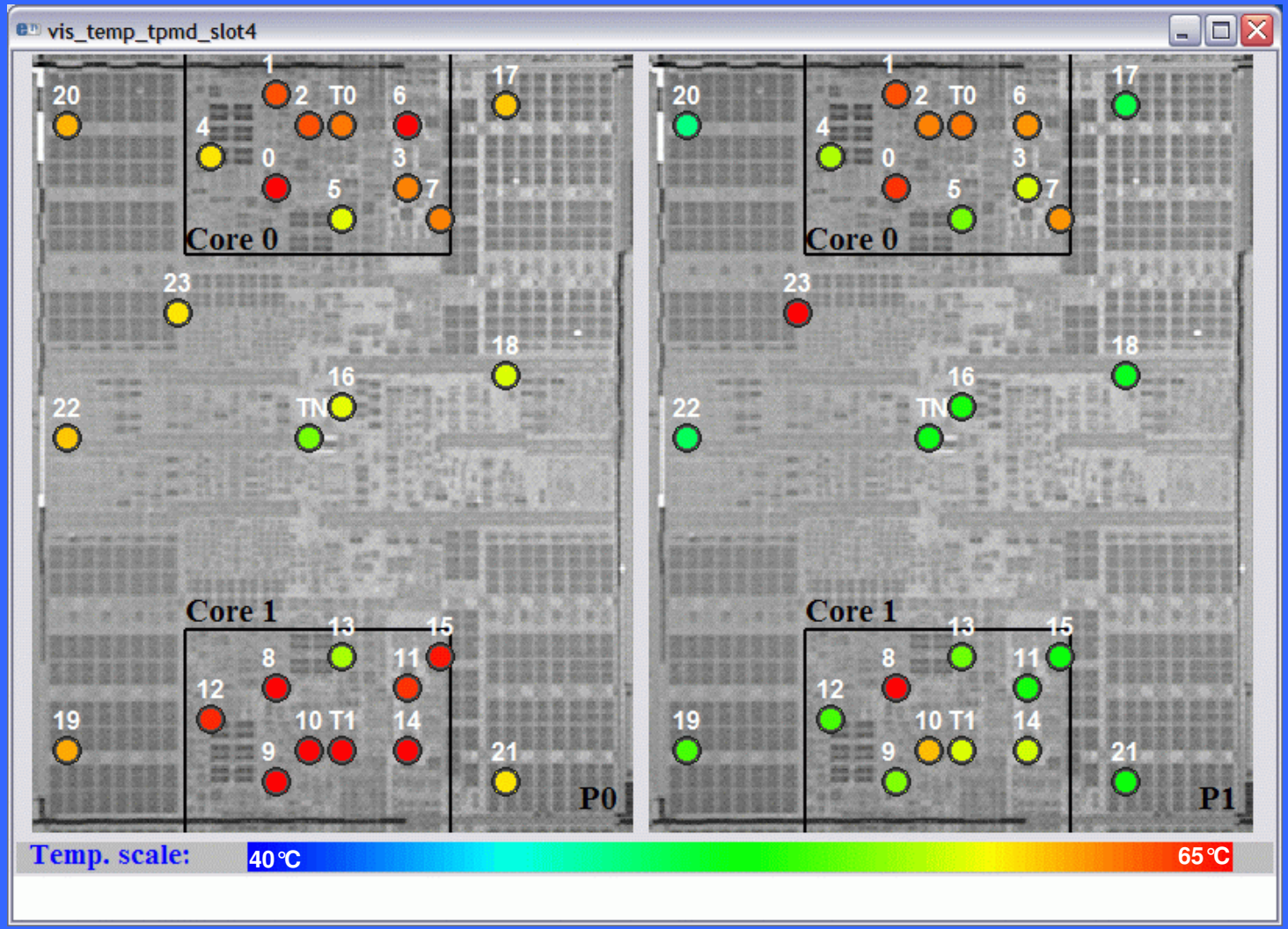
- **Finer Data Center Control and Management**
  - Policy-driven control allows the customer to direct behavior of the system
  - Guarantee customer-configured power, thermal, and performance targets
  - Reduce or eliminate under-provisioning found in many data centers
  - Reduce or avoid capital construction costs in new and existing data centers
  - Provide critical event and comprehensive system usage information
  - Enable integration into larger scope system and datacenter management frameworks

- **Lower Operating Costs with Improved Performance**
  - Dynamically maintain the power and temperature within prescribed limits
  - Reduce cost by simplifying facilities needed for power and cooling
  - Allows higher processor frequency and performance within a power budget than otherwise possible

- **Enhanced Availability**
  - Safely continue system operations under adverse thermal conditions and power availability
  - React to key component failure such as power supply and cooling faults (Oversubscription Protection)

- **Real-time, Measurement-based Control and System-level Optimization**
  - *Directly* measures voltage, current, and temperature to determine the characteristics of workloads.
  - Honors power measurement times to 1-millisecond accuracy.
  - Performs closed-loop feedback control to provide the optimal system solution.
  - Considers a holistic view of system power consumption, not just focus on the processor

Example | Real-Time Temperature Monitoring in POWER6 chip

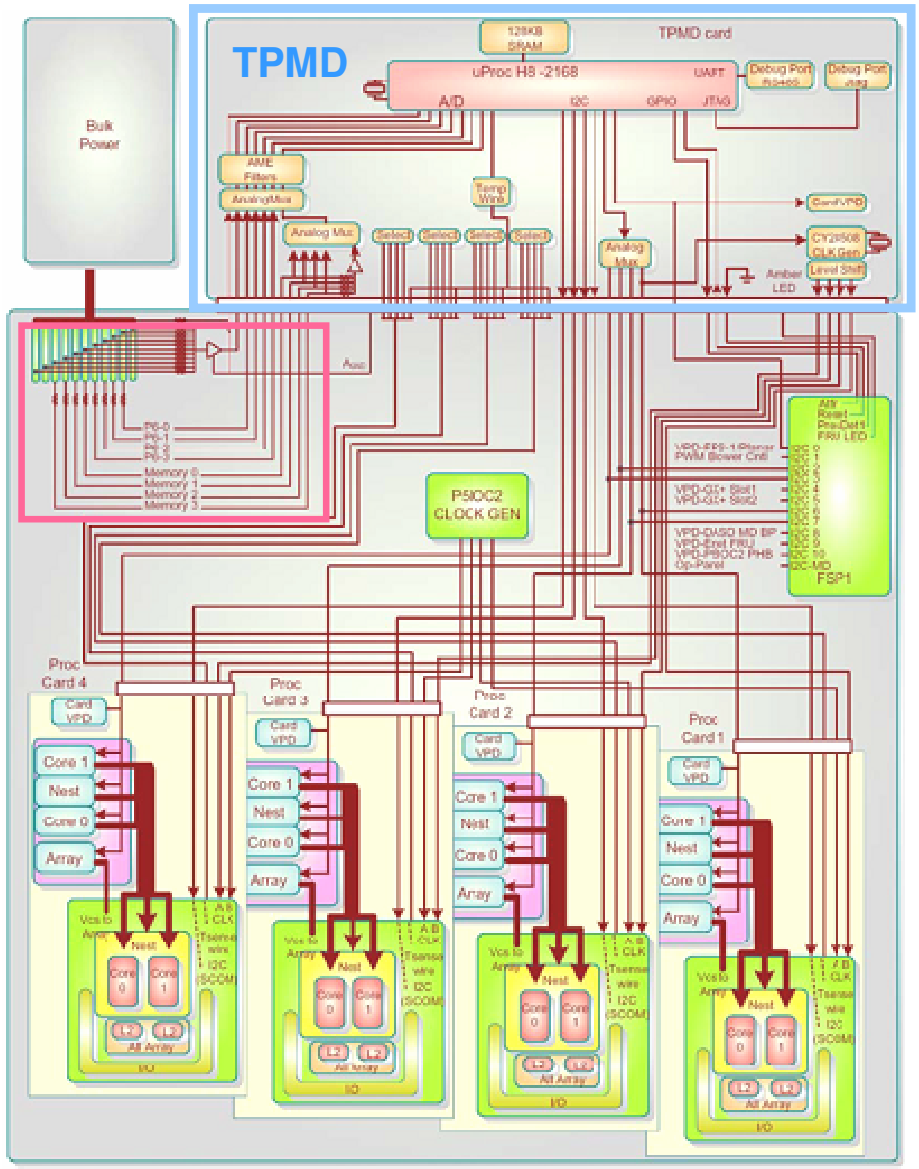


System Implementations | Power6 HV8 Overview

- **System Overview**
  - 8U Rack-mount or Tower
  - 256 GB of System Memory
  - 6 Standard-size *or* 12 small-form-factor SAS Disks
  - Linux-based Flexible Service Processor (FSP)
  - 4 Processor Cards, each featuring a Dual-core POWER6 Module
- **Optional Add-on "Thermal and Power Management Device" (TPMD)**
  - H8-2166 Microcontroller
  - ThreadX RTOS
  - 1 Millisecond Accuracy
  - I2C Communication Channels
  - Shared interfaces to VRMs
  - Dedicated interfaces to each P6
  - On-chip A/D Converters measure AME Circuitry
- **Primary Actuators**
  - Dynamic Voltage / Frequency Slewing
  - CPU Throttling
  - Memory Throttling
  - Memory Power-down

System Implementations | POWER6 HV8 Diagram

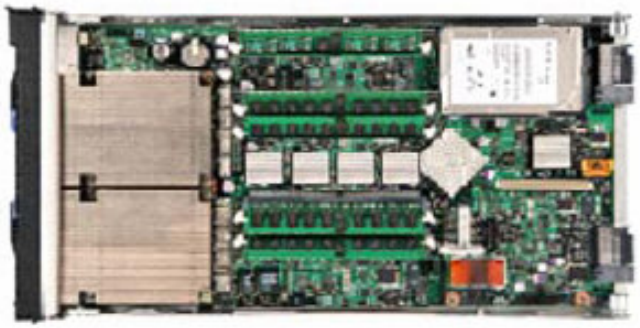
AME Circuits



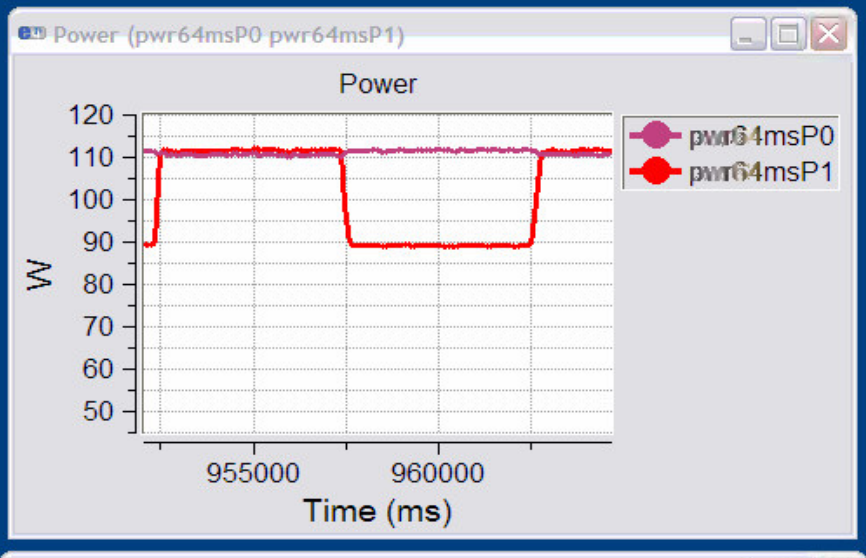
- **TPMD Card Highlights**
  - H8-2166 Microcontroller @ 31 Mhz
  - Dynamically Skewable Clock Generator
  - 80 pin connector
  - Dedicated IIC Bus & Attention line to FSP
  - Dedicated IIC Bus to each P6
  - P6 Thermistor Sense Lines
  - Bulk Current Sense
  - Bulk Voltage Sense
  - Processor Current Sense
  - Memory Current Sense
  - Shared IIC Bus to VRMs

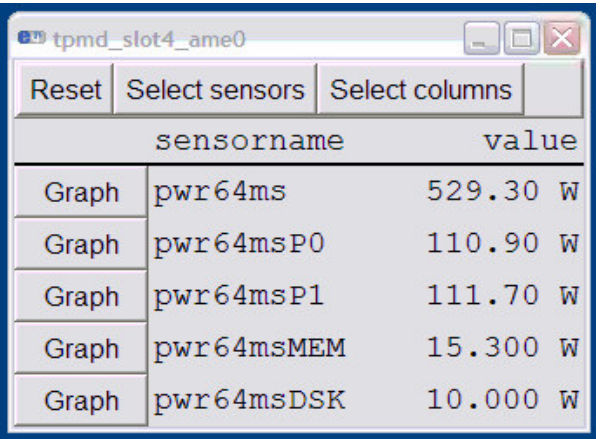
System Implementations | POWER6 Blade

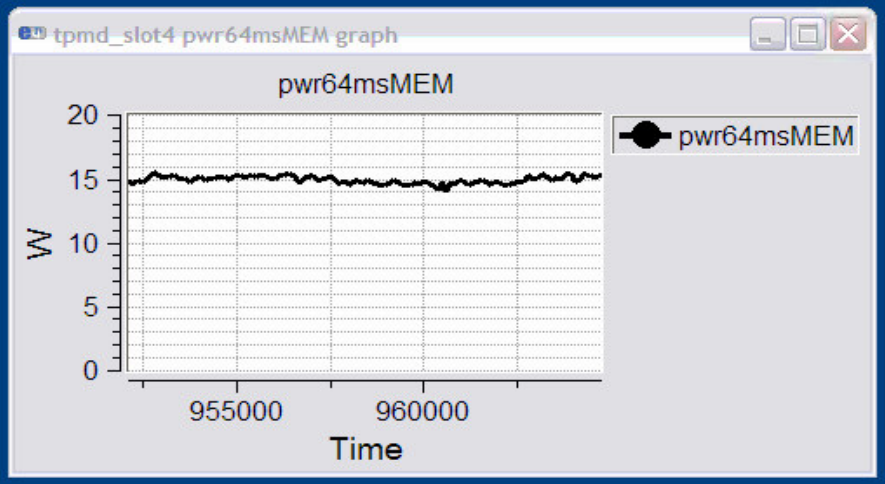
- **System Overview**
  - IBM BladeCenter Blade
  - 32 GB of System Memory (Base Blade)
  - 0-1 2.5” SAS Drives
  - Linux-based Flexible Service Processor (FSP)
  - 4-way: 2 dual-core P6 modules
- Mandatory **Embedded TPMD card for Oversubscription Support**
  - H8-2166 Microcontroller
  - ThreadX RTOS
  - 1 Millisecond Accuracy
  - I2C Communication Channels
  - Shared interfaces to VRMs
  - Dedicated interfaces to each P6
  - On-chip A/D Converters for AME Circuitry
  - Oversubscription Notification
- **Primary Actuators**
  - CPU Throttling
  - Memory Throttling
  - Memory Power-down
- **Very Similar to the HVx Design**

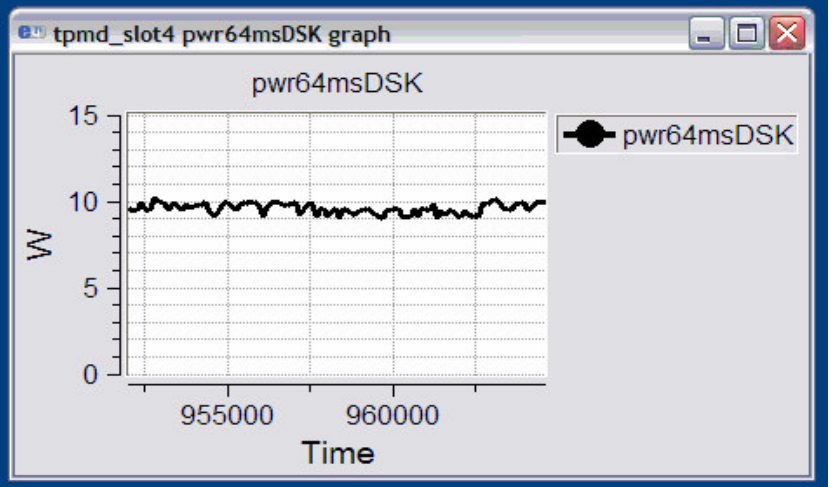


POWER6 Lab Demo | Power Measurement



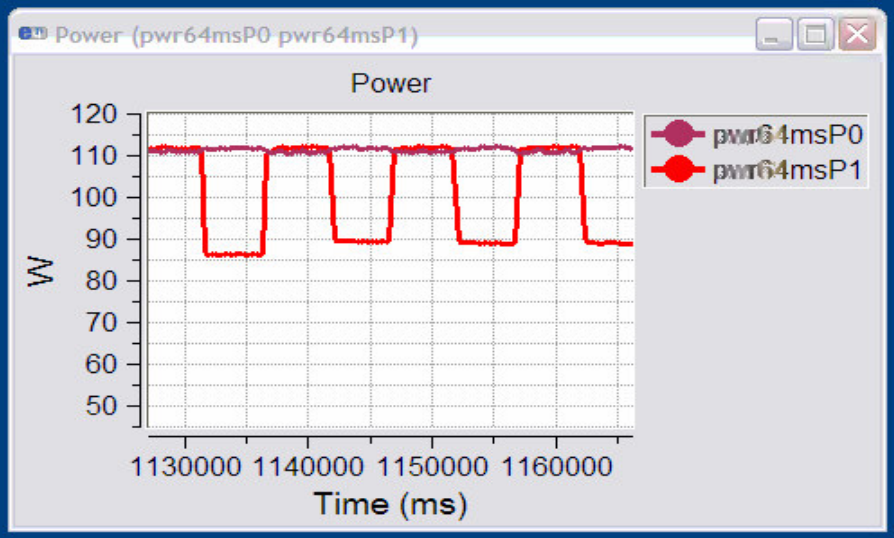


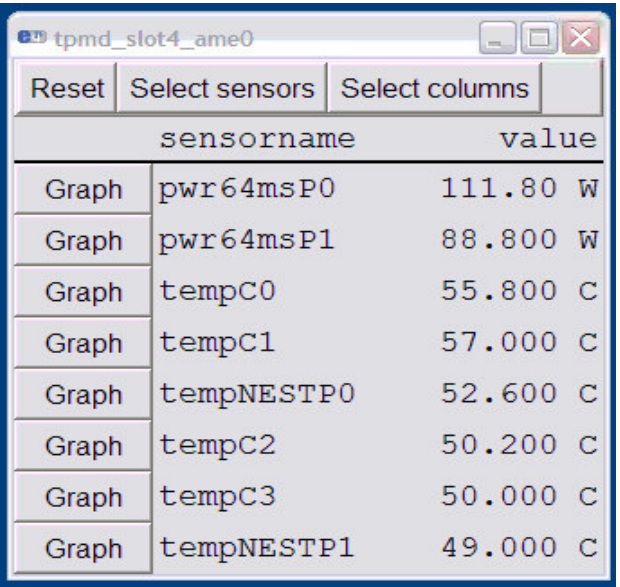


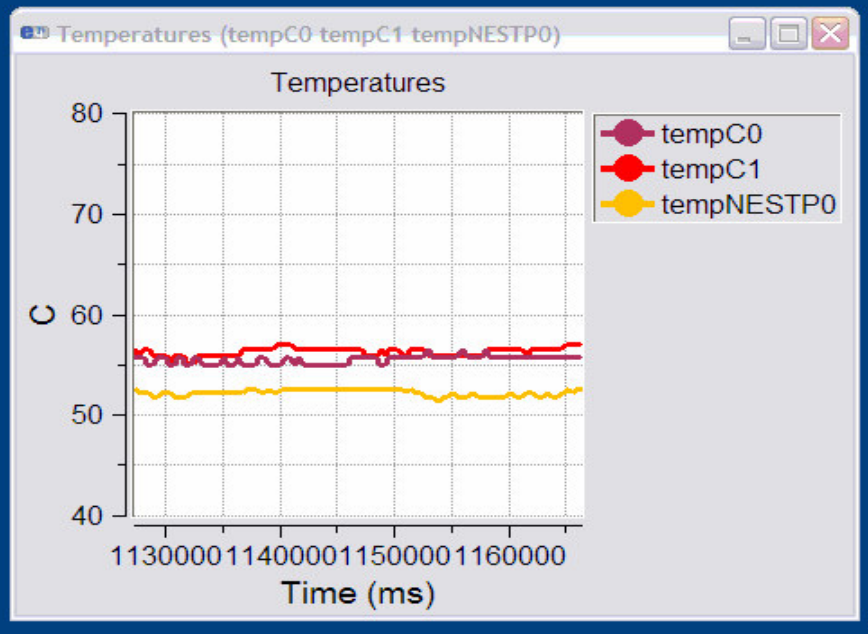


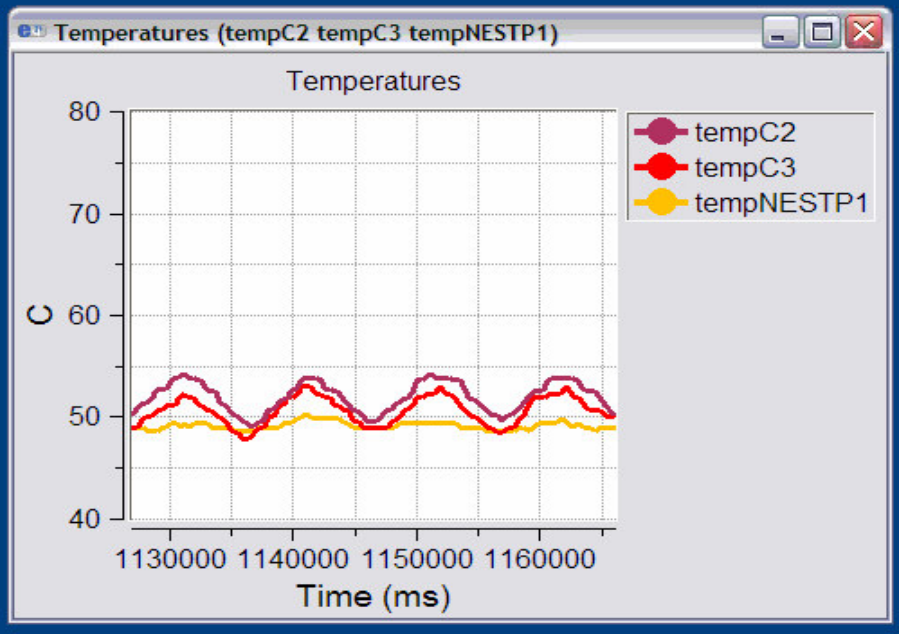
All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

POWER6 Lab Demo | Temperature Measurement

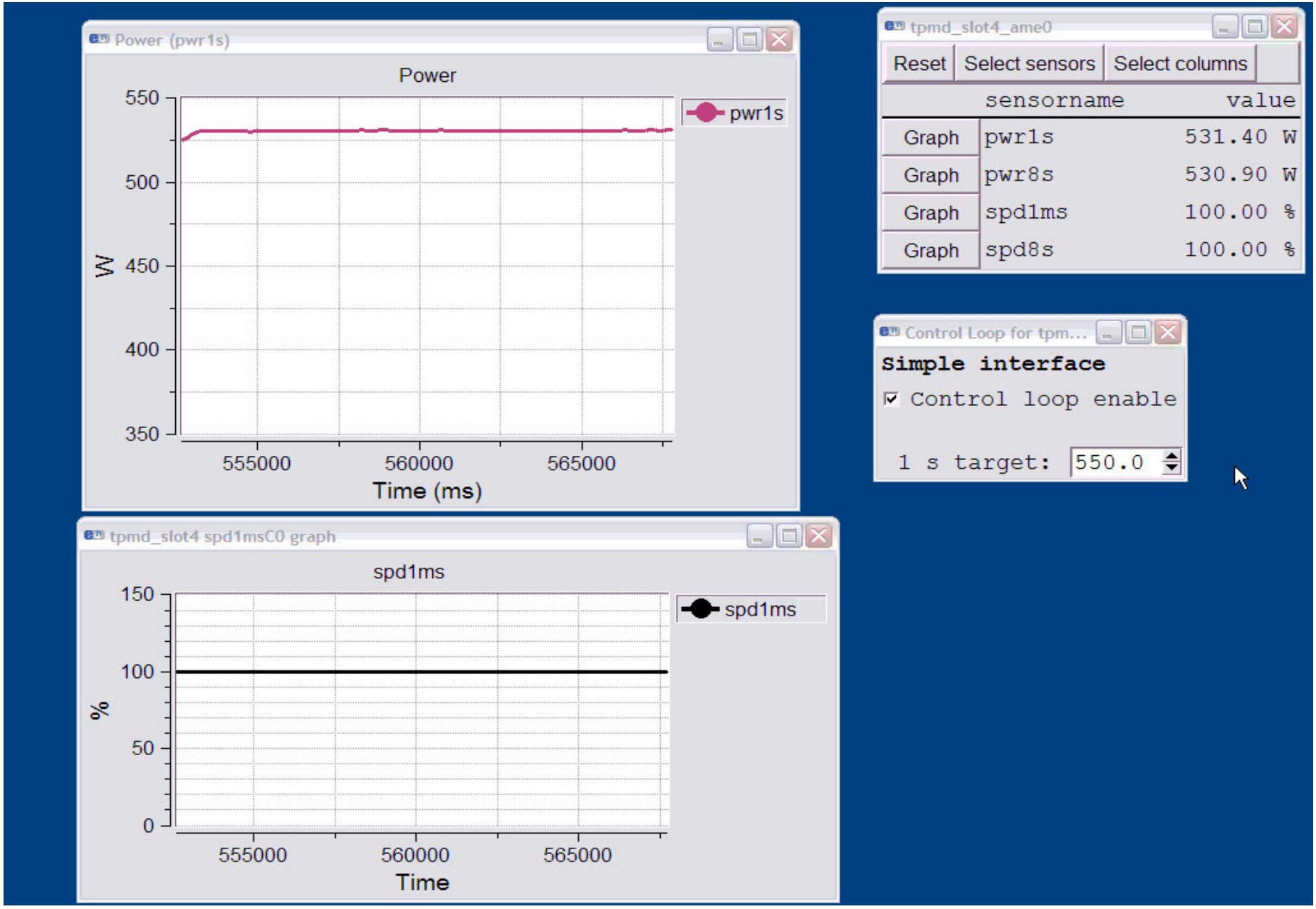








POWER6 Lab Demo | Power Capping



Vision | Energy Management Policies

PWR MGMT MODE	Description
Max Power Save	Minimum average energy, Benchmark & Regulation "mode"
Performance Aware Power Save	Set a performance floor and run in Max Power Save Mode
Power Capping	Clip performance based on VRM currents
Turbo	Increase power until thermal sensor or current sensor detect overload situation
Acoustic Optimization	Detect system temperatures and find minimum fan speed to meet max temp requirement
Workload Exploitation/Fix performance	Traditional mode we ship with extra capability to exploit workload power offset from Max power

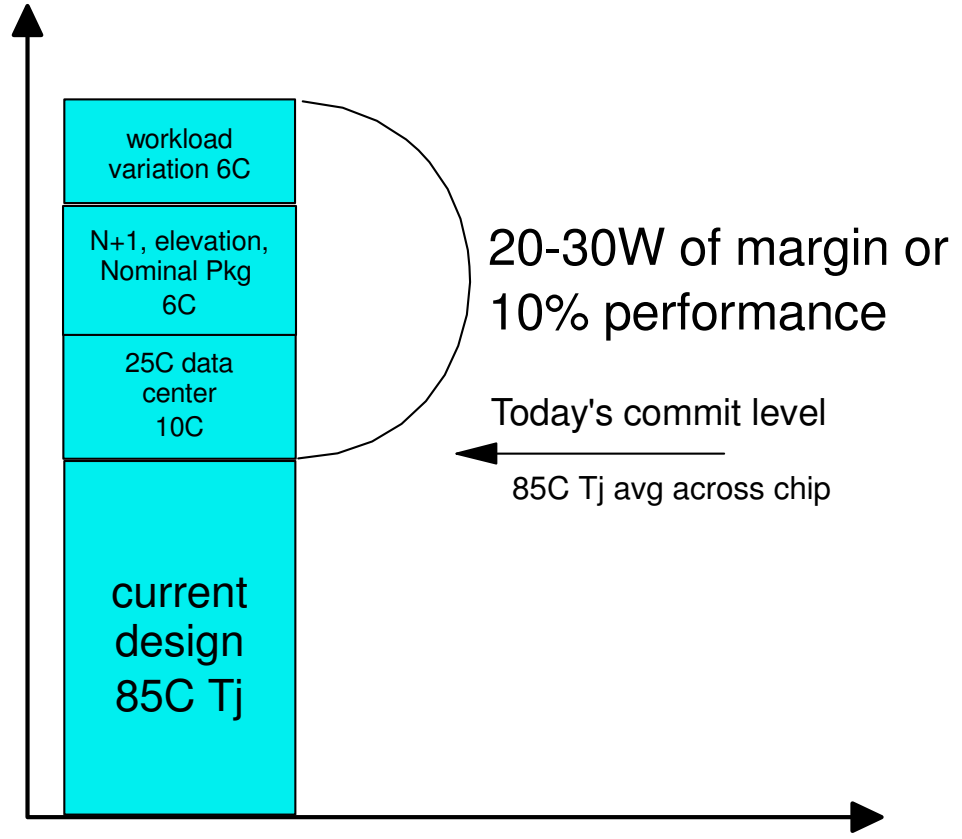
Energy management policies designed to enable clients to optimize utilization of their most precious resources

Vision | Enhance Performance (Turbo mode)

- **Create an "autonomic" system model that enables the computer to utilize all margin in the installed environment to optimize a customer valued parameter**
  - **- performance**
  - **- acoustics**

  Exploiting
  - all cooling capabilities
  - workload variation
  - environment offer
  - system manufacturing tolerance

- **The “autonomic” mechanisms of the new design still enable the traditional performance and risk trade-offs of traditional thermal management designs**



All statements regarding IBM future directions and intent are subject to change or withdrawal without notice and represent goals and objectives only. Any reliance on these Statements of General Direction is at the relying party's sole risk and will not create liability or obligation for IBM.

Special Notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.

Features and functions described in this document may vary by release level of the operating system. Contact your IBM or Business Partner representative for information on specific support limitations.

When referring to storage capacity, 1TB equals total GB divided by 1,000; accessible capacity may be less.

Revised September 26, 2006

Special Notices (Cont.)

The following terms are registered trademarks of International Business Machines Corporation in the United States and/or other countries: AIX, AIX/L, AIX/L(logo), alphaWorks, AS/400, BladeCenter, Blue Gene, Blue Lightning, C Set++, CICS, CICS/6000, ClusterProven, CT/2, DataHub, DataJoiner, DB2, DEEP BLUE, developerWorks, DirectTalk, Domino, DYNIX, DYNIX/ptx, e business(logo), e(logo)business, e(logo)server, Enterprise Storage Server, ESCON, FlashCopy, GDDM, i5/OS, IBM, IBM(logo), ibm.com, IBM Business Partner (logo), Informix, IntelliStation, IQ-Link, LANStreamer, LoadLeveler, Lotus, Lotus Notes, Lotusphere, Magstar, MediaStreamer, Micro Channel, MQSeries, Net.Data, Netfinity, NetView, Network Station, Notes, NUMA-Q, Operating System/2, Operating System/400, OS/2, OS/390, OS/400, Parallel Sysplex, PartnerLink, PartnerWorld, Passport Advantage, POWERparallel, Power PC 603, Power PC 604, PowerPC, PowerPC(logo), Predictive Failure Analysis, pSeries, PTX, ptx/ADMIN, RETAIN, RISC System/6000, RS/6000, RT Personal Computer, S/390, Scalable POWERparallel Systems, SecureWay, Sequent, ServerProven, SpaceBall, System/390, The Engines of e-business, THINK, Tivoli, Tivoli(logo), Tivoli Management Environment, Tivoli Ready(logo), TME, TotalStorage, TURBOWAYS, VisualAge, WebSphere, xSeries, z/OS, zSeries.

The following terms are trademarks of International Business Machines Corporation in the United States and/or other countries: Advanced Micro-Partitioning, AIX 5L, AIX PVMe, AS/400e, Chiphopper, Chipkill, Cloudscape, DB2 OLAP Server, DB2 Universal Database, DFDSM, DFSORT, DS4000, DS6000, DS8000, e-business(logo), e-business on demand, EnergyScale, eServer, Express Middleware, Express Portfolio, Express Servers, Express Servers and Storage, GigaProcessor, HACMP, HACMP/6000, IBM TotalStorage Proven, IBMLink, IMS, Intelligent Miner, iSeries, Micro-Partitioning, NUMACenter, On Demand Business logo, OpenPower, POWER, PowerExecutive, Power Architecture, Power Everywhere, Power Family, Power PC, PowerPC Architecture, PowerPC 603, PowerPC 603e, PowerPC 604, PowerPC 750, POWER2, POWER2 Architecture, POWER3, POWER4, POWER4+, POWER5, POWER5+, POWER6, POWER6+, Redbooks, Sequent (logo), SequentLINK, Server Advantage, ServeRAID, Service Director, SmoothStart, SP, System i, System i5, System p, System p5, System Storage, System z, System z9, S/390 Parallel Enterprise Server, Tivoli Enterprise, TME 10, TotalStorage Proven, Ultramedia, VideoCharger, Virtualization Engine, Visualization Data Explorer, X-Architecture, z/Architecture, z/9.

A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml.
UNIX is a registered trademark of The Open Group in the United States, other countries or both.
Linux is a registered trademark of Linus Torvalds in the United States, other countries or both.
Microsoft, Windows, Windows NT and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries or both.
Intel, Itanium, Pentium and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States, other countries or both.
AMD Opteron is a trademark of Advanced Micro Devices, Inc.
Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries or both.
TPC-C and TPC-H are trademarks of the Transaction Performance Processing Council (TPPC).
SPECint, SPECfp, SPECjbb, SPECweb, SPECjAppServer, SPEC OMP, SPECviewperf, SPECapc, SPEChpc, SPECjvm, SPECmail, SPECimap and SPECsfs are trademarks of the Standard Performance Evaluation Corp (SPEC).
NetBench is a registered trademark of Ziff Davis Media in the United States, other countries or both.
AltiVec is a trademark of Freescale Semiconductor, Inc.
Cell Broadband Engine is a trademark of Sony Computer Entertainment Inc.
Other company, product and service names may be trademarks or service marks of others.

Revised September 28, 2006