Dell Energy Smart Architecture (DESA) for 11G Rack and Tower Servers

By

John Jenne
Vijay Nijhawan
Robert Hormuth

Table of Contents

Tables

Figures

1 Introduction

Power and cooling in the Enterprise has become a top pain point for many data centers. Computation needs grow at an ever-increasing rate. Server OEMs integrate new component technologies that increase CPU cores, memory and storage capacities, network connections, and more to allow computation to keep up with demand. These technologies create significant power and cooling challenges.

Many customers face one or two primary challenges:

- Insufficient power to add new IT infrastructure
- Insufficient cooling to add new IT infrastructure

Dynamically balancing server performance to a given workload is an integral part of simplifying IT and easing the power and cooling challenge. Dell's Energy Smart Architecture (DESA) takes a new approach to dynamically manage system performance, thermals, and power at the platform level, which enables Dell customers to compute more while consuming less.

DESA's core tenets, built using Dell Energy Smart Technologies (DEST), are designed for efficiency, measurement, control, and reporting. Table 1 highlights the DESA core tenets and some of the underlying Dell Energy Smart Technologies.

Table 1. Dell Energy Smart Architecture (DESA) Core Tenets and Technologies

	Design	Measure	Control	Report
Dell Energy Smart Technologies (DEST)	Power Supplies • High efficiency • Right sized Voltage regulators • High efficiency • Switching regulators • Phase shedding o Processor VRs o Memory VRs Board design • Low loss PCBs • Low loss connectors Chassis • High airflow BIOS • Processor P-states • Processor C-states • Processor T-states • DDR3 frequencies • DDR3 CKE • FSB power management	Temperatures • Inlet (ambient) • Processor • Chipset • DIMMs • Power supplies Power • System-level • Metrics o Amperage o BTUs o Voltage o Watts Performance • Processor utilization • Memory throughput • I/O throughput	BIOS • Active Power Controller o OS-independent processor P-state manager • OS Enabled Processor Power Management • Power management profiles • Processor core disables • Host LOM port disables • Unused component disables • Unused controller disables	Power • Averages o Overall o 1 minute o 1 hour o 1 day • Peaks • Real-time

	Design	Measure	Control	Report
Dell Energy Smart Technologies (DEST)	Thermal • Optimized heat sinks • Independent fan control • Fan de-population • PWM fans • DIMM heat spreaders • DIMMs with temperature sensors • DIMM closed loop thermal monitoring Components • Low voltage processors • Low voltage DIMMs • 2Gb DRAMs (DIMMs) • UDIMMs • QR x8 DIMMs • Green HDDs • SSD HDDs • Low power fans		Firmware • Non-linear fan curve • Adaptive thermal algorithm • Closed loop thermal monitoring • Power capping • Management LOM port disables • IDLE memory control • Static memory phase shedding • Power inventory and budget • AC staggered power on Open Manage • Remote power ON/OFF	Mechanisms • Tables • Graphing Alarms/alerts

2 Implementation

DESA is primarily an out-of-band implementation that provides the ability for real-time monitoring and control along with platform tuning. DESA uses its extensive characterization of the platform to optimize platform level tradeoffs. DESA's out-of-band implementation enables the use of advanced silicon features and enhances technologies such as Intel® Dynamic Power Technology. Intel Dynamic Power Technology is a collection of silicon features that aide in power management. DESA takes these base features and adds Dell enhancements. Figure 1 represents a high-level view of the DESA depicting set points, control algorithms, control, and measurement points with reporting.

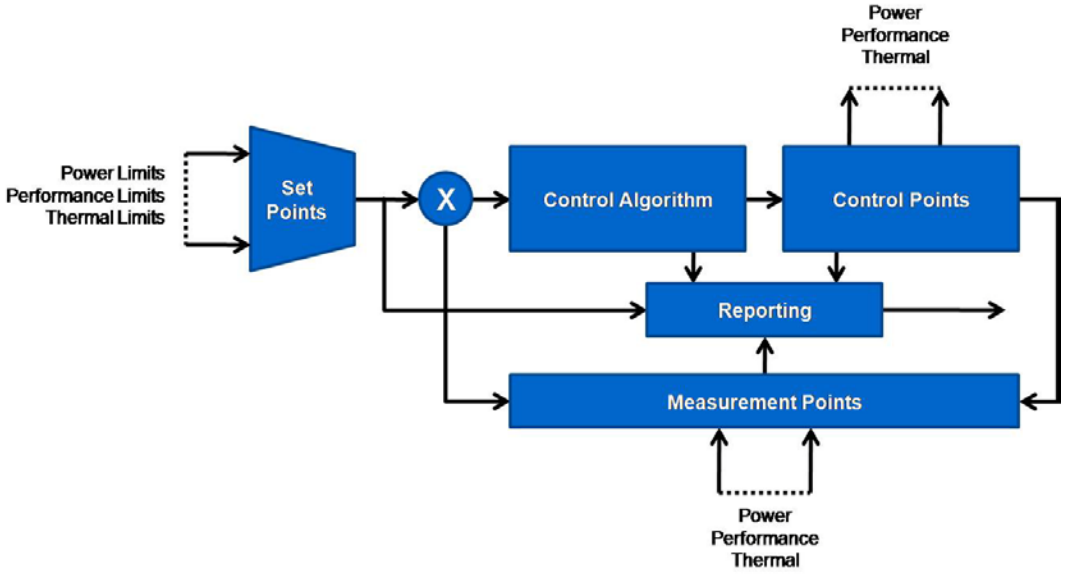

Figure 1. Dell Energy Smart Architecture Control Process

2.1 Core Tenets

2.1.1 Design

DESA optimizes performance and power through a system-wide approach using efficient design principles across electrical, mechanical, and thermal. Combining this with intelligent component selection sets the bar for power efficiency levels.

Electrical contributions to power efficiency include highly optimized power supply designs, highly efficient switching voltage regulator designs that include static and dynamic phase shedding capabilities, and low loss printed circuit board (PCB) layouts and connectors.

Mechanically, DEST includes technologies such as optimized venting that provides significant air flow through the platform, thus minimizing fan speeds.

Thermal technologies include a high number of thermal zones, low power fans, and adaptive fan speed algorithms.

Other DESA design technologies include configuration options for low-power components such as processors, memory, and hard drives.

2.1.2 Measurement

Real timer measurements are made of power, thermal, and performance fed into the DESA control algorithms. These measures help the DESA make decisions to optimize performance per watt of the overall system. Temperature measurements include ambient, processor, memory PSU, and other components. System and various component power use, with component performance levels and fan speed, are also monitored. In addition to the many monitoring capabilities, DEST also provides industry leading monitoring accuracy.

2.1.3 Control

DESA incorporates system firmware running on a high-performance base Integrated Dell Remote Access Controller (iDRAC) embedded in the system to provide an intelligent, centralized control mechanism. The iDRAC provides a real-time component that uses DESA's measurement technologies to analyze component and platform tradeoffs for determining performance per watt optimizations. For example, the DESA thermal algorithm may limit memory throughput instead of allowing fan speeds to reach upper power regions of the fan response curve.

DESA supports OS-based processor power management as well as Dell Active Power Controller (DAPC). DAPC is an OS-agnostic processor power manager that maximizes performance per watt by monitoring processor use and managing processor performance states to achieve the best performance per watt. DAPC provides consistent processor power management across all operating systems, with additional benefits such as Power Management from the moment the Server is powered on.

Additional benefits of DESA include support for enhanced applications such as platform or data center level-power capping that enables an end-user to define power limits to aid power management in the data center. By applying greater intelligence to the systems overall power use, DESA drives new levels of power efficiency.

System power capping enables enhanced power provisioning. Through DEST, power capping is enabled to enable you to specify a power threshold between the maximum and minimum potential power consumption. The user-specified power threshold is compared to the measured power use and the system level controls maintain server power consumption below this power threshold.

2.1.4 Reporting

DESA provides various reporting metrics (including power, utilization, and power limits) through in-band and out-of-band management interfaces. Reporting is provided through the OMSA GUI and CLI, IT Assistant, iDRAC GUI, racadm, and the IPMI Tool. Real-time processor performance state information is also provided to compatible operating systems and hypervisors to provide a tighter coupling between the platform and the OS. The OS, in turn, can make smarter decisions in terms of scheduling and SLA requirements by using platform performance, power, and thermal constraints in its algorithms.

2.2 Proven Results

DESA provides significant improvements to power consumption and performance per watt. Figure 2 demonstrates DESA's direct impact on Dell® PowerEdge™ R710 performance per watt improvements over its predecessor, the Dell PowerEdge 2950 III.

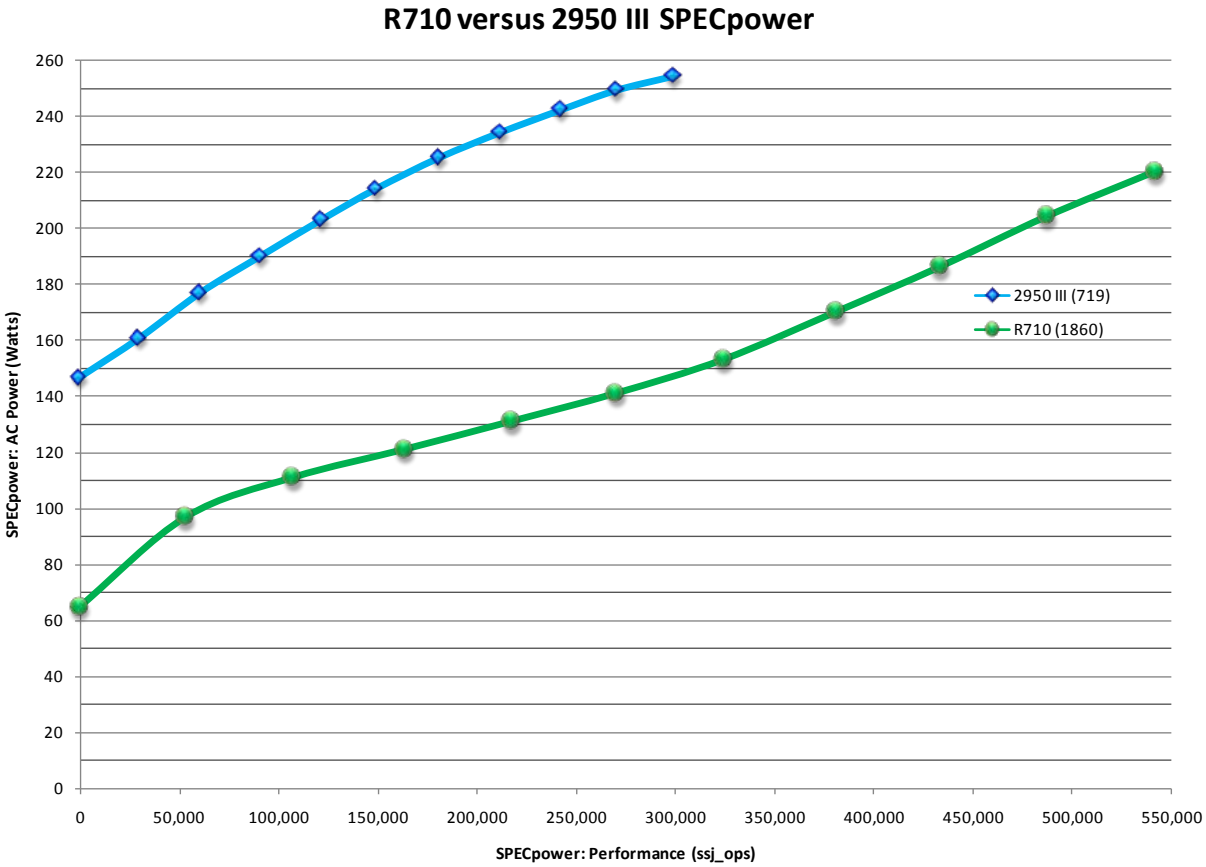

Figure 2. Dell Energy Smart Architecture Improvements to Performance per Watt

Results referenced can be found at http://www.spec.org/power_ssj2008/results/.

3 Design

3.1 Power Supplies (AC/DC)

Over-provisioned, inefficient server power supply units (PSUs) that draw more power than necessary are common in enterprise data centers. Dell PSUs are engineered for high efficiency and optimally sized for typical environments. Design enhancements enable these PSUs to provide higher efficiency than the PSUs of comparable previous-generation Dell PowerEdge servers.

3.1.1 High-Efficiency Power Supplies

The 11G server platforms' PSUs have significant efficiency improvements over their 9G and 10G counterparts. The 11G PSUs meet the stringent requirements of various industry certifications such as Climate Savers, 80 Plus, and Energy Star. Energy efficient PSUs enable Dell customers to reduce power and cooling cost and complexity.

Table 2. Power Supply Efficiency Improvements

Platform	PSU		PSU Efficiency				Climate Savers	
	Model #	Output Power	10%	20%	50%	100%	Silver	Gold
10G - PE 2950	7001452-J000 NPS750BB-1 A	750 W	78.8% 79.4%	87.0% 86.8%	91.0% 90.9%	89.3% 90.7%	✓	–
11G - PE R710/T610	NPS-885AB A A870P-00	870 W	77.7% 81.0%	87.1% 88.0%	91.5% 91.3%	91.1% 90.9%	✓	–
	C570A-SO A570P-00	570 W	81.2% 81.5%	88.9% 88.1%	92.5% 92.0%	91.6% 91.5%	–	✓
10G - PE 1950	DPS-670CB-1 A	670 W	74.3%	85.4%	90.7%	91.4%	✓	–
11G - PE R610	A717P-00 DPS-764AB A	717 W	77.2% 78.1%	85.6% 88.3%	90.6% 92.1%	90.0% 92.0%	✓	✓
	A502P-00 C502A-SO	502 W	81.6% 79.9%	88.4% 88.4%	92.1% 92.5%	91.5% 92.0%	–	✓ ✓

3.1.2 Right-Sized Power Supplies

Some 11G platforms offer a right-sized Energy Smart (ES) PSU for typical server configurations. The ES PSU provides some key advantages over the High Output (HO) PSU, which is sized for a maximum server configuration.

3.1.2.1 Improved Efficiency

With a lower overall wattage, the ES PSU's efficiency curve moves "to the left" compared to the HO PSU's efficiency curve. At lower system loads, the ES PSU operates at a higher efficiency than the HO PSU. The benefits are even greater for redundant PSU configurations where the

load is shared across PSUs. The benefits are also greater for light to medium system configurations that regularly operate at the lower ends of PSU efficiency curves.

Figure 3 shows the power consumption differences between ES and HO PSUs at multiple workload levels. This example is based on benchmark measurements for an R710 with dual 95 W processors, 6 x 8 GB 1066 MHz RDIMMs, and one 10 GbE card.

For light loads, PSUs operate at less than optimal efficiencies, whereas as the highest efficiency is achieved at medium to heavy loads.

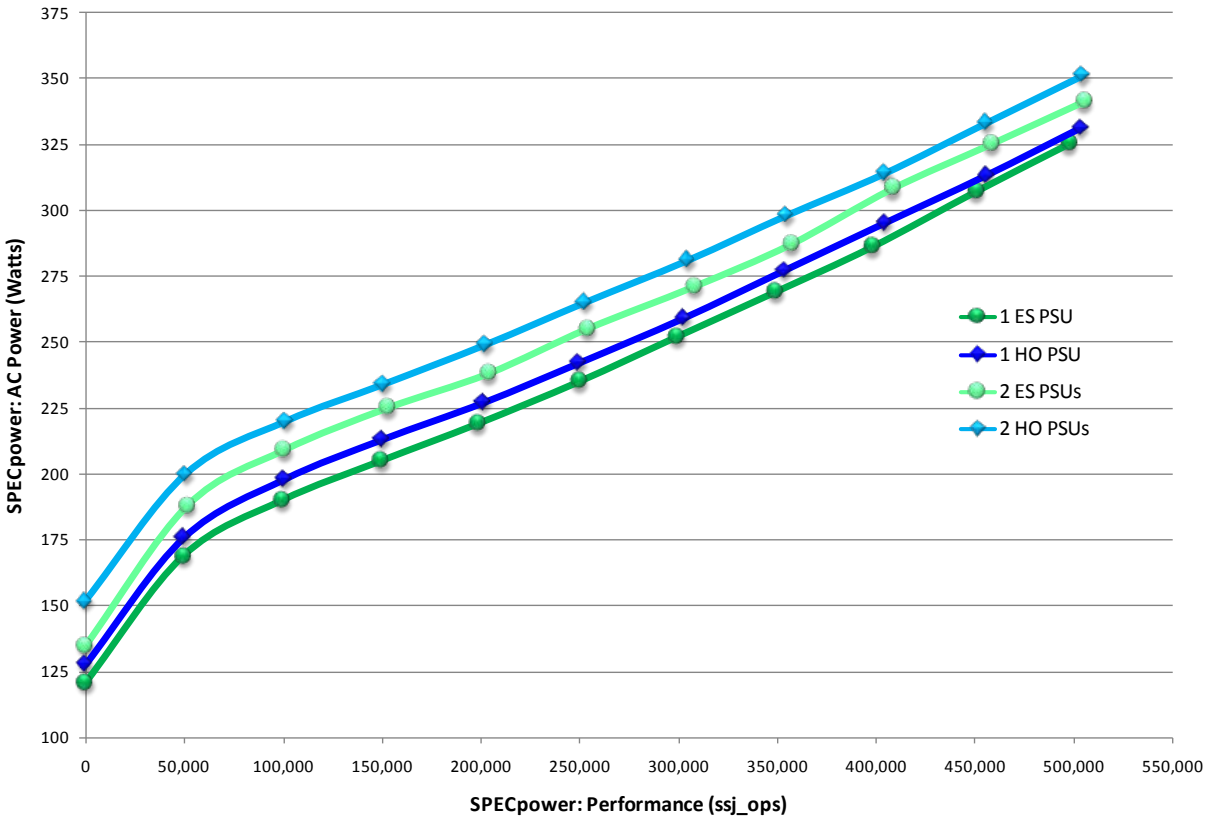

Figure 3. R710 PSU Efficiency Curves – 230VAC

3.1.2.2 Smaller Power Footprint

The Dell ES PSU meets the power footprint of an equivalent system from five years ago. The smaller power footprint enables customers to upgrade existing racks that are limited by power and cooling constraints. While DESA provides software-based power-capping features, many customers size power delivery, PSU, and cooling solutions based on the PSU regulatory label.

Due to the cost and power benefits of the ES PSU, 11G platforms that support ES PSUs default to the ES PSU in the online ordering tool.

3.2 Voltage Regulators (DC/DC)

While energy efficiency standards are focused on AC/DC efficiency of PSUs, that is only part of the overall power delivery solution for servers. The efficiency of the DC/DC design on the server printed circuit boards (PCBs) is equally important, especially because every watt on the DC/DC side is scaled by the AC/DC efficiency of the PSU. The DC/DC design uses voltage regulators

(VR) to generate the various voltage levels required by devices. The efficiency of the voltage conversion can have significant impacts to the server's power consumption.

The 11G platforms' design improvements significantly reduce power overhead associated with converting voltage levels. Table 3 shows voltage regulator power loss improvements over generations of Dell PowerEdge servers.

Table 3. Voltage Regulator Efficiency Improvements

Voltage Regulator Power Loss Reduction: 11G R710 vs. 9G 2950			
20% Load	**50% Load**	**80% Load**	**100% Load**
48%	43%	38%	38%

Note: This table assumes all voltage regulators are operating at the same load level at the same time.

3.2.1 High Efficiency Voltage Regulator Designs

Consider factors such as efficiency, cost, performance, and density when architecting a voltage regulator design. For most 11G platforms discussed in this document, efficiency was, by far, the highest priority and the key factor for selecting components and making design tradeoffs.

Linear regulators are popular for low-powered devices because they are easy to design with and are low cost. Linear regulators are extremely inefficient, however. As part of the server voltage regulator design every linear regulator was carefully scrutinized. If an opportunity to reduce power loss was found, a highly efficient switching regulator was considered based on cost per watt saved. Switching regulators are more complex circuits to design and higher cost, but the efficiency improvements are significant. The 11G platforms were designed with switching regulators on main voltage rails where previous generation platforms were designed with linear regulators.

Besides linear versus switching regulator tradeoffs, the voltage regulator design carefully selects voltage regulator controllers, MOSFETs, and other components to maximize the efficiency of the design. Optimized layout and enhanced PCB also helped reduce VR and PCB distribution losses.

3.2.2 Voltage Regulator Phase Shedding

The 11G platforms implemented multi-phase voltage regulator circuits for high-power devices such as processor and memory. Depending on the processor or memory configuration or power state, not all VR phases may be needed. A phase is a power stage in the voltage regulator. The higher the load requirement, the more phases are required to meet the load requirement. Reducing the number of active phases for lighter loads improves the VR efficiency curves and reduces the power loss of the VR solution.

3.2.2.1 Processor

The Intel Voltage Regulator Module (VRM) and Enterprise Voltage Regulator-Down (EVRD) 11.1 specification defines a Power State Indicator (PSI#) signal. The PSI# signal tells the voltage regulator when the processor has entered a lower power state. The 11G server VR solution uses PSI# to dynamically enable/disable processor VR phases.

The 11G platforms reduce the number of active VR phases to 1 when the PSI# signal is asserted.

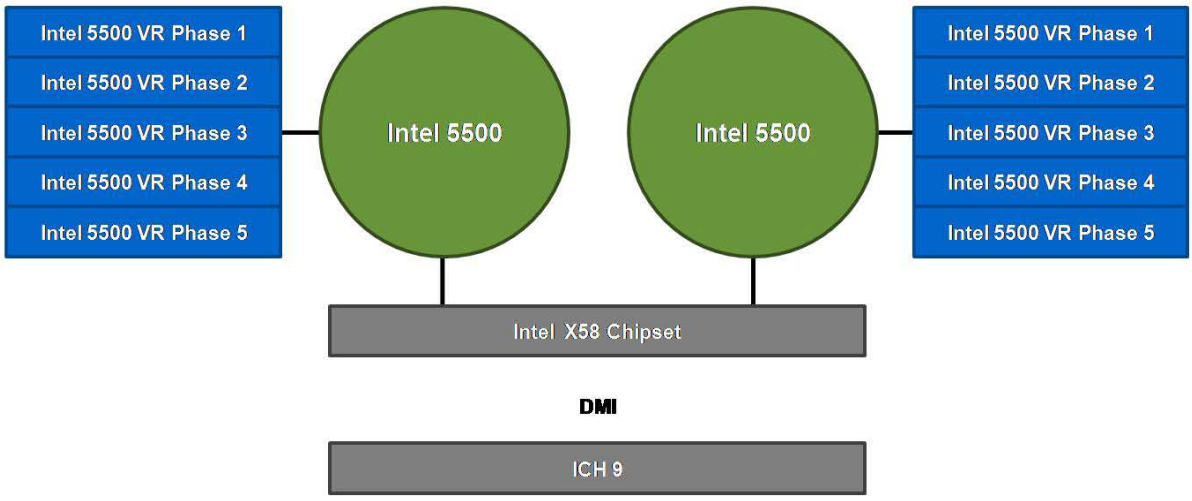

Figure 4. Processor VR Dynamic Phase Shedding Solution

3.2.2.2 Memory

Unlike the processor VR solution, the memory VR solution does not receive a low-power indicator from the memory controller. Therefore, the memory VR solution does not dynamically enable/disable phases based on load.

The memory VR solution supports the maximum DDR3 memory configuration. There is a wide range of memory configuration resulting in a wide range in power requirements. The 11G server platforms inventory the memory configuration at boot time and calculate the power requirements. The memory VR phases are then disabled to meet the memory power requirements.

The power inventory is described in detail in a later section, but the following is a short description with regards to memory VR static phase shedding. During the boot sequence, BIOS gathers an inventory of the memory configuration. This inventory is then passed to the iDRAC. The iDRAC owns the look-up tables to determine the power requirements for the memory configuration. The iDRAC uses the look-up results to determine if the memory VR solution can shed a phase. If a phase can be shed, the iDRAC disables the phase.

For lower power memory configurations, the memory VR solution enables only one phase.

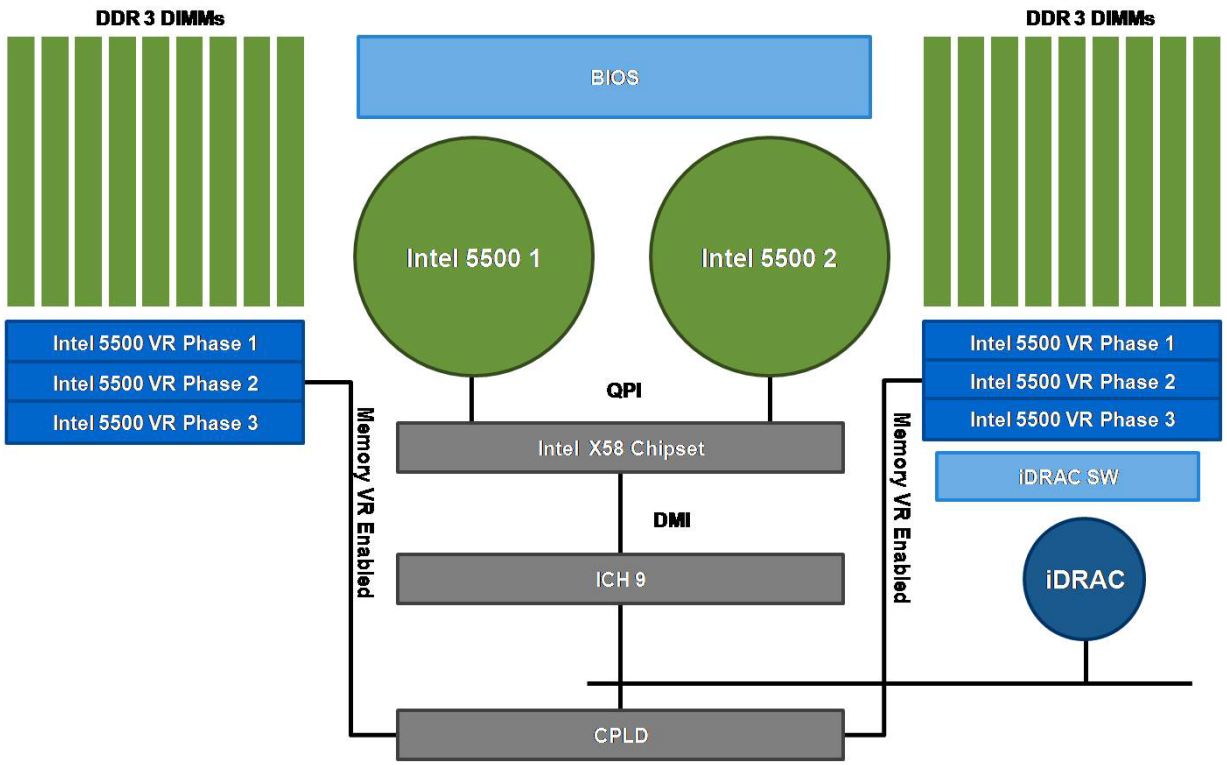

Figure 5. Memory VR Static Phase Shedding

3.2.3 DDR3L

JEDEC has finalized the DDR3L specification. The design goal is to lower the DDR3 1.5 V voltage down to 1.35 V. The 11G server platforms have designed flexible memory VR architecture to take advantage of DDR3L if it is supported by future processor releases.

The 11G server DDR3 VR design includes independent DDR3 VRs for each processor socket. For Intel based servers, the DDR3 VRs are Intel 11.1 compliant and support changing of the output voltage dynamically. The output voltage for Intel 11.1 compliant VRs is selected by voltage identification (VID) inputs. The VID lines for the processor VRs are directly controlled by the processor. The 11G platforms provide system level control of the DDR3 VR outputs by routing MEM_VID to general purpose input/output (GPIO) accessible by platform software.

While DDR3L specification targets 1.35 V, 11G server platforms can support down to 1.2 V should the industry adopt another voltage below 1.35 V.

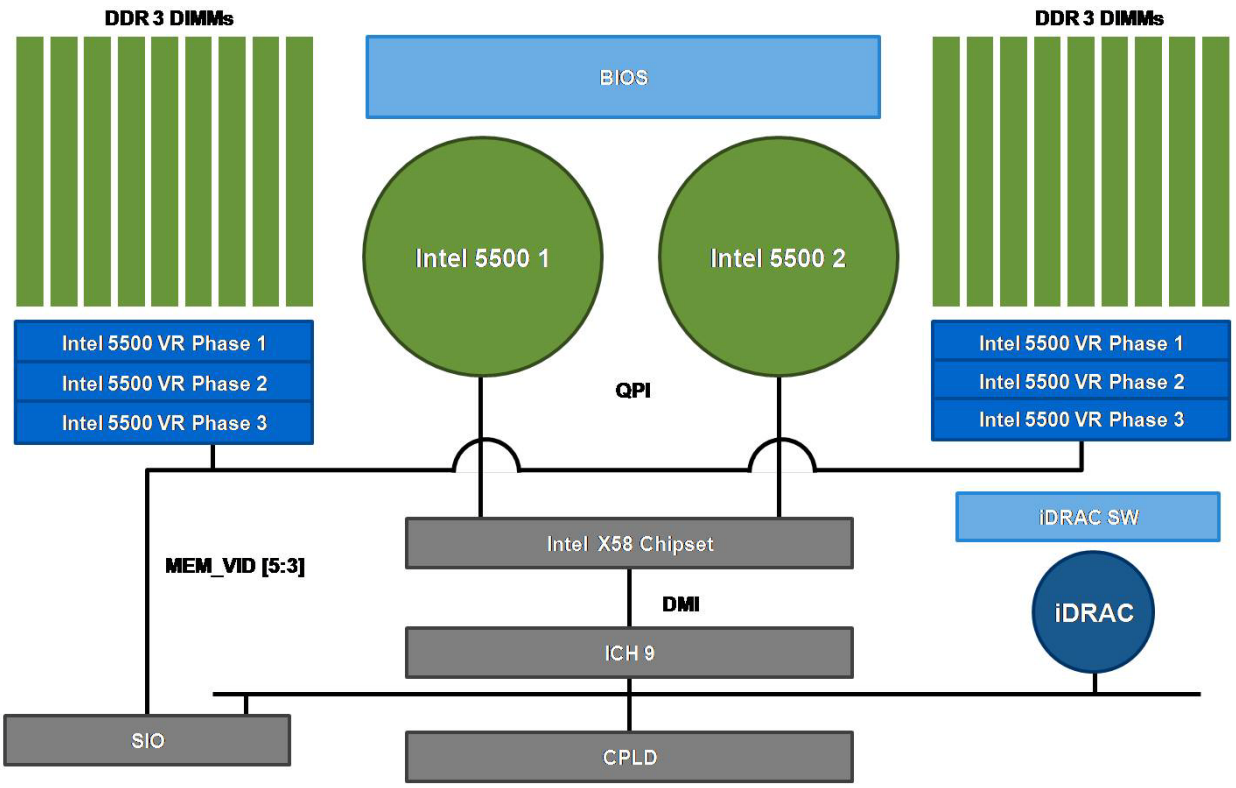

Figure 6. DDR3L Solution

3.2.4 PSU 12 Vaux

The 11G platforms changed the PSU auxiliary output voltage from 3.3 Vaux to 12 Vaux. The 12 Vaux allowed for VR efficiency optimizations for the auxiliary voltage rails on the PCB.

3.3 Board Design

In addition to the efficient design of the voltage regulators, focus was also put on the PCB design of 11G platforms. Improvements that reduced power loss include:

- Use of 2-ounce copper (Increasing the thickness of the copper improves its conduction, reducing power loss when compared to boards using thinner copper)
- Optimization of power routes
- Placement of connectors to reduce board routing and cable lengths

3.4 Thermal

The thermal solution in 11G server platforms has made a vast reduction in system power consumed by fans. Figure 7 compares fan power consumption of similarly configured PowerEdge servers:

- PowerEdge R710: 2x80W processors, 12GB DDR3, 1 HDD
- PowerEdge 2950: 2x80W processors, 16GB FBDIMM, 1 HDD

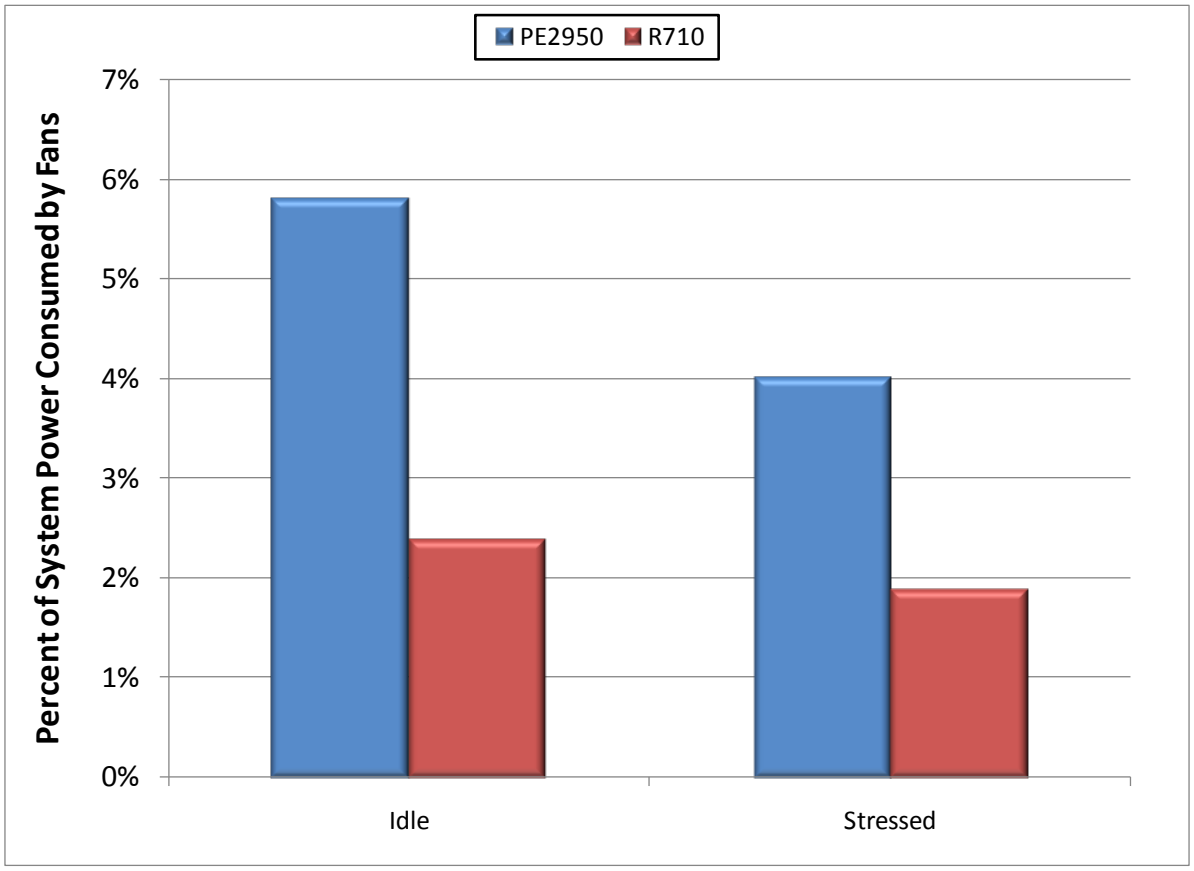

Figure 7. Generation-over-Generation 2U Fan Power Reduction

The fan power reduction is a result of many design enhancements covered in the following subsections.

See the Thermal Design of the Dell PowerEdge T610, R610, and R710 Servers white paper for more information.

3.4.1 Chassis Venting and Airflow

System air flow impedance was reduced through multiple mechanical design techniques. 3D venting provided additional venting area in the rear of the chassis and vented closeouts and an optimized backplane design also contributed to reducing air flow impedance.

Figure 8. 3D Venting Example on R610

Air bypass was also improved by sealing the fans to the chassis and tightly ducting the airflow.

3.4.2 Fan Zones

Fan zones refer to the section of the server that the fan or fans cool. Previous server generations often had multiple fans grouped together in a fan zone. Fans grouped together would all run at the same speed. Beginning with the 11G server platforms, there is only one fan per fan zone giving the system independent control of each fan connection. Independent fan control enables fine-tune adjustments by the thermal algorithm to only run fans as fast as absolutely needed, which conserves fan power.

3.4.3 Thermal Sensors

The 11G server platform has thermal sensor coverage of all major components within the system. By accurately monitoring key components, the thermal algorithm can optimize fan speeds. Table 4 highlights some of the key thermal sensors.

Table 4. Key Thermal Sensors

Sensor		Description
Type	**Name**	
Temperature	Input air temperature	Ambient temperature of air entering system
	Processor die temperature	Die temperature of each processor installed
	DIMM temperature	Temperature of each DIMM installed
	IOH temperature	Temperature for I/O hub
	Planar temperature	Temperature on planar
	PSU temperature	Temperature of each power supply installed
System Feedback	Fan tachometer	Tachometer for each fan installed
	Chassis intrusion	Detects chassis cover removal

The 11G server platforms optimized the location of temperature sensors to improve accuracy. As a result, the system fans run slower reducing fan power consumption.

3.4.4 Fans

3.4.4.1 Power Efficiency

Dell partnered with fan technology leaders to drive aggressive fan efficiency goals. As a result, 11G platforms have highly efficient fans that reduce fan power consumption. For example, the R710 60 mm fan reduced the maximum power consumption by 36% over its predecessor, the PE2950.

3.4.4.2 Fan De-Population

Some 11G platforms support fan de-population based on processor configuration. Removing a fan when the number of processors installed is less than the number of available processor sockets reduces fan power consumption and system cost. Table 6 provides an example of fan population options for three popular 11G server platforms.

Table 5. Example Fan Configurations

System	1 Processor	2 Processors	Redundant
R710	4	5	N + 1
R610	5	6	N + 1
T610	2	2	2N

3.4.4.3 Fan-Off Function

A fan-off function is supported by fans in 1U server platforms. Dell's AutoCool feature, which provides system and PSU cooling when in standby, is the main use case for the fan-off function. The 1U PSUs do not have an integrated fan and rely on system fans for cooling. The AutoCool algorithm, implemented in iDRAC firmware, only turns on the AutoCool fan in response to a PSU thermal event. Under normal conditions the fan is turned off when the platform is off or in standby reducing power consumption and acoustics.

3.4.5 Heat Sinks/Spreaders

3.4.5.1 Processor Heat Sink

The 11G servers were designed with custom heat sink solutions that reduce fan speeds and balance airflow in the chassis. Improvements were made to processor heat sink conduction heat transfer by using high-performance thermal interface materials

3.4.5.2 RDIMM Full DIMM Heat Spreader

The 11G server platforms support FDHS on all RDIMMs 4 GB and above to improve cooling, which also reduces fan power and noise level.

3.4.5.3 IOH Heat Sink

Custom-designed IOH heat sinks are used to reduce airflow needs for IOH cooling.

3.5 Low Power Components

DESA supports optional low-power components to enable Dell 11G server customers to customize their configuration for additional power savings or performance per watt optimizations. The follow sections highlight some of the low-power components supported on 11G server platforms.

3.5.1 Processor

Processor vendors offer lower voltage SKUs that provide the same performance and functionality at lower power levels. For example, in Intel's X5500 series, the L5530 is a 60 W variant of the 80 W E5530. The 11G server platforms support and validate low voltage processors.

3.5.2 Memory

3.5.2.1 DIMM Suppliers

In addition to cost, reliability, and business history, DIMM suppliers are evaluated with an increased focus on power consumption.

3.5.2.2 DIMM Configurations

Many considerations went into the 11G server supported memory matrix including cost, capacity points, and performance. Power consumption was also a major factor that led to the definition of power-optimized DIMM configurations.

3.5.2.3 Quad-Rank DIMMs

The 11G platforms support quad-rank DIMMs that can provide a significant cost and power savings over dual-rank equivalents.

3.5.2.4 UDIMMs

In additional to registered DIMMs (RDIMMs), 11G platforms also support unbuffered DIMMs (UDIMMs). UDIMMs can save up to 1.25 W per DIMM. Technical limitations in the UDIMM architecture restrict capacity and DIMM count.

3.5.3 HDDs

The 11G server platforms support both green hard disk drives (HDDs) and solid state drives (SSDs). These technologies provide significant power reduction as highlighted in Table 6.

Table 6. Storage Power Savings

Drive	Operating Power Limit	Power Reduction
3.5" SAS 10K	25.2	–
3.5" SAS 10K (GREEN)	14.6	-42.06%
3.5" SAS 5.4K	17	–
3.5" SAS 5.4K (GREEN)	10	-41.18%
2.5” SATA HDD (5V only)	3.75	–
2.5” SATA SSD (5V only)	3.25	-13.33%

4 Measure

On 11G server platforms, iDRAC firmware collects temperature, fan speed, and power measurements every two seconds. Server management uses these measurements to adjust various controls accordingly.

4.1 Temperature

As highlighted in the previous section, multiple temperature sensors are monitored throughout the server. These include temperature sensors for processors, chipset, DIMMs, power supplies, and inlet (ambient air).

4.2 Fan Speed

As described in previous sections, all 11G server platforms support independent control of all system fans and fan tachometer signals are available for monitoring fan speed. iDRAC firmware collects the fan speed for all fans and rotors (1U servers have dual-rotor fans).

4.3 Power

The 11G power supplies (PSUs) support IPMM commands over a PMBus™ interface between the iDRAC (BMC) and the PSU(s).

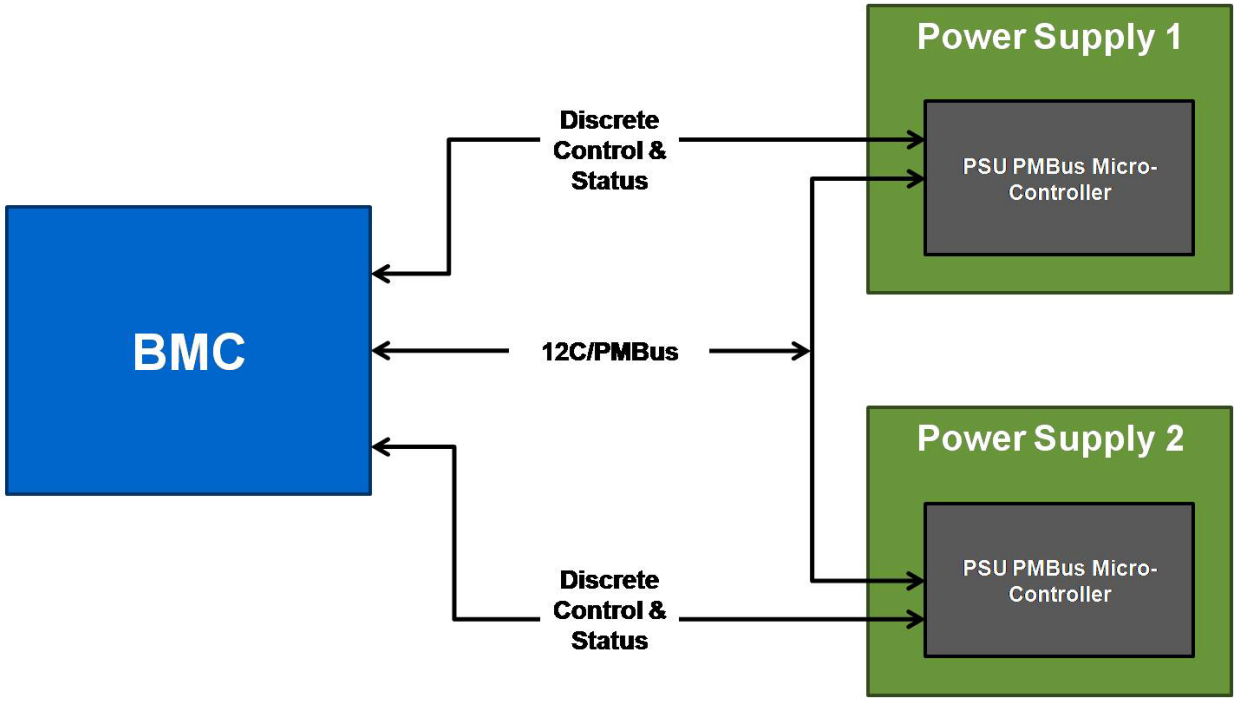

Figure 9. PSU PMBus Interface

4.3.1 System Power

System power is a measurement of AC power that the system's power supplies consume in one to five seconds. Each power supply is monitored for system power. The actual hardware readings are sampled multiple times over a fixed interval and then averaged before being reported. This results in accurate representations of actual power consumption and avoids the reporting of momentary spikes or dips.

4.3.2 System Current

To collect the system current draw an amperage sensor reading is collected from each manageable power supply installed in the system. The system current indicates the number of AC amperes the power supply draws from the circuit/grid into which the system is plugged. When there are multiple power supplies in the system, the power load is shared across all of them. The load across power supplies is near equal, but, most likely, not identical.

4.3.3 System Input Voltage

System input voltage is the AC voltage level. Power supply units support voltage input ranges of 90–264 VAC and 47–63 Hz.

5 Control

5.1 Processor and Memory Controller Power Reduction Features

The Advanced Configuration and Power Interface (ACPI) Specification defines processor power management states: C-states, P-states, and T-states. Processor vendors have unique implementations of these standard processor power management states. There are no dependencies between C-states, P-states, and T-states.

Processors supported on 11G server platforms provide significant improvements to power management capabilities. For example, Intel marketing collateral claims that the Xeon® 5500 series provides five times as many operating states, five times reduction in idle power, and five times faster transitions to and from low-power states compared to the first Intel quad-core server processor.

5.1.1 C-states

Processor C-states are power and thermal management states that can be implemented at the thread, core, and package levels. C-states are sleep states for the processor package and cores so no instruction execution occurs. Greater power savings come with deeper C-states and, therefore, increased latency to return to execution state. C-states are essential for reducing idle power consumption. The Intel® 5500 Chipset processor supports C3, Package C3, C6, and Package C6 states. When a user turns on the C states in BIOS setup, BIOS exposes an ACPI interface to OS which enables OS to exercise these states when it is idle.

The number of C-states supported may vary by processor vendor or processor SKU. The 11G platforms provide BIOS support for all C-states supported by the processor. C-state transitions are controlled by the operating system.

In addition to the C3 and C6 states, 11G servers also support the C1E state. The C1E state enables the CPU to transition to the lowest performance state when the OS thread enters a `cli hlt` or `monitor/mwait` loop. BIOS setup provides a setup option to turn C1E on and off.

Dell PowerEdge servers are optimized for performance per watt and set C states and the C1E state to **on** by default in BIOS setup. Customers focused on performance might choose to disable these options.

5.1.2 P-states

Unlike C-states, processor performance states (P-states) are execution states that provide different levels of power consumption and performance capability. Processor vendors have implemented P-states as processor frequency and voltage pairs. Processor power consumption decreases with the processor's operating frequency and voltage. P-states are essential for reducing power consumption when processors are not in high use.

The number of P-states supported might vary by processor vendor or processor SKU. The 11G platforms provide BIOS support for all P-states supported by the processor. P-state transitions are typically controlled by the operating system, but 11G platforms also support a feature called DAPC in which the platform controls processor P-states. Both forms of P-state control make performance versus power tradeoffs.

5.1.3 Memory Clock Gating (Clock Enable or CKE)

The 11G server platforms have intelligent memory controllers that support memory clock gating (CKE) to conserve power when the DIMMs are not accessed. DIMMs enter a low-power, self-refresh mode that significantly reduces power consumption. Previous memory controller hubs (MCHs) had this feature, and it is also used widely in the mobile segment. Memory Clock Gating in 11G Servers becomes functional when the processor enters the package C6 state.

5.2 Dynamic Power Management

5.2.1 Processor Power Management

5.2.1.1 OS Control

The 11G platforms continue supporting OS-based processor P-state management. OS-based processor P-state management is dependent on OS support. When enabled, the operating system determines processor use by analyzing the scheduled processes, threads, and more. If the OS determines that the processor is lightly used, it reduces the processor P-state. If the OS determines that the processor is heavily used, it increases the processor P-state. In this mode, BIOS exposes the possible P states to the OS using the processor power management device objects as described in ACPI 2.0b specification. At the end of POST, BIOS leaves the P-state of the processor at a maximum non-turbo mode and enables the OS to manage the P-states through the ACPI processor performance state objects.

5.2.1.2 Dell Active Power Controller

DAPC is a BIOS implementation of the processor P-state manager. DAPC replaces the OS implementation of the processor P-state manager. Dell's BIOS-based implementation provides significant advantages over the OS-based implementation. These advantages include:

- OS-agnostic
  - Consistent behavior and power savings regardless of OS
  - Support for processor power management even if not supported by OS
  - Provides processor power management for hypervisors or virtual machine monitor (VMMs)
- Additional power savings without sacrificing performance

Figure 10 demonstrates the significant power savings provided by DAPC.

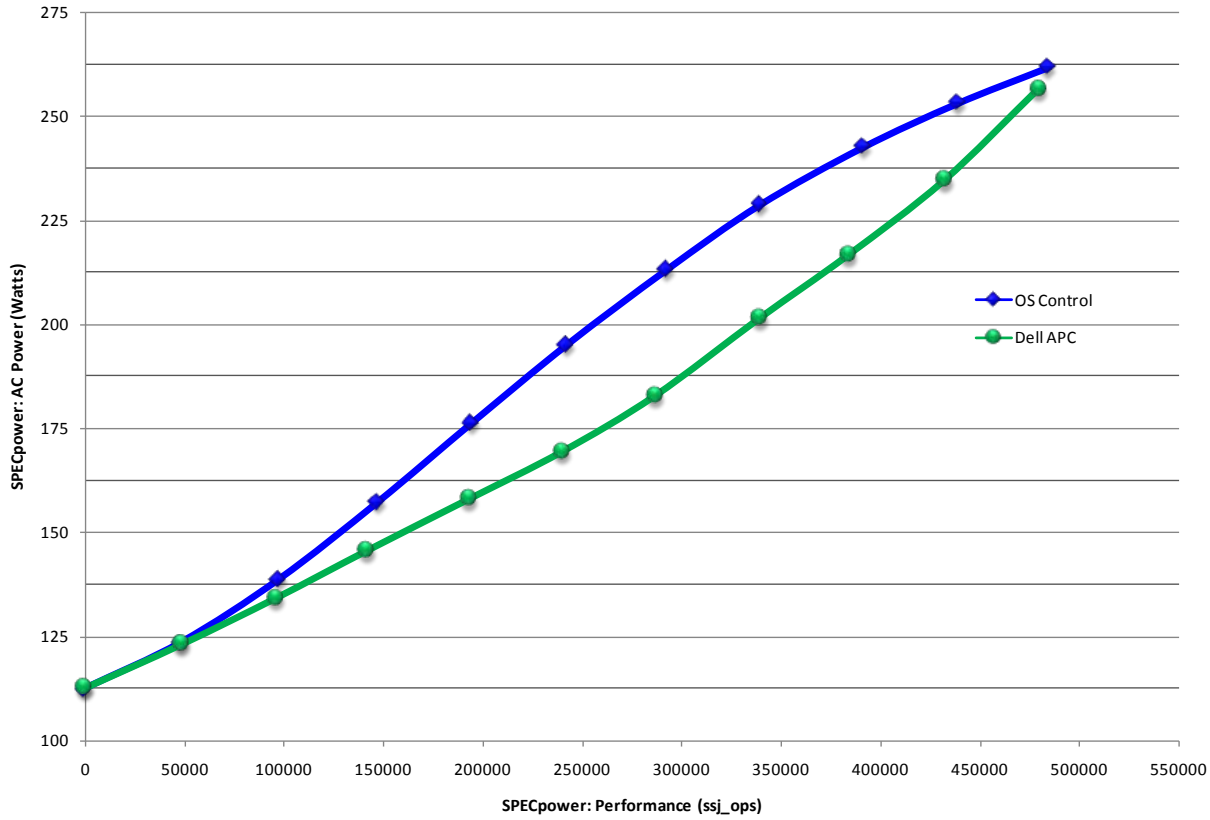

Figure 10. OS Control Versus Dell APC

5.2.1.2.1 Algorithm

In DAPC mode, P-State Management uses a periodic SMI to monitor processor use and make the decision to change P-state of the processor. The Enhanced Intel SpeedStep® Technology (EIST) feature enables software to place processors in various performance states (P-states). P-state changes cause the processor to operate in various core-speed ratio and voltage levels. BIOS has used various architectural registers of the processor in the APC algorithm

DAPC uses processor thread use as a primary indicator for power management. While use is a good indicator of the current load in the system, it doesn't capture the efficiency with which a processor is executing a workload. For many workloads, a processor can stall while waiting for the completion of a processor-independent operation such as memory and I/O accesses, without providing any performance benefit by running at higher frequency.

BIOS provides a means to control power consumption by controlling core frequency. It decreases core frequency during periods of low workload and raises it during high loads. Core frequency is, in the simplest terms, in direct proportion to performance. For any given set of tasks performance is in inverse proportion to system use, so that as performance increases use decreases. Because core frequency and performance are in direct proportion, core frequency and use are in inverse proportion. Therefore, core ratio can be substituted for core frequency.

The periodic System Management Interrupt (SMI) cycle time provides the base frequency for operating efficiency sampling. It also provides the base for weighted averaging and decay. The design goal is to choose a cycle to minimize sampling intrusion and maximize detection of short-term loading changes.

Run-Time Processing of the Dell APC:

- Count Calculation
- Current Use Calculation
- Target Ratio Calculation
- Target Ratio Update
- Exit

The detailed description of each section is beyond the scope of this document. Even though P-state transitions are efficient on CPU silicon today, some cost and care must be taken to ensure that changes in frequency and voltage do not lead to excessive power/performance overhead (primarily System Management Mode (SMM) overhead and small Phased Locked Loop (PLL) lock time).

5.2.1.2.2 Sideband and OS/ACPI Reporting

In operating systems and hypervisors there is a need to provide guaranteed performance for virtual machines (VMs). Systems in which the performance/watt optimization mode is enabled (using an OS-based driver) rely on knowing the precise performance state (P-state) of the CPU. Current ACPI specification defines specific methods (_PSS, _PCT, _PPC, _PSD) for BIOS to report these P-states and current max P-state to the OS.

This works well when the OS/hypervisors are managing the CPU P-states through their own driver. However, when the P-states are managed outside of OS context, like using BIOS or BMC in a server, as in DAPC mode, BIOS cannot report the P-state ACPI objects for OS/hypervisor to consume because that directly conflicts with the BIOS/BMC based P-state manager.

Dell defines new ACPI objects that enable reporting of P-state information, but that do not allow any control aspect for P-states to the OS/hypervisors. With these ACPI objects, OS/hypervisor can determine which P-states are available in the installed CPUs and also the P-state currently set as the Max-P-state, ensuring performance of the virtual machines (VMs). Additionally, there is no mechanism to communicate the state of the CPU to the operating system when the service processor asserts PROC_HOT signal to CPU to cover thermal/power scenarios in server designs. The ACPI method PPPC as defined in the following section enables reporting of this state to OS.

5.2.2 Link Power Management

The 11G BIOS platform supports the QPI L1 power management state.

5.2.3 LOM Disable

Customers might not want to use the LOM ports if they do not need all of the available LOM ports or if they prefer a different network controller vendor.

The 11G server platforms support two different LOM disable modes.

- PCI_DISABLE
  - Disables the PCIe interface of the LOM for power saving. Not accessible by host for data network traffic.
  - Side band management interface remains active and available for server management network traffic.
- PLAY_DEAD
  - Lowest power mode.
  - All interfaces disabled and LOM is completely unavailable for data or management network traffic.

Table 7 highlights the BIOS and iDRAC settings and the resulting LOM mode.

Table 7. 11G LOM Modes

BIOS Data LOM Enable	iDRAC Management LOM Mode	LOM Mode
Enable	Shared	Normal
Enable	Dedicated	Normal
Disable	Shared	PCIE_DISABLE
Disable	Dedicated	PLAY_DEAD

5.2.4 Core Disable

Processors on 11G platforms support processors with multiple cores. BIOS provides setup options that enable you to configure the number of active cores per processor socket. The number of cores can vary based on the specific processor. You can select one, two, or all physical cores to be enabled in four-core processors. 11G servers have anywhere from 4-8/8+ cores. All cores are enabled by default.

Core disable has multiple uses:

- Performance improvements
  - Improve likelihood of Turbo Mode for applications with few operating threads
  - Greater cache available per core (less cache thrashing in shared caches)
- Conserve power (Disabling core provides greater power savings than C-states)
- Reduce software licensing costs (Per core software licensing can cause a significant cost burden for new server deployments. Cores can be enabled as needed, thus deferring costs)

5.2.5 Power Profiles

The 11G server platforms introduce the concept of power profiles. A power profile is a group of platform settings optimized for a particular behavior to simplify the power management configuration settings for Dell customers. Table 8 highlights the supported power profiles.

Table 8. 11G Power Profiles

Power Profile	Platform Setting	
Maximum Performance	CPU Power and Performance Management Fan Power and Performance Management Memory Power and Performance Management	= Maximum Performance = Maximum Performance = Maximum Performance
OS Control	CPU Power and Performance Management Fan Power and Performance Management Memory Power and Performance Management	= OS DBPM = Minimum Power = Maximum Performance
Active Power Controller	CPU Power and Performance Management Fan Power and Performance Management Memory Power and Performance Management	= System DBPM = Minimum Power = Maximum Performance
Custom	CPU Power and Performance Management Fan Power and Performance Management Memory Power and Performance Management	= See Section 5.2.6.1 = See Section 5.2.6.2 = See Section 5.2.6.3

5.2.6 Custom Power Profile

The Custom Power Profile enables the user to define individual settings rather than providing a default profile.

The following sections describe the individual settings.

5.2.6.1 CPU Power and Performance Management

The CPU Power and Performance Management setting combines the selection of processor static P-states and P-state management. Table 9 describes the setting values.

Table 9. CPU Power and Performance Management Settings

Value	Definition
Minimum Power	The processor P-state is statically set to the lowest supported P-state. DBPM is disabled.
Maximum Performance	The processor P-state is statically set to the highest supported P-state. DBPM is disabled.
OS DBPM	OS DBPM is enabled and all supported processor P-states are provided to the OS in the ACPI table.
System DBPM	System DBPM is enabled and all supported processor P-states are available.

5.2.6.2 Fan Power and Performance Management

The Fan Power and Performance Management setting is used to optimize DDR3 thermal algorithms for performance or acoustics/power.

Table 10. Fan Power and Performance Management Settings

Value	Definition
Minimum Power	Increase fan speed in response to DDR3 temperatures. Fan speeds are capped at a specified level in respect to fan speed requests from the memory subsystem to reduce power consumption. DDR3 throughput is throttled to maintain temperatures when the fan speed is at the capped level.
Maximum Performance	Uses a combination of increased fan speed and dynamically throttling DDR3 throughput in response to DDR3 temperatures.

5.2.6.3 Memory Power and Performance Management

The Memory Power and Performance Management setting is used to configure the DDR3 frequency to meet workload requirements. The maximum DDR3 frequency is always dependent on the installed processor model and memory configuration. You can specify a lower DDR3 frequency, however.

DDR3 frequency has a significant impact on power consumption. Workloads may not require the higher bandwidths or lower latencies available at higher DDR3 frequencies.

Table 11. Memory Power and Performance Management Settings

Value	Definition
Minimum Power	Lowest supported DDR3 frequency. Lowest supported DDR3 frequency for Xeon 5500 series is 800MHz.
Maximum Power	Highest supported DDR3 frequency. Highest supported DDR3 frequency for Xeon 5500 series 1333MHz.
Specific Memory Frequency (varies)	Selectable DDR3 frequency. Xeon 5500 series supports 800MHz, 1067MHz, and 1333MHz.

The logical abstraction of DDR3 frequency, minimum power, and maximum performance, is also used by Dell Management Console as a means to provide one-to-many memory settings.

5.2.6.4 Power Profile Default

Dell PowerEdge Servers are optimized for performance per watt. Therefore, Power Management defaults to Dell Active Power Controller, which enables C states and C1E.

Customers who are focused on performance might want to set the following options in BIOS setup:

- Maximum Performance mode in the Power Profile settings of the Power Management menu.
- Disable C states and C1E under the Processor menu.

5.2.7 Power Inventory and Budget

The 11G server platforms implement a power inventory. Rack and tower servers support a wide range of configurations. Power inventory software calculates the potential power consumption for a given configuration.

5.2.8 Algorithm

When AC power is present at the PSU(s), the PSUs' DC auxiliary rails automatically turn **ON**. The iDRAC boots and then collects wattage information for the installed PSU(s) using PMBus. Once the main power turns ON, the iDRAC accesses the storage enclosure processor (SEP) through I2C. The iDRAC collects the hard drive information. The iDRAC continues normal operation and waits for the BIOS to provide additional power inventory information.

The BIOS inventories the installed processors and then performs a memory inventory by accessing the DIMMs' Serial Presence Detect (SPDs). BIOS then performs PCIe discovery. For power inventory the BIOS collects the type and quantity of the installed PCIe cards.

Near the end of POST the BIOS passes the collected power inventory information on processor(s), memory, and PCIe cards to the iDRAC firmware and waits for a response.

The iDRAC firmware then calculates the maximum power estimate. The maximum system power estimate is then compared to the PSU capacity. If the maximum system power estimate is more than the PSU wattage, the iDRAC generates system alert messages. The iDRAC then generates a response to BIOS.

BIOS uses the response from iDRAC to determine if the system is allowed to boot normally, in a throttled state, or not at all.

The system inventory scheme is shown in Figure 11 and Figure 12.

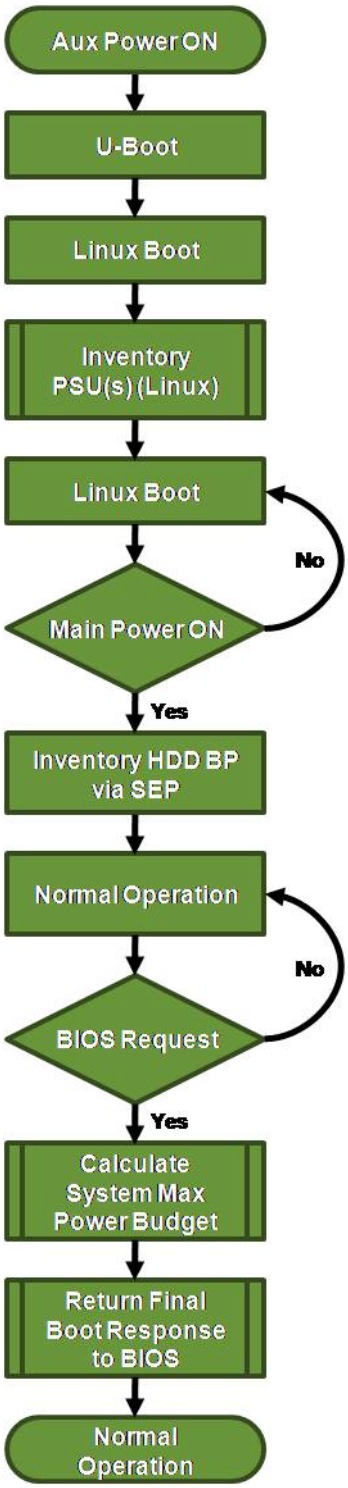

Figure 11. iDRAC System Power Inventory

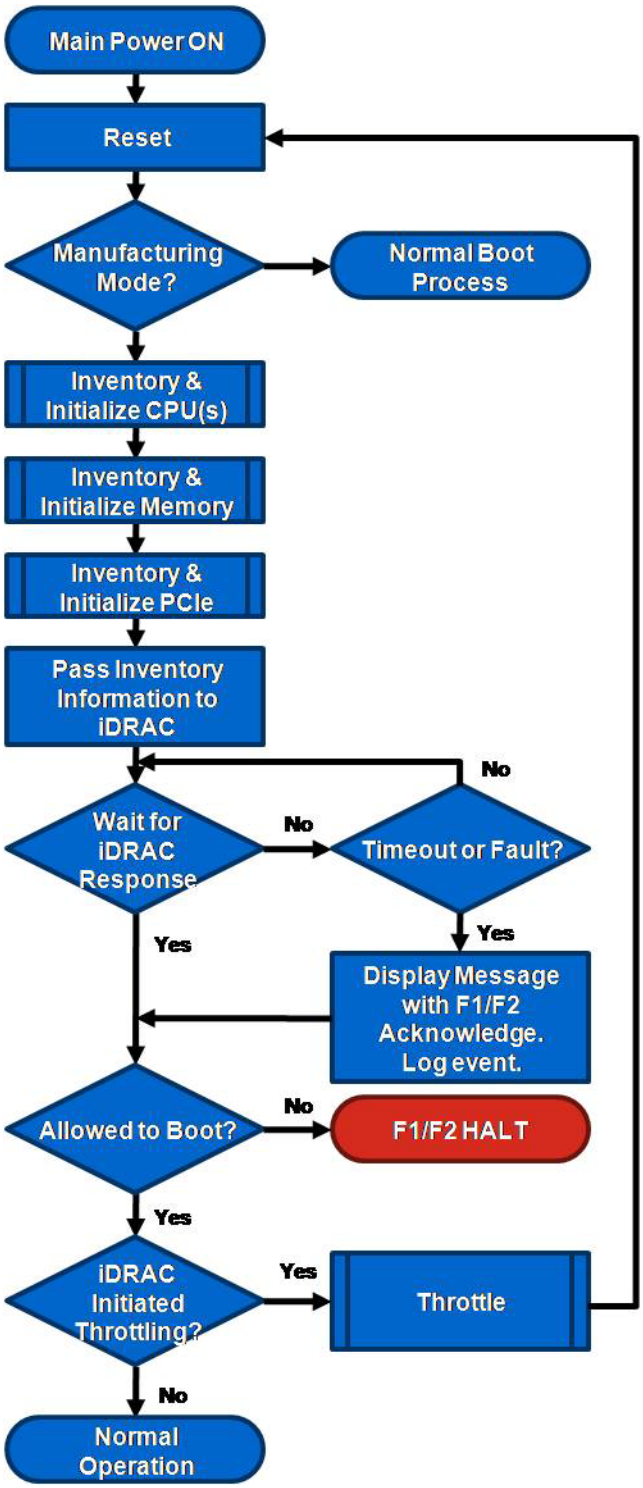

Figure 12. BIOS System Power Inventory

5.2.9 Use Cases

5.2.9.1 PSU Validation

A platform with a low-capacity ES PSU installed could be configured (processors, memory, etc.) beyond the output wattage capability of the ES PSU. While Dell ordering tools and factories ensure that system configurations that exceed the ES PSU are not ordered and shipped to the customer, the system configuration could be upgraded later by the end user. By performing a power inventory, the system can identify when it is over-configured so that it can alert the user through BIOS, system event log (SEL), and LCD messages.

5.2.9.2 DDR3 Memory Static VR Phase Shedding

As described previously, VR phase shedding helps improve the VR efficiency curves. The power inventory and budget determines the power requirements for the installed memory configuration. If the memory configuration can be supported with less memory VR phases, iDRAC firmware disables VR phases to optimized efficiency.

5.2.9.3 DDR3L

When processors or memory controllers that support DDR3L become available, the system software (BIOS, iDRAC) updates to provide system support. To determine how to configure the DDR3 VR voltages, the power inventory scheme determines whether to install standard DDR3 or DDR3L DIMMs.

At boot, BIOS gathers SPD data from the installed DIMMs and determines if the DIMMs are low voltage capable. If low-voltage capable, system software adjusts the DDR3 VR output voltage accordingly. DDR3L DIMMs are backward compatible with the standard 1.5 V. If DDR3 and DDR3L DIMMs are mixed in a system, the VRs are configured for the standard 1.5 V.

5.3 Thermal Management

See the Thermal Design of the Dell PowerEdge T610, R610, and R710 Servers white paper for more information.

5.3.1 Thermal Algorithm Inputs

The number, type, capacity, and speed of components are collected by the BIOS and reported to the iDRAC, which uses this information to determine the minimum fan speeds required for a given ambient temperature. Without this information, the fans would need to run at speeds to cool a worst-case maximum configuration even for lightly configured systems. This penalizes most configurations with higher airflow and power consumption.

5.3.2 Closed Loop Thermal Management

The 11G server platforms have implemented a closed loop thermal management solution that greatly reduces fan speeds. The thermal management algorithm monitors thermal sensors on critical components to accurately determine cooling requirements. These thermal sensors are polled every five seconds. By monitoring the temperature of critical components and using the independent fan PWM controls, the thermal algorithm efficiently provides air flow based on component need.

5.3.3 Optimizations

The thermal algorithm on 11G server platforms has two mechanisms, throttling and fan speed, to control processor and memory temperatures (BIOS settings are detailed in Section 5.2.6.2). The following sections show how to optimize thermal algorithms for performance or power.

5.3.3.1 Performance Optimized

When the thermal algorithm is optimized for performance, fan speed is increased as required to control memory temperatures. Memory throttling is only used as an exception when fans are unable to control memory temperatures alone.

In this mode, memory performance is maintained at the expense of acoustics and power consumption.

Thermal margin on the processor is increased to allow for turbo mode operation.

5.3.3.2 Power Optimized

When the thermal algorithm is optimized for power, a combination of fan speed and throttling can control memory temperatures. The power optimized thermal algorithm defines three memory throttle levels used with fan speed to manage memory temperatures.

5.3.4 AutoCool

The 11G server platforms improve upon the AutoCool thermal solution first delivered with the 9G server platforms. As densities within servers and racks increase, the heat dissipated when servers are in standby becomes a concern. AutoCool provides cooling as needed to minimize acoustics and power consumption when the server is in standby.

The 11G 1U servers do not have a PSU with an integrated fan. The 1U servers rely on system fans to cool the PSU. The iDRAC monitors the PSU temperature and turns on the AutoCool fan powered by the auxiliary 12 V as needed. The 1U fans support a fan-off function critical for minimizing the AutoCool acoustics and power consumption.

The 11G 2U servers have a PSU with an integrated fan. An embedded microcontroller within the 2U PSU monitors the PSU temperature and controls the PSU fan. While the PSU fan does not support the fan-off function, the fan speed necessary for cooling the PSU in standby is minimal.

5.4 Other Power Management Features

5.4.1 Power Capping

The 11G rack and tower server platforms are the first Dell platforms to support a power management feature called power capping. Power capping enables a user to select a power cap threshold from a defined range. The power capping algorithm uses processor and memory throttling mechanisms to reduce the operating power envelope. Starting mid-December 2009, processor throttling will limit P-states and use T-states and memory throttling will limit memory throughput.

Power capping only throttles the server, thus impacting performance only after the specified threshold is reached. Power capping does not impact server performance below the specified cap threshold.

The specified threshold might be exceeded for short periods of time until the system can detect and respond. Over time, the power consumption averages below the defined threshold.

Figure 12 shows an example of the impact of a power cap on the system power level.

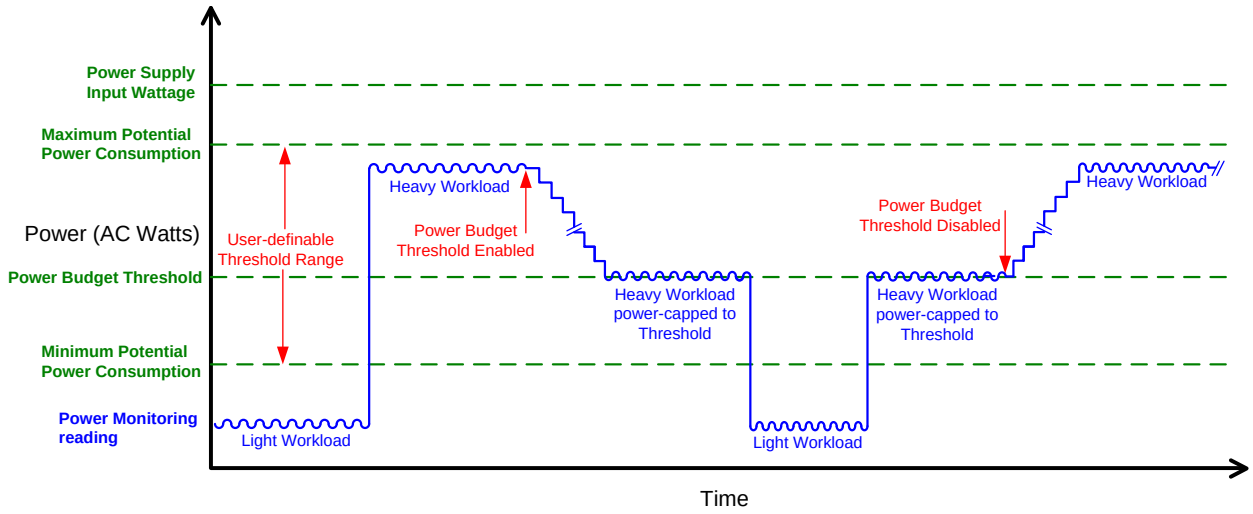

Figure 13. Dell Energy Smart Technology—Power Cap Example

The following iDRAC GUI screenshot shows how you can enable power capping and specify the threshold at which power capping begins throttling the system.

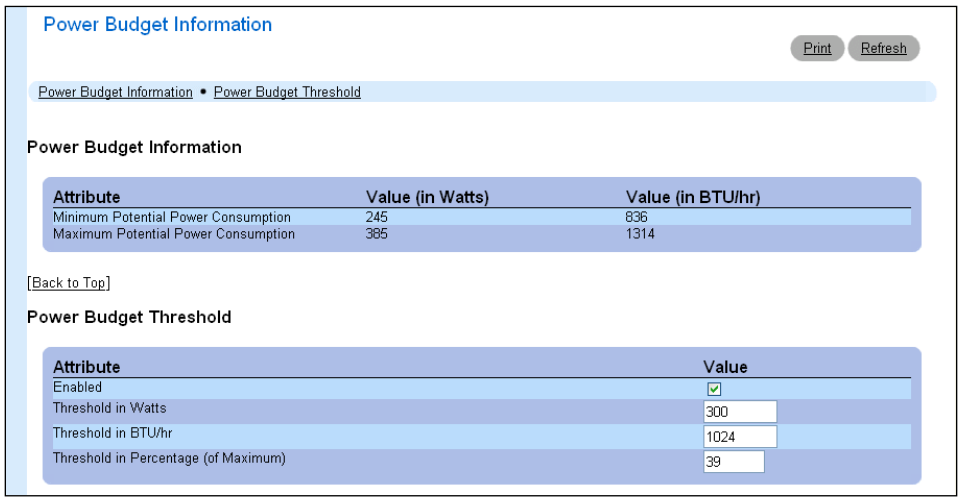

Figure 14. iDRAC GUI Power Budget Page

5.4.1.1 Algorithm

After the iDRAC has booted, it begins a polling loop. The first step in the loop is to read the power consumption from the PSU using PMBus. The iDRAC then verifies that a previous attempt to control processor P-state/T-state was successful. This step is necessary due to a variety of processor power management modes and OS power management support.

The iDRAC then compares the power measurement high and low thresholds calculated from the user-defined power cap. The high threshold provides an indicator that the power consumption is

approaching the power cap and that throttling needs to be increased. The low threshold provides an indicator that the power consumption has been reduced to the point where throttling can be reduced.

If the high threshold is exceeded the iDRAC increases throttling. If power consumption is below the low threshold, the iDRAC decreases throttling. If the power consumption is between the high and low thresholds, the iDRAC maintains the current throttle levels.

The following flow chart shows the overall flow of the user power cap algorithm.

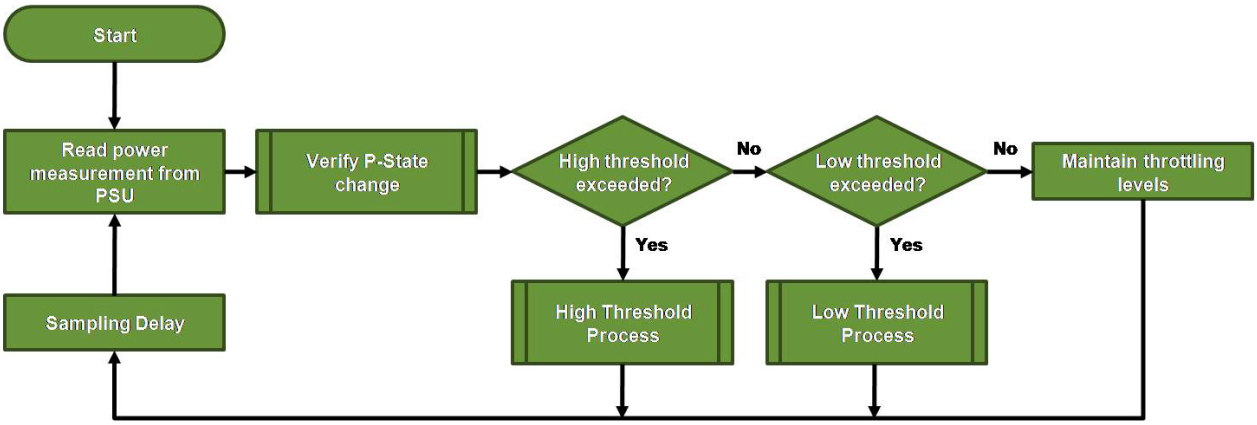

Figure 15. User-Defined Power Cap: Top Level Flow Chart

5.4.2 AC Power Staggered Startup

On previous server generations, the previous state option would bring all systems up simultaneously after an AC loss causes a significant current spike in the datacenter. The AC spike creates an unnecessary challenge for datacenter's to handle. The 11G server platforms provide an AC recovery model that supports a random power ON time.

BIOS maintained the setup option **AC Power Recovery** from previous generations, but added a new setup option called **AC Power Recovery Delay** that is tied to the AC Recovery Mode.

Table 12. 11G AC Power Staggered Startup Options

AC Power Recovery	AC Power Recovery Delay	Description
Off	N/A	System remains off.
On	Immediate	System turns on immediately.
	Random	System turns on after a random delay (30 - 240 seconds).
	User Defined	System turns on after a user defined delay (30 - 240 seconds).
Last, Performed Only if Last Action Was On	Immediate	When set to Immediate, there is no delay for power-up.
	Random	When set to Random, the system creates a random delay (30s to 240s) for power-up
	User Defined	When set to User Defined, the system delays power-up by that amount. System-supported user-defined power-up delay range is from 30s to 240s

5.4.3 Remote Power Control

11G server platforms include remote power control through iDRAC interfaces. Table 13 describes the available controls.

Table 13. Remote Power Control

Control	Description
Power On System	Turns ON the server's power (the equivalent of pressing the power button when the server power is OFF).
Power Off System	Turns OFF the server's power.
NMI	Generates a Non-Masking Interrupt (NMI) to halt system operation.
Graceful Shutdown	Shuts down the system.
Reset System	Resets the system without powering off. (warm boot)
Power Cycle System	Powers off and then reboots the system. (cold boot)

6 Reporting

6.1 Power

DESA in 11G server platforms reports multiple power measurements and includes graphing capabilities. Figure 16 is an example of the power measurements provided in the iDRAC GUI. The following sections explore each of these measurements in detail.

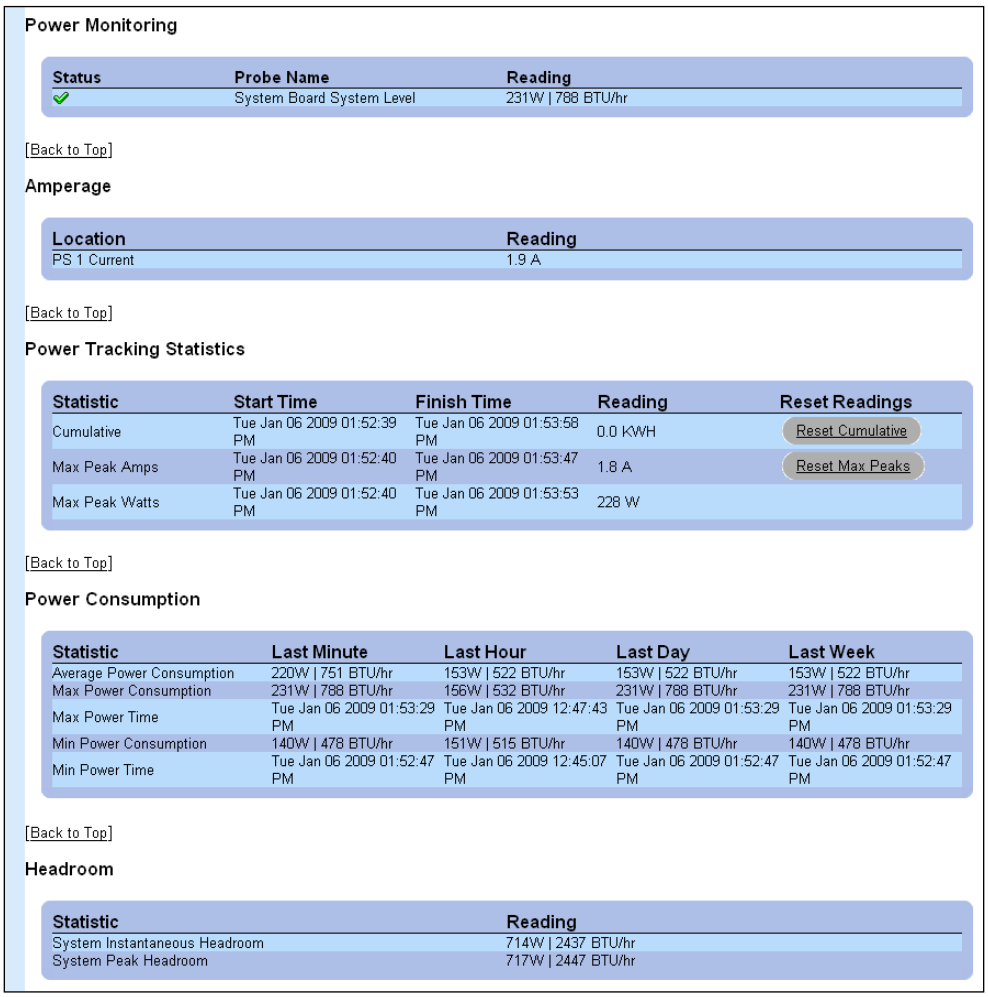

Figure 16. iDRAC GUI—Power Monitoring Page

6.1.1 Measurements

6.1.1.1 Power Monitoring

The power monitoring section of the iDRAC **Power Monitoring** GUI page provides power measurements (AC watts) for the entire system. System board system level is the probe name that provides this measurement.

The sensor provides a one-minute moving average of the system power consumption. Power measurements are read from the PSU(s) every two seconds. For a single PSU configuration, the power measurement is simply a reading from the one PSU. For a redundant PSU

configuration, the power measurement sample is the sum of power measurement readings from both installed PSUs.

6.1.1.2 Amperage

The amperage section of the iDRAC **Power Monitoring** GUI page provides at the wall current measurements for the entire system. A location name is provided for each of the PSUs installed. If a PSU is not installed, the location name is not displayed.

- PS 1 Current Current associated with PSU slot 1
- PS 2 Current Current associated with PSU slot 2

The sensor provides a one-minute moving average of the system current consumption per PSU. Current measurements are read from the PSU(s) every two seconds. The current is displayed per PSU. The total system current is the sum of the current for all installed PSUs.

The 11G PSUs support current sharing. For redundant PSU configurations, the system load is shared across PSUs. If a PSU is hot-added, the PSUs dynamically adjust the load across the redundant PSUs. The PSU specification has defined that the load sharing tolerance can differ by as much as 10% of the maximum output current. Therefore, under normal operation, load sharing is not guaranteed to be distributed equally across the PSUs.

The PSUs support embedded microcontrollers with which the system communicates for PSU details and power monitoring. The auxiliary voltage rail powers off the PSU controller and has the ability to draw auxiliary power from the other PSU if its PSU does not have AC power. This means that even if one of the two PSUs is not plugged in, current measurements are provided for both PSUs. The power for one of the PSUs is extremely low because it reports the auxiliary power that the microcontroller consumes. In this scenario, redundancy is reported as lost and the system load is supported by a single PSU. Under these conditions, the load sharing tolerance requirement described above is not applicable.

6.1.1.3 Power Tracking Statistics

6.1.1.3.1 Cumulative

The cumulative power tracking statistic is a running summation of one-minute average samples. This statistic pulls the one-minute average from the power monitoring system board system level measurement described in a previous section. The summation is then converted to kilowatt-hour (KWH). This statistic is refreshed every minute.

The start time associated with the cumulative power tracking statistic is the time stamp of when the statistic was last reset. The finish time is the current system/iDRAC time. The IPMI command for the cumulative power tracking statistic does not return a timestamp. A separate IPMI command is available to get the current system/iDRAC time.

6.1.1.3.2 Maximum Peak Amps

The maximum peak amps power tracking statistic provides the peak one-minute average current for the system. This statistic leverages the amperage statistics previously described. The amperage statistic provides one-minute current data per PSU. The maximum peak amps statistic is a system level peak, however. Redundant PSU configurations are the sum of the one-minute average currents for both PSUs.

The start time associated with the maximum peak amps power tracking statistic is the time stamp of when the statistic was last reset. The finish time is a time stamp of when the peak occurred.

6.1.1.3.3 Maximum Peak Watts

The maximum peak watts power-tracking statistic provides the peak one-minute average power for the system. This statistic leverages the one-minute average from the power monitoring system board system level measurement.

The start time associated with the maximum peak watts power-tracking statistic is the time stamp of when the statistic was last reset. The finish time is a time stamp of when the peak occurred.

6.1.1.3.4 Average Power Consumption

The average power consumption over the last minute is the average of 30 x 2-second sample, recalculated with every sample.

The average power consumption for the last hour is the average of 4 x 15-minute sample, the last day is the average of 96 x 15-minute sample, and the last week is the average of 672 x 15-minute sample. These values are recalculated every 15 minutes. A 15-minute sample is the average of 450 x 2-second sample.

6.1.1.3.5 Maximum Power Consumption

The maximum power consumption is the highest two-second power sample in the time interval specified (e.g., in the last minute, last hour, etc.), not since the last AC power cycle.

6.1.1.3.6 Maximum Power Time

The maximum power time is a time stamp for when the maximum power consumption occurred.

6.1.1.3.7 Minimum Power Consumption

The minimum power consumption is the lowest two-second power sample since the last reset of peaks.

6.1.1.3.8 Minimum Power Time

The minimum power time is a time stamp for when the minimum power consumption occurred.

6.1.1.4 Headroom

The headroom values displayed in the iDRAC GUI are new to 11G platforms. These values are based on one-minute samples and not maintained through iDRAC resets.

6.1.1.4.1 System Instantaneous Headroom

The system instantaneous headroom is the AC wattage of the PSU minus the reading for the system board system level sensor.

6.1.1.4.2 System Peak Headroom

The system instantaneous headroom is the AC wattage of the PSU minus maximum peak watts as reported by the sensor.

6.1.2 Graphing

DESA also supports reporting power measurements graphically to provide the user a visual indication of power consumption versus time of day. Figure 17 shows the graphing capabilities of the iDRAC GUI.

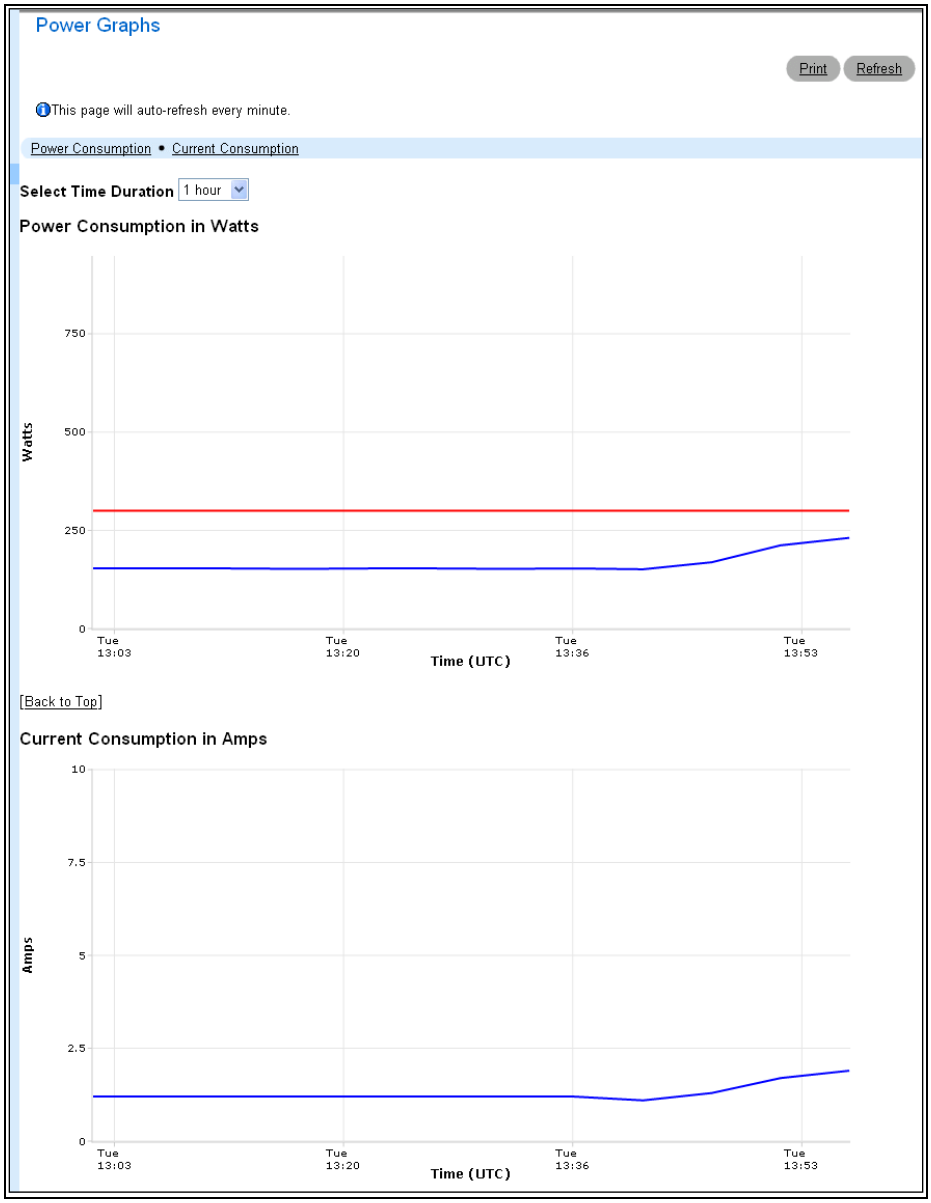

Figure 17. iDRAC GUI—Power Monitoring Graphs

6.2 Performance

DESA supports reporting key server performance measurements. Dell's Management Console (DMC) uses standard Windows® operating system and Linux® operating system performance counters to monitor server performance. When considered with other reports such as power and thermal, the user has a tremendous amount of information to characterize the system. Table 14 shows the performance metrics reported in the DMC.

Table 14. Performance Metrics

Component	Description
CPU	%Kernel Utilization Time
	%Processor Utilization Time
	%User Utilization Time
Logical Disk	Logical Disk Free Space
	Logical Disk IO/Sec
Memory	% Page File Usage
	Available Memory
	Pages IO/Sec
Network	Incoming Bytes/Sec
	Incoming Packets/Sec
	Outgoing Bytes/Sec
	Outgoing Packets/Sec
Physical Disk	Average Access Time
	Physical Disk IO/Sec
System	Context Switches/Sec
	Processor Queue Length

6.3 Thermal

DESA reporting includes thermal measurements such as ambient temperature, fan speeds, and fan health status. The following iDRAC GUI screenshots are examples of the information provided.

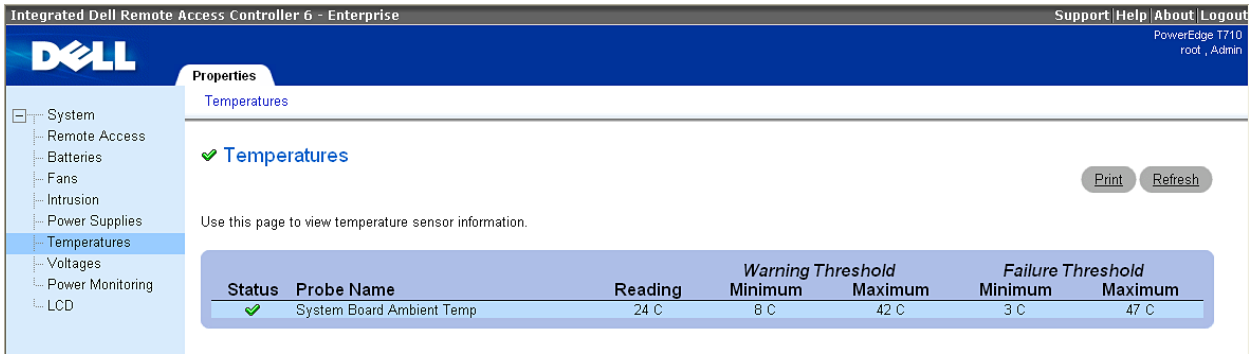

Figure 18. Temperature Sensor

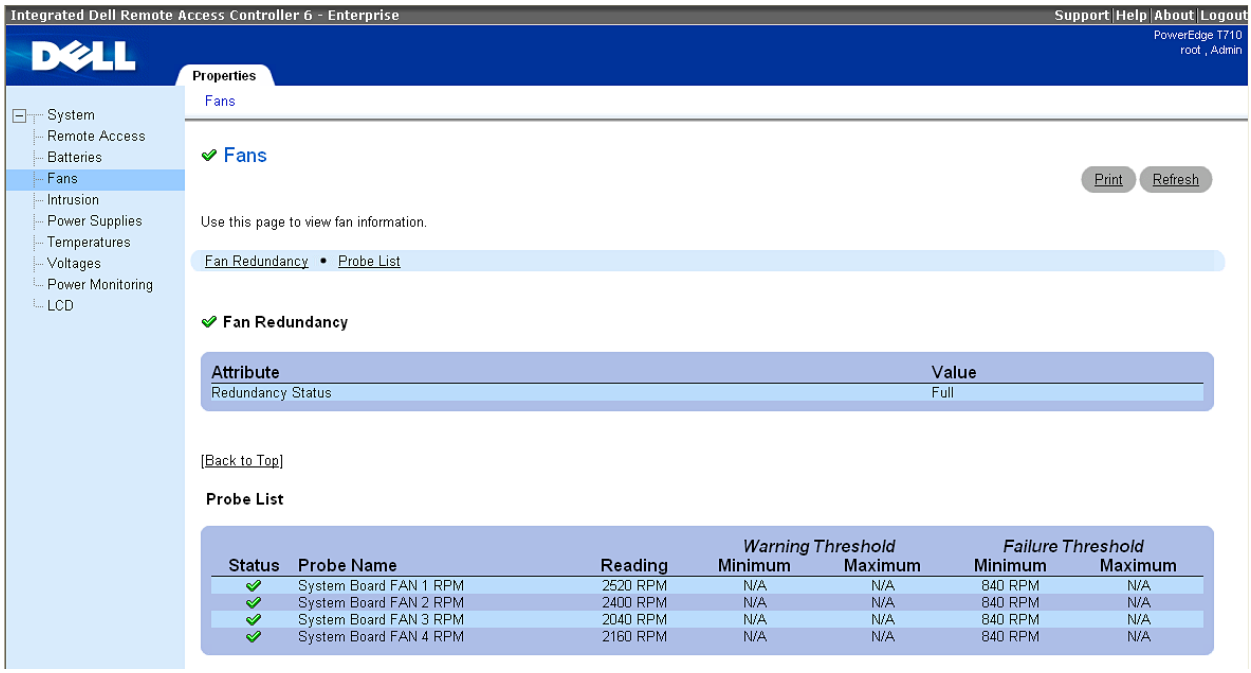

Figure 19. Fan Status