

The Lenovo logo is displayed in white text on a black rectangular background.

# Energy Efficiency Features of Lenovo ThinkSystem Servers

---

**Describes how ThinkSystem servers are designed to be energy efficient**

---

**Introduces key efficiency features**

---

**Explains ways you can configure your server to be most efficient**

---

**Introduces energy efficiency beyond the server**

**Robert R. Wolford**



# Abstract

As electrical costs continue to rise, customers demand decreasing power usage and increasing efficiency on servers. Extracting as much power savings as possible translates into many “dials and knobs” used to enable and tune the power features. These settings can quickly overwhelm a server administrator. As a result, the server is either not optimized for efficiency or worse, both efficiency and performance suffer.

This paper describes the energy efficiency and power saving features available on the Lenovo® ThinkSystem™ servers. Also discussed are tools available to monitor and cap power, tools for predicting power consumption, calculating TCO savings realized from power savings, and finally a reference to system power states.

The paper is intended for Lenovo ThinkSystem server administrators. Before implementing any items described in the paper, the user should have a comfortable understanding of the server hardware, the server configuration tools (e.g. UEFI F1 Setup, oneCLI), and configuring the target operating system.

At Lenovo Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at:

<http://lenovopress.com>

# Contents

|   |    |
|---|----|
| Introduction . . . . .                            | 3  |
| A balanced approach . . . . .                     | 3  |
| Relative influence of power features . . . . .    | 5  |
| 18 server features for power efficiency . . . . . | 6  |
| Summary: The role of efficiency in TCO . . . . .  | 20 |
| Additional resources. . . . .                     | 22 |
| About the Author . . . . .                        | 32 |
| Notices . . . . .                                 | 33 |
| Trademarks . . . . .                              | 34 |

# Introduction

Several years ago, not much thought was given to how much power servers consumed. With plentiful energy resources and low electrical rates, server characteristics such as low cost, high performance, and high reliability were considered more important than improving the power efficiency of servers.

That all changed when, during 2004 and 2005, the price of electricity began to rise dramatically, as shown in Figure 1. From 2004 to 2008, the price jumped by an additional 30%. Even after 2008, the dramatic jump leveled off but did not drop back to the previous levels. It seems that higher electrical rates are here to stay.

This dramatic increase in a short period of time sparked the desire of data center managers to develop more efficient computing solutions.

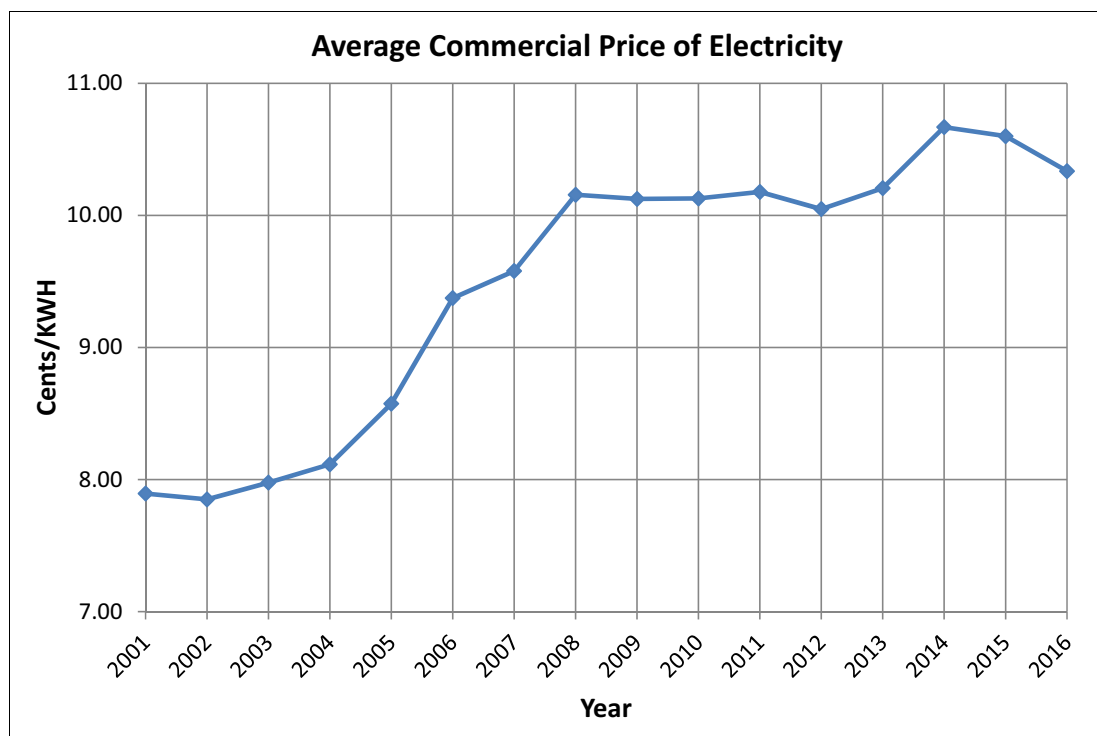


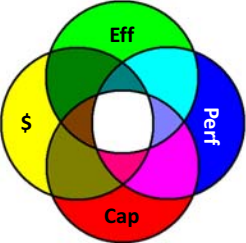
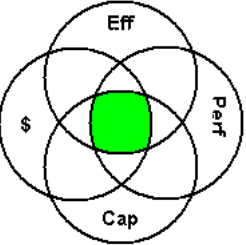
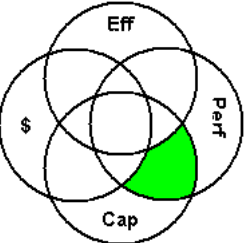
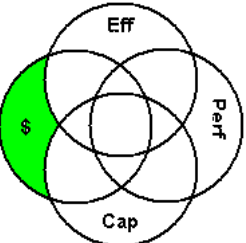
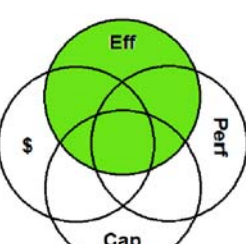
Figure 1 Average commercial price of electricity<sup>1</sup>

## A balanced approach

Designing high power efficiency into a server requires a balanced approach. If peak power efficiency were the only goal, a server can be designed with considerable efficiency. However, the server would be large, have slow absolute performance and high latency, little expansion capability, it would be expensive, and have few reliability, availability, and serviceability (RAS) features. Designing a server that is desirable requires the consideration of efficiency, performance, capability, and cost. These are related as shown in Table 1 on page 4.

<sup>1</sup> U.S. Energy Information Administration (EIA), <http://www.eia.gov>

Table 1 A balanced approach

|  |  |
|--|--|
| <p>Legend:</p> <ul style="list-style-type: none"> <li>▶ \$ Cost</li> <li>▶ Eff Efficiency</li> <li>▶ Perf Performance</li> <li>▶ Cap Capabilities</li> </ul> |    |
| <p>Some server offerings strike a balance among all four areas</p>   |    |
| <p>Others focus on a combination of two areas</p>  |   |
| <p>Still others focus on cost optimization</p>   |  |
| <p>The focus of this paper is to describe the energy efficiency features built into ThinkSystem servers.</p>   |  |

# Relative influence of power features

Increases in power savings and efficiency at the system level are a combined effect of many individual features. This is illustrated in Figure 2. The benefit of the power saving features varies depending on the utilization of the server.

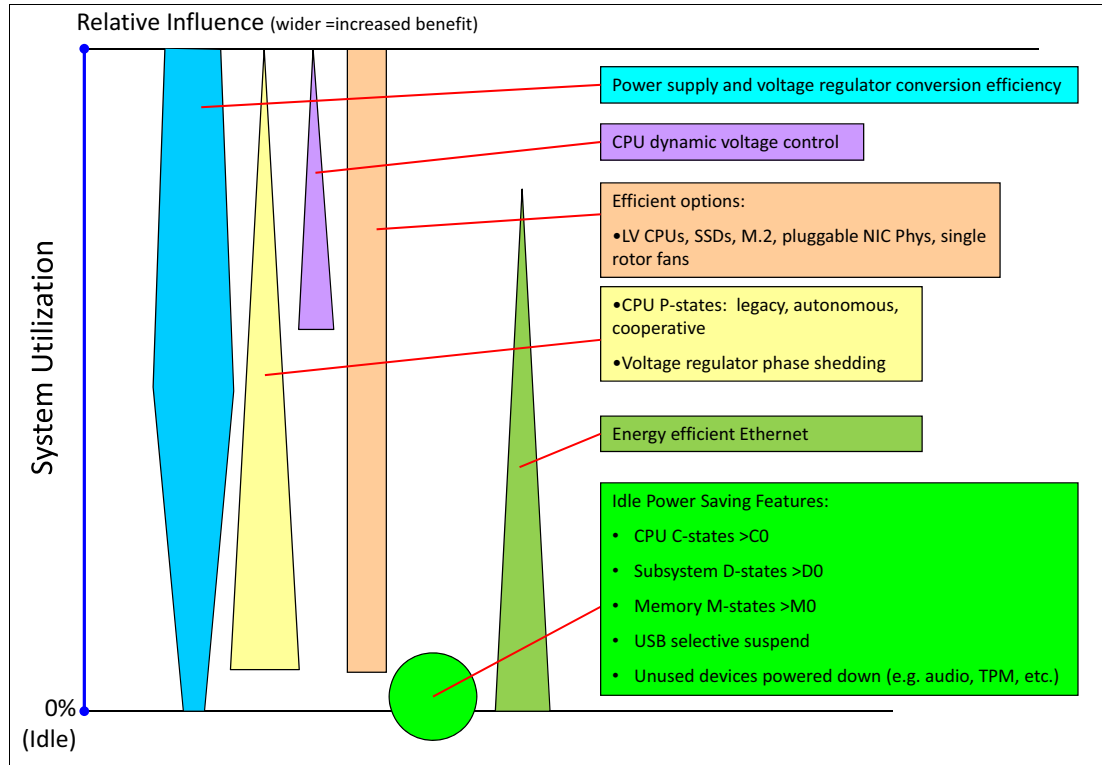


Figure 2 The effects of power savings and efficiency features

Figure 2 illustrates the relative influence of many power saving features. The vertical axis represents system utilization ranging from 0% (idle) to 100% (maximum utilization). The width of each polygon at any utilization level represents the relative benefit of each group of power saving features.

For example, at 50% utilization, the power supply and voltage regulator device (VRD) efficiency have a large influence on overall system efficiency. This is because the blue (left-most) polygon is wide at the 50% utilization point. In contrast is the green (right-most) energy efficient Ethernet, which has little benefit at 50% utilization. As another example, the idle (green circle) power saving features have no benefit at 50% utilization.

It is important to understand the portion of the utilization curve where the server operates. In this manner, it is possible to understand which power saving features are influencing the overall performance or watt efficiency of the server for the target workload. For example, if a server is running VMware and spends 95% of the time at 60-80% utilization, then the features that save power at less than 60% or greater than 80% are not as important. This is illustrated in Figure 3 on page 6.

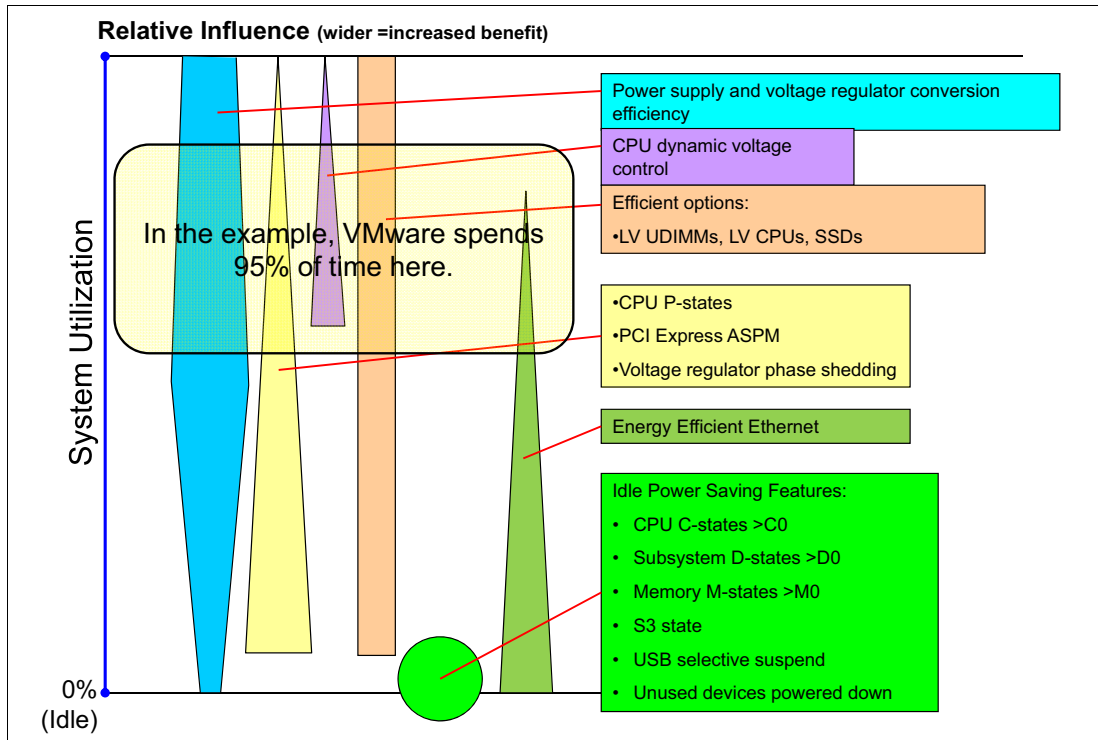


Figure 3 VMware workload example

In the following sections, we explore the individual power features illustrated in Figure 2 on page 5 and Figure 3.

## 18 server features for power efficiency

ThinkSystem servers contain many features that are designed to save power, improve efficiency, monitor power, and impose power limits where necessary. Here are 18 of them:

1. "80 PLUS Titanium server power supply" on page 7
2. "Active or standby power supplies" on page 8
3. "Voltage regulator efficiency" on page 9
4. "Efficient options" on page 9
5. "Energy-Efficient Ethernet (EEE)" on page 10
6. "USB selective suspend" on page 10
7. "Electrical safety protection" on page 11
8. "Unused devices" on page 12
9. "Energy efficient thermal design" on page 12
10. "Energy efficient turbo mode" on page 12
11. "CPU uncore frequency scaling" on page 13
12. "Power bias and performance bias" on page 13
13. "CPU P-state control" on page 14
14. "Power metering" on page 14
15. "Power capping" on page 17
16. "Energy Star certification" on page 19
17. "Higher temperature limits" on page 19
18. "PDUs and power distribution" on page 19

# 1. 80 PLUS Titanium server power supply

Higher efficiency in the bulk power supply of the server means less heat output from the server. Any AC power that is not directly converted to bulk 12V power is dissipated as heat or is consumed as part of the AC-to-DC conversion process. Such efficiency can have a dramatic effect on overall power consumption of the server and also the data center.

80 PLUS is an industry organization that rates power supply efficiencies and grades them based on the percentage conversion at specific power loads. Ratings include 80 PLUS Platinum and 80 PLUS Titanium. For details, including the requirements for each level, are at this web site:

<http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx>



Consider the following example:

A server contains a 750W power supply unit (PSU) and the server is consuming 375W DC. The PSU will be operating at 50% load (375W or 750W). Referring to Table 2, if the PSU is 80 PLUS certified the conversion efficiency will be 80%. That means that the power draw at the AC line cord will be 469W. 375W gets delivered to the server components, and 94W is dissipated as heat.

That 94W of heat has to be removed from the server. This is typically done with a combination of cooling fans in the server and the cooling infrastructure in the data center by the use of computer room air conditioners (CRACs), heat exchangers, air handlers, etc. The data center overhead (that is, the power delivered to the data center that is not consumed by the IT equipment itself) can easily reach double the power that is dissipated as heat from the power supply. Therefore, for just one server with an 80 PLUS rated PSU, the total power overhead because of PSU efficiency can reach 188W.

One of the best ways to reduce the power overhead is to improve the efficiency of the server PSU, as this has a cascading effect up through the data center for both power and cooling. In this example, if an 80 PLUS Titanium PSU is used instead of the 80 PLUS power supply, the 50% efficiency jumps to 96% (refer to the efficiency comparison in Table 2). Working through the same calculations, the input power to the server is lowered to 391W of which 16W is dissipated as heat. At the data center level, the 188W is reduced to 32W, a dramatic reduction of 156W!

Table 2 80 Plus minimum efficiencies<sup>2</sup>

| Power supply load<br>(Percent of rated load) | 80 PLUS Titanium<br>Minimum efficiency  | 80 PLUS<br>Minimum efficiency<br>(Industry standard)  |
|--|---|--|
| 10% load                                     | 90%   | Not rated  |
| 20% load                                     | 94%   | 80%  |
| 50% load                                     | 96%   | 80%  |
| 100% load                                    | 91%   | 80%  |

<sup>2</sup> Source: <http://www.plugloadsolutions.com/80PlusPowerSupplies.aspx>

## 2. Active or standby power supplies

The concept of active or standby power supplies is where one of the available power supplies in a server is automatically and dynamically turned off so as to save energy. This feature is sometimes referred to as Zero Output, Active Standby, or Smart Redundancy

ThinkSystem servers support redundant power supplies. If a power supply fails, the remaining PSUs provide the power and prevent the server from powering down unexpectedly. When all installed PSUs are operating with no errors, the total power load is evenly distributed among each PSU installed. For example, with two PSUs installed, each PSU provides half of the power that the server consumes. Installing two PSUs is beneficial for fault tolerance. However, there is one disadvantage. When both PSUs are delivering power, the efficiency of each PSU can be reduced. This is because the typical PSU efficiency versus load curve is bell-shaped. Consider the following example:

If a server is consuming half of the rated power of the PSU for which the efficiency curve is shown in the blue solid line in Figure 4, then the PSU will be operating at the peak of its efficiency curve (point 1 on the efficiency curve). If a second PSU is then installed for redundancy, the total server power load is divided equally between the two PSUs. In that case, the load on each PSU is reduced to 25% (point 2 on the efficiency curve). At that point, the efficiency is lower than at the 50% load point. This is illustrated with the orange boxes (boxes 1 and 2) and lines in the efficiency curve plot.

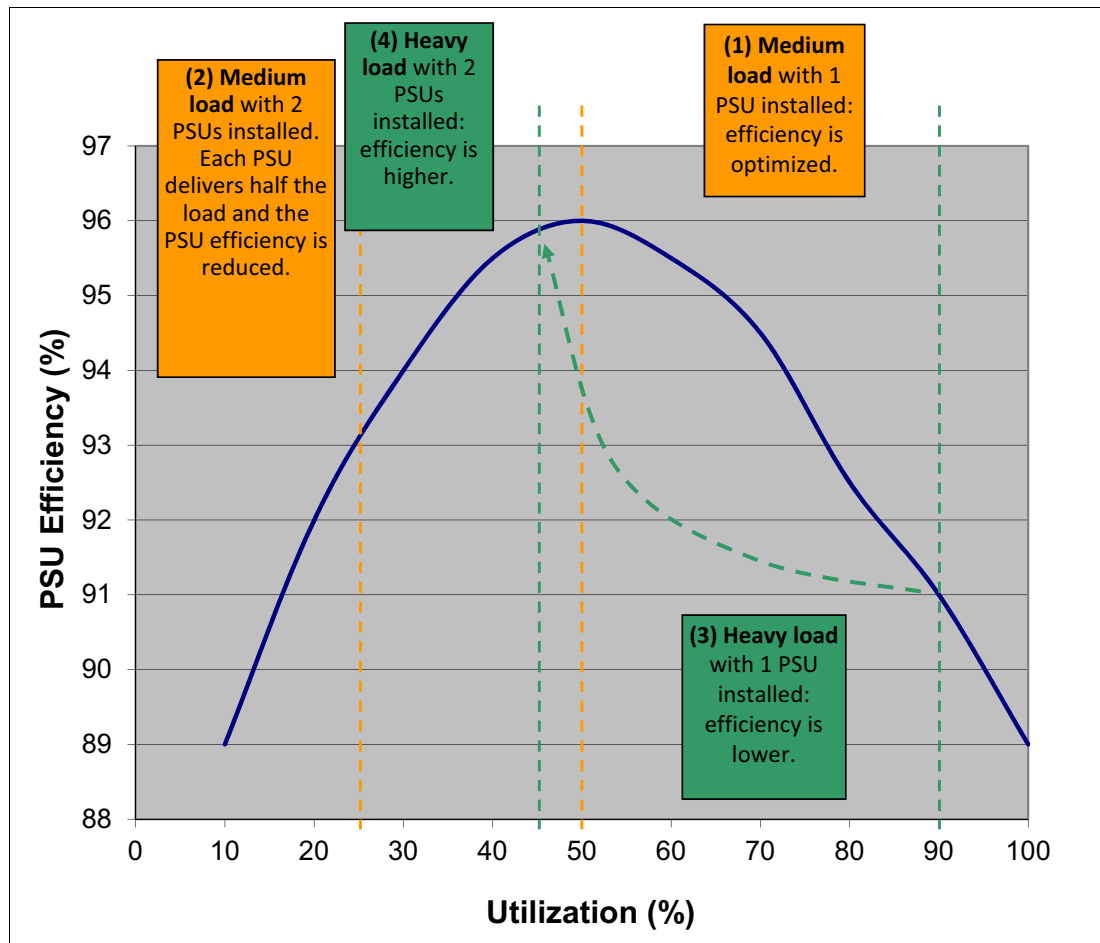


Figure 4 Operating points on the power supply efficiency curve



To combat this problem, ThinkSystem servers are designed to indicate to the PSUs when the server power load is low enough such that one of the installed PSUs can be placed into a low power, standby state. In this manner, the remaining power supply delivers the entire load and efficiency is boosted. Of course, under heavy loads, it is more beneficial to leave both PSUs in an active state.

With heavy loads, dividing the load equally between both PSUs will actually result in an efficiency increase because the load point on the PSU efficiency curve shifts from the extreme right to the middle portion of the curve. This is illustrated with the green boxes (points 3 and 4) and lines in the efficiency curve plot. Overall, by intelligently managing the power that each PSU provides, the PSU can operate in a more efficient region of its efficiency curve.

### 3. Voltage regulator efficiency

There are several voltage regulatory devices (VRDs) in the system that convert 12V to the target voltage needed for each subsystem. Just like with the system power supplies, the goal is to make each VRD highly efficient so that minimal power is lost and dissipated as heat. The VRDs incorporate several features to minimize power loss and maximize efficiency:

- ▶ Portions of the VRD are dynamically turned on and off based on the load.
- ▶ The frequency that each VRD operates in is optimized to reduce switching losses.
- ▶ The use of inefficient linear regulators is minimized.
- ▶ The motherboard layout is optimized to reduce power losses between the VRD and the component it is driving.
- ▶ VRD components are selected that exhibit minimal inherent power losses.

### 4. Efficient options

Some options that can be installed in servers are inherently more efficient than others. They can contain more efficient components, have solid state media versus spinning media, or contain fewer overall parts that consume power. ThinkSystem servers offer a variety of options that allow a user to choose between minimum power consumption, maximum performance, and a balance between the two. Energy efficient options that are available on ThinkSystem servers include:

- ▶ LV CPUs: LV (low voltage) CPUs generally operate at lower power for a given performance level.
- ▶ DDR4 DIMMs: DDR4 DIMMs can save up to 25% power compared to predecessor DDR3 DIMMs.
- ▶ SSDs and M.2: Having no spinning parts helps solid state drives (SSD and M.2) to achieve lower power and quicker engagement of power management compared to traditional spinning hard disk drives (HDDs).
- ▶ Lower power I/O adapters: Not every customer wants 10 Gb Ethernet or 16 Gb Fibre Channel connections. Many data centers cannot accommodate the higher power associated with high bandwidth adapters. Lenovo offers a variety of I/O options to configure the server for a good balance between performance and power consumption.
- ▶ Pluggable NIC Phys: The physical layer (Phy) of network interface cards (NICs) typically consume the highest portion of power for the total NIC solution. Many ThinkSystem server system boards include pluggable Phy modules. By making the various network interface physical layers pluggable, customers are not forced into a NIC solution that is soldered down on the motherboard. An additional benefit is that power is only consumed for the NIC interface that the customer actually wants to use.

- ▶ **Single rotor fans:** Several ThinkSystem servers include single rotor fans. Single rotor fans include 1 motor. Compared to dual rotor fans which have 2 motors, single rotor fans consume less power for most workloads.

In addition to the energy efficient options we have described, ThinkSystem is also designed so that higher power subsystems are not soldered down on the motherboard. For example, every user might not need a RAID 10 storage controller. The base storage controller on the motherboard is embedded in the core chipset. In this manner, no extra power is wasted on a higher feature storage controller if that function is not needed.

## 5. Energy-Efficient Ethernet (EEE)

Energy-Efficient Ethernet (EEE)<sup>3</sup> operates on the physical layer of Ethernet transmitters. When EEE is enabled and no data is being sent over the Ethernet link, the link is placed into a low power sleep state. A low power idle (LPI) indication signal is sent periodically to the transmit chip, instructing it to turn off for a specified period of time. If data is ready to be transmitted, a normal idle signal is sent to the transmitter to wake it up before the data is sent. Note that, when EEE is used, the Ethernet receive link always remains active even when the transmit link is in sleep mode.

Figure 5 shows the configuration panel for selecting the Energy-Efficient Ethernet (EEE) option under Windows for a Broadcom Ethernet controller.

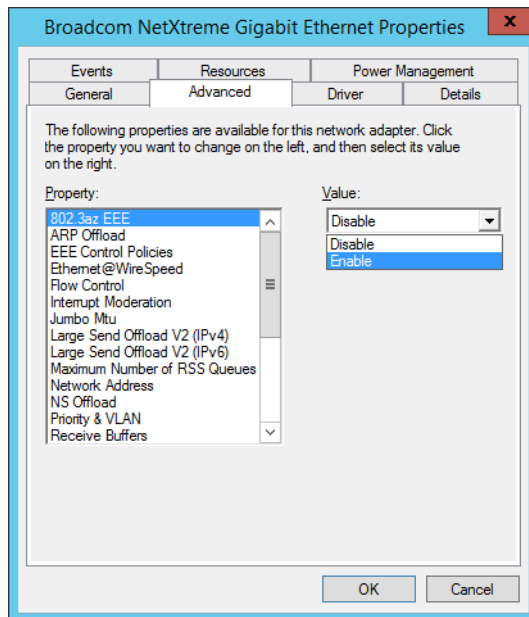


Figure 5 EEE configuration panel for a Broadcom controller in Windows Server 2012 R2

## 6. USB selective suspend

When USB selective suspend is enabled, a USB hub can suspend operation of individual USB ports based on the amount of activity on a port. This can be useful for USB devices that are used intermittently, such as keyboards, mice, printers, etc. When unused USB devices are suspended, the CPU cores spend more time in the deep power saving states (for example CPU C3, C6) when the OS is idle. This is because the CPU cores no longer have to

<sup>3</sup> The Energy-Efficient Ethernet (EEE) defines higher Ethernet networking standards to reduce power consumption. Defined by the Institute of Electrical and Electronics Engineers (IEEE).

intermittently process the transfer schedule from the USB hub. With today's modern CPUs, the savings from USB selective suspend can be on the order of 10W when a server is idle.

Figure 6 on page 11 shows the Control Panel Power Options window for enabling and disabling USB selective suspend under a typical Windows OS.

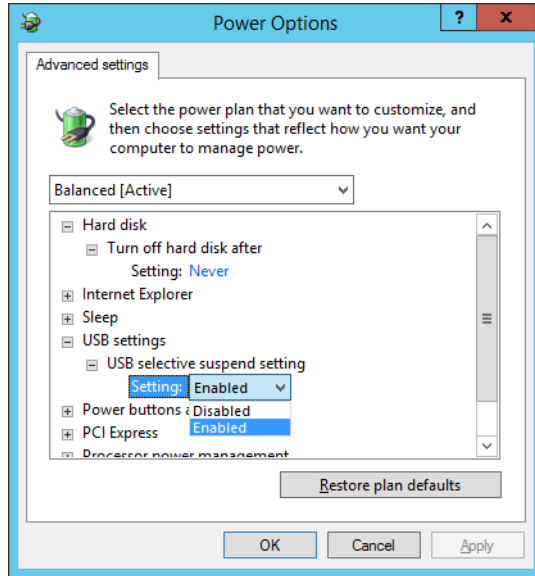


Figure 6 USB selective suspend setting in the Control Panel Power Options window

## 7. Electrical safety protection

Modern servers consume large amounts of power. When the server is running normally, not a second thought is given to the amount of power running through the server. However, when things go awry, it is valuable to have protections in place for both the server and the end user.

ThinkSystem servers have several layers of protection on the electrical delivery circuits:

- ▶ AC Input power fuse
- ▶ AC Input ground fault current interrupt (GFCI) circuit
- ▶ Inrush current limiting for power supplies (PSUs) and voltage regulator devices (VRDs)
- ▶ Over-voltage protection for PSUs and VRDs
- ▶ Overcurrent protection on the PSUs and all major power rails in the server
- ▶ Fast acting short circuit protection for the PSUs and VRDs.
- ▶ Output isolation for each PSU output
- ▶ Thermal protection for each PSU and high power VRDs
- ▶ The bulk voltage rail from the PSUs is split into several rails with each one having a lower power rating. In this manner, large amount of current and power are not present everywhere on the motherboard.
- ▶ Self-extinguishing materials used in the construction of the electrical circuits.
- ▶ Air inlet & exhaust holes on the server are sized such that fingers cannot contact the internal electric components, pins, and circuit traces.
- ▶ Lenovo XClarity™ Energy Manager (LEXM) can be used to monitor for abnormal spikes in server power consumption.

With the addition of the safety features listed above, some small power losses are incurred. However, ThinkSystem products minimize these power losses by the use of advanced circuitry that is optimized for light and heavy loads. ThinkSystem products offer superior user safety and electrical protection while maintaining optimal electrical efficiency.

## 8. Unused devices

Unused devices that are left powered on in a server waste power and act as small heaters. The data center is tasked with removing waste heat from those devices that are doing no useful work. Essentially, any unused device that is doing no work has 0% efficiency and adds to the power overhead of the entire server. To combat this issue, unused devices embedded in ThinkSystem servers are either powered down or placed into a very low power state. This is done automatically during the power-on-self-test (POST) or dynamically, at runtime. Examples of devices for which power is intelligently managed include:

- ▶ CPU cores
- ▶ Memory channels and DIMMs
- ▶ PCI Express ports
- ▶ QPI links
- ▶ SATA and SAS storage controllers
- ▶ Network controllers
- ▶ Serial ports
- ▶ USB controllers
- ▶ VRDs
- ▶ Trusted Platform Module (TPM)

## 9. Energy efficient thermal design

Earlier versions of system level thermal control algorithms were fairly rudimentary, as they simply used an ambient inlet temperature sensor as the only proxy to fan speed control methods. More recently, there have been sophisticated advances in not only the algorithms used, but also the efficient partitioning of the varying subsystems, allowing ThinkSystem to reach near perfect optimization. To better appreciate these advances, one must first understand the need. Cooling on server designs in the early 2000's consumed as much as 10+% of total system power, which clearly is a big piece of the power budget. With industry focus on energy consumption, continued efforts have been made to implement more sophisticated controls, which have now yielded solutions that consume as little as 2% of total system power while running applications. The added intelligence in using power sense circuits, ambient temperature, thermal sensors, configuration data, and the zoned cooling approach has given rise to these significant system level improvements.

In addition to system level advances, there is a growing movement to improve algorithms and controls beyond those of the server, targeting a more holistic solution at the rack and room level integration strategies. Some of the newer ThinkSystem flagship products now offer environmental health awareness consoles, where hot air recirculation can be detected, and thus optimized at the rack level. Along with these feedback mechanisms, improved user feedback and data center integration controls are being made available.

## 10. Energy efficient turbo mode

Historically turbo operation on Intel CPUs was a binary function. The maximum available turbo frequency was engaged immediately when the OS requested additional performance. The downside to the binary approach is that the CPU can thrash in and out of turbo mode,

often over a short period of time. The transitions into and out of turbo mode consume some incremental power. Constantly switching in and out of turbo mode was not energy efficient.

Later generations of CPUs included a power optimized turbo mode, in which a short delay transpired before the CPU was granted maximum turbo frequency. In this manner, short and sporadic turbo requests from the OS were filtered out and turbo ran more efficiently.

The latest generation of Intel CPUs takes turbo efficiency one step further. ThinkSystem products use this new mode called energy efficient turbo. When energy efficient turbo mode is enabled, the OS might request the maximum turbo frequency. However, the actual turbo frequency that the CPU is set to is proportionally adjusted based on the duration of the turbo request. In addition, the memory usage of the OS is also monitored. If the OS is using memory heavily and the CPU core performance is limited by the available memory resources, turbo frequency is reduced until more memory load dissipates and more memory resources become available.

## 11. CPU uncore frequency scaling

CPU P-states are used to run the CPU cores at different frequencies based on workload demand (see “P-states” on page 26). Prior to the ThinkSystem and System x® M5 servers, all of the CPU cores in a CPU package ran at the same frequency, regardless of whether a workload was running on one CPU core or all CPU cores. Similarly, the CPU package *uncore* (non-core related, for example, QPI links and miscellaneous logic) ran at a fixed ratio of the CPU frequency.

With ThinkSystem servers, the uncore runs at a speed that is independent of the CPU cores. This can save power for workloads where the ratio between CPU core demand and uncore demand is not always a fixed ratio. For example, a workload that runs on CPU cores and memory in the same package might require high CPU core frequencies, but low QPI link frequencies. Conversely, a moderate workload that utilizes memory on multiple CPU packages requires higher QPI frequencies when the workload is accessing memory on a remote CPU socket.

## 12. Power bias and performance bias

The power and performance bias setting controls how aggressively the CPU is power managed and placed into turbo. As the bias is adjusted towards harnessing performance, turbo is engaged quicker and the power management features will be engaged less. This has the overall effect of increasing performance and decreasing latency. However, it also increases power.

Conversely, adjusting the bias towards power will cause turbo to be engaged less and the power management features to be engaged more. Performance and latency can increase, but power savings will increase. Refer to Figure 4 on page 8 for the influence of the bias setting.

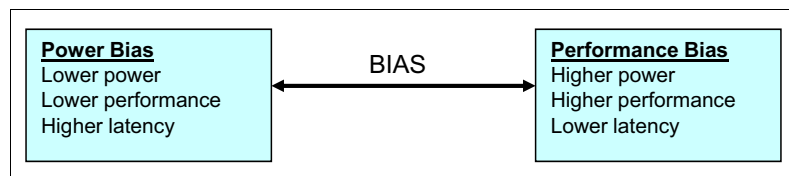


Figure 7 Influence of the bias setting

## 13. CPU P-state control

ThinkSystem servers include several methods of controlling the CPU frequency (P-states) based on a customer's preference of efficiency or performance. The choices and a short description of each are included below:

- ▶ **Autonomous:** With the Autonomous selection, the P-states are completely controlled by system hardware. From the operating system's (OS) perspective, there are no P-states and the OS always thinks the CPU is running at its rated frequency. Autonomous is useful when peak operating efficiency is desired.
- ▶ **Legacy:** When Legacy is selected, the CPU P-states will be presented to the OS and the OS power management (OSPM) will directly control which P-state is selected. Legacy provides slightly lower latencies than Autonomous and can help improve performance. Legacy is also useful on older operating systems that don't support Cooperative mode.
- ▶ **Cooperative:** Cooperative mode is a combination of Autonomous and Legacy modes. Like with Autonomous mode, the P-states are still controlled by the system hardware. However, in Cooperative mode, the OS is aware that different CPU operating frequencies exist and it can provide hints to the system hardware for the minimum, desired, and maximum P-states.

The system hardware does not always have to honor the OS requested P-state though. The final P-state chosen is influenced by the **Power bias and Performance bias** setting and also the current condition of the CPU (power draw, temperature, number of active cores). Newer operating systems (for example, Windows Server 2016, Linux kernel 4.2 and higher) are required to take advantage of cooperative P-states. If Cooperative is selected on an older OS that doesn't support it, the system will fall back to Autonomous mode.

- ▶ **None:** When P-states are set to None, P-states are completely disabled and the CPUs run at either their rated frequency or in turbo mode (if turbo is enabled). None is the best choice when the absolute lowest latency and highest performance are desired. The trade-off is that power consumption is higher.

For additional information on P-states, refer to "System power states" on page 22.

## 14. Power metering

ThinkSystem power metering offers powerful and detailed data collection and analysis capability without the need for cumbersome external power metering outfitted in either PDUs or with an in-series power device added to the rack or upstream power delivery infrastructure. Power metering is available for AC input power, CPU power, and memory power.

Power meter data can be obtained in many ways on ThinkSystem servers:

- ▶ **XClarity Controller (XCC):** Power meter data can be read using the GUI or the command line (CLI). Examples of the web interface as shown in Figure 8 on page 15 and Figure 9 on page 15.

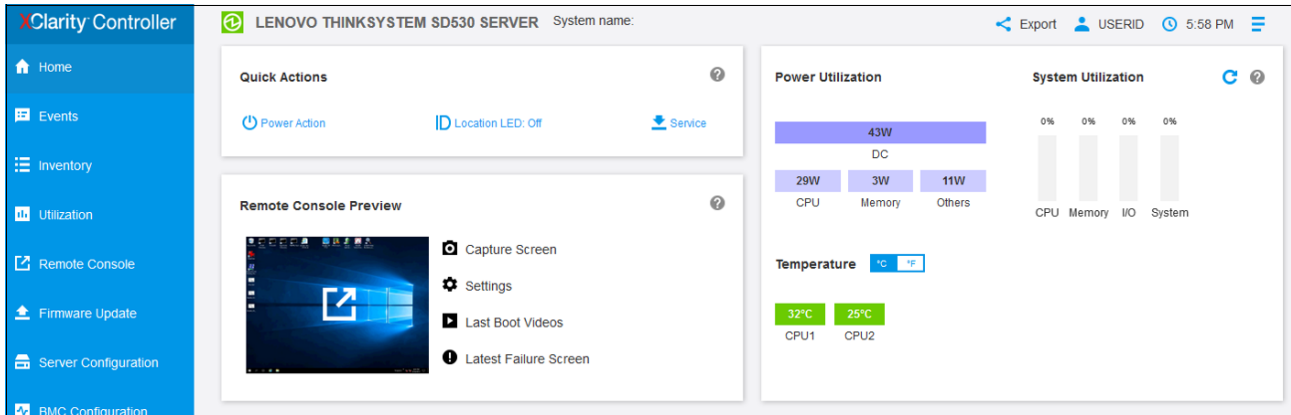


Figure 8 Power consumption metrics

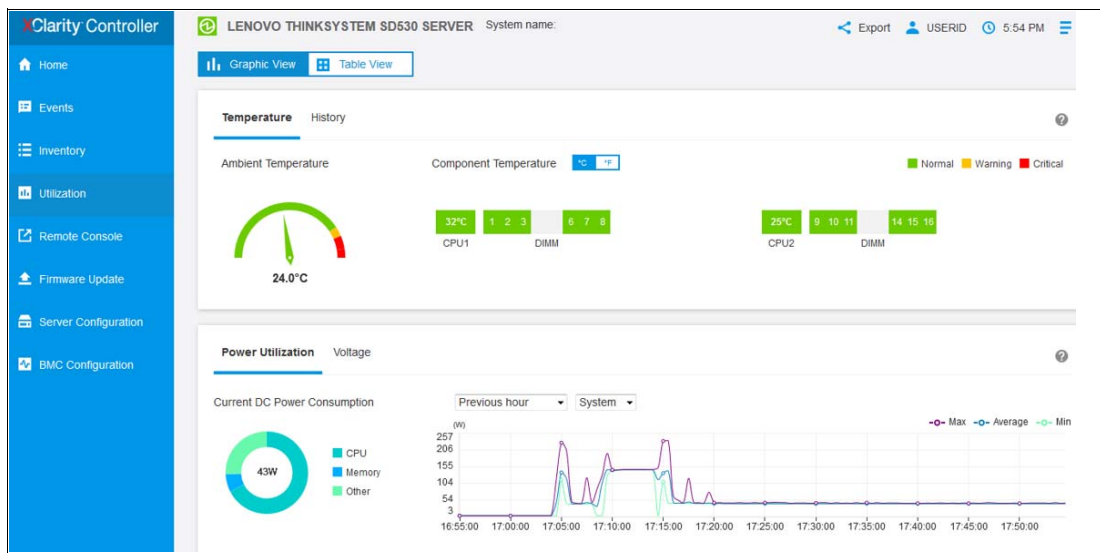


Figure 9 Power consumption history

- System Management Module (SMM) of the ThinkSystem D2 Enclosure and Modular Enclosure used with the SD530 servers: Power meter data can be read using the GUI or the command line (CLI). Examples of the web interface as shown in Figure 10 on page 16.

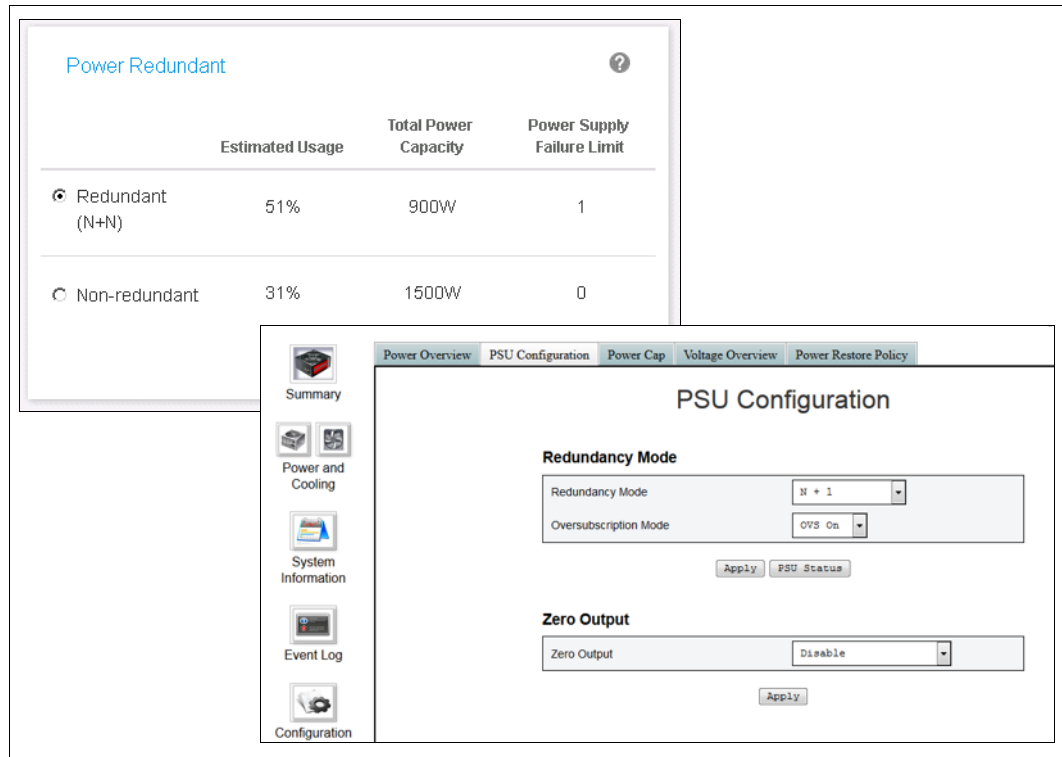


Figure 10 Power policies (SMM in the ThinkSystem D2 Enclosure and Modular Enclosure)

- ▶ **Node Manager 3.0 Intelligent Platform Management Interface (IPMI):** In addition, ThinkSystem servers support the Node Manager IPMI interface. Here, users can issue Node Manager IPMI commands through the XCC systems management controller on the server. An example follows. For full details, refer to the user guide for your server and the latest Node Manager specification.

– NM Power Monitoring Command Example:

Get Node Manager Statistics (command code 0xc8)

Example - Get Global System Power Statistics:

Request:

```
ipmitool -H $XCC_IP -U USERID -P PASSWORD -b 0x00 -t 0x2c raw 0x2E 0xC8 0x57 0x01 0x00 0x01 0x00 0x00
```

Response:

```
57 01 00 38 00 04 00 41 00 39 00 ec 56 f7 53 5a 86 00 00 50
System wattage =0x38 =56W AC (min =4W , max=65W, avg=57W)
```



- ▶ *Data Center Manageability Interface (DCMI)*: Also, ThinkSystem servers support the DCMI. DCMI is an industry standard interface specification that is management software neutral, providing monitoring and control functions that might otherwise be exposed through standard management software interfaces. For further details, see the XCC user guide for your server and the latest DCMI specification.
  - DCMI Get Power Reading Command Example
 

Request:

```
ipmitool -H $XCC_IP -U USERID -P PASSWORD raw 0x2c 0x02 0xdc 0x01 0x00 0x00
```

Response:

```
dc 39 00 38 00 3b 00 39 00 e3 6f 0a 39 e8 03 00 00 40
System wattage =0x39 =57W AC (min=56W, max=59W, avg=57W)
```
- ▶ *Upward Integration Modules (UIM)*: Power meter data can be read by using the Lenovo XClarity Integrator for Microsoft System Center, VMware vCenter.
 

For more information on the XClarity Integrators, see these web pages:

  - Lenovo XClarity Integrator for Microsoft System Center:
 

<https://support.lenovo.com/us/en/solutions/LNVO-MSUIM>
  - Lenovo XClarity Integrator for VMware vCenter:
 

<https://support.lenovo.com/us/en/solutions/LNVO-VMWARE>
  - Lenovo XClarity Administrator:
 

<https://lenovopress.com/tips1200-lenovo-xclarity-administrator>

## 15. Power capping

Similar to power metering capabilities, ThinkSystem servers also support power capping. Power capping can be used to limit the maximum power that a server consumes. It is beneficial in curbing random spikes or surges in power, thereby allowing rack and data center power limits to be maintained.

Server power capping can be controlled in many ways. These include using the GUI or the CLI of the XCC, the Node Manager 3.0 IPMI interface, the DCMI interface, or the Lenovo Upward Integration Modules (UIMs).

- ▶ *XClarity Controller (XCC)*: Figure 11 shows the power capping GUI interface using XCC.



Figure 11 XCC GUI for power capping

- ▶ *SMM*: Figure 12 shows the power capping GUI interface using the System Management Module of the ThinkSystem D2 Enclosure.

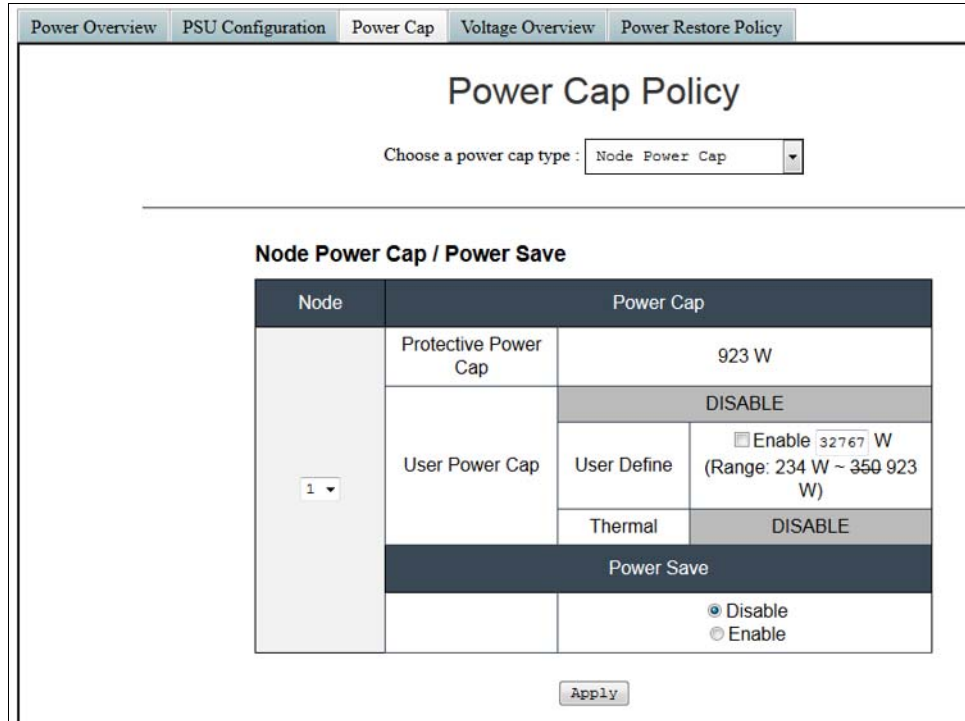


Figure 12 SMM GUI for power capping

- ▶ *Node Manager 3.0 Intelligent Platform Management Interface (IPMI)*: The following is an example of how to set a power cap using the Node Manager 3.0 IPMI interface.

- Set NM Power Cap Command Example

Set Node Manager Policy (command code 0xc1)

Example – Set CPU domain power limit =100W with a policy ID=0x60

Request:

```
ipmitool -H $XCC_IP -U USERID -P PASSWORD -b 0x00 -t 0x2c raw 0x2E 0xC1 0x57
0x01 0x00 0x11 0x60 0x10 0x00 0x64 0x00 0xE8 0x03 0x00 0x00 0x00 0x05
0x00
```

Response:

```
57 01 00
```

- ▶ *Data Center Manageability Interface (DCMI)*: A power cap can be set using the DCMI IPMI interface

- Set DCMI Power Cap Command Example

Set AC power limit =100W

Request:

```
ipmitool -H $XCC_IP -U USERID -P PASSWORD raw 0x2c 0x04 0xdc 0x00 0x00 0x00
0x00 0x64 0x00 0xe8 0x03 0x00 0x00 0x00 0x00 0x05 0x00
```

Response:

```
dc
```

- ▶ *Upward Integration Modules (UIM)*: enables you to set power capping capabilities in conjunction with Microsoft System Center and VMware vSphere. For details, refer to the user guide for your server and the latest UIM documentation.

## 16. Energy Star certification

Energy Star sets a minimum set of standards that are required to declare that a server is energy efficient and certified to be Energy Star compliant. It provides a simple and effective way to ensure that a server is energy efficient. Lenovo makes design choices to ensure that many ThinkSystem servers are Energy Star compliant and certified. Refer to the following links for additional details:



[http://www.lenovo.com/social\\_responsibility/us/en/energy/](http://www.lenovo.com/social_responsibility/us/en/energy/)

<http://www.energystar.gov/productfinder/product/certified-enterprise-servers/results>

## 17. Higher temperature limits

In 2012, the American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE)<sup>4</sup> moved to expand IT equipment server design guidelines, to include two additional classes supporting up to 40°C (103°F) and 45°C (114°F), respectively. These classes were specified in response to regulatory pressures in parts of Europe and Asia, which allows the server hardware to operate at higher temperatures, effectively allowing clients to take advantage of free-air cooling schemes that do not require chillers.

Aside from free air cooling schemes and the inherent benefits of them, typical data centers use air handlers and chillers to remove heat. Even in these environments, it has been demonstrated that raising the inlet temperature to IT equipment can reduce overall data center power consumption, as it allows the air handlers and chillers to run at lower and more optimal set points. The savings can be significant, although there is an inflection point where the IT power draw begins to exceed the savings at the room level, which might adversely affect true net savings. Data from previous deployments suggests that the optimal temperature is most likely in the 25°C to 27°C range for most solutions.

In concert with these industry changes, the ASHRAE book, *Thermal Guidelines for Data Processing Environments*<sup>5</sup>, specifically deals with the questions and concerns about this topic. Lenovo ThinkSystem clearly recognizes the benefits with regard to this shift in the industry, and has responded by moving to support the new ASHRAE Class A3 standard, which supports up to a 40°C (103°F) ambient inlet, which lessens or eliminates the need for chillers.

## 18. PDUs and power distribution

Advancements in Lenovo PDUs have not only improved usability and fault tolerance features; they have also yielded higher overall efficiency and improved power control capabilities. ThinkSystem switched and monitored PDUs support extensive metering, data logging, and threshold alarms. In addition, individual outlet power control is supported. Unneeded equipment can be completely powered down to eliminate phantom loads. Lenovo PDUs typically are rated at 95% or higher efficiency under normal conditions. Finally, Lenovo PDUs protect against many power delivery problems, including those shown in Table 3.

<sup>4</sup> American Society of Heating, Refrigeration and Air Conditioning Engineers (ASHRAE). For more information, see <http://www.ashrae.org>

<sup>5</sup> *Thermal Guidelines for Data Processing Environments*, from ASHRAE  
<https://www.ashrae.org/resources--publications/publication-updates>

Table 3 PDU power protection features

| Feature              | Description   |
|----------------------|---|
| Power failure        | Also known as <i>blackout</i> . Voltage is turned off for an extended time period   |
| Power sag            | Also known as <i>brownout</i> . Voltage is reduced for a relatively short time.   |
| Power surge          | Also called <i>power spikes</i> . These usually last for a brief period and can be caused by lightning, motor startup, or other sources of noise on the same line.  |
| Under-voltage        | Under voltage occurs when the power company delivers less voltage than specified. For example, nominal U.S. voltage is 120V. But sometimes the power company might only deliver 100V. The under-voltage is sometimes evident when lights dim.   |
| Over-voltage         | Over-voltage occurs when the power company delivers more voltage than specified.  |
| Harmonic distortion  | When harmonic distortion is present, the waveform appears as a non-ideal sine wave on an oscilloscope. Harmonic distortion can be generated from a load.  |
| Line noise           | Line noise appears as extra static riding on top of the base waveform. Double conversion is used to overcome line noise.  |
| Switching transients | Switching transients appear as glitches or fast radical changes on the smooth sine wave. They occur in both the voltage and current waveforms. A good example is a grocery store where many refrigeration units are used. Double conversion is used to overcome switching transients. |
| Frequency variation  | Deviations from the base line frequency (for example, 60Hz in U.S.). Double conversion fixes this.  |

For more information on Lenovo PDUs and UPS units, see these web page:

- ▶ PDU and UPS Technical References:  
<https://support.lenovo.com/au/en/documents/1nvo-powinf>
- ▶ Lenovo Resource Library (select the product literature for the desired PDU or UPS):  
<http://www3.lenovo.com/us/en/data-center/server-library/>

## Summary: The role of efficiency in TCO

All of the power management and efficiency features described in this paper are designed to lower the total cost of ownership (TCO) for the server solution in the data center, specifically the operating costs. Of the solutions that follow, each contributes to lowering the total power consumption, making it easier to control the impact of power consumption in your data center:

- ▶ 80 PLUS Titanium power supplies
- ▶ Advanced liquid cooling
- ▶ Support for higher operating temperatures
- ▶ Intelligent power management
- ▶ Ease-of-use attributes

In assessing the total cost of ownership and the benefits of an energy efficient solution, several areas need to be considered, including:

- ▶ Hardware and software costs
- ▶ The duration of time that each server node is powered on and off

- ▶ Server wattage when powered on and running the customer workload
- ▶ Server wattage when powered off
- ▶ Available workload capacity of each server node
- ▶ Electrical rate at the data center location
- ▶ Power factor surcharges and penalties
- ▶ Total heat load for the data center servers operating at full workload capacity
- ▶ Cooling infrastructure topology and efficiency
- ▶ Administrative costs of a standard solution versus an intelligent solution

Lenovo has many resources available to help our clients understand the TCO associated with ThinkSystem solutions. Contact a Lenovo server representative for more information.

Links related to data center energy efficiency and assessments:

- ▶ Lenovo Server TCO Calculator:  
<http://alinean.com/tool-examples-lenovo-server-tco-calculator/>
- ▶ Lenovo data center information:  
<http://www.lenovopartnernetwork.com/datacenter>

## Additional resources

This section offers further information about topics related to power and energy efficiency:

- ▶ “System power states”
- ▶ “Software considerations” on page 29
- ▶ “Lenovo Capacity Planner” on page 30
- ▶ “Lenovo XClarity Energy Manager” on page 31
- ▶ “Efficiency definitions” on page 31

## System power states

Several power states exist in ThinkSystem servers. The entire server or individual subsystems in the server can be placed into different power states to reduce power consumption and optimize efficiency. Some of the power state transitions are initiated by the user. Others are initiated automatically if they are enabled. Some power states are used to save power when the server is idle, while others are used to increase efficiency when it is running.

The following sections provide an explanation of each group of power states utilized on ThinkSystem servers. Following that, the relationship and hierarchy of the power states are explained.

### G-states

G-states are *global server states* that define the operational state of the entire server. As the number of the G-state increases, there is additional power saved. However, as shown in Table 4, the latency to move back to G0 state also increases.

The user typically initiates G-state transitions. For example, when an OS is shutdown, the server is moved from G0 state to G2 state.

Table 4 G-states

| G-State | Can Applications Run? | Relative Wake up Latency | OS Reboot Required | Comments   |
|---------|-----------------------|--------------------------|--------------------|--|
| G0      | Yes                   | None                     | No                 | System is fully on but some components might be in a power savings state.  |
| G1      | No                    | Short to medium          | No                 | Standby or hibernate mode under Windows. See S-states.   |
| G2      | No                    | Long                     | Yes                | System is in a soft-off state. For example, the power switch was pressed. System draws power from AUX rail of power delivery circuit. Power supply might or might not be switched off. |
| G3      | No                    | Longest                  | Yes                | AC power is removed. OS is not loaded. Server is only receiving power from backup batteries for RTC, CMOS, and possibly RAID data.   |

----- Higher Power ----- ^  
 ----- Higher Latency ----- <<

## S-states

S-states define the *sleep state* of the entire server. Table 5 describes the various sleep states. S-states can be initiated either by the user, using an inactivity timer on the OS, or with a higher level workload management software.

**Note:** Even though the ACPI specification defines the S2 and S4 states, these states are not enabled in ThinkSystem servers.

Table 5 S-states

| S-State | G-State | BIOS reboot required | OS reboot required | Relative power | Relative latency | Comments   |
|---------|---------|----------------------|--------------------|----------------|------------------|--|
| S0      | G0      | No                   | No                 | 6X             | 0                | System is fully on but some components might be in a power savings state.  |
| S1      | G1      | No                   | No                 | 2.5X           | 1%               | Also known as <i>Idle</i> , <i>Standby</i> —if S3 not supported. Typically, when the OS is idle, it will halt the CPU and blank the monitor to save power. No power rails are switched off. This state might go away on future servers.      |
| S2      | G1No    | No                   | No                 | N/A            | N/A              | CPU caches are powered down. No known server or OS supports this state.  |
| S3      | G1      | No                   | No                 | 1.1X           | 10%              | Also known as “Standby”, “Suspend-to-RAM”. The state of the chipset registers is saved to system memory and memory is placed in a low-power self-refresh state. To preserve the memory contents, power is supplied to the DRAMs in S3 state. |
| S4      | G1      | Yes                  | No                 | X              | 90%              | Also known as “Hibernate”, “Suspend-to-disk”. The state of the OS (all memory contents and chip registers) is saved to a file on the HDD and the server is placed in a soft-off state. This state is not used by ThinkSystem servers.        |
| S5      | G2      | Yes                  | Yes                | X              | 100%             | Server is in a soft off state. When turned back on, the server must completely re-initialize with POST and the OS.   |

Higher Power ^^^^

Higher Latency <<<<

Just as with G-states, higher numbered sleep states save more power but there is additional latency when the system transitions back to S0 state. The middle state, S3, offers a good compromise between power savings and latency.

## C-states

C-states are *CPU idle power saving states*. C-states higher than C0 only become active when a CPU core is idle for a period of time. If a process is running on a CPU core, the core is always in C0 state. If hyper threading is enabled, the C-state resolves down to the physical core. For example, if one hyper thread is active and another hyper thread is idle on the same core, the core will remain in C0 state.

C-states can operate on each core separately or the entire CPU package. The CPU package is the physical chip in which the CPU cores reside. It includes the CPU cores, caches, memory controllers, PCI Express interfaces, and miscellaneous logic. The non-CPU core hardware inside the package is commonly referred to as the *uncore*.

Core C-states transitions are driven by interrupts or the OS scheduler with MWAIT commands. The number of cores in C3 or C6 also impacts the maximum turbo frequency that is available. If maximum peak performance is desired, enable all of the CPU C-states.

Package C-state transitions are autonomous. No OS awareness or user intervention is required. The package C-state is equal to the lowest numbered C-state that any of the CPU cores is in at that point in time. Additional logic inside the CPU package monitors all of the CPU cores and places the package into the appropriate C-state.

Note that CPU C-states do not directly map to ACPI C-states. The reason for this is historical. ACPI C-states range from C0 to C3. At the time when they were defined, there were no CPUs that supported the C6 state. So the mapping was 1:1 (ACPI C0 =CPU C0, ACPI C1=CPU C1, etc.). Newer CPUs, however, support the C6 state. Depending on the class of CPU used, the ACPI to CPU C-state mapping can vary above ACPI C1 state. Typically, to get the maximum power savings, the highest numbered ACPI state will map to the highest numbered CPU C-state. Some examples are shown in Table 6.

Table 6 C-state mapping

| ACPI C-state | CPU C-State for Intel E5-2600 and E5-2600 v2 | CPU C-State for Intel Xeon Scalable Family |
|--------------|--|--|
| C0           | C0   | C0   |
| C1           | C1   | C1   |
| C2           | C3   | C6   |
| C3           | C6   | Not used                                   |

Shown in Table 7 on page 25 is a description of each core and package C-state.



Table 7 G-states

| C-state | CPU core state   | CPU core power / latency approximation <sup>a</sup> | CPU package state  | CPU package power and latency approximation <sup>b</sup> |
|---------|--|---|--|--|
| C0      | Core is fully on and executing code.<br>L1 cache is coherent.<br>Core power is on.                           | 100% at Pn / 0nS                                    | At least one core is in C0 state.  | 100% / 0nS   |
| C1      | Core is halted.<br>L1 cache is coherent.<br>Core power is on.  | 30% / 5uS   | NA –core only state  | NA   |
| C1E     | NA –package only state   | NA  | At least on core is in C1 state and all others are in C1 or a higher numbered C-state.<br>All cores are running at lowest frequency.<br>VRD 2 switches to minimal voltage state.<br>PLL is on.<br>CPU package will process bus snoops.                                     | 50% / ~5uS   |
| C3      | Core is halted.<br>L1 cache is flushed to last level cache.<br>All core clocks stopped<br>Core power is on.  | 10% / 50uS  | At least one core is in C3 state and all others are in C3 or a higher numbered C-state.<br>VRD 2 is in minimal voltage state.<br>PLL is off.<br>Memory is placed in self-refresh.<br>L3 shared cache retains context but is inaccessible.<br>CPU package is not snoopable. | 25% / ~50uS  |
| C6      | <ul style="list-style-type: none"> <li>▶ L1 cache is flushed to LLC.</li> <li>▶ Core power is off</li> </ul> | 0% / 100uS  | <ul style="list-style-type: none"> <li>▶ All cores are in C6 state.</li> <li>▶ Same power saving features as package C3 plus some additional uncore savings.</li> </ul>  | 10% / ~100uS   |

----- Higher Power ----->>>

<<<----- Higher Latency -----

a. The number of C-states and the specific power savings associated with each C-state is dependent on the specific type and SKU of CPU installed.  
 b. Voltage regulator device (VRD).

## P-states

P-states are defined as the *CPU performance states*. Each CPU core supports multiple P-states and each P-state corresponds to a specific frequency (refer to Table 8). P0 is the fastest frequency. P<sub>n</sub> (where *n* is the maximum numbered P-state for the installed CPU) is the slowest frequency. Note, that P0 can run above the rated frequency for short periods of time if turbo mode is enabled. The exact turbo frequency for P0 and the amount of time the core runs at the turbo frequency is controlled autonomously in hardware.

Like core C-states, P-states are controlled by the OS (OS) scheduler. The OS scheduler places a CPU core in a specific P-state depending on the amount of performance needed to complete the current task. For example, if a 2GHz CPU core only needs to run at 1GHz to complete a task, the OS scheduler will place the CPU into a higher numbered P-state (slower frequency).

Each CPU core can be placed in a different P-state. Multiple threads on one core (for example, hyper threading) are resolved to a single P-state. P-states are only valid when the CPU core is in the C0 state. P-states are sometimes referred to as dynamic voltage and frequency scaling (DVFS) or enhanced Intel SpeedStep technology (EIST).

Table 8 P-states

| P-state          | CPU Frequency Approximation (100% =rated CPU frequency) | Description   |
|------------------|---|---|
| P0               | 100 to ~130% (with turbo)                               | CPU can run at the rated frequency indefinitely or at a turbo frequency greater than the rated frequency for short periods of time.<br>Note, turbo is an opportunistic feature. The turbo frequency and the time that turbo can be sustained is not guaranteed. |
| P1               | ~90 to 95%  | Intermediate P-state.   |
| :                |   |   |
| P <sub>n-1</sub> | ~45 to 60%  | Intermediate P-state.   |
| P <sub>n</sub>   | 1200 MHz  | Minimum frequency that CPU core can execute code.   |

The exact frequency breakdown for the P-states varies with the rated frequency and power of the specific CPU SKU used.

In addition to controlling the core frequency, P-states also indirectly control the voltage level of the VRD that is supplying power to the CPU cores. As the core frequency is reduced from its maximum value, the VRD voltage is automatically reduced down to a certain point. Eventually, the VRD will be operating at the minimum voltage that the CPU cores can tolerate. If the core frequency is lowered beyond this point, the VRD voltage will remain at the minimum voltage. This is illustrated in Figure 13 on page 27.

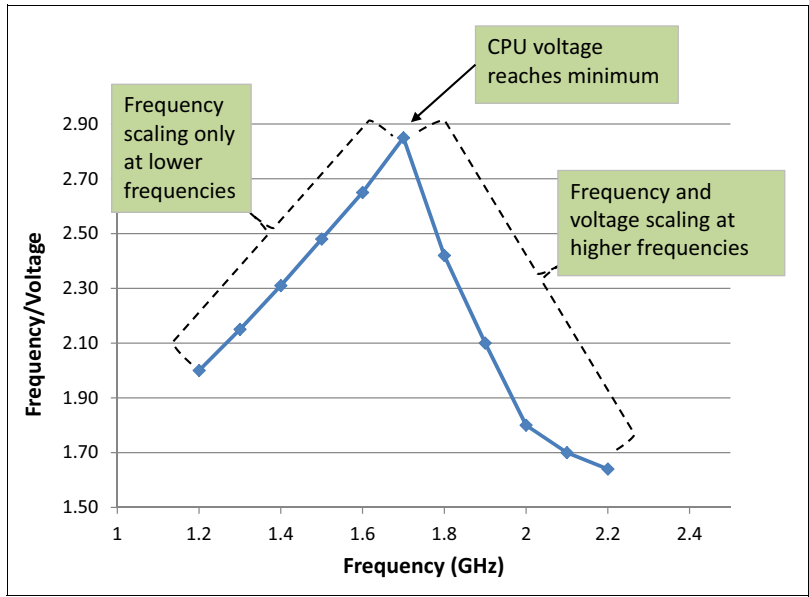


Figure 13 The effect of VRD voltage

Typically, the most efficient operating point is at the peak of the curve.

### D-states

D-states are *subsystem power saving states*. They are applicable to devices such as LAN, SAS, and USB. The OS can automatically transition to different D-states after a period of time or when requested by a device driver. All D-states occur when the server is in S0 state. Refer to Table 9 for a description of each D-state.

Table 9 D-states

| D-state           | Device power | Device context | Description  |
|-------------------|--------------|----------------|--|
| D0                | On           | Active         | Device is fully on. All devices support D0 by default even they do not implement the PCI Power Management specification.                       |
| D1                | On           | Active         | Immediate power state. Lower power consumption than D0. Exact power saving details are device specific.  |
| D2                | On           | Active         | Immediate power state. Lower power consumption than D1. Exact power saving details are device specific.  |
| D3 hot (ACPI D2)  | On           | Lost           | Power to device is left on but the device is placed in a low power state. Device is unresponsive to bus requests.                              |
| D3 cold (ACPI D3) | Off          | Lost           | Power to device is completely removed. All devices support D3 by default even if they do not implement the PCI Power Management specification. |

----- Higher Power -----  
 <<<----- Higher Latency -----

## M-states

M-states control the *memory power savings*. The memory controller automatically transitions memory to the M1 or M2 state when the memory is idle for a period of time. M-states are only defined when the server is in S0 state. Table 10 on page 28 shows a description of each M-state as well as the relative power and latency associated with each one.

Table 10 M-states

| M-state | Power / latency approximation | Description   |
|---------|-------------------------------|---|
| M0      | 100% at idle / 0              | Normal mode of operation  |
| M1      | 80% / 30nS                    | Lower power CKE mode. Rank power down.  |
| M2      | 30% / 10uS                    | Self-refresh. Operates on all DIMMs connected to a memory channel in a CPU package. |

## Relationships among the power states

Figure 14 shows an overview of the relationship among the power states of the server.

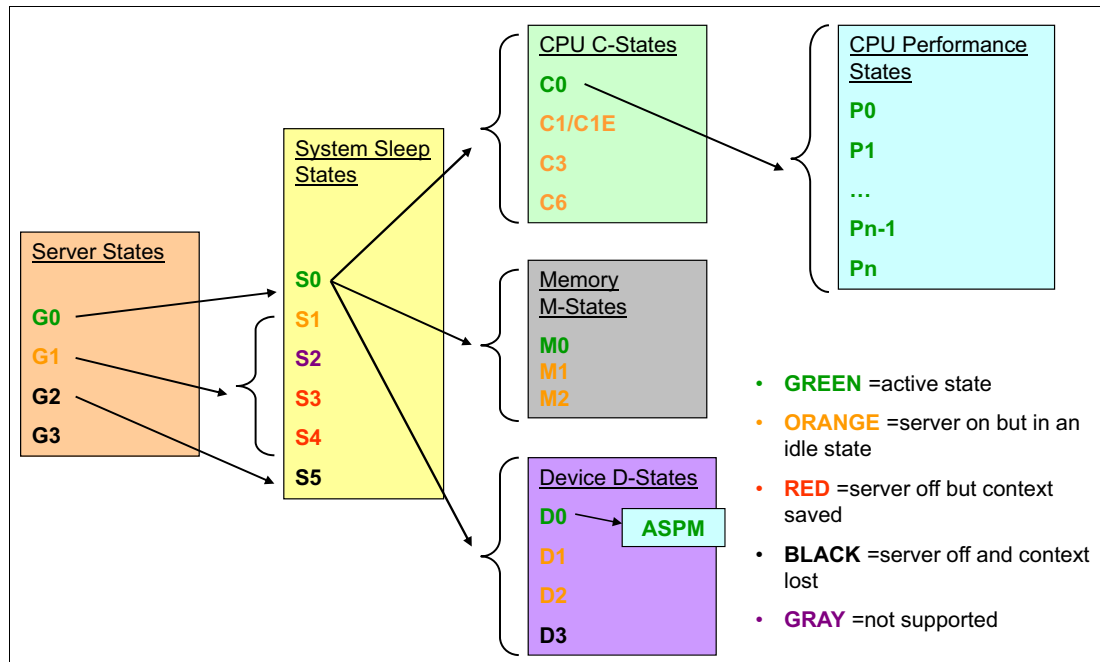


Figure 14 Power state interaction

There is a hierarchy among the power states. At the highest level, the G-states represent the overall state of the server. The G-states map to the S-states (system sleep states). Progressing to the right Figure 14, there are subsystem power states that represent the current state of the CPU, memory, and subsystem devices. As shown by the arrows, certain power states cannot be entered if higher level power states are not active. For example, for a CPU core to be in P1 state, the CPU core also has to be in C0 state, the system has to be in S0 system sleep state, and the overall server has to be in G0 state.

With any of the power savings states, there is a trade-off between power savings and latency. For example, enabling the CPU C6 state allows CPU cores to be completely turned off, which saves power. However, because the CPU cores are powered down, it takes additional time to restore their state when they transition back to the C0 state. If maximum overall performance is desired, all power saving states can be disabled. This will minimize latencies to transition into and out of the power states, but at the same time power will be increased dramatically. At

the other extreme, if power settings are optimized for maximum power savings, performance can suffer due to long latencies. For most applications, the default system settings offer a good balance between performance and efficiency. If necessary, the defaults can be changed if increased performance or power savings are desired.

For additional information about system power states, refer to the ACPI, Advanced Configuration and Power Interface (<http://www.acpi.info/>).

## Software considerations

This section describes methods for adjusting the power usage in your environment:

- ▶ Unified Extensible Firmware Interface (UEFI) and OneCLI
- ▶ OS profiles
- ▶ Steady state turbo mode.

### UEFI and OneCLI

ThinkSystem servers contains many UEFI settings that can be changed either by entering setup (press F1 at POST) or with OneCLI. Carefully consider making changes to these settings for achieving optimal efficiency. For additional information about each UEFI setting, refer to the User's Guide for the ThinkSystem server of interest. We suggest that users who are unfamiliar with the low-level settings chose one of the preset operating modes:

- ▶ *Minimal Power mode* strives to minimize the absolute power consumption of the system while it is operating. The trade-off is that performance might be reduced in this mode depending on the application that is running
- ▶ *Efficiency – Favor Power mode* maximizes the performance/watt efficiency with a bias towards power savings. It provides the best features for reducing power and increasing performance in applications where maximum bus speeds are not critical.
- ▶ *Efficiency – Favor Performance mode* optimizes the performance/watt efficiency with a bias towards performance. In “Efficiency – Favor Performance” mode, bus speeds are derated as they are in “Efficiency –Favor Power” mode. “Efficiency –Favor Performance” mode is the default mode.
- ▶ *Maximum Performance mode* will maximize the absolute performance of the system with little regard to power consumption. Things like fan speed and heat output of the system might increase in addition to power consumption. Efficiency of the system might go down in this mode but the absolute performance can go up depending on the benchmark that is run.
- ▶ *Custom* settings can be determined, which enable you to individually modify any of the low-level settings that are preset in the four preset modes listed here.

For more information, see the ThinkSystem Information Center:

<http://thinksystem.lenovofiles.com/help/index.jsp>

### OS power profiles

Similar to UEFI settings, operating systems also have profiles and settings that allow a user to adjust for maximum power savings, efficient operation, or maximum performance. Changing low-level settings for a plan is only recommended for advanced users who are familiar with each of the low-level settings. For most users, the defaults for the preset mode are fine. Refer to the documentation for your specific OS for more information.

An example of a power plan selection pane under typical Microsoft Windows operating systems is shown in Figure 15 on page 30.

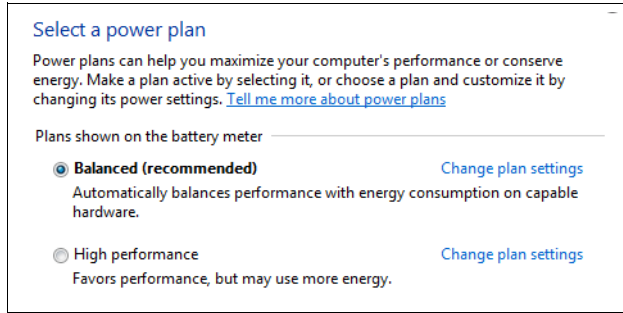


Figure 15 Windows OS power plans

When comparing the UEFI versus OS power management and efficiency settings, think of the UEFI settings as controlling the outer limits of the power management and efficiency settings. The OS operates within the limits set by UEFI. In addition, the OS can place further restrictions on the power management and efficiency of the server.

## Lenovo Capacity Planner

Using the Capacity Planner, the worst case and typical power consumption for a specific server configuration can be determined before a purchase is made.

<https://datacentersupport.lenovo.com/us/en/products/solutions-and-software/software/lenovo-capacity-planner/solutions/ht504651>

Figure 16 on page 30 shows the Lenovo Capacity Planner.

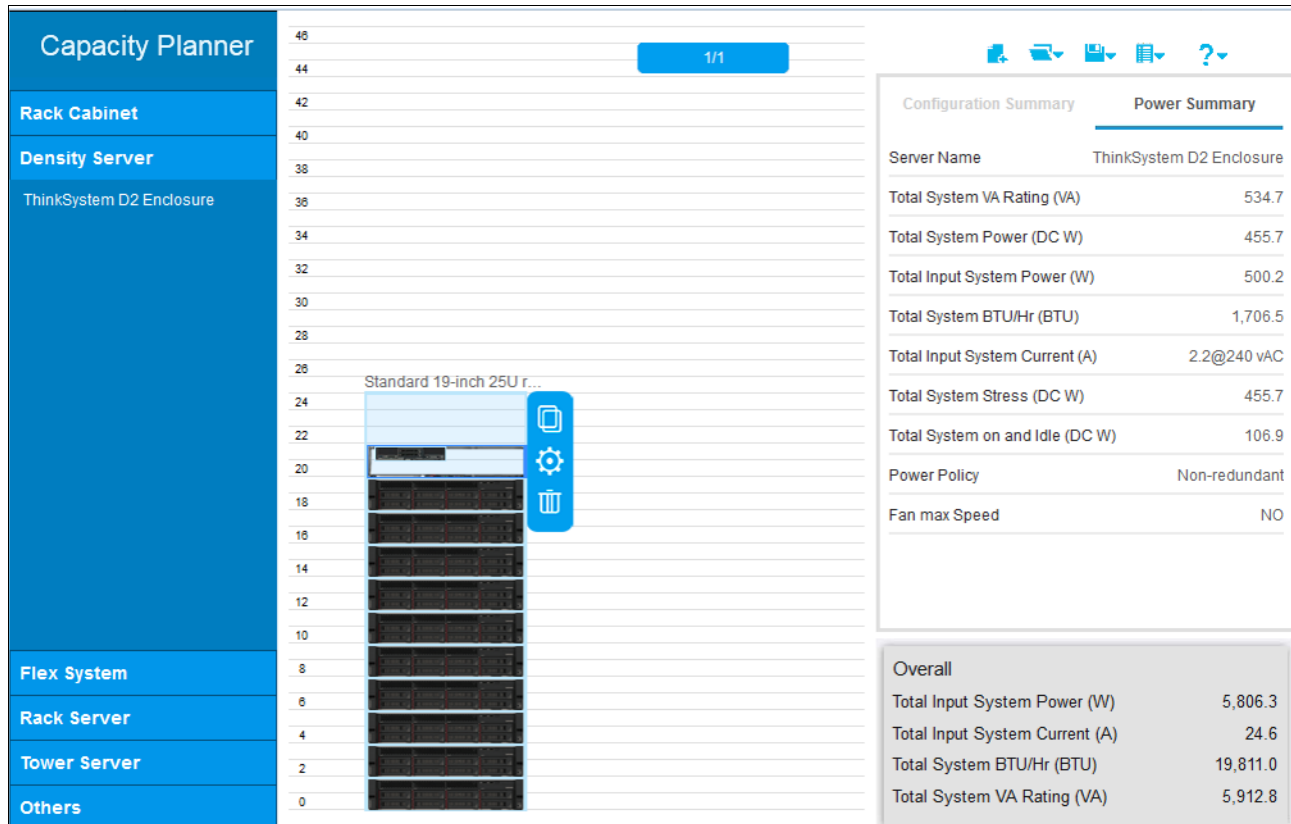


Figure 16 Lenovo Capacity Planner

# Lenovo XClarity Energy Manager

Lenovo XClarity Energy Manager (LXEM) provide a dashboard for monitoring server power, energy, and temperature metrics. Features include:

- ▶ Data center hierarchy management -physical and logical groups
- ▶ Real-time power, temperature, and utilization monitoring
- ▶ Individual or group power caps
- ▶ Power, temperature, and utilization trending analysis
- ▶ Energy optimization analysis
- ▶ Emergency power reduction mode
- ▶ Agentless management
- ▶ Supports end-points from multiple OEMs

Figure 17 on page 31 shows the XClarity Energy Manager interface.

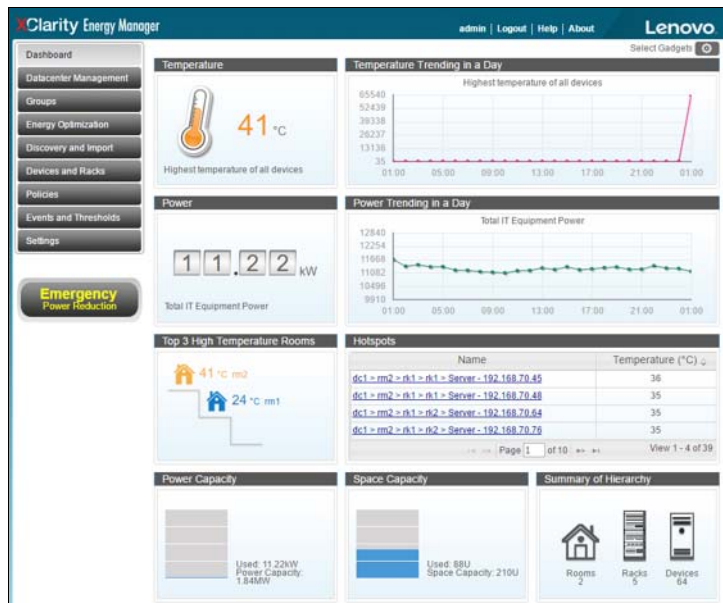


Figure 17 XClarity Energy Manager

## Efficiency definitions

When measuring the efficiency of a server, there are three ways to accomplish this.

Electrical conversion efficiency (ECE) measures how much power is lost to convert one power level to another (e.g. AC-to-DC or DC-to-DC conversion). If a power supply converts 220V AC to 12V DC and it is 95% efficient for a 500W load, 5% of the input power is converted to heat and is typically dissipated with a fan built into the power supply. In this example, 526W AC is required, 500W is delivered to the load, and 26W is dissipated as heat. Power supply and VRD efficiency has improved dramatically in recent years but no electrical circuit is ideal and some power is always dissipated.

$$ECE = \text{Power out} / \text{Power In}$$

Power usage effectiveness (PUE) measures how much power is lost in the data center to relative to actual IT equipment power. The overall PUE depends on how close to the true compute power load that the power out measurement is taken and also what ancillary loads are included in the calculation (e.g. lights, humidification, UPS, CRACs, chillers, etc.)

**PUE** = Total Facility Power / IT Equipment Power = 1 / data center Efficiency

Performance/watt efficiency (P/W E) is defined as how much performance can be achieved for every watt of power consumed.

**P/W E** =  $\sum$  Performance /  $\sum$  Power

P/W E focuses on the server, chassis, and rack efficiency. By comparison, ECE or PUE can be extended to the data center level or power station level.

## About the Author

**Robert (Bob) R. Wolford** is a Senior Engineering Staff Member for power management and efficiency at Lenovo. He covers all technical aspects associated with power metering, management, and efficiency. Previous assignments have included roles as a lead systems engineer for workstation products, video subsystems building block owner, signal quality and timing analysis engineer, gate array designer, and product engineering. In addition, he did a stint as a Technical Sales Specialist for System x. Bob has 11 issued patents and 3 pending patent applications. He enjoys working with young people during National Engineers Week, and at local science fairs. He holds a Bachelor of Science degree in Electrical Engineering with Distinction from the Pennsylvania State University.



# Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.  
1009 Think Place - Building One  
Morrisville, NC 27560  
U.S.A.  
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on April 25, 2019.

Send us your comments via the **Rate & Provide Feedback** form found at <http://lenovopress.com/1p0780>

## Trademarks

Lenovo, the Lenovo logo, and For Those Who Do are trademarks or registered trademarks of Lenovo in the United States, other countries, or both. These and other Lenovo trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by Lenovo at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of Lenovo trademarks is available on the Web at <http://www.lenovo.com/legal/copytrade.html>.

The following terms are trademarks of Lenovo in the United States, other countries, or both:

Lenovo®  
Lenovo XClarity™

Lenovo(logo)®  
System x®

ThinkSystem™

The following terms are trademarks of other companies:

Intel, Intel SpeedStep, Xeon, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows Server, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.