

# Power efficiency and power management in HP ProLiant servers

## Technology brief

Introduction.....	2
Built-in power efficiencies in ProLiant servers.....	2
Optimizing internal cooling and fan power with Sea of Sensors.....	2
Active Cool fans in HP BladeSystem enclosures.....	3
High-efficiency DC power regulation and phase shedding.....	4
Efficient Common Slot Power Supplies.....	4
Managing ProLiant power consumption using configurable power management options.....	6
Memory Interleaving and memory speed.....	6
QPI Link Power Management.....	6
Minimum processor idle power state (C-states).....	7
PCIe Generation 2 Support.....	7
HP Power Regulator.....	7
Power Profiles.....	7
Power efficiency versus server performance.....	8
Power monitoring and control of ProLiant servers.....	8
Monitoring server power use with iLO, Onboard Administrator, and Insight Control.....	8
HP Dynamic Power Capping.....	9
Power consumption for ProLiant servers by generation.....	10
Conclusion.....	11
For more information.....	12



# Introduction

In today's data center, any improvement in server power consumption can pay large dividends. The less power consumed by servers, storage and networking devices, the less heat generated, and the lower the losses for power distribution. This also reduces the cost of heat extraction at the end of the power life cycle. Lowering server power consumption, along with controlling that consumption, also allows you to deploy more servers per rack and still keep within a data center's overall power budget.

Beginning with the ProLiant G6 servers, HP engineers have spent a great deal of effort in research, design and development to make ProLiant servers as power-efficient as possible. They have accomplished it by rethinking and re-engineering every key element in the server platform—from power supplies to cooling technology to the monitoring and control of system processors. This technology brief describes the individual power savings technologies that we have incorporated into ProLiant servers. It also reviews the built-in ProLiant power monitoring and control functions. ProLiant power efficiency and ProLiant power monitoring and control are parts of HP's overall power management strategy that includes servers, the server rack and the data center.

## Built-in power efficiencies in ProLiant servers

ProLiant G6 and G7 servers use many different technologies to increase their overall power efficiency—from more efficient cooling to better power supplies to more sophisticated DC power regulation. These built-in power efficiencies result in lower power consumption without sacrificing system performance.

### Optimizing internal cooling and fan power with Sea of Sensors

Starting with G6, ProLiant servers use a more sophisticated system to monitor and control server cooling. This technology is popularly referred to as the "Sea of Sensors", and it consists of several different subsystems working together to achieve significantly improved server cooling efficiency.

#### **Thermal Sensors**

ProLiant servers use up to 64 individual sensor inputs to obtain the temperature at all points inside the system. This includes inputs from as many as 20 temperature sensors on the system board—as well as separate temperature sensors for DIMMs, the processors, and disk drives. In fact, more than 64 physical sensors may be present in the system since some components have multiple sensors. The CPUs, for example, can contain eight separate internal sensors that are combined into a single sensor input. This Sea of Sensors provides a much more accurate view of the temperatures at all points inside the server, including the hot spots that require additional cooling.

#### **System Fans and Fan Control**

To take advantage of these sensors, today's ProLiant servers use more fans (up to 12 for some systems, but typically 4 – 8) than older systems. Each fan in the server may have the outputs from multiple sensors mapped to it. Similarly, a given temperature sensor may also map to multiple fans. The result is a finely tuned system that provides more fan cooling only where it is needed without wasting power where additional cooling is not required.

#### **The PID control algorithm**

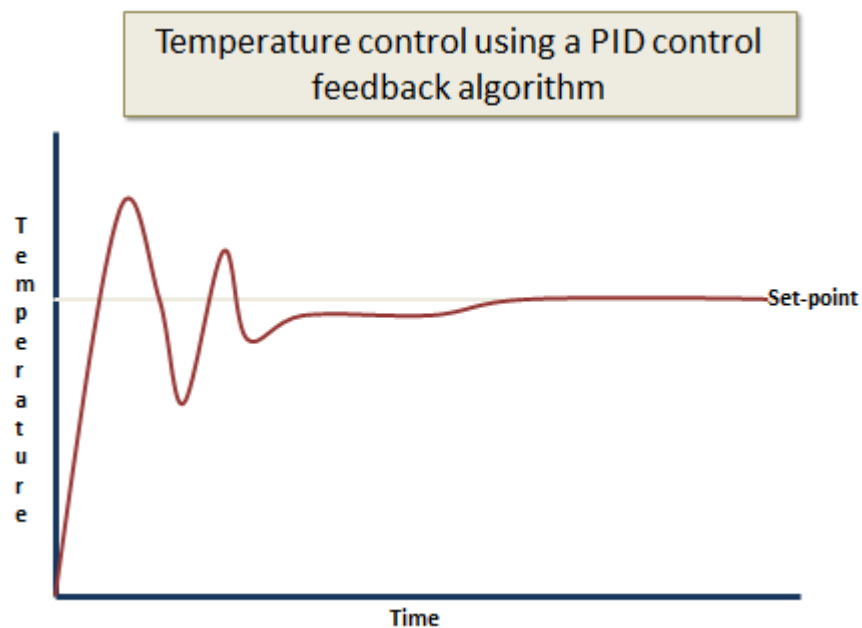
In addition to providing more granular fan control, the new iLO-based fan control system also uses a proportional-integral-derivative (PID) control feedback algorithm to control the speed for each fan. Unlike the older system, which simply set fan speed based on a temperature reading, the PID algorithm continuously adjusts fan speeds to try to maintain a particular set-point temperature

within each given area of the server. It does this using a formula that takes into account three separate factors.

- Measured temperature compared to a target temperature, or set point. This set point can be different for each sensor location, depending on desired operating temperature for the components it is measuring.
- The history of temperature change based on previous measurements
- Predicted future temperature values, based on the current rate of change

Figure 1 illustrates how the PID algorithm quickly and efficiently manages to a set-point temperature. This approach delivers more efficient and timely temperature control that saves on fan power consumption by not overcooling parts of the server.

**Figure 1.** The PID algorithm manages to a set-point temperature.



### Power savings with Sea of Sensors

The innovations in the Sea of Sensors technology work together to deliver significant savings on cooling power consumption in ProLiant servers. As an example, a typically configured ProLiant DL380 G6 or G7 server saves about 2 watts of power at idle and up to 24 watts at higher server utilization when compared to previous server generations. HP has Implemented Sea of Sensors technology in all of the ProLiant, including the ML, DL and SL lines as well as the BladeSystem enclosures.

### Active Cool fans in HP BladeSystem enclosures

In addition to the Sea of Sensors technology, HP uses its proprietary Active Cool fan technology to deliver additional power savings and cooling efficiencies in c3000 and c7000 BladeSystem enclosures. Active Cool fans use ducted fan technology combined with a high-performance motor and impeller. These fans deliver high airflow at high backpressure while using significantly less

power than a traditional blower-style fan solution. These fans can cool 16 server blades using as little as 150 watts of power—about 50% of the power typically required with other implementations.

## High-efficiency DC power regulation and phase shedding

DC voltage regulators on the server system board convert the 12 V dc power supplied by the power supply into the 5v, 3v, and other lower voltages used by the various system components. The DC voltage regulators in ProLiant G6 and later servers use high quality components to deliver peak efficiency, maintaining greater than 90% efficiency over a broad range of power demands. This results in an 8-point gain in DC power efficiency over previous generations of servers.

ProLiant servers also use phase shedding technology to increase the efficiency of power delivery. Switching DC power regulation uses multiples pulses, or phases, to deliver power to the components on the system board. With phase shedding the system adjusts the number of DC power phases delivered to a particular subsystem or component to match its power requirements. This produces additional increases in power efficiency beyond those provided by high quality voltage regulators. ProLiant servers use phase shedding to provide more efficient power delivery to the memory subsystems by adjusting the number phases delivered based on the number of DIMMs installed. G6 and later systems also use dynamic phase shedding. Dynamic phase shedding continuously adjusts the number of DC phases delivered to the system processors based on their real-time power consumption.

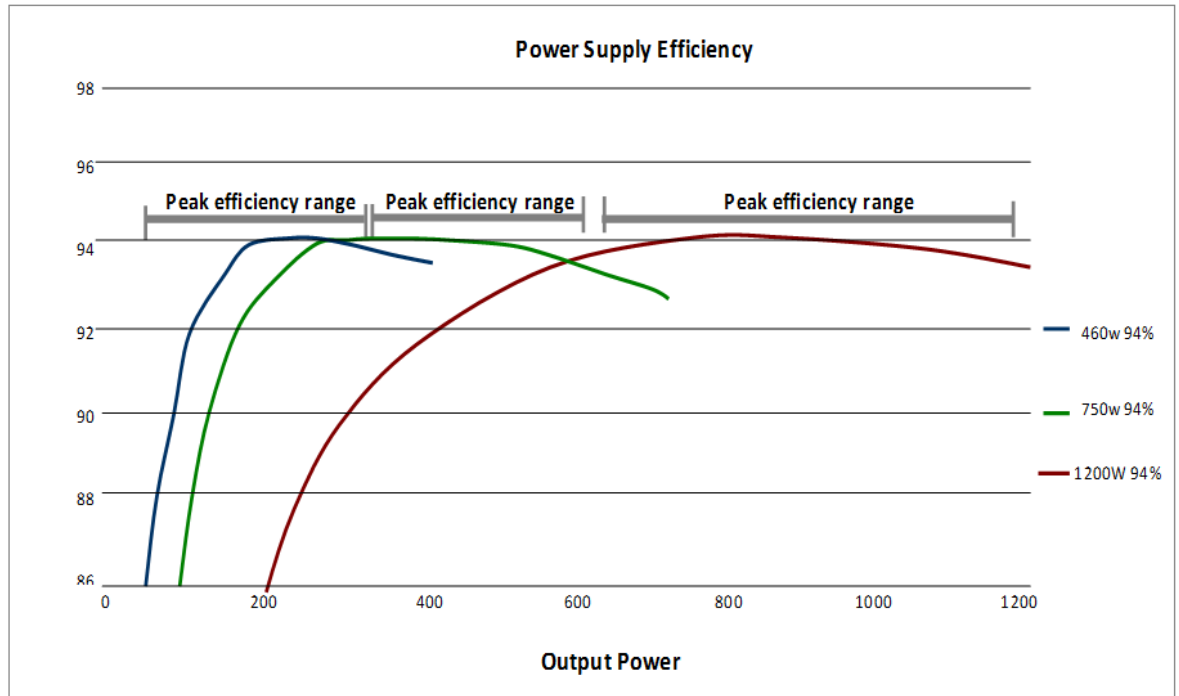
## Efficient Common Slot Power Supplies

ProLiant servers use the HP common slot (CS) power supplies. These power supplies provide a single mechanical form factor that allows you to use them interchangeably across multiple server platforms. HP Platinum CS power supplies also work with HP Intelligent Power Discovery (IPD) technology. IPD automates the discovery and management of your data center's power and IT equipment connectivity (topology). You can find more information about IPD technology in the "HP power and cooling technologies for the data center" technology brief. It is available at <http://h20000.www2.hp.com/bc/docs/support/SupportManual/c02018535/c02018535.pdf>.

### Right-sizing your power supply

HP CS power supplies provide you with the opportunity to "right-size" your power supply for a given server configuration and achieve maximum power efficiency. Power supplies are typically most efficient at converting input AC into the DC output used by the server components when they are operating at about 50% of their maximum output capacity. HP produces three common slot power supplies with maximum output power of 460 W, 750 W, and 1200 W, respectively. Figure 2 shows the efficiency curves for each these power supplies. Using CS power supplies, you can select a power supply that will operate at maximum efficiency for the planned server power load. A single 750 W power supply would be the optimal choice for a server that has an average power load of 350 W, because it would be 92% efficient in converting AC input to DC output at that load. Installing a 1200 W capable power supply in that server would be a poor choice, because it would only operate at 88% conversion efficiency.

**Figure 2.** A graph of the efficiency curves for Platinum CS power supplies illustrates their peak efficiency ranges.



### Silver, Gold and Platinum Power Supplies

In addition to offering three different wattages of CS power supplies, HP also produces these supplies at different levels of 80 PLUS<sup>®</sup> certification. The 80 PLUS program is an industry standard performance specification that requires power supplies to be 80% or greater energy efficient at several different power loads. The program also includes several different levels of certification based on increasing power supply efficiency. 80 PLUS Silver power supplies are required to be 88% efficient at 50% load, while 80 PLUS Gold supplies must be 92% efficient. 80 PLUS Platinum supplies must be 94% efficient. HP offers CS power supplies at multiple 80 PLUS certification levels, allowing you to choose the maximum level of power-supply efficiency that you require.

### Improving redundant power supply operation using high-efficiency mode

Using redundant power supplies delivers increased reliability, but it may result in decreased power efficiency. In typical servers, if all redundant power supplies are online simultaneously, the power drawn from each supply will be less. If this lowered power draw goes below 50% of the supply's maximum capability, power-supply efficiency declines. (see Figure 2).

To remedy this situation, the CS power supplies in ProLiant ML and DL servers support both a "balanced" and a "high-efficiency" operating mode. In high-efficiency mode, one of the redundant power supplies is kept in a standby state. This allows the remaining supply to support the full power load and operate at higher power efficiency in most situations. Special circuitry ensures that the standby power supply can be brought online quickly enough to maintain server power if the primary power supply fails. Additionally, for each server you can also choose which power supply (even or odd) is the standby supply. This allows you to balance the power draws across circuits in those installations where you have redundant power circuits to each rack. You can set the power supply mode using the ROM-based Setup Utility (RBSU) in ProLiant servers.

With HP BladeSystem servers, HP Dynamic Power Saver performs a similar function by allowing the enclosure to operate at the highest possible power efficiency. When Dynamic Power Saver is active

for a c-Class blade enclosure, the Onboard Administrator for the enclosure increases or decreases the number of power supplies that are active in order to keep them working at their most efficient level while still maintaining redundant operation.

## Managing ProLiant power consumption using configurable power management options

Beginning with G6 ProLiant servers, you can configure certain system hardware settings that lower overall system power consumption. Unlike the built-in power efficiencies, changing these hardware settings can save you power but may involve a performance trade-off.

You can configure each of the following hardware settings using the RBSU.

### Memory Interleaving and memory speed

Memory interleaving refers to the way the system maps its memory addresses to the physical memory locations in the memory channels and DIMMs. Typically, consecutive system memory addresses are staggered across the DIMM ranks and across memory channels in the following manner:

- **Rank Interleaving.** Every consecutive memory cache line (64 bits) is mapped to a different DIMM rank.
- **Channel Interleaving.** Every consecutive memory cache line is mapped to a different memory channel.

This distribution of logical system memory addresses enables the memory controller to parallelize access better across memory channels and DIMMs when retrieving blocks of data from the system memory. The result is usually better memory throughput for the system. Interleaving, however, results in greater DIMM power consumption. This is because it increases the probability that more DIMMs will remain in their most active state for longer periods.

ProLiant servers allow you to turn off rank interleaving or turn off both Rank and Channel interleaving. With interleaving disabled, you increase the probability that a given DIMM will be inactive for a longer period and therefore go into a lower power state. In a typical operating environment, disabling interleaving can save approximately 1 watt of power consumption per DIMM, but it may affect memory performance by 5% to 10%.

You can also save power by lowering the maximum memory bus frequency for a server. Our G6 and G7 ProLiant servers have a maximum possible memory bus speed of 1333 MT/s, and the next generation servers will probably reach 1600 MT/s. Normally the system automatically configures the bus speed to maximum speed supportable for the installed memory configuration. But the RBSU allows you to set the maximum memory bus frequency to 1066 MT/s or 800 MT/s. This decreases the power consumed by the memory subsystem. But it may also decrease memory throughput, depending on your application environment.

### QPI Link Power Management

Quick Path Interconnect (QPI) links are the high-speed connections that Intel-based processors use to move data between one another in multi-processor systems. RBSU allows you to enable or disable power management of these links. When you enable QPI power management, the system saves power by putting the QPI links into a low power state when they are not active. Enabling this feature can save power with minimal impact on system performance.

## Minimum processor idle power state (C-states)

C-states represent different power-saving states for idle processors or individual processor cores. The ProLiant RBSU allows you to choose the lowest power C-State for an individual processor at idle as well as the lowest power C-state for the entire processor when all cores are idle. The C-states represent how many processor and processor core functions the system shuts down during idle cycles and, therefore, how much power is saved.

For Intel processors, C6 is the lowest power C-State and C1E is the highest. The server OS determines whether you can idle an individual core, while the system BIOS will only allow you to idle the entire processor if all cores are in their idle C-state. You can also choose to disable C-States entirely, in which case the processor and processor cores remain at full power at all times. Unlike Performance States (P-States), which specify lower power states for a running processor, C-States only control the power consumption of idle processors. Setting the minimum processor idle power to a given C-State will lower overall power consumption for processors that spend significant portions of time in the idle state. However, this increases latency when processor operation resumes.

## PCIe Generation 2 Support

This feature allows you to disable support for PCIe Generation 2.0. If disabled, all PCIe devices in the system will run at the PCIe 1.0 transfer speed of 2.5 gigabits per second (Gb/s). This includes PCIe 2.0 capable devices, which can run at 5.0 Gb/s. Disabling PCIe Gen. 2.0 support can reduce system power consumption. It can affect performance for those applications in the system that require greater than 2.5 Gb/s PCIe throughput.

## HP Power Regulator

HP Power Regulator is a standard feature on most HP ProLiant servers. We have designed it to optimize processor power consumption based on server activity. In its default Dynamic Power Savings mode, Power Regulator monitors processor activity. It uses this information to adjust the processor's power consumption continuously to the lowest level that supports the server's workload without significantly affecting performance. It does this by changing the processor performance state (P-state), which controls the frequency and voltage at which the processors are running. Power Regulator also supports several other modes of operation. In Static High Performance Mode, it keeps the system processors in their maximum P-state at all times. In Static Low Power Mode, it keeps the processors in a low P-state at all times, saving more power but potentially affecting system performance. You can find more information on Power Regulator in the "Power Regulator for ProLiant servers" technology brief. Find it at

<http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00300430/c00300430.pdf>

## Power Profiles

HP Power Profiles are pre-defined groups of settings for the ProLiant power management options discussed in the previous section. You can choose one these Power Profiles through RBSU. It then configures the corresponding power management options appropriately.

- **Balanced Power and Performance Mode.** This mode enables power management options that have negligible impact on system performance. This is the default HP Power Profile setting.
- **Maximum Performance.** This Power Profile setting delivers maximum performance without regard to system power usage.
- **Minimum Power Usage.** This setting minimizes system power usage by enabling power reduction mechanisms that may affect performance.

Table 1 compares the individual power management options and their settings for each of the Power Profiles.

**Table 1.** Power Management Settings for HP Power Profiles

Power Management Option	Maximum Performance	Balanced Power & Performance	Minimum Power Usage
Intel QPI Link Power Management (AMD HT link power mode is controlled by C-States)	Disabled	Enabled	Enabled
Memory Interleave	Full interleave	Full interleave	Disabled
Memory Speed	Auto	Auto	800 MT/s
PCI Express Generation 2 Support	Force PCI-E Generation 2	Force PCI-E Generation 2	Force PCI-E Generation 1
Minimum Processor Idle Core Power State	No C-States	C6	C6
Minimum Processor Idle Package Power State	No C-States	C6	C6
Power Regulator	Static High Performance	Dynamic	Static Low Power

## Power efficiency versus server performance

Using the power management options to lower server power consumption doesn't necessarily reduce overall system performance. The performance of a system is dependent on many variables. For example, lowering the maximum memory speed will affect the performance of the memory subsystem, but it might not affect the overall performance in an application environment that is I/O-bound. Testing a system in its intended application configuration is the best way to determine which power management options to use without affecting overall performance.

## Power monitoring and control of ProLiant servers

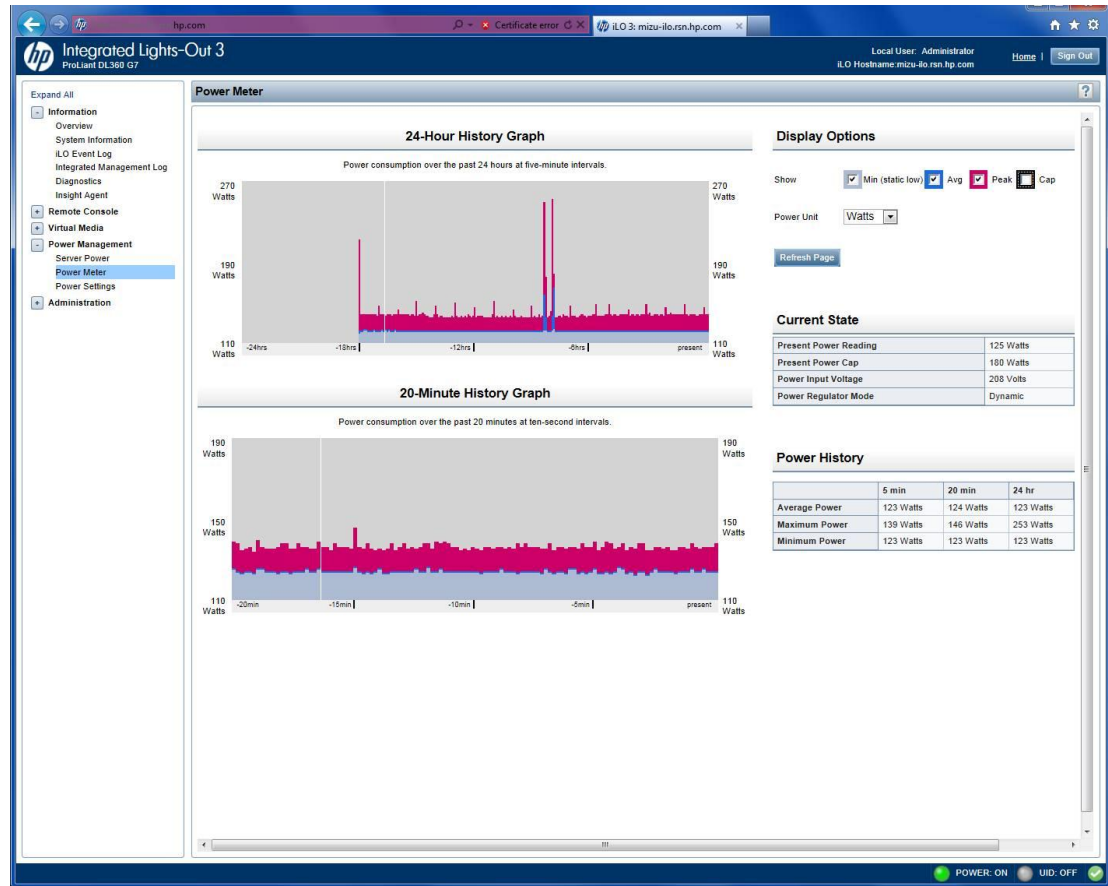
In addition to the power efficiencies and low-level power management options built into every ProLiant server, we also provide higher level power monitoring and control functions for each server. These functions help you monitor and manage ProLiant power consumption within the rack or enclosure.

### Monitoring server power use with iLO, Onboard Administrator, and Insight Control

Today's ProLiant servers monitor and report their power consumption using HP's power metering technology. The iLO management processor in each server continuously measures and records the server's power usage, which you can view using the server management interface (Figure 3). On HP BladeSystem servers, the Onboard Administrator performs the power monitoring function.



**Figure 3.** This screenshot shows server power consumption monitoring using the iLO management processor.



HP Insight Control provides another level of power monitoring and reporting functionality beyond that of iLO and the Onboard Administrator. Using Insight Control, which integrates with HP Systems Insight Manager, you can monitor the power consumption of a server over long periods and can maintain years of peak and average power consumption data.

You can use the power consumption data for a server to help determine power optimization strategies. For example, you can use the average power consumption of a server in setting appropriate power capping levels. This data is also helpful in determining whether a different sized common slot power supply would provide greater power conversion efficiency. Additionally, you can use power consumption data to help determine the power savings provided by the various power management options for a server in a given application environment.

## HP Dynamic Power Capping

We designed HP Dynamic Power Capping to help you maximize your data center power and cooling capacity by eliminating the need to overprovision power and cooling at the rack level. By removing the guesswork as to how much power a server might consume, Dynamic Power Capping allows you to fit more servers into a limited power and cooling capacity.

We have incorporated HP Dynamic Power Capping technology into our G6 and later ProLiant. Dynamic Power Capping uses sophisticated monitoring and control circuitry to safely limit maximum server power consumption to a level that you set. Because it is hardware-based, Dynamic Power Capping can control server power consumption quickly enough to ensure that sudden surges in

power demand from the servers will not trip circuit breakers in the HP Power Distribution Units (PDUs). You can use the power consumption information gathered with Insight Control to set appropriate power caps to individual servers using iLO or to groups of servers through Insight Control. Setting power caps to the peak observed power consumption of the server(s) places a defined limit on power consumption without affecting performance. This frees up additional power capacity that you can then reallocate to additional servers.

HP Enclosure Dynamic Power Capping expands the power capping concept by combining Dynamic Power Capping on individual blade servers with a control algorithm running in the BladeSystem Onboard Administrator. The algorithm automatically adjusts blade caps up and down in real time as workload changes. You can then set a total enclosure power cap that is lower than it otherwise could be using manual caps on each blade, while still not affecting the overall performance of the blades in the enclosure.

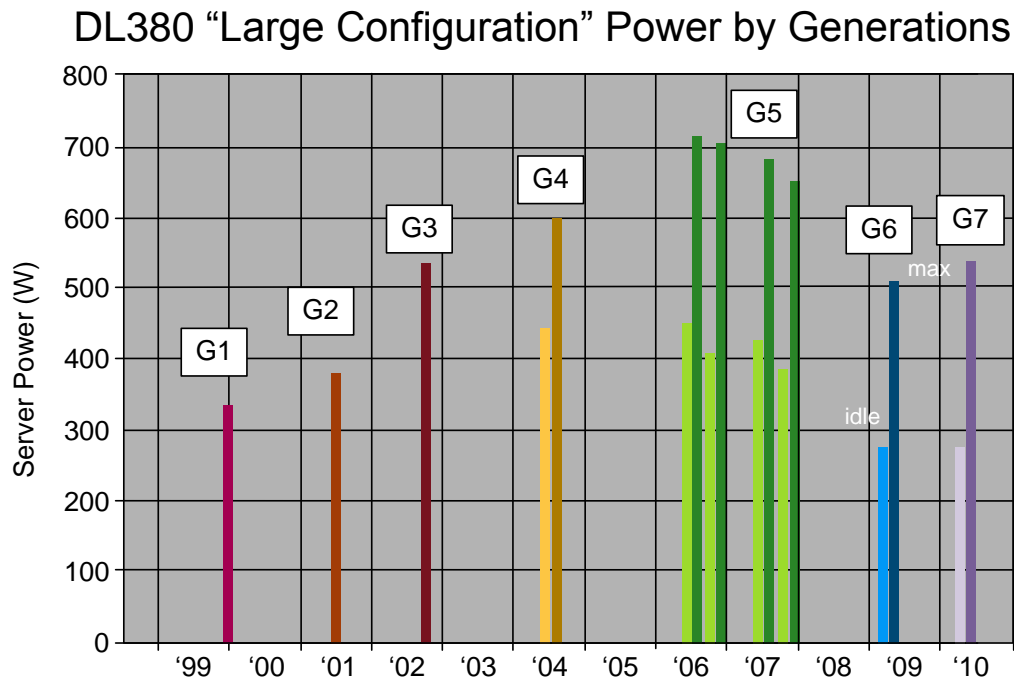
You can use Dynamic Power Capping to limit the maximum power consumption of a ProLiant server. It is not intended to improve the power efficiency of the individual server. Instead, Dynamic Power Capping helps you optimize data center provisioning by eliminating the need to overprovision power and cooling at the rack or data center level. For a more detailed discussion of HP Power Capping technology, read the "HP Power Capping and Dynamic Power Capping for ProLiant Servers" technology brief. Find it at

<http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01549455/c01549455.pdf>.

## Power consumption for ProLiant servers by generation

All of the technologies we have discussed in this paper contribute to making the current generation HP ProLiant servers much more power efficient and manageable when compared to previous generations. Figure 4 shows the calculated idle and maximum power for "large configuration" ProLiant DL380 servers across the generations. Although the ProLiant DL380 G6 and G7 servers have significantly better performance than a similarly configured DL380 G5, they consume nearly 200 W less at maximum power.

**Figure 4.** This graph illustrates the power consumption of the ProLiant DL380 across generations. The darker color shades for the G4-G7 lines show maximum power; the lighter colors show idle power levels.



## Conclusion

Making HP ProLiant servers more power efficient is a complex, interdisciplinary effort. As this paper illustrates, HP has achieved this goal by taking a comprehensive approach to the challenge of increasing server power efficiency. HP engineers have re-thought and re-engineered all aspects of power usage in the server. This includes the efficiency of power conversion as well as the optimal use of power by the processors and memory. It also includes the efficient removal of heat from the server and better monitoring and management of overall power consumption. HP engineers continue to innovate wherever possible to keep each new generation of ProLiant servers at the forefront in overall power efficiency as well as power monitoring and control.

Maximizing server power efficiency, combined with monitoring and controlling server power consumption, are key parts of HP's overall power management strategy. It extends beyond the server to the entire data center, and includes power distribution as well as cooling infrastructure. Improving power management is an ongoing effort—one that we believe is extremely important for the modern data center.

## For more information

Visit the URLs listed below if you need additional information.

Resource description	Web address
HP Common Slot power supply technology	<a href="http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA3-2676ENW.pdf">http://h20195.www2.hp.com/V2/GetPDF.aspx/4AA3-2676ENW.pdf</a>
80 PLUS program	<a href="http://www.80plus.org/">http://www.80plus.org/</a>
Power Regulator for ProLiant servers Technology brief, 5 <sup>th</sup> edition	<a href="http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00300430/c00300430.pdf">http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00300430/c00300430.pdf</a>
HP power and cooling technologies for the data center Technology brief, 2nd edition	<a href="http://h20000.www2.hp.com/bc/docs/support/SupportManual/c02018535/c02018535.pdf">http://h20000.www2.hp.com/bc/docs/support/SupportManual/c02018535/c02018535.pdf</a>
HP Power Capping and Dynamic Power Capping for ProLiant Servers Technology brief, 2 <sup>nd</sup> edition	<a href="http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01549455/c01549455.pdf">http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01549455/c01549455.pdf</a>

Send comments about this paper to [TechCom@HP.com](mailto:TechCom@HP.com)



Follow us on Twitter: <http://twitter.com/ISSGeekatHP>.

© Copyright 2012 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

80 PLUS is a registered trademark of Ecos Consulting Inc.

TC1109811, January 2012

