HPE Gen10 Servers Intelligent System Tuning

Abstract

This white paper is intended for server administrators to describe how HPE's Intelligent Server Tuning features are used to improve performance in HPE server environments.

Part Number: 882269-001
Published: July 2017
Edition: 1

Notices

Acknowledgments

Contents

Introduction

For the past several years, server-class customers have seen processor-based performance increase generation over generation. This is due in a large part to increases in core counts and more efficient instruction set architectures. Unlike the preceding decades, the base frequency of the CPU has stayed rather stable and only the number of cores has increased. Processor vendors, realizing that not all workloads benefit from increased core counts, introduced features that allow the processor to run opportunistically at higher frequencies when these extra cores or the power to run them are not being utilized.

Although these opportunistic frequency upsides can increase performance, they also introduce an unwanted side effect. Frequency shifting itself introduces computation jitter, or non-determinism, and undesirable latency. Jitter and the latency associated with it create problems for several customer segments. For example, high-frequency traders, who rely on time-sensitive transactions, cannot tolerate the microseconds of delay that can be added randomly to a trade caused by a frequency shift. These delays over time can cost a trader upwards of millions of dollars. Servers running RTOS (real-time operating systems) to control critical functionality also cannot tolerate random latencies that happen when opportunistic-frequency features are left enabled.

The current trend for latency-sensitive customers is to disable the features that normally would result in increased application performance. A trade executes faster if the processor runs faster, but if it comes at the cost of random delay, the benefit of increased performance is lost.

Starting with Gen10 servers using Intel Xeon Scaleable Processors and iLO 5, HPE has introduced a new feature that allows customers to achieve both frequency upside and low latency. Processor Jitter Control allows the customer to remove or reduce the jitter caused by opportunistic frequency management resulting in better latency response and higher throughput performance.

NOTE:

- Gen10 servers using AMD processors are not supported.
- An iLO Advanced license is required to use this feature.

Sources of jitter within the processor

The processor introduces jitter any time it executes changes in operating frequency. Several possible reasons exist for a processor to dynamically change frequencies during run time. Some of the sources that request frequency changes are driven by software, while others are driven by the processor itself.

P-states and power management

P-states are predefined performance states that are made available by the processor for software to control how much performance the processor is capable of delivering so that it can manage the power-performance efficiency of the platform. Performance states are mapped to a specific frequency at which the processor is capable of operating. Power management software instructs the processor to change P-states (frequency) in order to save power when processor utilization (demand) is low. A processor often offers several different P-states over a range of operating frequencies.

Turbo Boost

Intel's Turbo Boost allows the processor to run at higher frequencies than the base frequency guaranteed by its specification, assuming it follows certain conditions. The conditions include the amount of heat being dissipated, the temperature of the part, and the number of cores active (enabled and not idle.) When a workload is run on these processors while Turbo Boost is enabled, the processor will opportunistically switch between frequencies in an attempt to achieve the highest possible performance. But as the demands of the workload change, so can the frequencies. When frequencies change, we get frequency jitter as well as a small amount of latency that occurs, which is required to electrically change frequencies.

The net effect of having Turbo Boost enabled is that while the processor attempts to provide the maximum amount of performance within its limits, it often does so by changing frequency often.

C-states

C-states are predefined power-saving states that the processor offers to power management software to use when the operating system idles a processor core. The operating system puts the processor into one of a number of C-states that are made available. The deeper the C-state, the more power that is saved, but at the cost of longer exit latencies to return to the operating state. In an attempt to save power, C-states on Intel processors also lower the frequency of the processor. Upon exiting a C-state, the processor, running at the lowest frequency available to the C-state, must perform an additional frequency shift to return to the previously requested P-state by power management software.

C-states are useful in saving power when the processor is not being utilized. However, entering and exiting these states introduce a large amount of jitter.

Power and thermal events

The processor, in an attempt to run within the constraints of its design, employs the use of frequency throttling in order to protect itself from thermal or overcurrent conditions.

Frequency throttling allows the processor to control how much the workload that is running on the host can introduce the stress that results in higher heat and current draw. Several factors can lead to high operating temperature or overcurrent events. Server ambient temperature, air flow, and other factors all play an important role in processor temperature.

An overcurrent can occur when the processor executes workloads that are capable of driving very high demand power-hungry resources within the processor itself. Overcurrent can also occur if Turbo Boost is enabled and the processor attempts to maximize the amount of performance when a particularly aggressive workload executes and the power that is available to be consumed is driven very high.

Special instructions (AVX)

Server processors offer special instructions that are capable of performing complex math at the cost of utilizing logic that is capable of driving very high power usage inside the processor itself.

If left unchecked, overcurrent throttling is required when these instructions eventually drive the processor to consume higher power. Instead of reactively throttling, processors typically proactively force cores to run at a lower frequency to limit the chances of extreme power excursions whenever those instructions are executed.

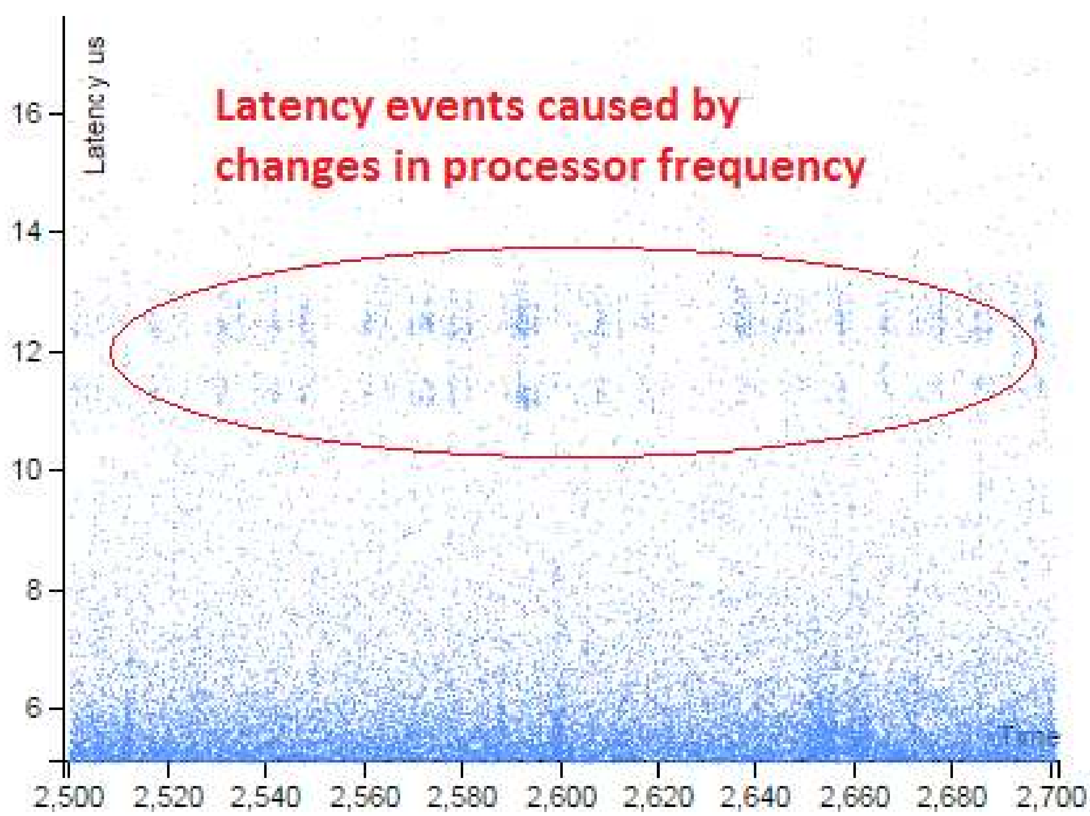
On Intel processors, the use of Advanced Vector Extensions (AVX) instructions cause the processor to limit the processor frequency automatically. Because these instructions cause the processor to automatically limit and potentially lower the frequency, their usage often introduces jitter.

Jitter and latency

Jitter and latency are directly related. Jitter induced by processor frequency changes introduces latency observed by a workload. When a processor executes a change in frequency, it goes through a process that causes thread execution to stop entirely before the processor is capable of running at the new chosen frequency. This process occurs regardless of whether the processor will shift to a faster or a slower frequency. The amount of time that the processor is stopped can vary, but is typically between 10 and 15 microseconds. For a workload that depends on processor execution, a change in frequency will then always introduce and additional 10-15 usec of latency.

Because frequency shifts are often asynchronous to application tasks running on the server, these latencies are random and thus non-deterministic from an application standpoint. Also, it is important to note that a processor that varies its frequency also creates a non-deterministic level of performance for the running applications. Software will execute slower as the frequency is lowered and vice versa. The difference in frequency itself also means that there is a variable amount of latency involved if an application depends on a certain amount of execution time.

Latency introduced by frequency changes can be illustrated by measuring latency when the processor is configured to allow for frequency shifts (i.e. Turbo Boost is enabled.) The figure below illustrates how latency spikes in the 10-15 usec range occur when the processor is allowed to shift frequency.



Processor jitter control

Processor Jitter Control is a feature that is hosted by platform firmware within HPE ProLiant Gen10 servers. It allows the user to tune servers to reduce or remove processor jitter either automatically or manually. Jitter control has three modes and can be configured via the RBSU (ROM Based Setup Utility) or via the RESTful interface. Jitter control can be disabled, or configured for auto-tuned or manual-tuned mode.

Auto-tuned mode

When Processor Jitter Control is configured to run in auto-tuned mode, HPE Server firmware disables the impact of power management and dynamically makes adjustments to the processor during run time in order to eliminate the occurrence of frequency shift induced jitter. The result of running in auto-tuned mode is that the processor will eventually run at the highest frequency that can be achieved where the processor stops making frequency changes in order to stay within its thermal, power, and core usage constraints. Auto-tuned mode lowers the frequency upon detection of frequency changes caused by the following sources:

- C-state transitions
- AVX induced transitions
- Turbo transitions (due to power, thermal, and core usage)
- Thermal throttling

NOTE:

When selecting auto-tuned mode via RBSU, C-state settings are also set to disabled. Most operating systems rely on BIOS reporting of support of C-states via the Advanced Power and Configuration Interface (ACPI). However, certain Linux distributions that load the intel_idle driver will ignore the ACPI reporting of C-state support. For Auto-tuned to function properly, the intel_idle driver must be disabled by adding `intel_idle.max_cstate=0` in the kernel line parameters.

Manual-tuned mode

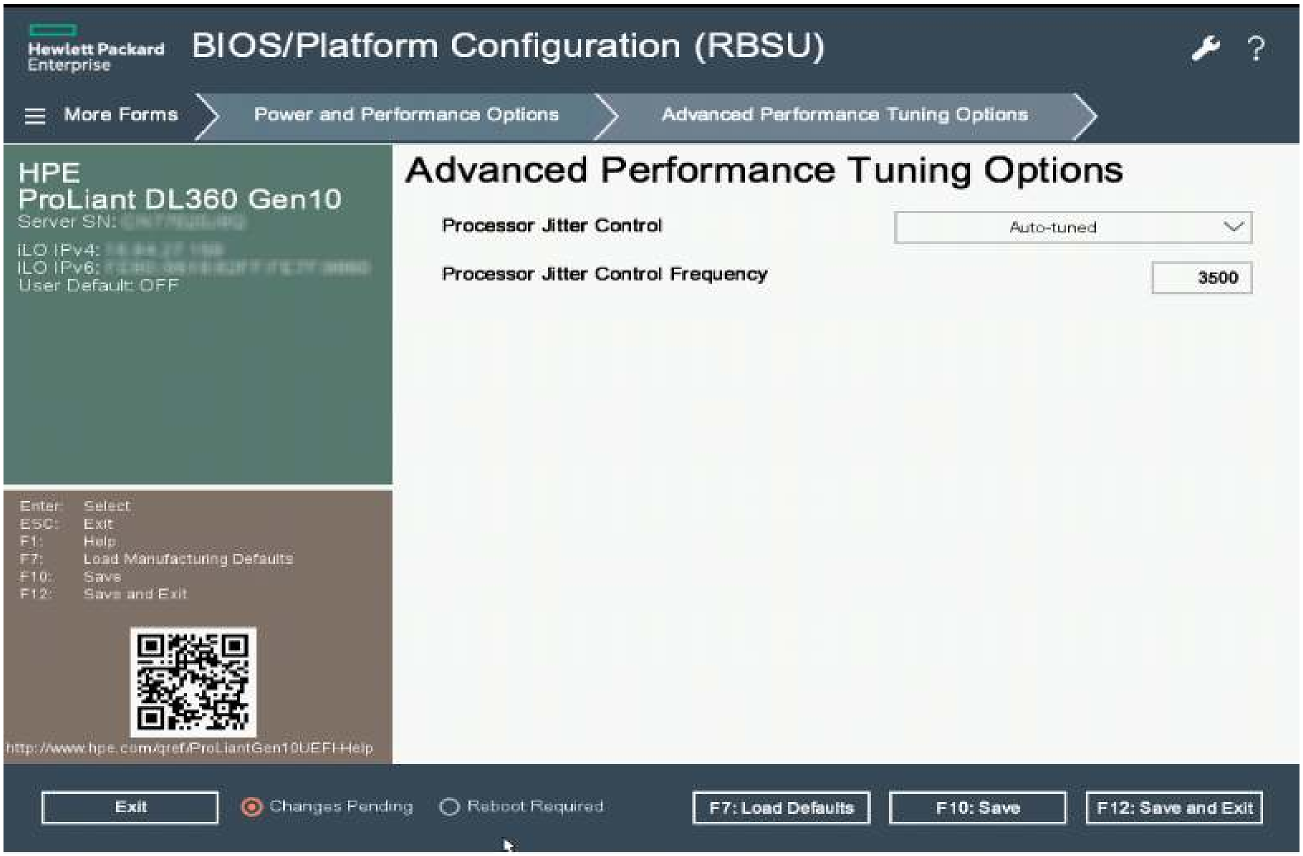
When Processor Jitter Control is configured to run in Manual-mode, the processor is configured to run at a user selectable frequency. In this mode, firmware does not lower the frequency dynamically even if processor frequency changes are detected. This mode is useful for users who desire to manually tune for jitter reduction and for those who wish to set a maximum operational frequency. Unlike in auto-tune mode, if a frequency change occurs below the programmed frequency, the server will not reduce the operating frequency permanently and the processor is allowed to return to the maximum frequency when the limiting constraints no longer exist.

Configuring processor jitter control via System Utilities

NOTE:

An iLO Advanced license or higher is required to use Processor Jitter Control.

Processor Jitter Control configuration options are located in System Utilities under **System Configuration** > **BIOS/Platform Configuration (RBSU)** > **Power and Performance Options** > **Advanced Performance Tuning Options**. The two options **Processor Jitter Control** and **Processor Jitter Control Frequency** are available when the server supports jitter smoothing and an iLO Advanced License or higher is installed.



The **Processor Jitter Control** option has three modes: **Disabled**, **Auto-tuned**, or **Manual-tuned**. Selecting **Auto-tuned** or **Manual-tuned** mode allows the user to also edit the **Processor Jitter Control Frequency** input option, which allows the user to select the desired target frequency for manual-tuned mode or the starting maximum frequency for auto-tuned mode. Frequency is entered in units of megahertz (Mhz) and the system firmware rounds up to the nearest frequency interval allowed by the processor. For example, Intel Xeon Server processors support frequency programming in intervals of 100 MHz. If a user inputs 2050 MHz, the resulting frequency will be 2100 MHz if supported by the installed processor.

Increasing performance with jitter smoothing

Jitter smoothing can be used to tune for best performance in workloads that are traditionally sensitive to latency, as well as in workloads that are impacted by excessive amounts of frequency shifting.

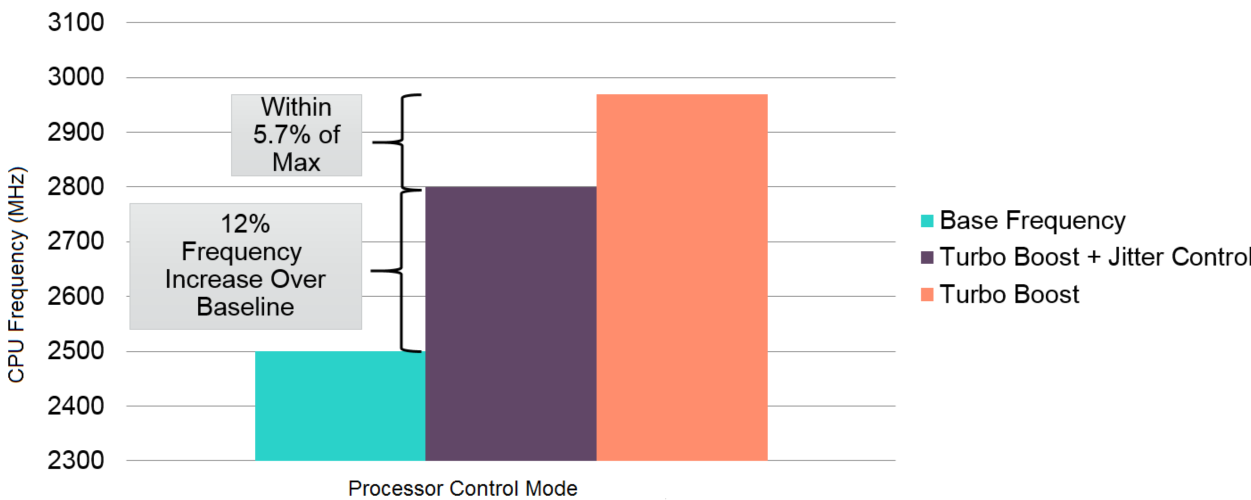
Users that traditionally tune servers for low latency and follow HPE's *Configuring and tuning HPE ProLiant Servers for low -latency applications* technical white paper are familiar with the practice of disabling P-state power management, C-states, and Turbo mode in order to eliminate jitter caused by frequency shifting by the processor. However, a significant amount of performance can be gained by leaving Turbo Boost enabled and also enabling Processor Jitter Control.

NOTE:

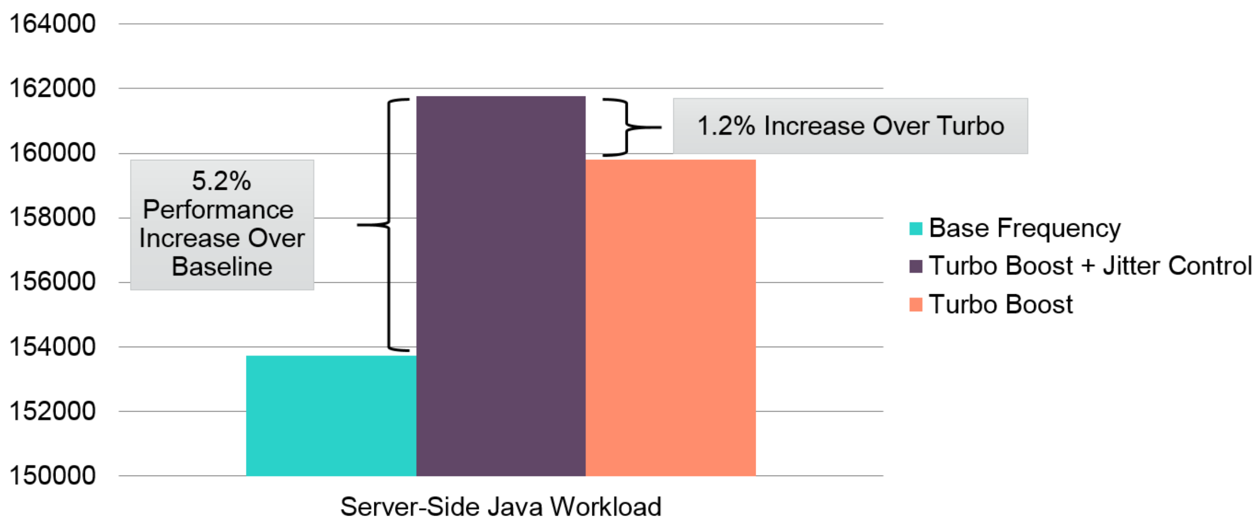
The *Configuring and tuning HPE ProLiant Servers for low -latency applications* technical white paper is available on the following website:

http://h20564.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c01804533&lang=en&cc=us

Although the frequencies between the base frequency of the processor and the maximum Turbo Boost mode frequency are not guaranteed, a workload is not likely to encounter limiting constraints if the processor is limited to a specific frequency within the Turbo Boost frequency range. Processor Jitter Control, when configured for Auto-mode, finds the frequency dynamically for any workload and environment in which the server is deployed. If the resulting Auto-tuned frequency is higher than the base frequency, gains in performance are achieved over the conservative practice of disabling Turbo Boost. The following image demonstrates the gains that can be achieved using Processor Jitter Control with an HPE ProLiant DL360 Gen10 with Intel® Xeon® Platinum 8180 Scalable Family Processor.

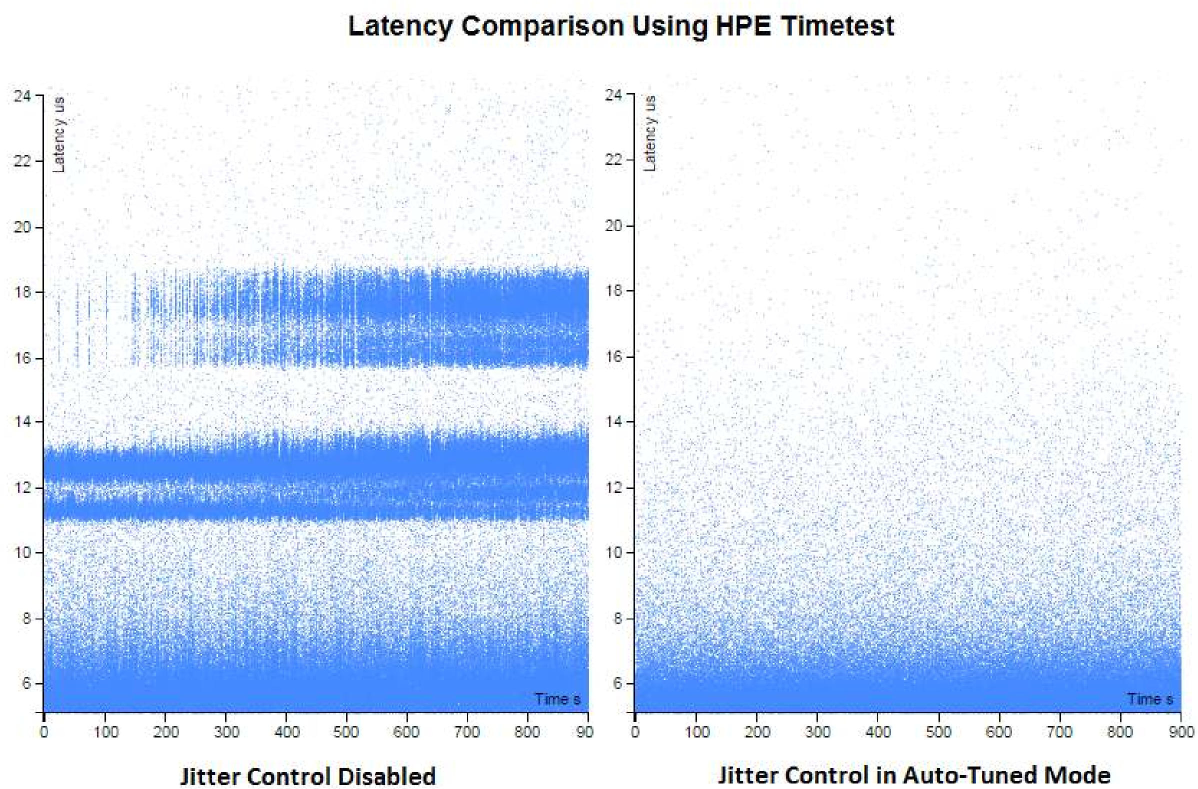


When utilizing Processor Jitter Control, a workload may even surpass the performance seen by enabling Turbo Boost alone. If a workload drives the processor to a state where Turbo is constantly making frequency shifts in order to maximize performance, but cannot stay at the higher frequencies for a substantial period of time, then the gains by attempting to run at higher frequencies are nullified, or, in some cases, performance may be reduced. When using the HPE ProLiant DL360 Gen10 with Intel® Xeon® Platinum 8180 Scalable Family Processor running a server-side Java workload, it is possible to see gains above using Turbo Boost alone.



Jitter smoothing is designed to reduce latency by limiting the causes for frequency changes. While throughput performance metrics are impacted by enabling Jitter Control, the removal of latency spikes is critical for workloads that either cannot tolerate latency or have costs associated with computational latencies. When using auto-tuned mode, jitter within the 10-15 usec range caused by frequency changes can be completely eliminated.

The following figures illustrate the latencies that can be measured by the HPE Timetest Tool. Without jitter control and with turbo mode enabled, timetest plots show a band of 10-15 usec latencies that are introduced due to processor frequency changes. After running the same workload with auto-tuned mode, those latencies disappear.



Websites

Intelligent System Tuning

http://www.hpe.com/info/ist

iLO websites

iLO 5

http://www.hpe.com/support/ilo-docs

iLO licensing

http://www.hpe.com/info/ilo/licensing

iLO mobile app

http://www.hpe.com/info/ilo/mobileapp

HPE ProLiant Gen8 servers

http://www.hpe.com/info/proliantgen8/docs

HPE ProLiant Gen9 servers

http://www.hpe.com/support/proliantgen9/docs

HPE ProLiant Gen10 servers

http://www.hpe.com/support/proliantgen10/docs

General websites

Hewlett Packard Enterprise Information Library

www.hpe.com/info/EIL

Single Point of Connectivity Knowledge (SPOCK) Storage compatibility matrix

www.hpe.com/storage/spock

Storage white papers and analyst reports

www.hpe.com/storage/whitepapers

Support and other resources

Accessing Hewlett Packard Enterprise Support

- For live assistance, go to the Contact Hewlett Packard Enterprise Worldwide website:

  http://www.hpe.com/assistance
- To access documentation and support services, go to the Hewlett Packard Enterprise Support Center website:

  http://www.hpe.com/support/hpesc

Information to collect

- Technical support registration number (if applicable)
- Product name, model or version, and serial number
- Operating system name and version
- Firmware version
- Error messages
- Product-specific reports and logs
- Add-on products or components
- Third-party products or components

Accessing updates

- Some software products provide a mechanism for accessing software updates through the product interface. Review your product documentation to identify the recommended software update method.
- To download product updates:

  Hewlett Packard Enterprise Support Center

  www.hpe.com/support/hpesc

  Hewlett Packard Enterprise Support Center: Software downloads

  www.hpe.com/support/downloads

  Software Depot

  www.hpe.com/support/softwaredepot
- To subscribe to eNewsletters and alerts:

  www.hpe.com/support/e-updates
- To view and update your entitlements, and to link your contracts and warranties with your profile, go to the Hewlett Packard Enterprise Support Center **More Information on Access to Support Materials** page:

  www.hpe.com/support/AccessToSupportMaterials

IMPORTANT:

Access to some updates might require product entitlement when accessed through the Hewlett Packard Enterprise Support Center. You must have an HPE Passport set up with relevant entitlements.

Customer self repair

Hewlett Packard Enterprise customer self repair (CSR) programs allow you to repair your product. If a CSR part needs to be replaced, it will be shipped directly to you so that you can install it at your convenience.

Some parts do not qualify for CSR. Your Hewlett Packard Enterprise authorized service provider will determine whether a repair can be accomplished by CSR.

For more information about CSR, contact your local service provider or go to the CSR website:

http://www.hpe.com/support/selfrepair

Remote support

Remote support is available with supported devices as part of your warranty or contractual support agreement. It provides intelligent event diagnosis, and automatic, secure submission of hardware event notifications to Hewlett Packard Enterprise, which will initiate a fast and accurate resolution based on your product's service level. Hewlett Packard Enterprise strongly recommends that you register your device for remote support.

If your product includes additional remote support details, use search to locate that information.

Remote support and Proactive Care information

HPE Get Connected

www.hpe.com/services/getconnected

HPE Proactive Care services

www.hpe.com/services/proactivecare

HPE Proactive Care service: Supported products list

www.hpe.com/services/proactivecaresupportedproducts

HPE Proactive Care advanced service: Supported products list

www.hpe.com/services/proactivecareadvancedsupportedproducts

Proactive Care customer information

Proactive Care central

www.hpe.com/services/proactivecarecentral

Proactive Care service activation

www.hpe.com/services/proactivecarecentralgetstarted

Warranty information

To view the warranty for your product or to view the *Safety and Compliance Information for Server, Storage, Power, Networking, and Rack Products* reference document, go to the Enterprise Safety and Compliance website:

www.hpe.com/support/Safety-Compliance-EnterpriseProducts

Additional warranty information

HPE ProLiant and x86 Servers and Options

www.hpe.com/support/ProLiantServers-Warranties

HPE Enterprise Servers

www.hpe.com/support/EnterpriseServers-Warranties

HPE Storage Products

www.hpe.com/support/Storage-Warranties

HPE Networking Products

www.hpe.com/support/Networking-Warranties

Regulatory information

To view the regulatory information for your product, view the *Safety and Compliance Information for Server, Storage, Power, Networking, and Rack Products*, available at the Hewlett Packard Enterprise Support Center:

www.hpe.com/support/Safety-Compliance-EnterpriseProducts

Additional regulatory information

Hewlett Packard Enterprise is committed to providing our customers with information about the chemical substances in our products as needed to comply with legal requirements such as REACH (Regulation EC No 1907/2006 of the European Parliament and the Council). A chemical information report for this product can be found at:

www.hpe.com/info/reach

For Hewlett Packard Enterprise product environmental and safety information and compliance data, including RoHS and REACH, see:

www.hpe.com/info/ecodata

For Hewlett Packard Enterprise environmental information, including company programs, product recycling, and energy efficiency, see:

www.hpe.com/info/environment

Documentation feedback

Hewlett Packard Enterprise is committed to providing documentation that meets your needs. To help us improve the documentation, send any errors, suggestions, or comments to Documentation Feedback (**docsfeedback@hpe.com**). When submitting your feedback, include the document title, part number, edition, and publication date located on the front cover of the document. For online help content, include the product name, product version, help edition, and publication date located on the legal notices page.