HP Z4 G6i and Z8 Fury G6i Desktop Workstation Performance



©2025 HP Development Company, L.P.

Table of Contents

Introduction 2
HP BIOS Performance Options 2
Microsoft Windows Performance Control Options 5
Turbo Frequencies 7
Boot Time 9
Memory Configuration 9
Physical Environment 10
Conclusion 10

©2026 HP Development Company, L.P.

Introduction

This paper provides detailed information about the performance control and optimizations available on the HP Z4 G6i and HP Z8 Fury G6i Desktop Workstations. HP provides performance control options located in the HP BIOS and in the Windows OS. These options enable users to customize the balance of system performance, acoustics, and energy consumption to suit individual environments and workflow preferences. Additionally, some of these options control a tradeoff in performance between specific types of workflows; optimizing performance for one workflow might come at the expense of the performance of another workflow.

The efficacy of performance control features varies, and is highly dependent on numerous factors including the hardware configuration, OS image, workload type, workload duration, and the physical environment the system is running in.

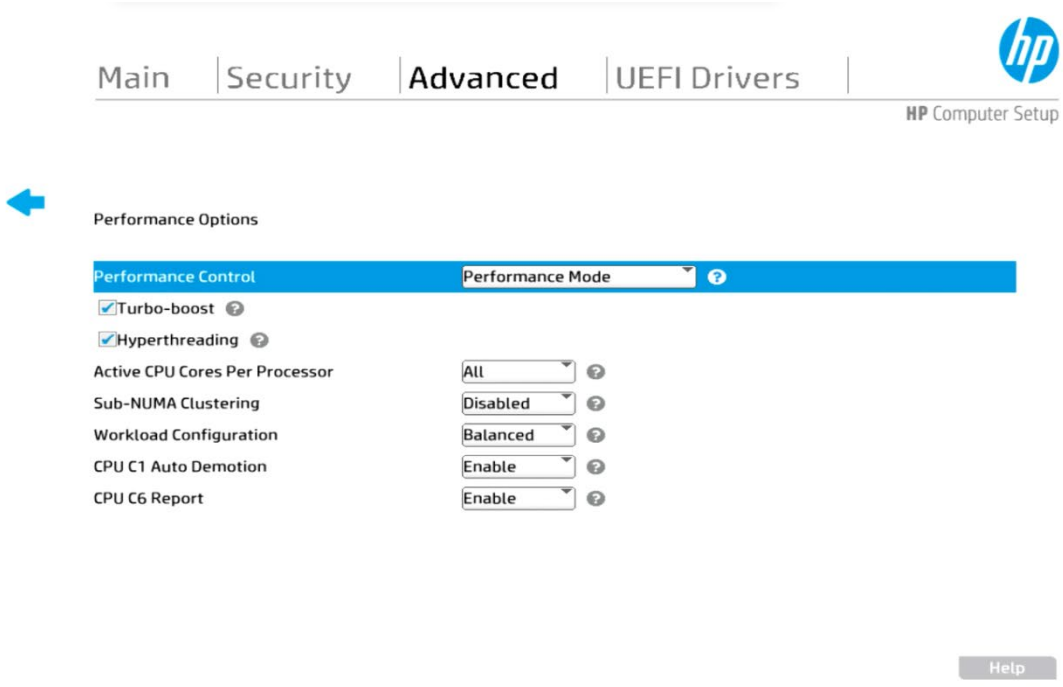
HP BIOS Performance Options

The HP BIOS has a menu named "Performance Options" containing performance control options. To access this menu:

1. Enter the "BIOS Setup" (F10) menu by pressing the F10 key while the system is booting. Press F10 repeatedly after the "HP" splash screen shows until you see a message in the bottom left that you are entering the F10 BIOS Setup.
2. In the BIOS Setup menu, navigate to Advanced->Performance Options.

Noteworthy performance options are explained below.

Figure 1: A screenshot of the Performance Options menu (BIOS v01.01.06).



©2026 HP Development Company, L.P.

Performance Control

Performance Control enables users to select a base performance operating mode according to preferences. The three available modes are Quiet Mode, Performance Mode, and Ultimate Performance Mode. They vary in power limits, cooling algorithms, and acoustics.

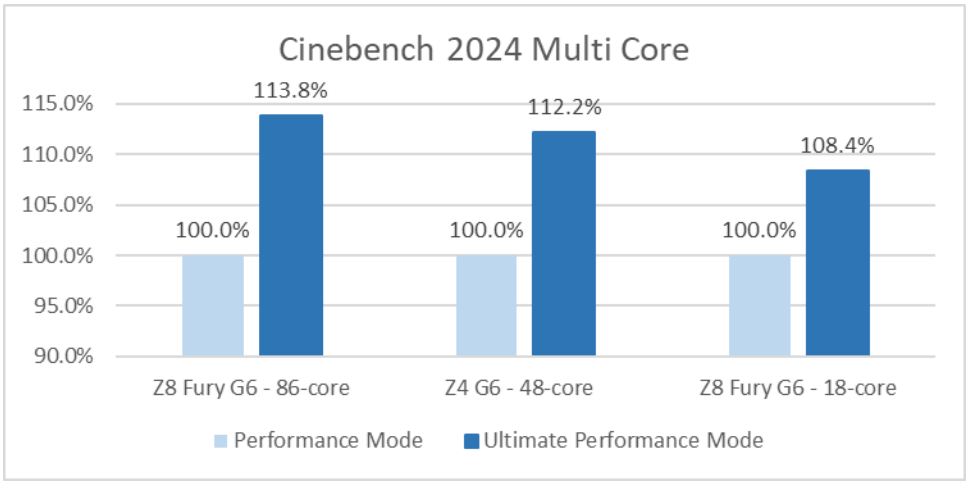
Ultimate Performance Mode: Provides best performance. In multi-threaded applications, Ultimate Performance Mode can show *up to a 13% improvement* in performance (see Figure 2). Acoustics may be higher due to the increased power.

Performance Mode: Optimizes performance with balanced acoustics. It is the default in Z4 G6i and Z8 Fury G6i workstations.

Quiet Mode: Optimizes acoustics with balanced performance. Note that depending on the hardware configuration, the fans can still ramp up significantly, e.g. in a configuration with multiple high-power GPUs.

The impact of these modes is more noticeable as the load on the system increases. Thus, multi-threaded workloads that utilize most or all the CPU's cores see more performance differentiation between the modes. Furthermore, CPUs with higher core counts and peak power use will see more differentiation. The graph below shows an example of performance differentiation between Performance Mode and Ultimate Performance Mode. Note that results vary by use case and config.

Figure 2: A graph showing the difference between Performance Mode and Ultimate Performance Mode across 2 platforms and 3 CPUs. This data was collected using the Cinebench 2024 Multi Core test.



Sub-NUMA Clustering (SNC)

SNC is a feature available on Intel® Xeon® processors. It groups together CPU cores, memory controllers, and slices of the last level cache (LLC). Within these clusters, cores can more efficiently access their associated sections of memory and the LLC. However, access to sections of memory and the LLC outside of the cluster becomes less efficient. These SNC clusters are exposed to the operating system as NUMA nodes. The following diagrams illustrate this concept on a high level.

©2026 HP Development Company, L.P.

Figure 3: A high-level diagram of the CPU architecture with SNC disabled.

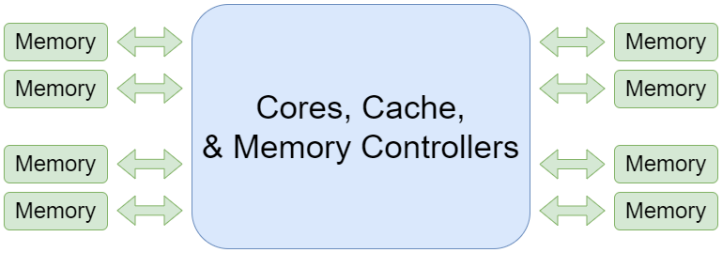
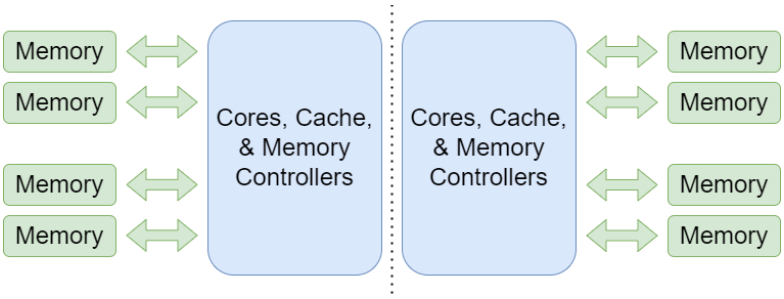


Figure 4: A high-level diagram of the CPU architecture with SNC enabled and 2 clusters.



The number of clusters SNC will create depends on the number of cores and memory controllers in the CPU and the CPU's architecture. Hence, in the HP BIOS, you can simply enable or disable SNC; the number of clusters is determined by the system. For the Intel® Xeon® 600 generation, only the 64-core and 86-core CPU will have SNC clusters created when SNC is enabled. For these CPUs, 2 clusters will be created. For Intel® Xeon® 600 CPUs with less than 64 cores, enabling SNC will not have an effect.

SNC can improve the performance of some workloads but can lower the performance of others. If a workload uses more cores than are present in one SNC cluster, it will cross SNC boundaries which incurs a performance hit; such a workload would likely perform better with SNC disabled. So, if a workload is heavily multi-threaded and scales to run on most cores, SNC should generally be disabled. If a workload uses a smaller number of cores or is explicitly NUMA-optimized, enabling SNC may improve performance.

CPU C1 Auto Demotion

Enabling this setting can improve system responsiveness. It allows the CPU to keep cores in a higher power state to reduce the latency of waking up cores. The power states for CPU cores are called C-states and range from C0 (actively executing instructions) to C6 (a deep sleep state). When this setting is enabled and the OS requests a low power state for a core (like C6), the CPU can "demote" this request to C1, which is a higher power state. Cores can wake from C1 almost instantaneously to start executing instructions, while there is more latency associated with waking a core from C6.

©2026 HP Development Company, L.P.

However, the more cores that are in C0 or C1, the lower the maximum possible turbo frequency[1]. To achieve max turbo frequencies, one would want to limit the number of cores in C0 and C1. So, disabling this setting can improve single-threaded performance by increasing turbo frequencies. **Beware that disabling this can noticeably affect system responsiveness, such as when opening apps, moving windows, and copying files.** See more info in the section titled "Turbo Frequencies".

[1]*Applicable to Intel® Xeon® 600 Processors for Workstations.*

CPU C6 Report

Enabling this setting enables the C6 state for cores. It is highly recommended to keep this enabled. If disabled, all cores will be seen as active for the purposes of the max turbo frequency calculation. Hence, lightly threaded and single-threaded performance will suffer due to lower turbo frequencies. Disabling this setting can also significantly increase CPU power use when the system is idle.

Microsoft Windows Performance Control Options

Power Plans

Power Plans vs. BIOS Performance Control mode

Note that Windows power plans are independent from the HP BIOS Performance Control mode—they do not conflict or overlap.

The HP BIOS Performance Control modes control behavior such as:

- CPU power limits
- System power limits
- Thermal limits
- Fan control settings and acoustics

Windows power plans control behavior such as:

- CPU performance states (P-states) and power states (C-states)
    - When to idle
    - When to exit idle
    - When to use max Turbo frequency vs. saving power
    - When to use the max memory and I/O frequencies vs. saving power
- Thread scheduling and priorities

The HP BIOS Performance Control mode tends to have more impact when the system is fully stressed (where acoustic or power limits may be reached), while power plans can have more impact during light usage, with many settings affecting idling behavior, and the tradeoff between energy savings and low latencies.

©2026 HP Development Company, L.P.

Power Plans vs. BIOS Power Options

While the BIOS Performance Control mode is distinct from the Windows power plan, there are other BIOS performance and power management options that affect P-states or C-states and effectively override the power plan. The F10 BIOS options that affect the P-states and C-states are:

- Performance Options->C1 Auto Demotion
    - When enabled, gives hardware (CPU) more control over C-States
- Performance Options->CPU C6 Report
    - Enables/disables C6 state
- Power Management Options->Runtime Power Management
    - Enables/disables P-States
- Power Management Options->Hardware P-States
    - When enabled, gives hardware (CPU) more control over P-States
- Power Management Options->Energy/Performance Bias Control
    - When enabled, gives hardware (CPU) more control over P-States

As you can see, these BIOS options are broad controls that either enable/disable P-States and C-States or influence how much control over those states the CPU has versus the OS. In comparison, power plans provide much more fine-grained control over when CPU cores should enter different P-States and C-States in response to application and user activity.

Note that a power plan's "advanced power settings" that are configurable in the Control Panel are a small subset of the settings contained within a power plan.

HP Power Plans

HP has created two custom Windows power plans for the HP Z4 G6i and HP Z8 Fury G6i—HP Optimized and HP Ultimate Performance. The recommended plan to use is "HP Ultimate Performance", which will maximize performance in most situations. It is similar to Microsoft's "Ultimate Performance" (which Microsoft makes available only on systems with Xeon® processors) but contains several optimizations that increase energy savings while maintaining maximum performance. Note that Microsoft's Ultimate Performance is very similar to Microsoft's High Performance plan, which is available on all systems regardless of processor type.

The HP Optimized plan finds the optimal balance between performance and energy savings. If you would like to increase energy savings, try this plan.

Note that for some light workloads involving single-threaded or lightly threaded tasks, HP Optimized may deliver higher performance than HP Ultimate Performance. By allowing more CPU cores to remain in a sleep state, HP Optimized enables higher turbo frequencies on the active cores. By comparison, HP Ultimate Performance generally performs better for heavily multi-threaded workloads, intensive multitasking, and overall system responsiveness.

©2026 HP Development Company, L.P.

Turbo Frequencies

In most real-world cases, you will not see an Intel® Xeon® CPU hit its max Turbo Boost frequency. Before exploring why, we should first note that there are two "max" Turbo frequencies. The following are Intel-provided descriptions for both Turbo frequencies.

1. **Intel® Turbo Boost Technology 2.0 (ITB 2.0)**: "Intel® Turbo Boost Technology 2.0 Frequency is the maximum single core frequency at which the processor is capable of operating using Intel® Turbo Boost Technology. Frequency is typically measured in gigahertz (GHz), or billion cycles per second."
2. **Intel® Turbo Boot Max Technology 3.0 (ITBM 3.0)**: "Intel® Turbo Boost Max Technology 3.0 identifies the best performing core(s) on a processor and provides increased performance on those cores through increasing frequency as needed by taking advantage of power and thermal headroom. Intel® Turbo Boost Max Technology 3.0 frequency is the clock frequency of the CPU when running in this mode."

In other words:

1. ITB 2.0 frequency is the max single core frequency any core is capable of.
2. ITBM 3.0 frequency is the max single core frequency that only a few select cores (aka "favored" cores) can achieve. Specifically, on Intel® Xeon® for Workstation CPUs, there are generally 2 cores at +0.2 GHz compared to ITB2.0. On some Intel® Xeon® for Workstation CPUs, there may be an *additional* 2 cores at +0.1 GHz compared to ITB2.0.

"Single core frequency" is the frequency when **only <u>one</u> core is in the C0 or C1 state** (for more on C-states, see the section "CPU C1 Auto Demotion"). The more cores that are in C0 or C1, the lower the maximum frequency. For the purposes of max turbo frequency, cores in C0 or C1 are considered "active cores". Below is an example of how the number of active cores correlates to the maximum turbo frequency.

Figure 5: A table showing number of active cores versus maximum turbo frequency for the 64-core Intel® Xeon® 696X, when the CPU is executing non-AVX (SSE) instructions.

Number of Active Cores	Maximum Core Frequency
1 to 2	4.8 GHz (ITBM 3.0); 4.6 GHz (ITB 2.0)
3 to 4	4.7 GHz (ITBM 3.0); 4.4 GHz (ITB 2.0)
5 to 16	4.2 GHz
17 to 40	4.1 GHz
41 to 52	3.7 GHz
53 to 64	3.5 GHz

©2026 HP Development Company, L.P.

The operating system and other tools can read this table directly from the CPU.

Note that the operating system will generally automatically schedule CPU-intensive tasks on the favored cores first (the cores capable of ITBM 3.0 frequencies). However, when more than four cores are active, the favored cores' frequencies end up being limited to the same max turbo frequency the rest of the cores are limited to.

Turbo Frequency on Intel® Xeon® 600-series 64-Core and 86-Core

Intel® Xeon® 600-series CPUs with 64 and 86 cores have additional nuance in how their turbo frequencies are calculated. These CPUs have their cores split across two dies, which each have their own implicit turbo table. If all active cores are on one die (e.g., 24 cores on die 1), the frequencies are lower than if the active cores are split between both dies (e.g., 12 cores on each die). The combined turbo table for the entire CPU (Figure 5) assumes the cores are spread evenly across the dies. As of the release of the G6i HP Z Desktop Workstations, Windows is designed to schedule threads evenly across the 2 dies; however, some applications make their own scheduling decisions and might not schedule optimally.

Instruction Sets and Turbo Frequencies

The maximum CPU frequencies are also dependent on the specific instruction set being executed. Various instruction sets—including non-AVX (SSE), AVX-2, AVX-512, and AMX—each correspond to their own turbo frequency table. Among these, non-AVX instructions have the highest maximum turbo frequency, while AMX instructions have the lowest. As a result, different applications may run at varying frequencies depending on which instruction set they utilize.

Maximizing Turbo Frequency

Here are some tips on how to increase your single- or low-core frequency by influencing the number of cores in C0/C1:

1. Ensure CPU C6 Report is enabled in the HP BIOS (see relevant section earlier in this document).
2. Reduce the number of apps that are open. Minimize open windows not currently in use, so the OS knows the app is idle.
3. Keep your system free of unnecessary apps and background programs.
4. Experiment with different power plans (Windows Search->Choose a Power Plan). You may find that a Balanced power plan will keep more cores in a sleep state than a Performance power plan; this may help single- or low-core performance but will likely come at the expense of responsiveness and multi-core performance.
5. Consider disabling CPU C1 Auto Demotion in the HP BIOS. Note that this will likely come with the tradeoff of lower system responsiveness (see relevant section earlier in this document).

©2026 HP Development Company, L.P.

6. Visit https://www.hp.com/drivers. Make sure all the required Intel® drivers for your workstation have been installed. This may include:
    a. Intel® Chipset Driver or Chipset Device Software
    b. Intel® Processor Power Management (PPM) Package (if available)
    c. Intel® Innovation Platform Framework (IPF) Driver (if available)
    d. Intel® Dynamic Tuning Technology (DTT) Driver (if available)
7. Install the latest BIOS for your workstation. Again, visit https://www.hp.com/drivers for more information.

Boot Time

There are several phases involved when booting the system. HP has optimized the parts of the boot process controlled by the HP BIOS to reduce the boot time. However, significant portions of the boot process are inherent to the CPU and OS being used, and so are dependent on the implementations provided by the respective vendor.

If you are experiencing a longer boot time than usual, it is likely due to your system performing "*extended* memory training" (this is above and beyond the usual memory training). Every time you swap hardware within the system, such as a GPU, network card, DIMM, etc., the system will perform a process called extended memory training on the next boot. This is a process that initializes RAM to ensure stable and optimal operation. This will take considerably longer than a usual boot.

Additionally, the BIOS runs a memory health check once a month. As a result of this check, the BIOS may determine the system needs to perform extended memory training. The most this can happen is once a month.

Regarding standard boots (with no extended memory training): the more memory a system has, the longer the boot will take. Every 256 GB of memory adds roughly 10 seconds of boot time to a standard boot.

Improving Boot Time

One setting you can change to improve boot time is disabling Network (PXE) Boot. This can be found in F10 BIOS Setup->Advanced->Boot Options->Network (PXE) Boot. This can save up to 10 seconds per network port.

Memory Configuration

Memory Capacity

Always make sure you have enough memory capacity for your workflows. If you do not have enough memory, the system will be forced to swap data to and from the mass storage devices, which can be dramatically slower than DRAM memory.

Memory Channels

The number of DIMMs and the order in which they are installed can also have an impact on system performance. Each system has several memory channels, each of which can be utilized simultaneously:

©2026 HP Development Company, L.P.

- **Z4 G6i**: 4 memory channels
- **Z8 Fury G6i**: 8 memory channels

Because memory channels can be used in parallel, memory should be distributed across as many channels as possible for best performance. E.g. on a Z8 Fury G6i, 8x 16 GB DIMMs will generally perform better than 2x 64 GB DIMMs, though they both total to a 128 GB memory capacity.

Each memory channel can support up to two DIMMs. However, when two DIMMs are installed on the same channel (referred to as *2 DIMMs per channel*, or 2DPC), the channel's operating frequency is reduced due to electrical signal integrity limits. For example, DDR5-6400 memory operates at 6400 MHz with one DIMM per channel (1DPC), but at 5200 MHz with 2DPC.

Although 2DPC reduces memory clock speed, it enables memory interleaving within the channel. This allows the system to alternate access between the two DIMMs, which can improve throughput for workloads that heavily access memory or require very large memory capacity.

As a general rule, you should always populate every memory channel with one DIMM before adding a second DIMM to any channel. Use 2DPC only after all channels are filled, and primarily for workloads that are memory intensive or require additional capacity, e.g. large dataset processing or simulation.

See the Z4 G6i and Z8 Fury G6i architecture whitepapers for more details on memory capabilities and load orders.

CPU Cores for Memory

For best performance with memory intensive workflows, the CPU also needs to have enough cores to efficiently transfer data to and from the memory. In addition to cores doing other computations, you should have at least two cores for each memory channel in use. For example, if your workflows will heavily utilize 10 cores, and you have 4 memory channels in use, it is recommended to purchase a CPU with at least 18 cores to prevent memory and CPU bottlenecks.

Physical Environment

Beyond the performance control options, it's important to note that the system's physical (ambient) environment can impact its performance. Systems residing in cooler environments will perform better than those residing in warmer environments.

This is due to processors being more power efficient at lower temperatures. At a given power level, a processor can operate at a higher frequency (i.e., execute more instructions per second) when its temperature is lower.

Before comparing performance test data between systems, ensure the systems' physical environments are similar.

Conclusion

HP's performance control options allow users to prioritize what matters most to them based on workload, performance needs, and acoustic preferences on HP Z Desktop Workstations.

©2026 HP Development Company, L.P.

HP recommends that customers consult this paper to assess their system's performance, make improvements tailored to their workflows, and guide decisions on purchasing HP Z Desktop Workstations..

©2026 HP Development Company, L.P.

For use with general public.

© Copyright 2026 HP 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.

Microsoft and Windows are trademarks of the Microsoft group of companies. All other trademarks are the property of their respective owners.

4AA8-5565ENW APRIL 2026