Workload Tuning Guide for AMD EPYC™ 7002 Series Processor Based Servers

Application Note

Publication #	**56745** Revision: **0.80**
Issue Date:	**November 2019**

Advanced Micro Devices

Contents

Chapter 1 Introduction 6

Chapter 2 BIOS Options and Their Benefits 7

2.1 Infinity Fabric Settings 7

2.1.1 Link Speed 7

2.1.2 Dynamic Link Width Management (DLWM) 7

2.1.3 Power States 8

2.1.4 C-States 8

2.2 NUMA and Memory Settings 9

2.2.1 ACPI SLIT and SRAT 9

2.2.2 NUMA Nodes per Socket (NPS) 10

2.2.3 Memory Clock Speed 10

2.2.4 Transparent Secure Memory Encryption (TSME) 11

2.3 Power Efficiency Settings 12

2.3.1 Core Clock Dynamic Power Management (CCLK DPM) 12

2.3.2 Power vs Performance Determinism Settings 12

2.3.3 Processor Cooling and Power Dissipation Limit Settings 12

2.3.4 ACPI—Collaborative Processor Performance Control (CPCC) 13

2.4 Processor Core Settings 14

2.4.1 Cache Prefetchers 14

2.4.2 Symmetric Multithreading (SMT) Settings 14

2.4.3 Core Boost Frequency Settings 14

2.5 I/O Settings 15

2.5.1 APIC Settings 15

2.5.2 SR-IOV Settings 15

2.5.3 PCIe Ten Bit Tag 16

2.5.4 Preferred I/O Settings 16

2.5.5 Input-Output Memory Management Unit (IOMMU) Settings 16

Chapter 3 BIOS Setting Selection by Workload 17
3.1 General-Purpose Workloads 17
3.1.1 Infinity Fabric Settings 17
3.1.2 NUMA and Memory Settings 17
3.1.3 Power Efficiency Settings 18
3.1.4 Processor Core Settings 18
3.1.5 I/O Settings 19
3.2 Virtualization and Containers 19
3.2.1 Infinity Fabric Settings 19
3.2.2 NUMA and Memory Settings 20
3.2.3 Power Efficiency Settings 20
3.2.4 Processor Core Settings 20
3.2.5 I/O Settings 21
3.3 Database and Analytics 21
3.3.1 Infinity Fabric Settings 21
3.3.2 NUMA and Memory Settings 21
3.3.3 Power Efficiency Settings 22
3.3.4 Processor Core Settings 22
3.3.5 I/O Settings 22
3.4 HPC and Telco Settings 23
3.4.1 Infinity Fabric Settings 23
3.4.2 NUMA and Memory Settings 23
3.4.3 Power Efficiency Settings 24
3.4.4 Processor Core Settings 24
3.4.5 I/O Settings 25

Revision History

Date	Revision	Description
November 2019	0.80	Added *default* to cells that were blank. Minor clarifications throughout.
October 2019	0.75	Initial release.

Chapter 1 Introduction

Default BIOS options are generally selected to produce the best overall performance for typical workloads. However, typical will differ from end-user to end-user. Because of that, defaults may not be best for your specific workload. AMD has tested various workloads in our labs, and this document discusses BIOS options that we feel are relevant for both maximum performance and performance-per-watt (power efficiency). You will find in chapter two a list of various BIOS options and the potential benefit of modifying each one. In chapter three, you will find example workloads and recommended BIOS settings. These BIOS settings are not a one-size fits all, as your specific workload is most certainly not going to be identical to one of these synthetic benchmarks.

Not all platforms support all the listed BIOS controls. Please contact your platform vendor if a needed control is not visible.

Chapter 2 BIOS Options and Their Benefits

2.1 Infinity Fabric Settings

2.1.1 Link Speed

Benefit: Setting this to a lower speed can save uncore power that can be used to increase core frequency or reduce overall power. It will also decrease cross socket bandwidth and increase cross socket latency.

Setting	Options
xGMI Link Max Speed	• 10.667 Gbps • 13 Gbps • 16 Gbps • 18 Gbps

2.1.2 Dynamic Link Width Management (DLWM)

Benefit: xGMI Dynamic Link Width Management saves power during periods of low socket-to-socket data traffic by reducing the number of active xGMI lanes per link from 16 to 8. However, under certain scenarios, involving low bandwidth but latency sensitive traffic, the transition from a low power xGMI state to a full power xGMI state can adversely impact latency. Setting xGMI Link Width Control to manual and specifying a Force Link Width eliminates any such latency jitter. Applications that are known to be insensitive to both socket-to-socket bandwidth and latency can set a forced link width of eight (or two on certain platforms) to save power, which can divert more power to the cores for boost.

Setting	Options
xGMI Link Width Control	• Auto: Do not show Max Link Width and Force Link Width control options • Manual: Show Max Link Width and Force Link Width control options
xGMI Max Link Width	• 0: Max width x8, min width x8 (x2 on certain platforms) • 1: Max width x16, min width x8 (x2 on certain platforms)
xGMI Max Link Width Control	• Auto: Do not show xGMI Max Link Width control • Manual: Show xGMI Max Link Width control

Setting	Options
xGMI Force Link Width Enable	• Unforce: Use automatic xGMI Link Width selection • Force: Use link width specified by xGMI Force Link Width
xGMI Force Link Width	• 0: Use width x2 (not supported on all platforms) • 1: Use width x8 • 2: Use width x16

2.1.3 Power States

Benefit: Enable or disable Algorithm Performance Boost (APB). In the default state, the Infinity Fabric selects between a full-power and low-power fabric clock and memory clock based on fabric and memory usage. However, under certain scenarios, involving low bandwidth but latency-sensitive traffic (and memory latency checkers), the transition from low power to full power can adversely impact latency. Setting APBDIS to 1 (to disable APB) and specifying a fixed Infinity Fabric P-state of 0 will force the Infinity Fabric and memory controllers into full-power mode, eliminating any such latency jitter. Certain CPU OPNs and memory population options result in a scenario where setting a fixed Infinity Fabric P-state of 1 will reduce memory latency at the expense of memory bandwidth. This setting may benefit applications known to be sensitive to memory latency.

Setting	Options
APB Disable (APBDIS)	• 0: Dynamically switch Infinity Fabric P-state based on link usage • 1: Enable fixed Infinity Fabric P-state control
Fixes SOC P-State	• P0: Highest-performing Infinity Fabric P-state • P1: Next-highest-performing Infinity Fabric P-state • P2: Next highest-performing Infinity Fabric P-state • P3: Minimum Infinity Fabric power P-state

2.1.4 C-States

Benefit: Much like CPU cores, the Infinity Fabric can go into lower power states while idle. However, there will be delay changing back to full-power mode causing some latency jitter. In a low latency workload, or one with bursty I/O, one could disable this feature to achieve more performance with the tradeoff of higher power consumption.

Setting	Options
DF C-states	• Disabled: do not allow Infinity Fabric to go to a low-power state when the processor has entered Cx states • Enabled: allow Infinity Fabric to go to a low-power state when the processor has entered Cx states

2.2 NUMA and Memory Settings

2.2.1 ACPI SLIT and SRAT

Benefit: Controls automatic or manual generation of distance information in the ACPI System Locality Information Table (SLIT) and NUMA proximity domains in the System Resource Affinity Table (SRAT). Some operating systems and hypervisors do not perform L3 aware scheduling and some workloads will benefit from having the L3 declared as a NUMA domain. For 2P systems, the remote socket distance can affect memory allocation decisions. Setting this to a high value >= 32 (32 recommended) may improve scheduling of lightly threaded workloads. Setting this to a lower value < 32 (22 recommended) may improve scheduling of heavily threaded workloads. If a workload spans two sockets, AMD recommends setting the distance to < 32 and if the workload can be confined to a socket, you should set it to 32.

Setting	Options
ACPI SRAT L3 Cache as NUMA Domain	• Disable: Do not report each L3 cache as a NUMA domain to the OS • Enable: Report each L3 cache as a NUMA domain to the OS
ACPI SLIT Distance Control	• Auto: Use default remote and same socket distances • Manual: Enable remote and same socket distance controls
ACPI SLIT Remote Relative Distance	• Near: Let BIOS select default values that describe remote cores as relatively close to each local core • Far: Let BIOS select default values that describe remote cores as relatively far away to each local core
ACPI SLIT <various> Distance	• Enabled when ACPI SLIT Distance Control set to manual • Each item can be set to a value from 10–255 decimal to specify the distance for domains relative to each other

2.2.2 NUMA Nodes per Socket (NPS)

Benefit: This setting enables a trade-off between minimizing local memory latency for NUMA-aware or highly parallelizable workloads vs. maximizing per-core memory bandwidth for non-NUMA-friendly workloads. NPS2 and/or NPS4 may not be an option on certain OPNs or with certain memory populations. The default configuration (one NUMA domain per socket) is recommended for most workloads. NPS4 is recommended for HPC and other highly parallel workloads. When using 200 Gbps network adapters, NPS2 may be preferred to provide a compromise between memory latency and memory bandwidth for the NIC. This setting is independent of ACPI SRAT L3 Cache as NUMA Domain in section 2.2.1. When ACPI SRAT L3 Cache as NUMA Domain is enabled, this setting now determines the memory interleaving granularity. With NPS1, all eight memory channels are interleaved. With NPS2, every four channels are interleaved with each other. With NPS4, every pair of channels is interleaved.

Setting	Options
NUMA Node per Socket	• NPS0: Interleave memory accesses across all channels in both sockets (not recommended) • NPS1: Interleave memory accesses across all eight channels in each socket, report one NUMA node per socket (unless L3 Cache as NUMA is enabled) • NPS2: Interleave memory accesses across groups of four channels (ABCD and EFGH) in each socket, report two NUMA nodes per socket (unless L3 Cache as NUMA is enabled) • NPS4: Interleave memory accesses across pairs of two channels (AB, CD, EF and GH) in each socket, report four NUMA nodes per socket (unless L3 Cache as NUMA is enabled)

2.2.3 Memory Clock Speed

Benefit: By default, the BIOS for EPYC 7002 Series processors will run at the maximum allowable clock frequency by the platform. This configuration results in the maximum memory bandwidth for the processor, but in some cases, it may not be the lowest latency. The Infinity Fabric will have a maximum speed of 1467 MHz (lower in some platforms), resulting in a single clock penalty to transfer data from the memory channels onto the Infinity Fabric to progress through the SoC. To achieve the lowest latency, you can set the memory frequency to be equal to the Infinity Fabric speed. Lowering the memory clock speed also results in power savings in the memory controller, thus allowing the rest of the SoC to consume more power potentially resulting in a performance boost elsewhere, depending on the workload.

Setting	Options
Memory Clock Speed	• Auto: Determine maximum memory speed based on SPD information from populated DIMMs and from platform memory speed support. • Values 400 MHz–1600 MHz: run DRAM memory clock at the specified speed (note DRAM memory clock is half of the DDR rate)

2.2.4 Transparent Secure Memory Encryption (TSME)

Benefit: TSME provides hardware memory encryption of all data stored on system DIMMs. This encryption is invisible to the OS. The impact of this encryption is 5 ns–7 ns of additional memory latency.

Setting	Options
TSME	• Auto/Disabled: Disable transparent secure memory encryption • Enabled: Enable transparent secure memory encryption

2.3 Power Efficiency Settings

2.3.1 Core Clock Dynamic Power Management (CCLK DPM)

Benefit: When enabled, the SoC Efficiency mode will maximize the Performance-per-Watt by opportunistically reducing the core clocks via a dynamic power management algorithm. This internal algorithm to maximize the Performance-per-Watt is targeted at throughput-based server workloads that exhibit a stable load below the SoC maximum capabilities. The default, auto, maximizes performance of the SoC.

Setting	Options
EfficiencyModeEn	• Auto: Optimize core clock dynamic power management for performance • Enabled: Optimize core clock dynamic power management for power efficiency

2.3.2 Power vs Performance Determinism Settings

Benefit: The determinism slider selects between uniform performance across identically configured systems in a datacenter (Performance setting) or maximum performance of any individual system but with varying performance across the datacenter (Power setting). If setting this to Performance, ensure that cTDP and PPL are set to the same value (see Processor Cooling and Power Dissipation Limit Settings for more details). The default (Auto) setting for most OPNs will be Performance Determinism mode.

Setting	Options
Determinism Control	• Auto: Don’t show the control • Manual: Enable the Determinism Slider control
Determinism Slider	• Auto: Determined by OPN fusing • Power: Ensure maximum performance levels for each CPU in a large population of identically configured CPUs by throttling CPUs only when they reach the same cTDP • Performance: Ensure consistent performance levels across a large population of identically configured CPUs by throttling some CPUs to operate at a lower power level

2.3.3 Processor Cooling and Power Dissipation Limit Settings

Benefit: Configurable Thermal Design Power (cTDP) allows the user to modify the platform CPU cooling limit and the Package Power Limit (PPL) allows the user to modify the CPU Power Dissipation Limit. Many platforms will configure cTDP to the maximum supported by the installed CPU. For example, an EPYC 7502 part has a default TDP of

180 W but has a cTDP maximum of 200 W. Most platforms also configure the PPL to the same value as the cTDP. If Performance Determinism is desired (see Power vs Performance Determinism Settings), these two values must be set to the same value. Otherwise, the user can set PPL to a value lower than cTDP to reduce system operating power. The CPU will control CPU boost to keep socket power dissipation at or below the specified Package Power Limit.

Setting	Options
cTDP Control	• Manual: Set customized configurable TDP • Auto: Use platform & OPN default TDP
cTDP	• Values 85–280: Set configurable TDP (in Watts)
Package Power Limit Control	• Manual: Set customized Package Power Limit (PPL) • Auto: Use platform & OPN default Package Power Limit
Package Power Limit	• Values 85–280: Set Package Power Limit (in Watts)

2.3.4 ACPI—Collaborative Processor Performance Control (CPCC)

Benefit: CPCC was introduced with ACPI 5.0 as a mode to communicate performance between an operating system and the hardware. This mode can be used to allow the OS to control when and how much turbo can be applied in an effort to maintain energy efficiency. Not all operating systems support CPCC, but Microsoft began support with Windows® 2016.

Setting	Options
CPCC	• Disabled: Disabled • Enabled: Allow OS to make performance/power optimization requests using ACPI CPPC

2.4 Processor Core Settings

2.4.1 Cache Prefetchers

Benefit: Most workloads will benefit from the L1 & L2 Stream Hardware prefetchers gathering data and keeping the core pipeline busy. There are however some workloads that are very random in nature and will actually obtain better overall performance by disabling one or both of the prefetchers. One example workload is SPECjbb 2015, a Java heavy workload. By default, both prefetchers are enabled.

Setting	Options
L1 Stream HW Prefetcher	• Disable: Disable prefetcher • Enable: Enable prefetcher
L2 Stream HW Prefetcher	• Disable: Disable prefetcher • Enable: Enable prefetcher

2.4.2 Symmetric Multithreading (SMT) Settings

Benefit: Some workloads, including many HPC ones, observe a performance neutral or even performance negative result when SMT is enabled. Some applications license by the hardware thread enabled, not just physical core. For those reasons, disabling SMT on your EPYC 7002 Series processor may be desirable. There are also some operating systems that have not enabled support for the x2APIC within the EPYC 7002 Series Processor, which is required to support beyond 255 threads. If you are running an operating system that does not support AMD's x2APIC implementation, and have two 64-core processors installed, you will need to disable SMT.

Setting	Options
SMT Control	• Disable: Single hardware thread per core • Auto: Two hardware threads per core

2.4.3 Core Boost Frequency Settings

Benefit: Some workloads do not need to be able to run at the maximum capable core frequency to achieve acceptable levels of performance. To obtain better power efficiency, there is the option of setting a maximum core boost frequency. This setting does not allow you to set a fixed frequency. It only limits the maximum boost frequency. If the BoostFmax is set to something higher than the boost algorithms allow, the SoC will not go beyond the allowable frequency that the algorithms support. Actual boost performance will be dependent on many factors and other settings mentioned in this document.

Setting	Options
BoostFmaxEn	• Manual: Use BoostFmax setting • Auto: Use default BoostFmax setting
BoostFmax	• Values 0x0–0xFFFFFFFF: f_{Max} frequency limit to apply to all cores in MHz

2.5 I/O Settings

2.5.1 APIC Settings

Benefit: AMD introduced an x2APIC in our EPYC 7002 Series processors for the first time. In general, interrupt delivery is faster when using x2APIC mode over the legacy xAPIC mode. However, not all operating systems support AMD's x2APIC implementation, so you need to check for support prior to enabling. If your operating system supports x2APIC mode, this mode is recommended even in a configuration with fewer than 256 logical processors.

Setting	Options
Local APIC Mode	• xAPIC: Use xAPIC, scales to only 255 hardware threads • x2APIC: Use x2APIC, scales beyond 255 hardware threads but not supported by some legacy OS versions • Auto: Use x2APIC only if 256 hardware threads in system, otherwise use xAPIC

2.5.2 SR-IOV Settings

Benefit: SR-IOV requires the enablement of PCIe® Alternative Routing-ID interpretation (ARI) on both root complexes and endpoints. ARI devices interpret the PCI address as an 8-bit function number instead of a 3-bit function number and the device number is implied to be 0.

Setting	Options
PCIe ARI Support [SRIOV]	• Disable: Disable Alternative Routing ID interpretation • Enable: Enable Alternative Routing ID interpretation

2.5.3 PCIe Ten Bit Tag

Benefit: With PCIe Gen 4, to achieve full bandwidth, an adapter should support 10-bit extended tags. This allows for a 3x increase over the previous number of non-posted requests, allowing the adapter to achieve more performance. Not all PCIe Gen 4 devices support 10-bit extended tags and can cause issues during boot. Disabling this feature will allow the server to boot if the adapter is having issues.

Setting	Options
PCIe Ten Bit Tag Support	• Disable: Disable PCIe 10-bit tags for all devices • Enable: Enable PCIe 10-bit tags for supported devices • Auto: Disabled

2.5.4 Preferred I/O Settings

Benefit: Allows devices on a single PCIe bus to obtain improved DMA write performance. See the *Preferred I/O Usage Guide*, #56570, for more details.

Setting	Options
Preferred I/O	• Manual: Enable Preferred I/O for the bus number specified by Preferred I/O Bus • Auto: Disabled
Preferred I/O Bus	• Values 00h–FFh: Specify the bus number for which device(s) you wish to enable preferred I/O

2.5.5 Input-Output Memory Management Unit (IOMMU) Settings

Benefit: The IOMMU provides several benefits and is required when using x2APIC. Enabling the IOMMU allows devices (such as the EPYC integrated SATA controller) to present separate IRQs for each attached device instead of one IRQ for the subsystem. The IOMMU also allows operating systems to provide additional protection for DMA capable I/O devices. If the IOMMU is believed to be causing a performance problem, leave it enabled in BIOS and disable it via OS options (e.g., iommu=pt on the Linux® kernel command line).

Setting	Options
IOMMU	• Disabled: Disable IOMMU • Enabled: Enable IOMMU

Chapter 3 BIOS Setting Selection by Workload

Use these guidelines for general-purpose workloads. In some cases, benchmarks are listed to better describe what workloads AMD used to obtain these settings. Default settings are used when labeled *default*.

3.1 General-Purpose Workloads

3.1.1 Infinity Fabric Settings

Infinity Fabric Settings	SPEC CPU® 2017	SPECjbb® 2015	SPECpower_ssj® 2008	I/O Intensive
xGMI Link Max Speed	*default*	*default*	*default*	*default*
xGMI Link Width Control	*default*	*default*	*default*	*default*
xGMI Max Link Width	*default*	*default*	*default*	*default*
xGMI Max Link Width Control	*default*	*default*	*default*	*default*
xGMI Force Link Width Enable	*default*	*default*	Enable	*default*
xGMI Force Link Width	*default*	*default*	x16	*default*
APBDIS	*default*	*default*	*default*	*default*
Fixed SOC P-State [see ABDIS]	*default*	*default*	*default*	*default*
DF C-States	*default*	*default*	*default*	*default*

3.1.2 NUMA and Memory Settings

NUMA and Memory Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
ACPI SRAT L3 Cache as NUMA Domain	*default*	*default*	*default*	*default*
ACPI SLIT Distance Control	*default*	*default*	*default*	*default*

NUMA and Memory Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
ACPI SLIT Remote Relative Distance	*default*	*default*	*default*	*default*
ACPI SLIT <various> Distance	*default*	*default*	*default*	*default*
NUMA Nodes per Socket (NPS)	*default*	4/2	4	*default*
Memory Clock Speed	*default*	1467	1200	*default*
TSME	Disabled	Disabled	Disabled	Disabled

3.1.3 Power Efficiency Settings

Power Efficiency Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
EfficiencyModeEn	*default*	*default*	*default*	*default*
Determinism Control	Enabled	Enabled	Enabled	Enabled
Determinism Slider	Power	Power	Power	Power
cTDP Control	Manual	Manual	Manual	*default*
cTDP	OPN Max	OPN Max	OPN Max	*default*
Package Power Limit Control	Manual	Manual	Manual	*default*
Package Power Limit	OPN Max	OPN Max	OPN Max	*default*
CPPC	*default*	*default*	Enabled	*default*

3.1.4 Processor Core Settings

Processor Core Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
L1 Stream HW Prefetcher	*default*	Disabled	Disabled	*default*
L2 Stream HW Prefetcher	*default*	Disabled	Disabled	*default*
SMT Control	Enabled	Enabled	Enabled	Enabled

Processor Core Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
BoostFmaxEn	*default*	*default*	1	*default*
BoostFmax	*default*	*default*	2500	*default*

3.1.5 I/O Settings

I/O Settings	SPEC CPU 2017	SPECjbb 2015	SPECpower_ssj 2008	I/O Intensive
Local APIC Mode	*default*	*default*	*default*	x2APIC
PCIe ARI Support [SRIOV]	*default*	*default*	*default*	*default*
PCIe Ten Bit Tag Support	*default*	*default*	*default*	Enabled (if supported by NIC)
Preferred I/O	*default*	*default*	*default*	*default*
Preferred I/O Bus	*default*	*default*	*default*	*default*
IOMMU	*default*	*default*	*default*	*default*

3.2 Virtualization and Containers

3.2.1 Infinity Fabric Settings

Infinity Fabric Settings	VMmark® 3	TPC®x-V	Containers
xGMI Link Max Speed	*default*	*default*	*default*
xGMI Link Width Control	*default*	*default*	*default*
xGMI Max Link Width	*default*	*default*	*default*
xGMI Max Link Width Control	*default*	*default*	*default*
xGMI Force Link Width Enable	*default*	*default*	*default*
xGMI Force Link Width	*default*	*default*	*default*
APBDIS	*default*	*default*	*default*
Fixed SOC P-State [see ABDIS]	*default*	*default*	*default*
DF C-states	*default*	*default*	*default*

3.2.2 NUMA and Memory Settings

NUMA and Memory Settings	VMmark 3	TPCx-V	Containers
ACPI SRAT L3 Cache as NUMA Domain	Enabled	Enabled	*default*
ACPI SLIT Distance Control	*default*	*default*	*default*
ACPI SLIT Remote Relative Distance	*default*	*default*	*default*
ACPI SLIT <various> distance	*default*	*default*	*default*
NUMA Nodes per Socket (NPS)	*default*	4	4
Memory Clock Speed	*default*	*default*	*default*
TSME	Disabled	Disabled	Disabled

3.2.3 Power Efficiency Settings

Power Efficiency Settings	VMmark 3	TPCx-V	Containers
EfficiencyModeEn	*default*	*default*	*default*
Determinism Control	Enabled	Enabled	Enabled
Determinism Slider	Power	Performance	Power
cTDP Control	Manual	Manual	Manual
cTDP	OPN Max	OPN Max	OPN Max
Package Power Limit Control	Manual	Manual	Manual
Package Power Limit	OPN Max	OPN Max	OPN Max
CPPC	*default*	*default*	*default*

3.2.4 Processor Core Settings

Processor Core Settings	VMmark 3	TPCx-V	Containers
L1 Stream HW Prefetcher	*default*	*default*	*default*
L2 Stream HW Prefetcher	*default*	*default*	*default*
SMT Control	Enabled	Enabled	Enabled
BoostFmaxEn	*default*	*default*	*default*
BoostFmax	*default*	*default*	*default*

3.2.5 I/O Settings

I/O Settings	VMmark 3	TPCx-V	Containers
Local APIC Mode	*default*	*default*	*default*
PCIe ARI Support [SRIOV]	*default*	*default*	*default*
PCIe Ten Bit Tag Support	*default*	*default*	*default*
Preferred I/O	*default*	*default*	*default*
Preferred I/O Bus	*default*	*default*	*default*
IOMMU	*default*	*default*	*default*

3.3 Database and Analytics

3.3.1 Infinity Fabric Settings

Infinity Fabric Settings	HammerDB	Hadoop	TPCx-IOT
xGMI Link Max Speed	*default*	*default*	*default*
xGMI Link Width Control	*default*	*default*	*default*
xGMI Max Link Width	*default*	*default*	*default*
xGMI Max Link Width Control	*default*	*default*	*default*
xGMI Force Link Width Enable	*default*	*default*	*default*
xGMI Force Link Width	*default*	*default*	*default*
APBDIS	*default*	*default*	*default*
Fixed SOC P-State [see ABDIS]	*default*	*default*	*default*
DF C-States	*default*	*default*	*default*

3.3.2 NUMA and Memory Settings

NUMA and Memory Settings	HammerDB	Hadoop	TPCx-IOT
ACPI SRAT L3 Cache as NUMA Domain	*default*	*default*	*default*
ACPI SLIT Distance Control	*default*	*default*	*default*
ACPI SLIT Remote Relative Distance	*default*	*default*	*default*

NUMA and Memory Settings	HammerDB	Hadoop	TPCx-IOT
ACPI SLIT <various> Distance	*default*	*default*	*default*
NUMA Nodes per Socket (NPS)	*default*	4	4
Memory Clock Speed	*default*	*default*	*default*
TSME	Disabled	Disabled	Disabled

3.3.3 Power Efficiency Settings

Power Efficiency Settings	HammerDB	Hadoop	TPCx-IOT
EfficiencyModeEn	*default*	*default*	*default*
Determinism Control	Enabled	Enabled	Enabled
Determinism Slider	Power	Power	Power
cTDP Control	Manual	Manual	Manual
cTDP	OPN Max	OPN Max	OPN Max
Package Power Limit Control	Manual	Manual	Manual
Package Power Limit	OPN Max	OPN Max	OPN Max
CPPC	*default*	*default*	*default*

3.3.4 Processor Core Settings

Processor Core Settings	HammerDB	Hadoop	TPCx-IOT
L1 Stream HW Prefetcher	*default*	*default*	*default*
L2 Stream HW Prefetcher	*default*	*default*	*default*
SMT Control	Enabled	Enabled	Enabled
BoostFmaxEn	*default*	*default*	*default*
BoostFmax	*default*	*default*	*default*

3.3.5 I/O Settings

I/O Settings	HammerDB	Hadoop	TPCx-IOT
Local APIC Mode	*default*	*default*	*default*
PCIe ARI Support [SRIOV]	*default*	*default*	*default*
PCIe Ten Bit Tag Support	*default*	*default*	*default*
Preferred I/O	*default*	*default*	*default*

I/O Settings	HammerDB	Hadoop	TPCx-IOT
Preferred I/O Bus	*default*	*default*	*default*
IOMMU	*default*	*default*	*default*

3.4 HPC and Telco Settings

3.4.1 Infinity Fabric Settings

Infinity Fabric Settings	HPC	OpenStack® NFV	OpenStack® for RealTime Kernel (NFV)	EDA
xGMI Link Max Speed	*default*	*default*	*default*	*default*
xGMI Link Width Control	Manual	*default*	*default*	*default*
xGMI Max Link Width	1 (x16)	*default*	*default*	*default*
xGMI Max Link Width Control	Manual	*default*	*default*	*default*
xGMI Force Link Width Enable	Enabled	*default*	Enabled	*default*
xGMI Force Link Width	x16	*default*	x16	*default*
APBDIS	1	*default*	*default*	*default*
Fixed SOC P-State [see ABDIS]	P0	*default*	*default*	*default*
DF C-States	*default*	*default*	*default*	*default*

3.4.2 NUMA and Memory Settings

NUMA and Memory Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
ACPI SRAT L3 Cache as NUMA Domain	*default*	*default*	*default*	*default*
ACPI SLIT Distance Control	Manual	*default*	*default*	*default*
ACPI SLIT Remote Relative Distance	22	*default*	*default*	*default*
ACPI SLIT <various> Distance	*default*	*default*	*default*	*default*

NUMA and Memory Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
NUMA Nodes per Socket (NPS)	4/2	2	2	4
Memory Clock Speed	CFD & Other Memory Bound: Auto Chem, Physics & Other SIMD Bound: Try 1467	*default*	*default*	*default*
TSME	Disabled	Disabled	Disabled	*default*

3.4.3 Power Efficiency Settings

Power Efficiency Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
EfficiencyModeEn	*default*	*default*	*default*	*default*
Determinism Control	Enabled	Enabled	Enabled	Enabled
Determinism Slider	Power	Power	Power	Power
cTDP Control	Manual	Manual	Manual	Manual
cTDP	OPN Max	OPN Max	OPN Max	OPN Max
Package Power Limit Control	Manual	Manual	Manual	Manual
Package Power Limit	OPN Max	OPN Max	OPN Max	OPN Max
CPPC	*default*	*default*	*default*	*default*

3.4.4 Processor Core Settings

Processor Core Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
L1 Stream HW Prefetcher	*default*	Enabled	Enabled	Enabled
L2 Stream HW Prefetcher	*default*	Enabled	Enabled	Enabled
SMT Control	Disabled	*default*	*default*	Disabled
BoostFmaxEn	*default*	Disabled	Disabled	*default*

Processor Core Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
BoostFmax	*default*	*default*	*default*	*default*

3.4.5 I/O Settings

I/O Settings	HPC	OpenStack NFV	OpenStack for RealTime Kernel (NFV)	EDA
Local APIC Mode	X2APIC	x2APIC	x2APIC	*default*
PCIe ARI Support [SRIOV]	*default*	Enabled	Enabled	*default*
PCIe 10-Bit Tag Support	*default*	Enabled	Enabled	*default*
Preferred I/O	Enabled if Only One IB NIC	Enabled if Only One NIC	Enabled if Only One NIC	*default*
Preferred I/O Bus	IB NIC Bus Number	NIC Bus Number	NIC Bus Number	*default*
IOMMU	Disabled	Enabled	Enabled	*default*