IBM appreciates the EPA's efforts in completing the second draft of the ENERGY STAR® Product Specification for Computer Servers Version 3 and the fact that the draft addresses IBM Draft one comments on issues with the direct attachment of APA chips and the need to adjust or eliminate idle power limitations for resilient servers.

IBM generally supports the full set of comments submitted by Information Technology Industry Council (ITI) and will not repeat those comments here. The ITI comments cover the majority of the specific questions and requests for comment made by EPA in the draft 2 document. We offer the following, additional, IBM specific comments.

Lines 61-65: Modify the resilient server definition. A new resilient server technical definition is proposed at the end of the paper under Appendix B. It is the same definition as provided in the ITI comments.

Lines 79-95: High Performance Computing (HPC) Definition: The definition changes proposed by ITI are important because HPC servers and systems are becoming increasingly sophisticated as new technologies are introduced and as their application is expanded to Artificial and Augmented Intelligence and Deep Learning applications. Higher performing processors and GPUs, when combined with advances in the speed and bandwidth of communications links, are able to perform HPC tasks with a smaller number of heterogeneous nodes, necessitating modifications in the structural definitions provided by the definition.

Lines 212-216: APA Definition: IBM supports the differentiation of expansion and integrated Auxiliary Processing Accelerators in the APA definition and the exclusion of integrated APAs from the requirements in the Version 3 server requirements. At this time and for the near future, integrated APAs will largely be confined to HPC systems. It is likely that integrated APAs will be incorporated into general cloud and enterprise servers over the next several years. EPA is encouraged to work with industry to determine the best approach to manage these systems in ENERGY STAR Version 4. One promising possibility is to design integrated APA systems such that the APA can be turned off through the BMC/FSP or operating system so that server performance and power measurements could be collected independent of the APAs.

IBM is also working with its GPU supplier to develop data on the idle power use of server GPUs. We are hopeful that the IBM partner will be able to provide data for the October 16, 2017 ITI submittal to EPA.

Lines 579-580: APA Idle Limit: Based on IBM's internal assessment of net generation GPU products and their ability to reduce idle power, IBM has reason to believe that the proposed 30 watt idle limit for GPU expansion cards is too conservative for server GPU products currently on or planned for the market.

IBM is working with its GPU supplier to develop data on the idle power use of server GPUs. We are hopeful that the IBM partner will be able to provide data for the October 16, 2017 ITI submittal to EPA.

Lines 241 to 256: Low-end, Typical and High-end Performance Configurations: IBM supports the open memory capacity requirements for the 3 test configurations that define the product family so that IBM is able to select the memory size and capacity that enables it to optimize the SERT active efficiency score for its products.

IBM also appreciates EPA's approach to certifying a server family by identifying the range of configurations which meet the ENERGY STAR active efficiency and idle power thresholds and that the range of qualified configurations can be a subset of the total configurations which can be created for a given server product family. Again, this will provide IBM flexibility in identifying and qualify ENERGY STAR certified product families.

Lines 412 to 413: Active state efficiency thresholds: IBM supports the use of SERT V2.0.0 for ENERGY STAR servers V3. Given the impact of the changes on the memory worklet scores on the active efficiency scores and because of the fact that resilient servers use large memory capacities, IBM encourages EPA to carefully assess the revised V2.0.0 active efficiency scores and assure that the selected active efficiency certification thresholds are set appropriately for resilient servers.

New Processor Announcements: IBM can be expected to announce server products using the Power9® processors sometime in the future. If available during the development time frame of ENERGY STAR server V3, IBM will try to provide EPA with an assessment of how active efficiency scores may change for servers built with the next generation Power processors.

Appendix B: Resilient Server

Changes in server technology require modifications to the resilient server definition as most, but not all, resilient servers will no longer use memory buffer chips. The "Proposed Resilient Server Definition" in column B in Table 1 represents a consensus definition developed by members of ITI and Green Grid. Column A provides the current Version 2 resilient server definition with annotations in red regarding changes for reference.

For previous generation of products, there were a limited subset of processors that could qualify a server in the resilient category. Under the new x86 processor family, the requisite RAS functions required to meet the resilient server definition are available in both the gold and platinum level of processors. However, the RAS functions available in the processor have to be enabled in the firmware and/or operating systems in the product family and the server has to possess the redundant, hot swappable systems and memory capabilities required to qualify as a resilient server. It is expected that these considerations will continue to limit resilient servers to a small percentage of the overall server market and that resilient systems will continue to have the higher power profile driven by redundant, hot swappable components and support systems and the higher power demanding RAS functions.

One addition to the definition that EPA could consider, that would both limit the number of processors which can certify resilient server systems and acknowledge the fact that servers capable of supporting high memory capacity require additional, higher levels of power than volume servers, would be to include a requirement that resilient servers be able to provide a maximum memory capacity of 1.5 Terabytes.

ENERGY STAR v3 (Draft 1//2)	Proposed Resilient Server Definition
Definition: A computer server designed with extensive Reliability, Availability, Serviceability (RAS) and scalability features integrated in the micro architecture of the system, CPU and chipset. For purposes of ENERGY STAR certification under this specification, a Resilient Server shall have the characteristics as described in Appendix B of this specification.	Definition: A computer server designed with extensive Reliability, Availability, Serviceability (RAS) and scalability features integrated in the micro architecture of the system, CPU and chipset. For purposes of ENERGY STAR certification under this specification, a Resilient Server shall have the following characteristics
Appendix B	
A. Processor RAS and Scalability- All of the following shall be supported: (1) Processor RAS: The processor must have capabilities to detect, correct, and contain data errors, as described by all of the following: (retained) (a) Error detection on L1 caches, directories and address translation buffers using parity protection; (retained) (b) Single bit error correction (or better) using ECC on caches that can contain modified data. Corrected data is delivered to the recipient (i.e., error correction is not used just for background scrubbing); (Modified slightly) (c) Error recovery and containment by means of (1) processor checkpoint retry and recovery, (2) data poison indication (tagging) and propagation, or (3) both. The mechanisms notify the OS or hypervisor to contain the error within a process or partition, thereby reducing the need for system reboots; and (retained – moved under System Recovery & Resiliency) (d) (1) Capable of autonomous error mitigation actions within processor hardware, such as disabling of the failing portions of a cache, (2) support for predictive failure analysis by notifying the OS, hypervisor, or service processor of the location and/or root cause of errors, or (3) both. (Deleted) (2) The processor technology used in resilient and scalable servers is designed to provide additional capability and functionality without additional chipsets, enabling them to be designed into systems with 4 or more processor sockets. ~~The processors have additional infrastructure to~~	Processor RAS: The processor must have capabilities to detect, correct, and contain data errors, as described by all of the following: 1. Error recovery by means of instruction retry for certain processor faults. 2. Error detection on L1 caches, directories and address translation buffers using parity protection; 3. Single bit error correction (or better) on caches that can contain modified data. Corrected data is delivered to the recipient as part of the request completion. System Recovery & Resiliency: No fewer than six of the following characteristics shall be present in the server: 1. Error recovery and containment by means of (1) data poison indication (tagging) and propagation which Includes mechanism to notify the OS or hypervisor to contain the error, thereby reducing the need for system reboots. (2) Containment of address/command errors by preventing possibly contaminated data from being committed to permanent storage. 2. The processor technology used in resilient and scalable servers is designed to provide additional capability and functionality without additional chipsets, enabling them to be designed into systems with 4 or more processor sockets. 3. Memory Mirroring: A portion of Available memory can be proactively partitioned such that a duplicate set may be utilized upon non-correctable memory errors. This can be implemented at the granularity of DIMMs or logical memory blocks. 4. Memory Sparing: A portion of available memory may be pre-allocated to a spare

~~support extra, built in processor busses to support the demand of larger systems.~~ (Modified) (3) The server provides high bandwidth I/O interfaces for connecting to external I/O expansion devices or remote I/O without reducing the number of processor sockets that can be connected together. These may be proprietary interfaces or standard interfaces such as PCIe. The high performance I/O controller to support these slots may be embedded within the main processor socket or on the system board (Deleted) B. Memory RAS and Scalability - All of the following capabilities and characteristics shall be present: (1) Provides memory fault detection and recovery through Extended ECC; (Deleted) (2) In x4 DIMMs, recovery from failure of two adjacent chips in the same rank; (Deleted) (3) Memory migration: Failing memory can be proactively de-allocated and data migrated to available memory. This can be implemented at the granularity of DIMMs or logical memory blocks. Alternatively, memory can also be mirrored; (Modified - addressed under #3) (4) Uses memory buffers for connection of higher speed processor -memory links to DIMMs attached to lower speed DDR channels. Memory buffer can be a separate, standalone buffer chip which is integrated on the system board, or integrated on custom-built memory cards. The use of the buffer chip is required for extended DIMM support; they allow larger memory capacity due to support for larger capacity DIMMs, more DIMM slots per memory channel, and higher memory bandwidth per memory channel than direct-attached DIMMs. The memory modules may also be custom- built, with the memory buffers and DRAM chips integrated on the same card; (Deleted) (5) Uses resilient links between processors and memory buffers with mechanisms to recover from transient errors on the link; and (Modified - addressed under #8) (6) Lane sparing in the processor-memory links. One or more spare lanes are available for lane failover in the event of permanent error. (Modified - addressed under #4)	function such that data may be migrated to the spare upon a perceived impending failure. (New) 5. Support for making additional resources available without the need for a system restart. This may be achieved either by processor (cores, memory, IO) on-lining support, or by dynamic allocation/deallocation of processor cores, memory and IO to a partition. 6. Support of redundant IO devices (storage controllers, networking controllers) 7. Has I/O adapters or storage devices that are hot-swappable 8. Identify failing Processor-to-Processor lane(s) and dynamically reduce the width of the link in order to use only non-failing lanes or provide a spare lane for failover without disruption. (New) 9. Capability to partition the system such that it enables running instances of the OS or hypervisor in separate partitions. Partition isolation is enforced by the platform and/or hypervisor and each partition is capable of independently booting.(New) 10. Uses memory buffers for connection of higher speed processor -memory links to DIMMs attached to lower speed DDR channels. Memory buffer can be a separate, standalone buffer chip which is integrated on the system board, or integrated on custom-built memory cards. .
C. Power Supply RAS: All PSUs installed or shipped with the server shall be redundant and concurrently maintainable. The redundant and	Power Supply RAS: All PSUs installed or shipped with the server shall be redundant and concurrently maintainable. The redundant and repairable

repairable components may also be housed within a single physical power supply, but must be repairable without requiring the system to be powered down. Support must be present to operate the system in degraded mode ~~when power delivery capability is degraded due to failures in the power supplies or input power loss~~. (Partially deleted)	components may also be housed within a single physical power supply, but must be repairable without requiring the system to be powered down. Support must be present to operate the system in degraded mode.
D. Thermal and Cooling RAS: All active cooling components, ~~such as fans or water-based cooling~~, shall be redundant and concurrently maintainable. The processor complex must have mechanisms to allow it to be throttled under thermal emergencies. Support must be present to operate the system in degraded mode when thermal emergencies are detected in system components. (Removed water-based cooling having redundant components)	Thermal and Cooling RAS: All active cooling components shall be redundant and concurrently maintainable. The processor complex must have mechanisms to allow it to be throttled under thermal emergencies. Support must be present to operate the system in degraded mode when thermal emergencies are detected in system components
E. System Resiliency: – no fewer than six of the following characteristics shall be present in the server: (Mostly addressed under 'System Recovery and Resiliency'; a few deleted) (1) Support of redundant storage controllers or redundant path to external storage; (Deleted – addressed under #6) (2) Redundant service processors; (Deleted – addressed under #6) (3) Redundant dc-dc regulator stages after the power supply outputs; (Deleted – addressed under #6) (4) The server hardware supports runtime processor de-allocation; (Deleted) (5) I/O adapters or hard drives are hot-swappable; ; (Deleted – addressed under #6) (6) Provides end to end bus error retry on processor to memory or processor to processor interconnects; (Deleted – addressed under #8) (7) Supports on-line expansion/retraction of hardware resources without the need for operating system reboot ("on-demand" features); (Deleted – addressed under #9) (8) Processor Socket migration: With hypervisor and/or OS assistance, tasks executing on a processor socket can be migrated to another processor socket without the need for the system to be restarted; (Deleted – addressed under #9) (9) Memory patrol or background scrubbing is enabled for proactive detection and correction of	

errors to reduce the likelihood of uncorrectable errors; and (Deleted) (10) Internal storage resiliency: Resilient systems have some form of RAID hardware in the base configuration, either through support on the system board or a dedicated slot for a RAID controller card for support of the server's internal drives. (Deleted)	
F. System Scalability – All of the following shall be present in the server: (1) Higher memory capacity: >=8 DDR3 or DDR4 DIMM Ports per socket, with resilient links between the processor socket and memory buffers; and (Deleted) (2) Greater I/O expandability: Larger base I/O infrastructure and support a higher number of I/O slots. Provide at least 32 dedicated PCIe Gen 2 lanes or equivalent I/O bandwidth, with at least one x16 slot or other dedicated interface to support external PCIe, proprietary I/O interface or other industry standard I/O interface (Deleted)	

Table 1: Proposal for a revised Resilient Server Definition