June 6, 2012

To: Robert J. Meyers
Project Manager, ENERGY STAR for Computer Servers
United States Environmental Protection Agency

Re: Hewlett-Packard Response to *ENERGY STAR® Program Requirements, Product Specification for Computer Servers*, Draft 2 of Version 2.0

From: Hewlett-Packard Company, Enterprise Group

This document may be published on the ENERGY STAR website.

Hewlett-Packard (HP) has a long-standing association with the *ENERGY STAR®* program and HP welcomes this opportunity to participate as a valued stakeholder in the process of creating Version 2.0 of the *ENERGY STAR® Program Requirements for Computer Servers*.

The consistent feedback that should be apparent in both this and previous replies is the need to enable ENERGY STAR partner companies to swiftly, accurately, economically and efficiently certify all server configurations that meet the ENERGY STAR technical criteria. It is also highly important that server manufacturers be enabled to do in-house certification testing and that there is no interruption of any lab certifications when version 2.0 is introduced. The following commentary provides the means to help meet these goals and to foster future improvements in server energy efficiency.

1. Qualifying Products Eligibility Criteria

The sections below discuss clarifications, issues, changes, and suggested solutions to creating Eligibility Criteria that can better help all eligible servers to receive their earned recognition and ENERGY STAR certification status.

1.1. Section 1, Definitions

1.1.1. Line 16. Defining a Computer Server as one that is targeted for "enterprise applications" means that the specification is not intended for servers that are designed for other purposes (even if they may be deployed with enterprise applications). This specification also needs to comprehend that Computer Servers that are designed for enterprise applications are often deployed into non-enterprise applications (e.g. High Performance Computing) and so an enterprise server deployed into a non-enterprise application should still be able to carry the ENERGY STAR certification.

1.1.2. Lines 70-84, Resilient Server definition. Based on our analysis of servers, we have identified system attributes which drive a higher level of base power use (please refer to Section 3 of this reply). As criteria for identifying a Resilient

Server, the recommendation is to require any qualifying server to meet the Reliability, Availability, Serviceability (RAS) Capabilities criterion (see Section 3) and also satisfy 2 of the remaining 3 features or attributes. The use of a specified number of "menu items" for RAS capabilities, high power processor socket attributes and the 6 main Resilient Server criteria, is necessitated because different servers and processor types will have different attributes depending on the proposed applications and system component capabilities. It is important to provide flexibility in conforming to the requirements while setting requirements that differentiate Managed Servers from Resilient Servers.

1.2. Section 2, Scope

1.2.1. Lines 265-278. HP supports the inclusion of multi-node servers for participation in ENERGY STAR version 2.0 for Computer Servers. It is not clear from the stated inclusions and exclusions whether multi-node servers are included. The Multi-node Server testing and certification process is very similar to that of Blade Servers and should have similar certification criteria.

1.2.2. Lines 265-278. HP supports the inclusion of Resilient Servers with similar qualification criteria to 3S/4S Computer Servers. It is not clear in this section if Resilient Servers are eligible or excluded.

1.3. Section 3, Qualification Criteria

1.3.1. Section 3.2 of the Draft specification uniformly discusses power supply units (PSUs) as being "in the chassis". A more generic and preferable approach would be to require that PSUs supplying power to a chassis must meet the requirements of Table 2. This change would support power supply solutions that are aggregated at a multi-server level, but might not be classifiable as either multi-node or blade servers.

1.3.2. Section 3.2.3 of the Draft specification needs to include dual-node and multi-node servers, which also use power supplies similar to pedestal, tower and/or rack mount servers.

1.3.3. Section 3.3.1 needs to include "or the operating system" when discussing allowable methods of enabling processor power management.

1.3.4. HP supports the decision to treat the qualification criteria of blade servers (and multi-node servers) much like the specification treats 3 and 4 socket servers.

1.3.5. Line 398. A significant issue continues to exist for the idle power criteria of two-socket servers that ship with only one processor. The ENERGY STAR program must refrain from dictating market choices. Customer demand and sales volumes for 2S/1P servers are very high, and 1S/1P servers have orders of magnitude smaller sales volumes and are deployed in different application environments. The ENERGY STAR program needs to reflect that reality and add to Table 3 a Category E for 2S/1P unmanaged servers and Category F for 2S/1P managed servers with the recommended Base Idle Power allowances shown below (based on empirical test results):

Table 3 Category	#P	Managed Server	Base Idle Power
E	2S/1P	No	85W
F	2S/1P	Yes	135W

1.3.6. Line 406. HP advocates keeping the version 1.1 idle power allowance for additional power supplies (20W per additional power supply). The slight increase in version 2.0 power supply efficiency requirements does not translate into significant wattage savings at idle.

1.3.7. Line 406. The idle power allowance of 0.75 watts per GB is a statistical median value for 4GB DIMMs at idle and not a value that guarantees that all similar DIMMs can pass an audit, so it would force customers to buy higher capacity DIMMs. A more prudent and statistically valid choice for a screening threshold would be 0.8 watts per GB.

1.4. Section 4, Standard Information Reporting Requirements

HP supports the ability to have a broader selection of processors and other system attributes grouped into a single "Product Family", so that fewer product families are needed to cover each server model. However, we assert that four test configurations are adequate to describe and certify a product family, instead of five. The "typical" configuration for the fifth test is arbitrarily chosen by each vendor doing the test, whereas the four corner tests will be similar from one vendor to the next.

1.5. Section 5, Standard Performance Data Measurement and Output Requirements

No issues at this time.

June 6, 2012

1.6. Section 6, Testing

1.6.1. HP requests that 3-phase AC-DC power supplies be comprehended in the test method. 3-phase power supplies can have superior energy efficiency and phase balance provides additional benefits for distribution power reliability.

1.6.2. HP requests that DC-DC power supplies be included that have input voltages that span the entire range of what is defined as "Low Voltage" by the National Electrical Code (e.g. <600V). Power distribution in the 360VDC-400VDC range is becoming more common, can provide some additional data center-level energy efficiency, and needs to be comprehended.

1.7. Section 7, Effective Date

The effective date for the version 2.0 specification should be at least 9 months after the release date of the specification, and should allow products shipped during that interim period to test and claim certification with either ENERGY STAR version 1.1 criteria or version 2.0 criteria.

1.8. Section 8 Considerations for Future Revisions

Using SERT data collected during version 2.0 submissions as the basis for choosing idle and active mode pass-fail criteria for version 3.0 would not accurately portray the energy efficiency profile of the entire server market. ENERGY STAR has a stated goal of choosing the top quartile of market performers. Since version 2.0 submissions only provide data from servers in the current top quartile, then any expectations set by only using that data would skew the reality of the market.

2. Draft Test Method

This feedback applies to the Draft Test Method document, dated May 2012

2.1. Lines 46-48. The Dc Server input voltage requirement of +/- 53 V dc is unnecessarily limiting. DC input voltages in the range of 360V-400V are not uncommon and should be added to the list of approved test methods.

2.2. Lines 46-48. This test method should also comprehend 3-phase power supplies.

2.3. Lines 158-162. The cost to fully populate a chassis of 16 fully-configured blades is going to be very high (several million dollars). The preferred approach would be to only require populating two identical blade servers in a blade enclosure. There is no direct comparison between blade and rack-mount servers, so a well-documented test with a few blades in the enclosure should provide adequate data transparency.

2.4. Lines 158-162. If a full rack set of tests is required, then we suggest that all but one of the blade servers be kept at a minimal configuration, and only one blade server be required to be changed to perform the "four corners" (plus typical) configuration testing for product families; which would vastly reduce both the time and parts costs for certifying each blade product family. The focus can then be on the one blade that changes.

2.5. Lines 158-172. Having to test both full blade enclosures and half-full blade enclosures will double the testing cost and provides no additional beneficial data.

2.6. Lines 161-172. This section specifies requirements for loading a blade server enclosure to one half its maximum population, yet there are no specifics in section 7 for actually testing this configuration. What is the reason and necessity for also testing a half loaded enclosure? As a more cost-effective alternative, testing a half-full chassis is much preferable to testing a full chassis. Since we won't be comparing blade servers directly to non-blade servers, there is no reason to test a full chassis.

2.7. Line 197. The 5-15 minute window of time to let a server boot to an idle state is overly prescriptive. It is unknown if 15 minutes is enough time for a large enclosure of blade servers, that might need to sequence the server start-up times in order to keep circuit breaker current below their rated levels. The sentence reads like it cannot be longer than 15 minutes, when it should be the undefined length of time needed to let the server(s) boot and all become ready to run applications.

3. Resilient Server Definition from HP

This section is provided to advise the ENERGY STAR program about Resilient Servers, while leveraging Hewlett-Packard's extensive experience in developing that class of computer systems.

1. **Use of high power/highly extensive processor sockets**: Some processors have a higher socket level power draw because of the following capabilities. To meet the requirement of resiliency, the processor would need to support 2 of the following 3:
   a. The processor technologies used in Resilient Servers are designed to provide additional capability and functionality without additional chipsets, enabling them to be designed into systems with more than 4 processor sockets. The processors have additional infrastructure to support extra, built-in processor busses to support the demand of larger systems.
   b. They provide high bandwidth I/O interfaces for connecting to external I/O expansion devices or remote I/O without reducing the number of 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.

   c. RAS Capabilities: The processor supports ECC on the second and higher level caches, parity on chip internal I/O paths, CRC on memory channel, memory RAS (ECC, chipkill, mirroring).

2. **Use of memory buffers**: 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.
3. **Advanced RAID support**: Resilient Servers support at least Level 5 RAID hardware in the base configuration, either through an internal controller or external arrays with RAID 5 capability.
4. **Greater I/O expandability**: Resilient Servers have larger base I/O infrastructure and support a higher number of I/O slots; a minimum of 5 for 2 socket Resilient Servers and greater than 5 for 4 socket Resilient Servers.
5. **RAS Capabilities**: Must have 6 of 11
   - Support of redundant storage controllers
   - Redundant and concurrently maintainable fans
   - Redundant Service Processors
   - Redundant DC-DC regulator stages after the power supply outputs
   - Architectural support of runtime processor de-allocation
   - I/O adapters or hard drives are hot-swappable
   - Contains 6 or more memory busses and 2 or more I/O busses per processor socket
   - Provides machine check architecture (i.e., both Fault Isolation and Resiliency)
   - Provides memory fault detection and system recovery though extended ECC and either DRAM chip sparing or mirrored memory
   - Provides end-to-end bus retry
   - Architecture supports on-line expansion/retraction of hardware resources without the need for operating system reboot ("on-demand" features)

Because of the higher power demand of the overall system based on the attributes above, a higher capacity fan and higher capacity power supply are required. This again drives higher overall system power use and generates a power profile which is significantly different from the power profile of lower power non-resilient servers. Figure 1 below provides a list of features and capabilities which drive a higher base power profile for a server and distinguish the "Resilient Server" category from the managed server category. Figure 2 below provides a list of components which are generally present in Resilient Server systems and estimates of their associated power adders.

Embedded Feature or Capability
Use of Memory buffers
Dedicated slots for additional RAID controller cards
Built-in chip I/O support for dedicated, high speed connectivity to external I/O and disk expansion enclosures.
Greater I/O expandability via larger number of I/O slots
Dedicated high performance I/O controller to manage the higher number of I/O slots. Can be a separate processor or embedded within the processor socket.
Higher power processors to support I/O scalability, integrated I/O controllers, expandability to larger socket-count systems, or multiple chips in each socket for higher core count
Require redundant power supplies
Higher capacity fan and higher capacity power supply due to higher power demand based on attributes above
Additional system management port (beyond 1)
RAS - ECC on second and higher level caches, parity on chip internal I/O paths, CRC on memory channel, processor checkpoint retry and recovery, memory RAS (ECC, chipkill, mirroring), runtime processor de-allocation

Figure 1: Additional Power Demands: Resilient Systems

Component Adders	Watts	Unit
Memory buffers	9.00	Per DIMM port
Memory	1.0	Per GB
Additional RAID controller card	15.00	Per additional adapter
Fan power adder	28.00	Watts
I/O adapters	Variable	Measured per adapter
Additional power supply	25.00	Watts

Figure 2: Additional Adders for Resilient Servers with Measured Power Values.

The current idle power limits for 1 and 2 processor socket servers are also biased towards limited feature, low cost, high volume servers. Such servers commonly have a reduced feature set, for example – no memory buffers, limited memory and I/O expansion capability given their smaller infrastructure and lower performance limits, and limited reliability, availability, and serviceability (RAS) features.

Graphing of the idle and maximum power of the current ENERGY STAR qualified two socket servers depict this bias. It should also be noted that it is likely that there is limited or no data for Resilient Servers in the current ENERGY STAR 1 and 2 processor socket datasets given the difficulty in qualifying these servers to the current idle criteria:

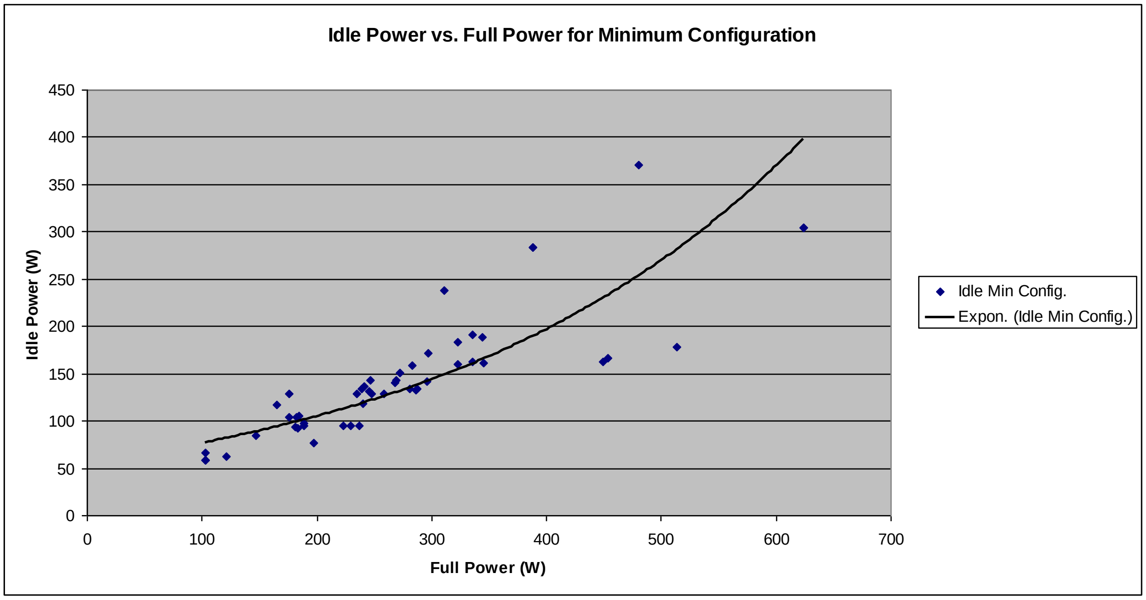

Figure 3: (2) Processor Socket Minimum Configurations: Idle to Full Power

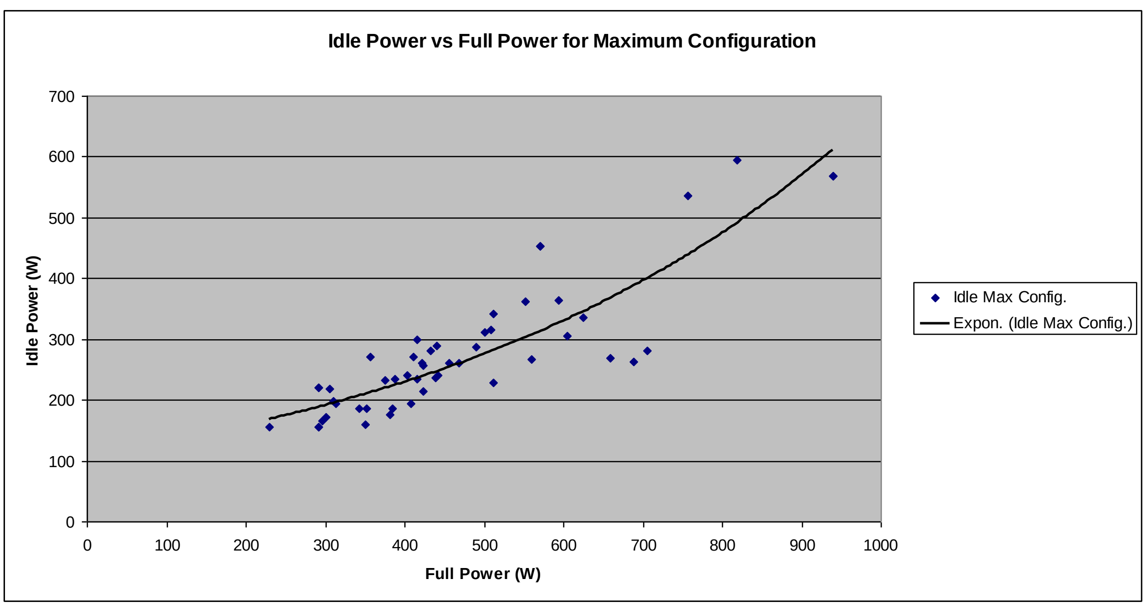

Figure 4: (2) Processor Socket Maximum Configurations: Idle to Full Power

June 6, 2012

The majority of the 2 processor socket products qualified have minimum configuration full power values of 400 W or less. Higher power systems are difficult to qualify to the idle criteria. However, analysis of the ratio of idle power to the maximum power, the data shows that higher power products are competitive with lower power products – they can reduce their energy use at idle by equivalent or better percentage as compared to the low power systems. This is depicted in the graphic below:

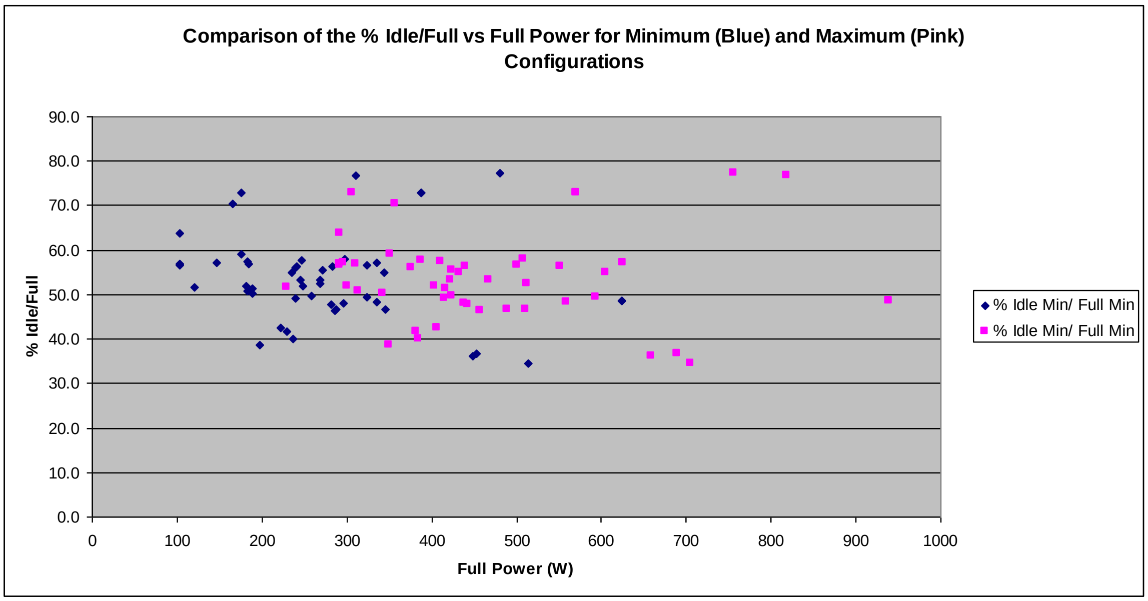

Figure 5: Comparison of Idle/Max Power Ratio for 2 Socket Servers

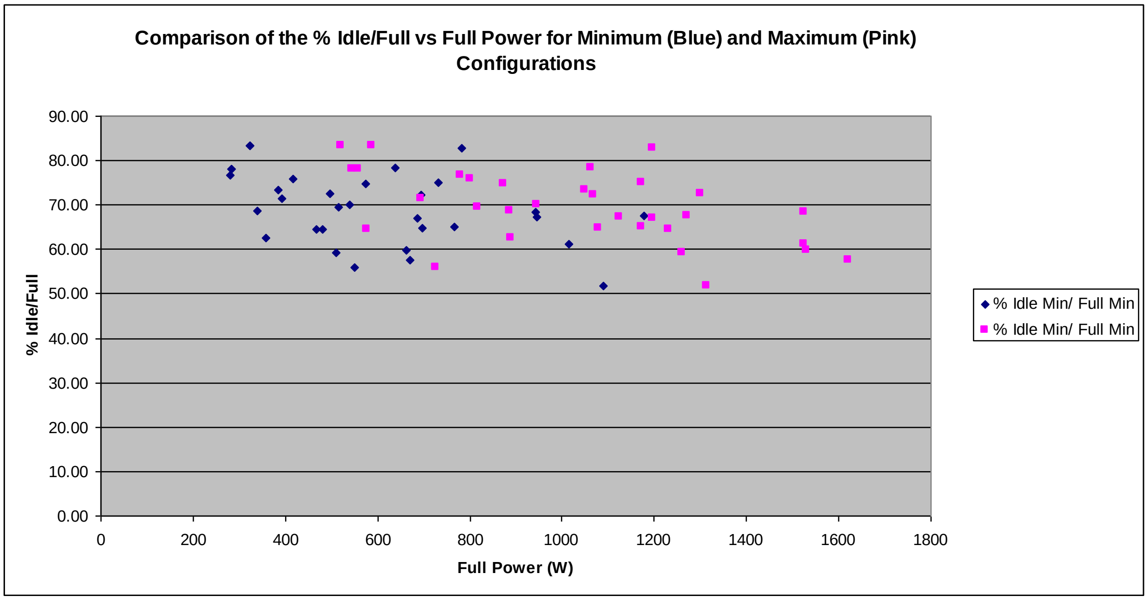

Figure 6: Comparison of Idle/Max Power Ratio for 4 Socket Servers

Similarly, a comparison of the Idle/Max Power Ratio for 4 socket servers (figure 6) shows a similar distribution of idle/max ratios, with some higher power servers exhibiting significant percentage reductions between idle and max power. The data for these systems suggests that if a power criterion is to be set for resilient and managed 4 processor socket systems and for two socket resilient systems that an idle/max ratio would be a preferred criterion.

Resilient Server Blades

Most of the same attributes contained on the mother board for rack mount resilient systems apply to resilient blades as well. The same scoring system should be used minus the allowance for fans and power supplies to delineate between resilient and volume blade servers.

Importance of the Resilient Server Category

Based on this analysis, we believe that in order for EPA to properly characterize the server market, it important to add a category for two and four processor socket Resilient Servers. There are material differences in the power profile between Managed Servers and Resilient Servers and the source of the difference, greater infrastructure and component power demands, can be clearly identified. Resilient Servers need the additional infrastructure to support the resiliency features identified. There is a place in the market for both types of systems, but the ENERGY STAR requirements need to recognize that both types of systems can deliver energy efficient computing.