Climate Savers Computing Initiative Feedback on Draft 4 Server Energy Star® Specification

Background

This document outlines the Climate Savers Computing Initiative (CSCI) feedback on the proposed requirements defined in the Draft 4 Server Energy Star® specification. The feedback is from the following companies in the CSCI AC/DC workgroup; Acer, Dell, Google, HP, Hitachi, Intel, LiteOn, Microsoft, Sun, and Supermicro. All feedback in this document was voted on and approved by the AC/DC workgroup.

This feedback is divided into 13 areas;

1. **Power factors at light loads**
2. **Additional Idle Power Allowances**
3. **Power Accuracy Reporting**
4. **Base System Idle Power Requirements for Category C Servers**
5. **One Processor-Two Socket Systems**
6. **AC Sensor Sampling Rates**
7. **Efficiency Requirements for Computer Server Power Supplies <500W**
8. **System idle definition**
9. **Maximum system definition**
10. **Dual node change to include n-nodes**
11. **Bandwidth requirement for testing procedure**
12. **Exclusion of pedestal servers from standardized data measurement**
13. **1000W threshold as delineation between low- and high-power PSU. - DONE**

1. **Power Factors at Light Loads:** For low power single and multi-output PSUs, there is a lower threshold for power savings at light loads due to fixed losses to operate basic PSU functions. Fixed losses have significant impact on light-load performance targets when compared to output loading as a percentage. For this reason, a lower threshold of 75W should be applied such that the power factor and efficiency requirements for PSUs operating at less than 75W are eliminated (Line 575).

2. **Additional Idle Power Allowances**: CSCI is pleased that Energy Star has raised the idle power allowances (line 633). However, CSCI feels the idle power allowances are still too strict and do not reflect power consumptions of current configuration components. CSCI proposes that additional components be included in Table 4: Additional Power Allowances for Extra Components (line 633). Those components are listed in the table below with the recommended additional power allowances for each component at idle.

Components	Additional Idle Power Allowances
External RAID Controller	18W
SAS	15W
Host Bus Adapter	8 W

DDR2	Make additional allowances for DDR2 memory greater than 2W/GB.

3. **Power Accuracy Reporting:** The current Draft 4 specification of ±10% accuracy for input power measurements is difficult to achieve over the entire PSU load range (line 889). For loads greater than 100W, power reporting can be maintained at ±10% accuracy as designated in the draft 4 specification. However CSCI recommends that the accuracy be changed to reflect the bounded limits for loads less than 100W for non-redundant PSUs as indicated in the table below.

Server input power range per power supply	Maximum allowed error
0W≤Pin<100W	±20 Watts per PSU
100W ≤Pin	±10%

   For Tier 2, Energy Star should consider tighter limits for light loads.

4. **Base System Idle Power Requirements for Category C Servers**: The proposed 100W idle power limit for standard dual servers is significantly more difficult to achieve compared to the 150W for managed systems. 25% of the listed systems in the Energy Star and SpecPower data sets do not meet the specified idle power limit. CSCI invites Energy Star to re-evaluate the data for standard dual processor systems and recalculate the idle power limit. When conducting a similar analysis of the Energy Star data set for Category C systems, CSCI found that only 17% of the systems passed with manufacturing margin.

   Table 1: Category C Summary at the end of this comments section summarizes the Energy Star and SpecPower data sets for this category. For analysis, an adjusted idle power was calculated by subtracting the adders from the measured power based on reported system configuration. One of the passing systems (#107) has an adjusted idle power of 47W assuming 3HDD adders. A dual process system with an idle power limit of 47W is highly unlikely and significantly skews the idle power passing limit. Of the 50 listed on SpecPower, CSCI found that only 4 systems have less than 100W. Given the concerns with the data set, CSCI recommends raising the idle power limit from 100W to 120W to reflect 25% of the systems passing the idle requirement.

   We have similar concerns for the category A and B systems as stated above for category C requirements. After reviewing the data and making allowances for manufacturing variability and idle power adders, less than 25% of these systems will meet the specification idle allowances in 2009.

5. **One Processor-Two Socket System:** CSCI recommends creating a separate category for 1P two socket systems. Currently the EPA forces the 1P/2S system into Category A (standard single installed processor) or category B (managed single installed processor) (line 632). The inclusion of 2 socket systems into the category A and B requirements makes it very difficult, if not impossible, for a 2

socket system with 1 installed processor to pass the 55W and 65W limits for standard and managed systems, respectively. Server customers purchase dual socket servers with one process installed for economies of scale in using the same system for one and two process compute needs. To encourage this configuration, Energy Star should designate a separate category for this configuration.

The EPA data set has 8 systems that are configured as 1 processor installed into a 2 socket system. The systems are listed in Table 2: EPA Data Set 1P/2 Socket System at the end of this document. All are 'managed' systems. System #22 is labeled as a standard processor in the data set. However, it has two installed PSUs which suggest it is a managed system. Of the 8 systems that are 1P/2S systems, only one system passes the idle power limit requirement. The passing system is #22. However we suspect the accuracy of the 53W idle measurement given the system has two installed PSUs each at 700W.

CSCI recommends that the Energy Star designate two new categories for the 1S/2P systems as designated below, Categories E and F. These new idle limits will meet Energy Star's 25% criteria for the supplied data set.

Computer Server System Type	Idle Power Limit (proposal #1)
Category E: Standard 1P/2S	95W
Category F: Managed 1P/2S	125W

6. **AC Sensor Sampling Rates**: The Draft 4 Server Specification for AC sensor sampling rates is specified such that the sampling requirements have a minimum of one sample per second (lines 895-896). CSCI does not understand the justification for this requirement given that this sampling requirement is not possible on systems under heavy utilization. In addition, this sampling requirement will increase the cost of systems installed with a more stable sensor. CSCI recommends eliminating this requirement for the following reasons:
   - The internal management control performing the polling of the sensors inside the system is performing many tasks. AC sensor sampling rates faster than 1 sample per second are not guaranteed in many systems.
   - There is no benefit to the end user for having a 1 sample per second polling rate and this is transparent to the users of the system. The key requirement should be a consistent averaging duration capability across systems.

   Instead, CSCI recommends that the key requirement is support for a rolling average of one minute for Tier 1 and 30 seconds for Tier 2. (lines 896-897).

7. **Efficiency Requirements for Computer Server Power Supplies <500W:** Both Climate Savers Computing Initiative and 80plus have aligned power supply efficiency requirements for single O/P PSU units for the bronze, silver, and gold levels. Energy Star's Server Specification V1.0 deviates from this alignment for single O/P PSUs less than 500W, choosing a specification that is separate from

this alignment. CSCI is concerned that this divergence will create industry confusion. CSCI recommends aligning with the PSU efficiency requirements as defined by the CSCI and 80plus (Line 574).

8. **System Idle definition**: CSCI recommends changing the idle definition (lines 424-427) to the following: "An operational state in which the operating system and other software have completed loading and the Computer Server is capable of completing workload transactions, but no active workload transactions are requested or pending by the system (i.e., the Computer Server is operational, but not processing any useful work)." This new idle definition would include idle states for all CPU, storage and networking devices.

9. **Maximum system definition:** CSCI recommends changing the maximum configuration definition (Lines 466-469) to the following: " Maximum Configuration: The highest performance system within a Product Family. The Maximum Configuration represents the highest configuration of power supplies, memory, hard drives, and I/O Devices, etc. available in the Product Family. The Maximum Configuration must represent the maximum power consumption possible within the Product Family and within the same enclosure."

10. **Dual node change to include n-nodes**: Change Dual-node to N-node in the Dual-Node server definition (lines 332-338) and in the active power requirements for dual-node servers in lines 770-777. In addition, the N-node server definition should include a designation that N-node servers do not have hot-swap server board capability. Blade servers should include a designation as having hot swap capability (lines 275-278). Allowing the hot swap capability designation will prevent confusion between n-node servers and blade servers.

11. **Bandwidth requirements for testing procedure:** CSCI recommends adding an input current measurement bandwidth range requirement to the testing procedures for all input power and power factor measurements. The recommended minimum and maximum required bandwidths are 3 kHz and 20 kHz respectively.

12. **Exclusion of pedestal servers from standardized data measurement**: Power monitoring requirement capabilities are used primarily by large data centers. 'Standard' and pedestal system servers should have an exemption from these power monitoring requirements since few if any have the AC input power monitoring capabilities. End users of pedestal systems do not use management console features. Including these monitoring requirements for this segment of systems only adds cost without the power savings advantage. In addition, pedestal servers typically use chipsets designed for desktop and workstations. Chipsets in these smaller server environments do not have features designed for data monitoring and measurements. Upgrading the chipsets would only add cost and power to the system. For these reasons, CSCI recommends excluding pedestal servers from the power monitoring requirement.

13. **1000W threshold as delineation between low- and high-power PSU**: CSCI supports the PSU efficiency requirements and power factor for high power PSU, >1000W (lines 574-575). However, CSCI does not believe that the 1000W barrier represents a natural delineation between low- and high-power PSUs. The 1200W rated output power represents a more natural delineation to separate efficiency requirements between low-and high-power PSU. CSCI recommends the following:
    1. Move the power requirements threshold from 1000W to 1200W. 1200W is a more natural delineation point in part because the industry standard C13 and C14 connectors are applicable up to 1200W. In addition, there are few power supplies between 1200 and 1500W, whereas there are many power supplies just above the 1000W threshold that would require the higher efficiency requirements without a discernable technical reason.
    2. Set the requirements at the draft 3≤1000W levels across any power supply rating for the tier 1 specification, then project the tier 2 specifications will target gold efficiency levels. Refer to the section below on data set analysis to see why we think this may be an option and still come close to meeting the EPA requirements of top 25% of systems passing Energy Star requirements.

Table 1: Category C Summary

	Cat C (dual proc, standard) (< 100W)							
	System #	Adjusted Idle Power	Non-adjusted Idle	CPUs	Memory	PSUs	HDD	Pass/Fail
EPA Data	107	47	71	2.5GHz, Quad Core 2x6MB L2 Cache, 1333MHz system bus	2 x 2GB	1	4	pass
	108	74	90	2.5GHz, Quad Core 2x6MB L2 Cache, 1333MHz system bus	2 x 2GB	1	3	pass
	109	80	96	2.5GHz, Quad Core 2x6MB L2 Cache, 1333MHz system bus	2 x 2GB	1	3	pass
	110	107	131	?	4 x 4GB	1	1	fail
	111	111	135	?	4 x 4GB	1	1	fail
	113	127	151	Quad-Core, 2.2GHz, 2MB L3 Cache	4 x 4GB	1	1	fail
	114	135	167	2x6MB L2 1333MHz	4 x 4GB	1	2	fail
	112	142	150	Quad-Core, 2.0GHz, 2MB L3 Cache	4 x 2GB	1	1	fail

	115	161	185	Quad-Core, 2.50GHz, 12MB L2 Cache	8x 2GB	1	1	fail
SpecPower Data	26	77	85	2.66GHz, 2x6MB L2 shared, 1333MHz system bus	2 x 4GB	1	1	Pass
	27	89	97	Quad-Core, 2.50GHz, 1333MHz System Bus	2 x 4GB	1	1	Pass
	28	93	101	2.5 GHz, 2x6 MB L2 shared, 1333 MHz system bus	2 x 4GB	1	1	fail (marginal)
	29	94	102	Quad-Core, 2.5 GHz, 2 x 6 MB L2 Cache, 1333 MHz system bus	2 x 4GB	1	1	fail (marginal)
	30	98	106	2.33 GHz, 2x6 MB L2 shared, 1333 MHz bus	2 x 4GB	1	1	fail (marginal)
	31	106	130	Quad-core 2.8GHz, 1600 MHZ FSB	4 x 4GB	1	1	Fail
	37	120						Fail
	33	124						Fail
	32	131						Fail
	34	131						Fail
	35	131						Fail
	36	136						Fail

	38	140						Fail
	45	179						Fail
	39	183						Fail
	41	188						Fail
	40	190						Fail
	42	195						Fail
	44	195						Fail
	43	212						Fail
	46	217						Fail

Total Systems = 30

Passing Systems = 5

% passing systems = 17%

Table 2: EPA Data Set 1P/2 Socket System

1 processor / 2 socket						
System #	Adjusted Idle Power	Non-adjusted Idle	CPUs	Memory	PSUs	Pass/Fail
10	64.2	72.2	?	4 x 2GB	1	fail (zero margin)
9	69.1	69.1	?	2 x 2GB	1	fail
13	73.5	81.5	HPT LV L5410 2.33G /50W / 1333	4 x 2GB	1	fail
18	135	163	3GHz, 2×1MB L2 Cache, 1000 MHz	4 x 1GB	2	fail
22	156	224	HPT LV L5410 2.33G /50W / 1333	4 x 2GB	2	fail
19	172	200	3GHz, 2×1MB L2 Cache, 1000 MHz	4 x 1GB	2	fail
20	172	200	3GHz, 2×1MB L2 Cache, 1000 MHz	4 x 1GB	2	Fail
22	25	53.1	2.00GHz, 12MB L2 Cache, 1333MHz	2 x 2GB	2	Pass