May 8, 2009

TO: Rebecca Duff
ICF International
1725 Eye Street, NW, Suite 1000
Washington, D.C. 20006
rduff@icfi.com

CC: Andrew Fanara
EPA
fanara.andrew@epa.gov

FROM: Chris Hankin
Sun Microsystems, Inc.
chris.hankin@sun.com
202-326-7522

Re: Comments by Sun Microsystems, Inc. on the Final Draft for the Energy Star Specification for Computer Servers

Dear Ms. Duff:

Thank you for the opportunity to provide comments on the EPA's Final Draft for the Energy Star specification for Computer Servers. Sun Microsystems appreciates the opportunities extended throughout this past year for inclusion in this process, and we look forward to continuing to help achieve a successful new specification.

We note that many of the suggestions provided by Sun and the industry in response to Draft 4 of this specification in March 2009 have been reviewed and included by the EPA in the final draft, but also note that problems remain. In particular, we are concerned that the deliberate decision by the EPA to ignore the energy efficiency benefits of 8-core microprocessors remains contradictory to the underlying intent of the specification, and delivers to the industry an imperfect and potentially misleading tool that does not achieve the desired benefits at the data center level.

The comments that follow are consistent with our most recent discussions, and are made with the purpose of achieving a specification that better achieves our mutual goals.

We look forward to discussing these points in more detail and to concluding the Tier 1 Energy Star for Servers specification.

Sincerely,

Chris Hankin
Sun Microsystems, Inc.
chris.hankin@sun.com
202-326-7522

Table of Contents

1.Introduction....4

2.Idle Power For Systems with 8 Core Microprocessors....5

3.Editorial Suggestions....5

4.Processor Utilization Accuracy....6

5.Power Supply Requirements 7

A) Power Accuracy (line 630)....7
B) Bandwidth Measurement for Input Current Testing....7
6.Test Equipment Accuracy (line 922)....8

7. Effective Date (line 841):....8

8.Comments on Energy Star Power and Performance Data Sheet and the QPI sheet....9

COMMENTS BY SUN MICROSYSTEMS, INC. ON THE EPA ENERGY STAR (FINAL DRAFT) PROGRAM REQUIREMENTS FOR COMPUTER SERVERS

1. Introduction

Sun commends the EPA on issuing the Final Draft of the Energy Star for Servers specification. This draft represents significant progress towards the goal of managing energy consumption in the data center. Sun applauds the open process that the EPA has followed, including the extensive dialog with the industry and the EPA's willingness to be available for detailed discussions. Sun appreciates the opportunity to meet with the EPA in one-on-one meetings and in industry conference calls, as well as EPA's outreach to the industry at various conferences and symposia.

Sun particularly commends the EPA's transparency throughout the process. The webinar conducted for the industry by the EPA on March 16, 2009 was very informative and educational and went a long way towards explaining the EPA's methodology for arriving at the details of the specification. The EPA's release of the idle power data spreadsheet, as well as the analysis methodology, has helped the industry understand the rationale behind the various aspects of the specification.

In particular, Sun appreciates the EPA's careful consideration of the following Sun proposals which the EPA has found sufficiently valuable to include in the final draft specification:

1. Elimination of the one second required interval for the sampling period
2. Clarification of the definition of I/O devices and I/O ports
3. The approach of granting idle power allowances for I/O devices on a technology neutral basis, qualified only by link speed and number of ports

While many of Sun's concerns from Draft 4 of the specification have been diligently reviewed, considered, and addressed by the EPA, Sun continues to have concerns about the following aspects of the Final Draft specification, which are detailed in this document:

1. Idle power for 8 core microprocessors
2. Specific concerns around power supplies
3. Test equipment accuracy
4. Concerns around the effective date
5. Comments on the power and performance data sheet

2. Idle Power For Systems with 8 Core Microprocessors

Customers size systems for particular workloads that need to be serviced in the data center, such as a total database transactional workload, or a peak web traffic workload, or a maximum HPC computational workload. System sizing is done on the basis of computational requirements (number of cores in the microprocessors), memory requirements (amount of DRAM in the system), storage requirements (the disk capacity) and networking/IO requirements (I/O and network bandwidth).

The data set analyzed by the EPA includes systems built primarily with 2-core and 4-core microprocessors. For systems built with 8-core microprocessors, the EPA has received data only from Sun, as Sun is currently the only vendor shipping with these highly innovative systems.

Vendors innovate for multi-core because of multiple reasons: it not only saves cost and makes the systems cheaper, but it also saves energy, as integrating a greater number of processing cores on a single socket burns less overall system-level energy than distributing those cores over multiple sockets. As such, the energy efficiency (performance per watt) of 8-core systems is greater than that of dual or quad-core systems, because fewer are needed to service a given quantum of workload. Yet, because the EPA grants idle power allowances only on the basis of socket count, not core count, 8-core systems, which are more energy efficient, are penalized, since their lower socket count restricts them to a lower idle power allowance.

The EPA has noted in the cover memo of the Final Draft: “EPA continues to believe that the best indicator of the base Idle level for Computer Systems is the number of discrete processors, not the total number of cores.”

Sun is very disappointed that the EPA, in spite of having repeatedly reviewed compelling and substantial evidence presented by Sun about the energy saving capabilities of systems with 8-core microprocessors, continues to labor under the misconception that the best indicator of the processing capacity of any system is the number of discrete processors and not the number of cores.

The unfortunate effect of the EPA's present stance on not recognizing the innovation of 8-core microprocessors, and the energy savings benefits thereof, will have a counterproductive effect on customers seeking to save energy in their data center through the deployment of Energy Star compliant servers. By penalizing rather than rewarding the innovation of reducing data center level power through the use of fewer servers with more processing cores per server, the EPA will create the conditions for the unintended consequence of increasing total power consumption at the data center level through the Energy Star for Servers program.

Sun believes that this approach will have inhibitory consequences to the credibility and acceptance of the Energy Star for Servers program as a useful criterion for customers to save overall power consumption in the data center.

3. Editorial Suggestions

Sun applauds the EPA on generalizing the notion of an I/O device and taking a technology neutral approach for allowances for I/O devices qualified only by link speeds and number of active ports. Sun suggests the following editorial clarifications to better articulate and reinforce this concept:

Request:

Sun requests that the definition of I/O devices (Line 318) be amended by deleting the word "networked" That sentence would now read "Devices which provide data input and output capability to the computer server from other devices." The rationale for this request is, if RAID and SAS controllers are examples of acceptable I/O devices, the other devices that they communicate with could be disk drives or JBOD storage servers, which are typically not considered networked as they do not communicate via Ethernet.

Request:

Sun agrees with the EPA's rationale which is articulated eloquently on lines 494 through 498 as follows: "EPA has modified the I/O Device allowances to be technology neutral and based only on the link speed and the number of active ports on the device. This technology neutral approach recognizes the variety of I/O technologies available in the marketplace, provides Computer Sever manufacturers greater design flexibility, and allows different I/O technologies to compete on energy efficiency."

Sun suggests that this text be moved from an editorial note to the normative section of the formal specification, as it helps clarify and reinforce the EPA's intent.

4. Processor Utilization Accuracy

The EPA's requirement for processor utilization measurements of up to 5% accuracy, using a particular algorithm, is extremely difficult to meet, and unnecessary for the following reasons:

1. Each processor and operating system that ships today already has built-in techniques to calculate processor utilization. These techniques and algorithms differ from processor to processor and from OS to OS, but they all yield reasonably accurate data for the purpose of decision making at the data center level. The imposition of a particular government specified formula for calculating processor utilization proposed in a draft specification dated April 24, 2009 and intended to become effective on May 15, 2009 is highly problematic, as it precludes any changes to the shipping systems. The imposition of a particular algorithm for calculating processor utilization also stifles innovation and improvement in measurement techniques.

2. There is continued innovation in power management at the microprocessor level which includes technologies like dynamic voltage scaling, dynamic frequency scaling, core power reduction, core disabling, cycle skipping, slower clocking, halt states, and several others. The algorithms for measuring processor utilization will continue to undergo ongoing innovation and refinement to account for these new and upcoming technologies. Any option to provide a defined algorithm to improve reporting accuracy will never be able to account for the variability and range in processor power management techniques.

3. The intent of making processor or system utilization available to the data center operators is to encourage them to track the use of their equipment and identify equipment that is not utilized or under-utilized. To enable this, the measurement only needs to be sufficiently accurate for the purpose of enabling decisions around the reprovisioning of workloads or the migration of

virtual machines. Existing CPU utilization measurement algorithms are already sufficiently accurate for this purpose. Many data centers today already rely on CPU utilization numbers as reported by currently shipping operating systems on currently shipping servers. They use this information successfully today to dynamically manage data center power consumption by reprovisioning workloads to minimize the under-utilization of machines. Requiring a specific accuracy criterion for CPU utilization will not provide any particular incremental value to customers, and will only increase the cost of the system due to the expensive additional micro-instrumentation required.

Request:

Change the reporting accuracy for processor utilization to +/- 25%. Do not mandate any particular algorithm or formula for calculating processor utilization.

5. Power Supply Requirements

A) Power Accuracy (line 630)

We note with appreciation that the EPA has changed the input power measurement requirements to +/- 10 Watts (instead of +/- 10%) for input power less than or equal to 100 Watts. In spite of this change, even a 10 Watt accuracy is very difficult to meet at loads less than 100 Watts. As the output load reduces the input current, the waveform degrades from being regular at 100% to irregular at 30%, making it difficult to define an RMS value.

When input power is low (below 30% load), the input current through the PFC sense resistor is small and irregular. The accuracy of measuring the voltage across the PFC resistor (as a representation of the current) is influenced by fixed errors and the point on the waveform where the measurement is taken.

Request:

Defer the requirement for any particular accuracy for loads below 100W until Tier 2.

B) Bandwidth Measurement for Input Current Testing

In order to maintain accurate results from test equipment when measuring input current a bandwidth range should be added to either the specification or the power supply test procedure referenced within the energy star specification.

Request:

Specify minimum and maximum required bandwidths of 3kHz and 20kHz respectively.

6. Test Equipment Accuracy (line 922)

On line 922, in Appendix A, there is a requirement for the power meter to have an accuracy of 0.01 Watt or better for power measurements of 10W or less.

- This requirement greatly increases the cost of the meter, making it more difficult for manufacturers to make the meters available to development staff.
- No server has an idle power of 10W. Since the smallest power that will be required to be measured in the EPA's idle power data set is 55W, there is no reason to require this level of accuracy at values under 10W.
- Given that the spec also indicates that power numbers should be rounded to the first decimal place, an accuracy to 2 decimal places is redundant.

Request:

Remove the requirement for the power meter to have an accuracy of 0.01 Watt or better for power measurements of 10W or less.

7. Effective Date (line 841):

In the Energy Policy Act of 2005 (Section 131 of PL109.58), which amends the Energy Policy and Conservation Act (42 USC 6294a. Sec. 324), the duties of the EPA administrator with respect to the Energy Star program are specified as follows:

> "The Administrator and the Secretary shall provide appropriate lead time (which shall be 270 days, unless the Agency or Department specifies otherwise) prior to the applicable effective date for a new or a significant revision to a product category, specification, or criterion, taking into account the timing requirements of the manufacturing, product marketing, and distribution process for the specific product addressed."

Since the Energy Star for Servers specification is a new specification, and server vendors need the appropriate lead time for manufacturing, marketing, and distributing products compliant with this new specification, we recommend that the EPA not deviate from the 270 day lead time notification suggested in the legislation above.

Request:

Sun recommends an effective date of 270 days following the publication of the Tier 1 specification for the reasons stated in the Energy Policy Act of 2005.

8. Comments on Energy Star Power and Performance Data Sheet and the QPI sheet

Sun's general comment is that the QPI sheet seems redundant with the P&P sheet as it has the exact

same information. There is no need to add cost and expense to the Energy Star qualification process by having two different data sheet format for the same information.

Lines 45, 46, 54 and 55 (P&P) and Section L.3 (QPI)

The EPA has required the publication of data in the Energy Star Power and Performance Data Sheet which many manufacturers do not publish. In particular, benchmark results for benchmarks standardized by SPEC and TPC permit the manufacturer to keep the results private. It is entirely optional for manufacturers to publicly list the results of benchmark tests on the web site of either SPEC or TPC.

While Sun agrees with the EPA on the need to publish power data at full loads, it does not necessarily follow that for a particular benchmark workload, the performance data at full load must also be published. Note also that the full load power data may be measured using any private, proprietary or non-standard benchmark in which case the public declaration of the benchmark score may not convey any useful information.

Request:

Make the reporting of full load performance data (Lines 45, 46, 54, 55, P&P, and Section L Line 3, QPI) optional, while keeping the reporting of full load power data public. Full load performance data, if necessary, can be shared privately by the vendor with the EPA.

Line 7 (P&P) and Section B Line 3 (QPI)

Request:

In addition to requiring the maximum number of processor sockets, the maximum number of processor cores supported by the system should also be a required declaration. Add a line for the vendor to report the maximum number of processing cores that the system can be configured with.

Lines 13, 16, 17 (P&P) and Section C.3 and C.6 (QPI)

Request:

Add the word "Single" in front of the words "Power Supply" to clarify that the data is per Power Supply in the server.

Line 13 (P&P) and Section C.3 (QPI)

Request:

Require the reporting of Voltage and Watts instead of just Watts.

Line 13+ (P&P) and Section C.3+ (QPI)

Request:

Add a new line to the table to require the reporting of the Standby Power (i.e. Power used by the Standby Rail).

Line 27 (P&P)

Request:

Delete the words "BMC or" from line 27, since the spec no longer covers blades and hence the requirement for a Blade Management Controller is moot.

Line 32 (P&P)

Request:

Change the word "limit" to "allowance" in line 32, since the phrase "Energy Star limit" is not defined anywhere in the normative specification.

Line 36, 47 and 56 (P&P)

Request:

Please clarify the assumptions necessary for estimating KWh per year (average utilization load of server, number of hours in the year that the server is expected to be used, etc.). This data only needs to be reported once, and not with each benchmark. Delete the requirement to report this data on lines 47 and 56.

Line 65 (P&P)

Request:

Change the column heading from "Shipped and Tested Enabled" to "Enabled at Ship Time."

Line 81 (P&P) and Section H.4 (QPI)

Request:

Any declaration of data collection protocols is meaningless unless also accompanied by the name of the MIB, schema, or command syntax against which the data collection can be done. We request clarification on the purpose of declaring just the data collection protocols, otherwise we request that these lines be deleted.

Line 85 (P&P)

Request:

Total power dissipated is always exactly equal to the total power consumed, since there is no net accumulation of energy inside the server over time. As such, the total power dissipated will always

vary with time since the power consumed varies with time as a function of workload fluctuations. Hence, it is impossible to report any fixed number in line 85. We suggest that this line be eliminated.

Line 86, 87, 88 (P&P)

Request:

The term "peak temperature" is undefined, although the data sheet suggests that it is 35 degrees Celsius. Note that only ASHRAE Class 3 equipment operates in ranges up to 35 degrees Celsius. Data center servers, which are ASHRAE Class 1 equipment, only need to operate to ASHRAE Class 1 guidelines, which go up to only 27 degrees Celsius. We suggest that the term "peak temperature" be defined as 27 degrees Celsius as has been done in Appendix A in the table on line 934.

Line 88 (P&P)

Request:

Airflow at minimum fan speed at peak temperature is an artificial and not particularly useful metric, because at peak temperature fans do not operate at their minimum speed. Hence the CFM reported in this line will never be reflective of real world conditions in the data center. We suggest that this line be eliminated.