May 24, 2010

Dear EPA,

Thank you for providing us the opportunity to comment early in the ENERGY STAR for Servers Specifications v2 specification development process. Please find enclosed Intel's feedback on the Draft 1, dated 4/9/10

Intel remains supportive of the US EPA's efforts to define energy efficiency goals and targets across the spectrum of computer products including the draft 1 proposal for ENERGY STAR for Computer Servers v2.0. We hope our input and our collaboration with industry stakeholders continue to benefit version 2 specification development including inclusion of SPEC's Server Efficiency Rating Tool (SERT™). We look forward to the webinar the EPA plans to hold on June 27th and future meetings with the ENERGY STAR team. Please contact us if there are areas where we may be able to improve these interactions.

We continue to work with our industry colleagues in Standard Performance Evaluation Council (SPEC), Climate Savers Computing Initiative (CSCI), The Green Grid (TGG), IT Information Council (ITI), Alliance for Telecommunications Industry Solutions (ATIS) and Storage Network Information Association (SNIA), in addition to supporting the ENERGY STAR for servers program to deliver increasing energy efficiency.

If you have any questions please feel free to contact myself or Henry L Wong, henry.l.wong@intel.com.

Sincerely,

Lorie Wigle
General Manager
Eco-Technology Program Office

## Summary

We have provided comments on the major sections highlighted in the ENERGY STAR® for Computer Server v2.0 Draft 1 document, dated 4/9/10. In the appendix we've included detailed notes and editorial suggestions for particular lines in the draft. We recommend holding subsequent reviews or workshops with industry groups such as Climate Savers Computing Initiative, The Green Grid, and SPEC to clarify and develop implementation details prior to EPA's next draft publication. The topics covered include definition and scope, active mode and idle, product family definitions, and real time monitors.

Given the aggressive schedule for version 2 of the server specification, we believe the priorities for evolution should be development and incorporation of SPEC's Server Efficiency Rating Tool™, inclusion of blade servers, and addressing the issues highlighted by the system manufacturers with the version 1 specification. We recommend against further increases in product scope, such as HPC clusters, or new methods of addressing idle limits. The number of ENERGY STAR compliant servers is still very limited. Instead of expanding server categories, our suggestion is to make qualifying for ENERGY STAR simpler. The main topics for version 2 are already challenging. We hope that ENERGY STAR and its key stakeholders realize the need to maintain focus on critical topics to improve the chances of meeting the timeline prescribed in the draft.

We hope these comments and recommendations will be useful to EPA's plans and targets for the ENERGY STAR for Computer Servers v2.0 specification. We welcome the opportunity to discuss these topics further as the team develops the details on the version 2 specification for servers.

# Commentary by Section

## Definitions and Scope

Volume Server

We believe that the existing 1-2S server definitions sufficiently cover the current description of a volume server.  A separate definition is not necessary.

Managed Server

We recommend staying with the current definition of a managed server.  Hardware event logging and remote power control, including reboot and restart independent of OS-based management are indeed characteristics of a server.   These features are not necessarily distinctive of a “managed” server, as other computing systems contain these features.  Intel does not believe these additional features would enhance the description of a managed server.

Fault Tolerant Server

Full hardware redundancy as currently described, is not required to be considered a fault tolerant server.   This class of server includes hardware redundancy and system architecture features that detect and circumvent hardware faults in the system.  Some systems may include spares or subsystem components or features that would serve as a fail-over to the original data or power paths.  For example, an I/O port fault can be detected and serviced by another I/O port or a lower speed port, without having a redundant system as described currently by ENERGY STAR v1.0.  The current description calls for a multi-node server which operates in lock-step, to allow for failover.  This is an example of but not a complete description of this category.  A prescriptive description should indicate duplicate or hardware redundancy in power delivery, cooling, compute, data storage (e.g. memory), and I/O. The category description should highlight tolerance to power delivery, cooling failure and compute availability.

Resilient Server

The draft description of a Resilient Server should be enhanced to identify specific examples and a clarification to the generic feature.  The Harvard Research Group (HRG) description of Availability Environment Classification (AEC) system is a good qualitative description of user expectations on availability.  Though this system provides a generic description of system availability, the groupings do not contain the quantitative detail required to base an ENERGY STAR category on. Grouping systems based on RAS would require more quantitative descriptors before they could be made applicable to categorize systems.  For RAS descriptors, we suggest further research with groups such as TGG, to better quantify the characteristics.  For resilient servers, we believe the feature list is sufficient but recommend adding the examples in our previous comments to provide clarity to the features.  We previously indicated that a resilient server should be a system that contains all or a significant number of these features:

- Memory Fault Detection and System Recovery: DRAM Chip Sparing, Extended ECC, Mirrored Memory
- Machine Check Architectures – Fault Isolation and Resiliency
- End to End Bus Retry

- Hot-swap components: I/O, hard drives and AC/DC power supplies
- Ability to perform on-line expansion and retraction of hardware resources without OS reboot - also referred to as “on-demand”
- (optional) Multiple physical banks of memory and I/O adapters

HPC Server

The number of memory controllers is not a key distinguishing factor for this type of server. HPC systems do contain memory and I/O fabrics that allow the system to expand its compute and memory structure to operate as a large scale system. Though the system may be comprised of similar components to a 1-4 socket server, the system hardware design and software prevent these systems to be evaluated as 1-4 socket server(s) and assumed to be simply an aggregate of individual pieces. As with multi-node servers, the ability and application to run as a single compute system will not be fairly represented by an active mode evaluation tool which is strictly targeted to support stand alone operation. We recommend clarifying these points and removing HPC and multi-node systems from the scope of the specification, until such time as an evaluation method can be best determined.

## Qualifying Products

High Performance Computing (HPC) Systems

We do not believe there have been significant changes in either the volume of shipments, available testing procedures, or energy profile information that support incorporating HPC systems at this time. Even though the base components may look similar, such as a rack 2P configuration vs. HPC system comprised of 2S motherboards, the total system integration emulates a multi-processor system, capable of operating on a logically unified memory image. The resulting energy profile is not a linear extrapolation of multiple 2S servers. In fact inter-platform coordination and logic may show higher idle for a single 2S motherboard, even though the multi-motherboard HPC system is more efficient than purchasing multiple 2S servers and developing an external means to aggregate the compute capabilities. It is also inaccurate and non-representative to test partial subsystems independent of the aggregate HPC system. The unique architecture and software stack may prevent the use of generic server benchmarks and tools. Therefore, we recommend that HPC systems be out of scope for this specification.

## Power Supply

Net Power Loss (NPL)
We agree and appreciate ENERGY STAR’s considerations of industry input in deciding to stay with industry standard efficiency methods and not convert to a NPL process.

Power Supply Efficiency
We understand the desire to aggressively pursue higher levels of efficiency. The levels and direction are consistent with the advances and capabilities in the industry. We agree with and recommend CSCI and 80Plus targets of gold levels of efficiency for single output and silver level for multi output internal power supplies.

## Active mode and Idle Specifications

Intel continues to fully support the development of the SPEC Server Efficiency Rating Tool (SERT™). The conversion of Version 2 to only disclose data from the tool and assess the market is a prudent choice in preparation of efficiency criteria for Version 3. We recommend that the Power Performance Data Sheet (PPDS) be updated to accommodate data collection for the performance parameters in SERT™, in addition to other performance parameters that may be needed to support or augment the use of the tool.

The use of maximum power or other performance parameters to adjust idle criteria is an intriguing option that may work with or be independent of SERT™ development. The methodology, however, would require data collection and sensitivity analysis across various platforms to determine its applicability. The actual performance benchmarks to determine maximum power would need to be evaluated to determine scalability and applicability across the different product categories. Given the timeline and data collection required for this investigation, we recommend that the version 2 timeframe be used to collect the needed performance and power data to assess the feasibility of this approach.

## Family Definition

We welcome and recognize the increased flexibility of configurations that would comprise a product family as proposed in draft1. The remaining restrictions on the processor, I/O, and power supply selection do not comprehend the platform sku's and the known power characteristics in server product lines. The power characteristics of these components and their contributions to system power are comprehended by system manufacturers to ensure compliance to system specifications and applicability to the manufacturer's power calculator. IT purchasers customize their selection within the product family highlighted by the board configuration (e.g. 2S or 4S), processor family, and form factor. Given the deterministic power characteristics, server manufacturers are able to define the minimum and maximum configuration that would comply with the ENERGY STAR criteria and be consistent with the family definition IT purchasers expect. This approach represents the full range of power and performance for the machine type (or with a defined subset of model numbers). Power information for a specific configuration can be determined using the power calculator that most manufacturers have available for their equipment.

Given the industry's capabilities, we recommend that the EPA's family table be changed as noted below:

| Base Component | Same Part Number Required for All Product Family Configurations | Same Technical and Power Specs Required in All Product Family Configurations | Quantity Required in All Product Family Configurations | |
|---|---|---|---|---|
| **Motherboard** | YES | YES | Same across family | |
| **Processor** | ~~YES~~<br>**NO** | ~~YES~~<br>**NO** | ~~Same across family~~<br>**May vary across product family** | Processor must all be from same **product generation, family or model line.** ~~Processors must all have the same core count and power specifications.~~<br>Processor **frequency** may vary within a product family |
| **Power Supply** | ~~YES~~<br>**NO** | ~~YES~~<br>**NO** | May vary across product family | |
| **I/O Device** | NO | ~~YES~~<br>**NO** | May vary across product family | |
| **HDD or SSD** | NO | NO | May vary across product family | HDD, SSD, and memory capacity may vary. If so, minimum, typical, and maximum must represent the full range of capacity options. |
| **Memory (DIMM)** | NO | NO | May vary across product family | |

The rationale behind these changes reflects the product sku's and deterministic behavior of the systems:

- Processor variations within the same processor model line generally vary with frequency and core count. Within a processor model the power characteristics and its impact to the system power is deterministic. The system power characteristics can be bounded by the maximum and minimum hardware definitions in the family by ensuring the maximum and minimum configurations are evaluated by the processor sku's that represent those conditions. All the variants of processors within the family are ensured to be within the range of these "book-end" configurations.

- Frequently, a server model is made available in depopulated configurations. These depopulated configurations do cause a difference in the overall power draw of a server. However, in many cases, both the depopulated variant of a server, and the fully populated variant of the same server, qualify for ENERGY STAR. For example, a two socket system may be sold in its fully populated configuration with two processors and 96GB of DRAM, and in a depopulated configuration with one installed processor and 48GB of DRAM. It is possible that both configurations qualify for ENERGY STAR. In this case, it should be permissible for a system manufacturer to qualify the depopulated version as the "minimum configuration" of the family and the fully populated version as the “maximum configuration” of the family. By bookending minimum and maximum configurations with server variations that include a variation in processor count and still qualify for ENERGY STAR, customers would be able to purchase desired configurations will also qualify. Hence, we recommend that the family definition be broadened to include variations in the processor count in situations where depopulated variants and fully populated variants of the same server model would otherwise independently qualify for ENERGY STAR on their own.

- I/O devices in the same server line can vary widely. Different I/O devices have different technical and power specifications. The variation in power characteristics are deterministic and should be supported by the system manufacturers’ minimum and maximum family configurations. The system developer should simply identify the range of I/O devices that would be represented by the minimum and maximum hardware configuration noted. As with the case with the current idle power specification, if the I/O device’s additive contribution to the system power is less than the allocation, these devices can be considered within the range compliant to ENERGY STAR.

- During the shipping life of a server model, the PSU model that is included in that server line is occasionally upgraded. PSU upgrades for a shipping server model happen because the PSU supplier may have made the original model of PSU obsolete. As long as the original and the upgraded PSUs meet the ENERGY STAR eligibility criteria for computer server power supplies, we recommend allowing these PSU variations within a single family definition. We recommend removing the requirement that (a) the same part number of PSU be required in all server configurations within a product family (because the server model may be shipped with both the original qualifying PSU type and the new qualifying PSU type), and (b) the same technical and power specifications for PSUs be required within a product family.

Intel recommends that the definitions for the maximum and minimum configurations remain **hardware based**. We **do not** recommend using active energy efficiency parameters to define the book-end configurations. This would create an indeterminate hardware definition and very difficult to translate power contributions within the configurations. In order to ascertain which configuration represents the maximum and minimum achievable active energy efficiency a detailed active energy versus configuration versus power use matrix would need to be tested. This complex model would be needed to determine the minimum and maximum active energy state and its associated configuration for that product family. The extent of characterization required to make this determination would negate any benefit of the product family categorization of products. We recommend staying with a hardware configuration based definition of maximum and minimum.

## Energy Efficient Ethernet

We appreciate the assessment and concur that the feature and underlying technology needs further analysis before being adopted as a requirement for the ENERGY STAR program.

## Real Time System Reporting

The rolling average concept and intervals do not mimic the use or requirements to optimize data center operations. For managed servers, we recommend maintaining a simplified approach of an ability to report ambient inlet temperature to the server and AC input power at a maximum of 1 minute intervals at +/-2°C and +/- 5% accuracy respectively.

There is also product demand for (unmanaged) systems that do need to conduct this reporting as there is no management function or activity that would be used for these systems. These are typically pedestal servers whose environmental status would not be used to change loading or operations of the environment. We should allow such systems to also be ENERGY STAR compliant and not require the reporting on unmanaged systems. The over head and energy consumed to enable these features which aren't going to be used, wastes compute resources, increases energy consumption, and increases system costs for little benefit.

Intel recommends dropping the AC power monitoring and reporting requirements on pedestal system servers with non-redundant power supply capabilities. AC power monitoring increases the system energy consumption and cost without providing benefit in these particular systems. Requiring power monitoring on these systems would also conflict with European Union's ErP Lot 6 requirements on EMI Class B systems. Briefly Lot 6 covers standby and off-mode losses of Energy Using Products (ErP). Beginning in 2010, systems in the off mode must draw less than 1 W of power in order to be sold in the European Union and less than ½ W of power by 2013. The reporting requirement would render these ENERGY STAR systems non-compliant to the European standards. Such systems could not be sold in Europe due to this conflict in specifications.

In Line 945 to 948, ENERGY STAR specifies the required measurement accuracy. As stated, the meter requirements equate to a power meters having accuracy of 0.1%. Line 947 requires 0.1W accuracy at 100W. Line 948 requires 1W accuracy if measuring 1000W. The most accurate power meters can only achieve ±0.2% accuracy when considering current, voltage, and power factor contribution. The low cost meters employed by most manufacturers today for ENERGY STAR- and SPECPower-compliance testing generally have an accuracy of ±0.5% when considering current, voltage, and power factor contributions. Intel recommends stating a meter accuracy of ±0.5% or 1W, whichever is larger, at the tested voltage, current, and power factor condition.

## Blades

For version 2, we recommend that blade servers be evaluated in similar manner as rack mount servers to reduce confusion (idle power limits for 1 and 2 socket servers; and processor power management enablement for three and four socket servers)

### Blade Chassis

With regards to EPA's proposal for qualifying a blade chassis, we recommend that EPA eliminate the Table 4 and 5 criteria for a blade chassis and replace it with the following set of requirements:

1. PSUs in the blade chassis should meet the computer server efficiency and power factor requirements.
2. The chassis should have variable speed fans.
3. The chassis should be capable of reporting power use and thermal information for the blade system.

This establishes the functional requirements for the chassis, without requiring extensive measurements of power use on the chassis. In addition, each manufacturer configures their blade chassis differently, with different percentages of the "overhead" power for fans, network, and hard drive components.

### Blade System Testing

The testing of individual blades should be augmented by appropriating the percentage of shared resources back to the figure of merit used for the blade. Determining the percentage of the shared resources can be done in a fully loaded chassis or partially loaded one. Since each portion of the loaded chassis is controlled by a fixed group of resources, partial loading can determine the shared resources amount. The key is to maximize the power range in that power segment of the system. This process is self correcting since manufacturers who under-load the power range would result in higher power allocation that would factor into their system assessments. Therefore, system manufacturers are able to choose the loading appropriate to the power partitioning of their chassis.

## Conclusion

Intel appreciates the opportunity to provide the EPA with the comments and recommendations for the draft of the ENERGY STAR Program Requirements for Computer Servers v2.0 specification. We hope you will include these considerations in the specifications later this year.

## Appendix

Line 221 – strike “and repetitively” -> “simultaneously run a single …”

Line 236 – there’s little value for the EPA to mandate disclosure of availability metrics. These characteristics and discussions are already occurring between the system manufacturer and customer.

### Active Mode {363-368}

Active State: The operational state in which the computer server is carrying out work in response to prior, current, or pending requests. Active state includes: (1) active processing, (2) time spent waiting due to data seeking/retrieval from memory, cache or internal/external storage, (3) writing updated records to internal/external storage, and (4) any housekeeping functions to preserve the integrity of the server like virus scans, backups, etc.

### Utilization

Recommend removing the word “computing” in line 398 so it reads “instantaneous processor activity”. Please note that logical operations and branches do not compute.