IBM COMMENTS TO EPA'S ENERGY STAR® "PROGRAM REQUIREMENTS FOR COMPUTER SERVERS DRAFT 2"

IBM appreciates the opportunity to comment on Draft 2 of the EPA ENERGY STAR® Computer Server Specification. The EPA continues to drive convergence on the specification requirements and Draft 2 makes meaningful progress toward establishing the framework and requirements for a final specification. IBM believes that the definitions, power supply efficiency requirements, standard information reporting proposal, and power and temperature measurement and reporting capability requirements are largely complete, requiring only minor modifications to be ready for publication in a final specification. IBM continues to have serious concerns with the workability and technical rationale for the qualifying products and the idle power requirements. Given the growing computing capability and the number of processor cores available on a single processor die, using up to 4 processor sockets without a form factor limitation, per the proposed qualifying criteria for rack mounted and tower servers, encompasses too broad a range of capability given the limitations of the SPECPower_ssj2008 test methodology. In addition, IBM continues to have very serious technical reservations about establishing an absolute or configuration-based idle power criteria given the complexity of Computer Server systems and the very large number of component configurations available within a given Computer Server model type. As the marketplace moves to more virtualized systems, customers will require more complex configurations than those submitted for SPECPOWER_ssj2008 results under the SPEC reporting process by manufacturers. The use of more complex server configurations will enable and facilitate the use of virtualization, the consolidation of workloads onto a smaller number of Computer Servers, and higher equipment utilization in the data center. This in turn will maximize the computing workload delivered per watt of energy used by the Computer Server. It is essential that the EPA not penalize or discourage the purchase of more complex configurations, as it is these systems which can combine the use of power saving functions to reduce power use at idle and virtualization capability to maximize the amount of work delivered by the Computer Server, which will transform energy use in the data center and drive results consistent with the best practice and state of the art technologies as depicted on page 10 , figure ES1, in the EPA report to Congress.[1]

IBM and industry groups such as ITI and Green Grid have provided the EPA with a workable configuration-independent alternative to an absolute idle power specification. This proposal can encourage the reduction in Computer Server power use when no workload is present and the development of the Computer Server level functionality that will be required to enable the implementation of software based, data center wide policies for power and workload management. It will ultimately be these policy vehicles which transform data center power use and management. If EPA chooses to set a standard for idle power as measured by SPECPOWER_ssj2008, IBM has provided a recommendation to limit the criteria for qualifying equipment to servers with up to 2 processors in a 1U/2U form factor to match the technical capabilities of the SPECPower_ssj2008 metric. IBM is working to develop and submit additional data to support these recommendations

[1] Report to Congress on Data Center and Server Efficiency, Public Law 109-431, U.S. Environmental Protection Agency ENERGY STAR® Program, August 2, 2007

and others that will be presented in the subsequent comments. IBM respectfully encourages the EPA to modify its proposals on these items to create a workable specification which will not only reduce power use for equipment when it is idled, but which will drive the equally important outcome of maximizing the use of installed equipment. This not only reduces the energy in the use phase, both for operation and from reduced space needs, but also by reducing the energy required to manufacture the equipment where one server does the work of 5, 10 or 100 individual servers.

SPECIFIC COMMENTS

1. SERVER DEFINITIONS

1.A: IBM continues to encourage the EPA to include the service processor as a defining characteristic of a Computer Server. In the ITI discussions among major Computer Server manufacturers including the main x86 processor manufacturers, it was agreed that a service processor was a defining system characteristic. The service processor manages the integration of the system components and network connectivity and differentiates a Computer Server from a Desktop Derived Server. Also, a service processor must always be powered on at some level to maintain network connectivity and maintain systems at a minimal readiness for serviceability and reliability purposes. Its presence drives a minimum idle power level well above the minimal watts achievable for workstations or Desktop Derived Servers.

1.A: The inclusion of "marketed and sold as a server" as a distinguishing characteristic of a server is logically inconsistent. This criteria is not a technical criteria, but rather a self definition; if a company designates something as a server, it is a server as long as it meets the other criteria. This criterion should be removed and replaced with the Dedicated Management Controller or service processor criteria to give the definition a strong technical basis.

1.B: The first sentence of the Blade Server definition should read "A **C**omputer **Server** consisting of, at a minimum…". The Blade Server definition needs to reference back to the Computer Server definition, as the Blade Server category is a subset of the Computer Server category.

1.I, J: The definitions for "High Redundancy Servers" and "Standard Redundancy Servers" should be changed to "Redundant Power Supply Servers" and "Single Power Supply Servers". This more accurately defines the distinction EPA is seeking to make with regard to the presence of redundant power supplies and their impact on system idle power. A server with a single power supply is, by definition, a server which is not redundant.

Consistent with the earlier comment, the Dedicated Management Controller or Service Processor should be removed as a distinguishing characteristic of "Redundant Power Supply Servers", as it is a universal characteristic of all Computer Servers, unless EPA

chooses not to designate the Dedicated Management Controller or Service Processor as a distinguishing characteristic of a Computer Server.

Both definitions should use the term "Computer Servers" rather than "server computers" or "computer servers" to definitively tie these definitions to the Computer Server definition.

Additional Comments on the Definitions:

1. Include definitions for high volume and medium and large server subcategories: IBM appreciates the EPA's addition of a specific statement regarding which Computer Servers are not considered qualifying products. However, as set forth in our comments to the EPA on May 9, 2008 and July 25, 2008 regarding previous drafts, IBM feels that the EPA needs to include a "Volume Server" and "Medium to Large Servers" category in the definitions. The additional definition will better delineate those servers which do and do not qualify under the proposed ENERGY STAR® specification and will complement the Blade Server definition. Inclusion of two or three distinct server categories to cover the "Volume" or "Small", "Medium" and "Large" Server categories[2] will help eliminate confusion with regard to which products can be qualified under ENERGY STAR®.
2. IBM supports the inclusion of DC-DC powered Computer Servers in the ENERGY STAR® Computer Server specification where qualifying products are powered from a direct current distribution system and utilize DC-DC power supplies. IBM provided its DC-DC power supply efficiency data to EPA on July 31, 2008. This data was not generated using the draft EPRI/ECOS Generalized Test Protocol released for comment on August 5, 2008, but did use standard, accepted industry test methods.

2. QUALIFYING PRODUCTS

The second sentence of the Qualifying Products paragraph should be changed to read "…is limited to ***C***omputer ***S***ervers and ***B***lade ***S***ervers capable of having at most ***two*** processor ***sockets in a 1U/2U form factor…***" …***Computer S***ervers with greater than ***2*** processor ***sockets and/a form factor above 2U*** are currently …"

"Sockets" should be added after "processor" in the definition of Qualifying Products to avoid any confusion between processor, processor cores, and processor sockets. As written, the Qualifying Product definition and the subsequent discussion of Idle Power in section 3.B, could be misconstrued to encompass those machines with only a single processor die with 4 processor cores. It is important that the EPA reference processor sockets in all its definitions.

IBM believes that qualifying Computer Server products for other than Blade Servers should be limited to Computer Servers with up to 2 processor sockets and a 1U or 2U form factor. As processor capabilities are expanded, with more processor cores available

[2] IBM Comments to "Energy Star® Specification Framework for Enterprise Computer Servers", Building Block #1: Definitions, August 31, 2007.

– current designs may include 4 processor cores per processor die – 4 processor systems can deliver 16 or more processor cores of processing power. These systems are capable of managing and utilizing significantly memory, storage and I/O resources, affecting idle power measurements and the utility of SPECPower_ssj2008 as a metric for these larger systems. Thus, IBM recommends that EPA limit the applicability of the specification to 1U/2U form factor systems with up to 2 processor sockets, as this is the range where SPECPower_ssj2008 provides a meaningful representation of the power use characteristics of the Computer Servers. As additional workloads are included in the SPECPower program, the work to develop the Tier 2 program requirements should expand the workloads and qualifying products accordingly.

If EPA were to utilize the Power Saving Function checklist criteria, with a reporting requirement for idle and maximum ranges for the range of configurations available for a Computer Server model, then use of the up to 4 processor socket qualifying criteria would be appropriate.

3. EFFICIENCY REQUIREMENTS

3.A: Power Supply Efficiency Requirements

IBM supports the power supply efficiency requirements proposed by the EPA. Segregating the power supplies by capacity has technical justification as power supply efficiency improves at all operating points with higher input power. Use of the converged CSC/ECOS 80+ power supply efficiency silver and gold standards sets achievable, technically sound efficiency requirements which set a stretch target but which are already reflected in design efforts for next generation Computer Server products.

IBM believes that the use of what was previously the ECOS 80+ Gold standard[3] as the basis for the efficiency standard and testing requirements is most appropriate for the full universe of Computer Server products due to the stated program emphasis on these products. In addition, the specification and testing procedure development process had extensive input from power supply and Computer Server manufacturers. Similarly, the less than 1000 watt specification using the ECOS 80+ silver standard, with a downward adjustment on the efficiency requirement at the 10% operating point, provides an achievable stretch target for these power supplies.

Based on the adjustments that the EPA has proposed for the efficiency requirements at the 10% operating point and the removal of the 10% operating point on multi-volt power supplies, IBM does not believe that there is any need for an exemption process for compliance with the efficiency requirement at the 10% operating point.

Notes: Multi-Voltage AC-DC Power Supplies

[3] February 2008 – Final Report on Efficient Power Suppliers for the Data Center and Enterprise Servers" ECOS 80+

Multi-voltage AC-DC power supplies are found on a limited number of "Redundant Power Supply" or "High Redundancy Servers" in the IBM product line. As the number of systems is limited and it is likely that these supplies will be phased out of the product line, IBM finds the proposed multi-voltage power supply efficiency standards to be achievable but with an appropriate stretch target. It is important, though, that the testing criteria utilize a 230 VAC input voltage, as this is the design point for the power supplies.

IBM supports the EPA's proposal that power supply efficiencies be reported without the inclusion of fan power use. Because of the diversity of fan configurations in Computer Servers, including the use, in many cases, of larger, more efficient fan systems to support both the cooling of the power supply, the processor system and other components, removal of the fan power from the efficiency calculation is important to provide a normalized comparison of power supply efficiencies.

3.B IDLE POWER

As stated in the introduction to these comments, IBM believes that the best available means to encourage the development and marketing of systems which minimize Computer Server power use when no workload is present is to set criteria for the availability of power saving functions on Computer Server models. The industry is working on developing data center management systems that will set datacenter wide policies to manage power use. One of the key building blocks of these management systems will be the equipment level power management functions. As such, it is appropriate for a Tier 1 specification to encourage the deployment of these power saving functions. IBM provided extensive comments to this effect in its July 25, 2008 comments which are provided with some noted edits (in italics) here.

The power drawn by a server when no workload is present is heavily dependent on the configuration: the number of processors, the quantity of memory, the number of active I/O points, etc., and the availability and enablement of power management functions on the server. The types of server which will have the lowest power level when working are often those which are least capable of running multiple operating systems and workload to drive higher utilization rates. Unlike a PC, there is no effective power limit or value that can be expected when a system is not doing work or has been put into sleep and hibernate mode, because of the tremendous variability in form and function among server configurations. IBM is very concerned that a focus on idle power will drive the wrong behaviors and purchasing decisions. The most efficient server will be the one that maximizes its utilization and the workload that it delivers and has the capability to minimize its power use when no workload is present.

Instead, we recommend that the EPA require that an ENERGY STAR server require a specified additional number of power management functions. Power6™ processors manufactured by IBM and many processors manufactured by others have functions which promote power saving when a processor is not doing work, although they may need to be enabled by the equipment manufacturers through firmware enablement or

other forms of hardware and software integration. In addition to processor level power management, here is a list of additional power management functions which the EPA should consider for inclusion in the proposed specification:

Processor Sleep;
Power Capping;
Dropping into lower static power modes as workload reduces;
Dynamic Voltage and Frequency Scaling;
Variable Speed Fan Control based on power or thermal readings;
Lower Power Memory States;
Lower Power I/O Interfaces;
Rack level or processor level liquid cooling; and
Export power and utilization to System administrators for action.

It should be noted that low power components, as opposed to those which can only reduce power use as workload decreases, such as memory and processor systems should not be required for an ENERGY STAR qualified system. Low power systems make trade-offs on performance capability in exchange for reduced power use. This is appropriate for some applications, but not others, and it is important for customers to be able to choose the type of system that they need to meet their performance, reliability and serviceability requirements. In fact, processors that best enable virtualization and consolidation tend to be those with higher power demands to drive the higher performance levels required to support virtualization.

To achieve an ENERGY STAR rating, the EPA can require that a server have a specified number of these capabilities enabled. *It is likely that there may be one or two power saving functions which provide 50-75% of the power saving capability – based on supplied data EPA could choose to require that those power saving functions be present on a Computer Server for that system to be qualified as an ENERGY STAR Computer Server.* In order to set the 25% threshold, the EPA can require all manufacturers of server equipment to provide a list of their current models and the attributes from the list which are enabled on their current equipment. EPA can then determine how many of these functions - four, five or six – need to be available to establish the "25% deployment threshold". The EPA may wish to work with groups like Green Grid and ITI to collect the data and ensure that the full universe of available products is analyzed to get an accurate representation of the current availability of these functions on the market. Also, it will be important to work with industry groups to develop educational and marketing messages to provide data on the efficacy of these functions in a data center environment and to encourage data center operators to enable these functions.

An additional benefit of this approach is that the EPA can resurvey the marketplace two years after the ENERGY STAR server spec takes effect and determine if new attributes are available which can be added to the list, and get the information needed to reset the 25% criteria. This provides EPA a simple and effective way to improve equipment efficiency, drive lower power use when no work is being done, and allow manufacturers to innovate across their product lines. By surveying and utilizing the current list of power

management technologies in the ENERGY STAR specification, EPA can encourage all manufacturers to make power management functions available on their servers.

IBM would like to respond to the three specific inhibitors that EPA identified with this approach in the "Power Saving Checklist" paragraph on page 9 of the draft.

1. **Savings through power management are difficult to characterize to determine programmatic savings and are user-dependant.** The reality is that the difference between power use at maximum workload and system idle is largely dependent on the presence of power saving functions in the components of a given server system. This is exemplified by the fact that older systems, n-2 generations from currently available equipment, exhibit differences between idle and maximum power of less than 30% and in most cases, 25% or less.[4] Hence, real reductions will be achieved where manufacturers make these functions available and data center operators utilize them. Google(TM) completed a study which estimates that implementation of dynamic voltage and frequency scaling on all servers in a data center can reduce overall IT power use by 20% [5]- a significant achievement and one whose value can be doubled if cooling delivery is matched to the IT heat generation. IBM has prepared a list of the functions and the expected range of power savings that can be achieved by each function and provided it with this submittal. This list illustrates the capability of these functions to reduce power use in the data center. If this list of available functions is combined with the reported idle power range for the Computer Server model (section 3B of draft 2), then customers can assess and compare the capability of competing models. As to the statement that utilization of these functions is user dependent, that is true whether the standard is set based on an absolute idle power measure or power saving function requirements. When measuring idle power, manufacturers will enable all the available power saving functions to their fullest capability during the performance test. Current practice at IBM is to ship many of its computer servers with these functions disabled and allow the data center operator to enable them if that is part of their business process. Either approach to driving reduce power use when no workload is present is dependent on the data center operator enabling and utilizing the available power saving functions.
2. **Addressing power saving features instead of the absolute watts drawn could overlook new energy saving technologies, unrelated to power management, which enter the market after the specification takes effect**. ITI and Green Grid members have reviewed the proposed list of power saving functions and agreed that this list is inclusive of current and planned (near future) functions. Given EPA's plan to update the specification within the next 12 to 24 months, the proposed list can be updated at the next rewrite of the specification. It is expected that EPA will approach the Computer Server specification similar to the V4 of the workstation specification – the next draft shortly after the previous draft was published. In addition, manufacturers will be reporting idle power ranges for the models on the Standard

[4] J. Chase, D. Anderson, P. Thakar, A. Vahdat, and R. Doyle, .Managing energy and server resources in hosting centers,. in Proceedings of the 18th Symposium on Operating Systems Principles (SOSP), Oct. 2001

[5] "Power Provisioning for a Warehouse Sized Computer"; Fan, Weber, and Barrosso; Proceedings of the ACM International Symposium on Computer Architecture, San Diego CA June, 2007

Information Reporting sheet. New developments will quickly be highlighted by manufacturers, and attendant data published to differentiate products in the marketplace.

3. **A prescriptive approach will become quickly outdated**. In its proposal, the EPA believes that specifics about power management features are best identified as a reporting requirement rather than a pass/fail requirement. By providing a broad list of known and planned power saving functions, and requiring specific, high value functions, the power saving list will have sufficient "shelf life" for the purposes of the Tier 1 specification. The EPA has already stated that they intend to begin work on the Tier 2 specification to incorporate performance criteria into the specification. The Tier 2 will either provide a new, performance based power methodology for assessing Computer Servers under the ENERGY STAR program or provide the opportunity to update the power saving function list to reflect all available technologies.

IBM continues to advocate for the power saving function proposal, due to the impact of configuration on the idle power and maximum power outputs and the increased capability of more heavily configured machines, particularly to virtualize workloads and drive higher equipment utilizations. The Power Saving Function approach makes the specification configuration independent, recognizes and points purchasers of the Computer Server to the functions they need to enable to reduce idle power use, and provides visibility to the idle power range of the equipment in the Standard Information Report. It also lays the groundwork and provides the functions needed to enable data center wide power provisioning policies that will serve to reduce energy use throughout the data center ecosystem.

Concerns with the Proposal for an Absolute Idle Power Criteria

A configuration dependent specification, particularly one which utilizes a base idle power plus adder approach similar to that used in the imaging specification, will depend on a product management process with the ability to make an ENERGY STAR® assessment on each configuration of each model. This will require extensive testing and data collection on each component to determine idle power and the ability of the manufacturer's configurator program to assess the conformance of each configuration with the specification. This will require several levels of data quality checks:

a. Verification of the idle power of each component or component type.
b. Verification that the model algorithms correctly calculate the idle power of the configuration and compare it to the ENERGY STAR® standard.
c. Program to do some Computer Server level power testing to verify that the configurator is correctly reporting the power use of the Computer Server products.

While this program is doable, there is significant cost to formalize the program, establish the algorithms to determine conformance with the ENERGY STAR® requirements and undertake the necessary testing and verification to satisfy EPA that the configurator results are representative of the performance of the actual Computer Server in the field (assuming that the data center operator enables all of the relevant power saving functions). Using a Power Saving Function criteria with reporting of idle power provides

a more easily managed and verified program which drives the same results of reducing overall Computer Server power use in the data center.

Comments on the EPA Idle Power Criteria

In the event that the EPA determines that the best approach for the specification is to set criteria based on an absolute idle power standard, then IBM supports the use of the simplified system proposed by the EPA in the Draft 2 document with the following modifications:

1. The criteria for qualifying products should be limited to Computer Servers with 2 processor sockets or less and a 1U or 2U form factor. The reasons for these limitations are discussed on pages 3 and 4 above.
2. Set the memory criteria for less than 4 DIMM slots (Low Memory Systems) and 4 or more DIMM slots (High Memory Systems). Currently, manufacturers are preparing to release a 16 GB DIMM. Demands for memory will increase in line with processor performance, so using the DIMM criteria will better represent actual system purchases and will make the specification more robust over time.
3. Reference "Processor Sockets" in the descriptors for the rows and columns of Tables 3 and 4.
4. Specifically indicate that idle power for Blade Servers will be assessed against the "multi-OS version of SPECPower_ssj2008", which is currently in the final stage of review by the SPEC organization. This version of SPECPower provides the workload and procedure appropriate for assessing Blade Servers.
5. Consider the use of LINPACK as the workload generator for the performance testing and idle and maximum power measurements. LINPACK is an accepted industry testing standard which scales well with increased numbers of processors and memory.
6. If the qualifying products continue to be defined as a Computer Server with 4 processor sockets or less, consider adding a more robust workload to the specification to properly exercise these larger systems such as a workload representative of a data serving environment or a large web environment. Work would need to be done by the industry to define the appropriate performance testing methodology to represent these large workloads for more highly configured systems.

3. B Standard Information Reporting Requirements

System Characteristics:
Per definition comments, server processor availability should be removed.

System Configurations:
Remove Installed Operating system, as system characteristics provide the supported operating systems.

Air Flow Rate Information

Remove the Delta T information, which is heavily dependent on configuration and workload and not really used for any meaningful determinations. Delta T issues will be evaluated in the context of the data center ecosystem.

4. Test Criteria:

As discussed above, multi-volt power supplies should be tested at 230 VAC.