IBM appreciates the opportunity to provide comments to USEPA on the ENERGY STAR® Program Requirements for Computer Servers Draft 1 Version 2.0. The initial draft of the Tier 2 requirements shows EPA's desire to add a performance component to the assessment of the servers under the ENERGY STAR® program. Power Management for IT equipment has three components: the amount of work that the equipment can deliver for each unit of energy supplied (its performance/power profile), its ability to maximize the amount of work it does (virtualizing workloads and maximizing system utilization) and its ability to reduce its power use when workload is not present. In the Tier 1 requirements, EPA primarily focused the requirements on only the idle power component of energy efficiency. Extending this to incorporate a performance criterion is an important next step in developing a criterion which truly identifies computer servers which deliver the most work per unit of energy supplied.

IBM supports EPA's plan to include blade systems in the Computer Server Tier 2 specification. Blade systems represent a significant portion of the computer server market and offer system efficiencies which can result in improved energy efficiency and performance. IBM is concerned, however, with EPA's proposal to set specific power standards for blade chassis. Given the diversity of blade chassis configurations within and between manufacturer's offerings, a power specific criterion does not provide a means to evaluate and differentiate chassis systems and will penalize manufacturers who install more overhead (fans, storage, and network connectivity) into their base chassis offering. IBM will present recommendations to require chassis systems to meet functional criteria to qualify for the ENERGY STAR designation.

IBM is very concerned that the sampling and reporting rates proposed for servers in the Tier 2, Draft 1 are unworkable. The collection and averaging frequencies collect an inordinate amount of data and report it on a time frame which is not representative of the response times in a data center. Data center operators may look at their temperature and power profiles on frequencies of ten or fifteen minutes. In a system which often has thousands of individual data points, anything more frequent is focused more on the "noise" within the system rather than on any meaningful trend in data that requires a response or action by the data center operator. In addition, generating and collecting that data requires significant IT support and provides little or no overall value to the operation of the data center. We encourage EPA to recognize that the intent of power and thermal reporting is to gather operational data that can be used to manage the data center and set reasonable, workable measurement accuracy and reporting frequency requirements for data collection.

IBM will offer specific recommendations to improve the product family criteria in a way which meets EPA's goal of identifying energy efficiency products while allowing industry to incorporate a broad range of configurations into a single product family to minimize testing requirements and complexities while providing data center operators the relevant information they need to select energy efficient products. As proposed in this draft, the product family proposal is little better than qualifying single product configurations and introduces unnecessary testing and reporting requirements.

Following are comments to the Tier 2, Draft 1 document and responses to specific questions or topics raised by EPA with the document. They are organized by topic and citations are provided to the document page and line numbers.

Page 5; L182-187: IBM does not believe that there is any value to adding a definition for a "volume server". There are no technical criteria which would clearly define or differentiate "volume servers" from other types of computer servers. Rather, "volume servers" are simply those computer servers which are not otherwise defined in the "Computer Servers Type" section.

Page 5; Line193-196: IBM does not object to the addition of the additional features in the definition of a "Managed Server". These features are typically available through the dedicated management controller. The benefit of keeping the definition at redundant power supplies and a dedicated management controller is that it allows for innovation and development of other "managed server" capabilities without setting specific feature requirements under the management controller.

Page 7 L236-244: Use of the Harvard Research Group (HRG) "Availability Environment Classification System" (AEC) or the IDC Server "Availability Levels" (AL) (another rating system similar to the HRG AEC). These systems provide a subjective delineation of systems availability which cannot be translated into specific, defined functions or attributes of a server system which can translate into specific definitions of product categories. A single product can attain the capabilities of a specific level of the HRG AEC or IDC AL by utilizing specific components or software settings. IBM encourages EPA to reject this suggestion.

P7 L240-42: The definition for fully fault tolerant server should remain as currently written and not add architecture components or attributes. A system which does not have full redundancy built in, but which also has architecture features to make it more reliable or close to a fully fault tolerant system is included under the resilient system category. Use of an architecture component for the fully fault tolerant server definition will only serve to obfuscate the definition and make it difficult or impossible to identify fully fault tolerant systems.

Page 7 L243-246: IBM agrees with the revised "resilient server" definition as proposed in the document (lines 223-235).

Page 7 L247-51: Multi-Node Servers: Modify the first sentence to read "…share an enclosure ***or are in separate enclosures that are interconnected by specialized hardware*** and one or more power... This addition is relevant, as IBM manufactures multi-node systems in which the nodes are in separate enclosures, but the system exhibits all the functions and characteristics of a dual node system.

Add "…in a single enclosure" to the end of item (i) under Multi-node servers.

Create an item (ii) Two enclosure Dual-node Server: A common multi-node server configuration consisting of two server nodes housed in separate, specially interconnected enclosures.

Pages 7, 8; L277-283: Computer Server Form Factors: IBM does not feel that there is a need to distinguish between pedestal and rack servers – from a component, power use, and performance characteristics these systems are interchangeable and operate to the same specifications. Addition of the distinction is not pertinent to energy use and efficiency and the addition of these definitions add no benefit.

Pages 8, 9; L320-331: Added Definitions for Motherboard, Processor, memory, hard drive, and SSD, as they are referenced in the product family guidelines.

1. For the mother board definition, insert: "...the motherboard "***typically"*** includes connectors…" In some cases, the motherboard may not have the identified connectors.
2. Expand the processor definition by adding this sentence at the end of the current definition. "***The typical CPU is a single microchip in a single physical package to be installed on the server motherboard via a socket or direct solder attachment. The CPU package may include one or more processor cores."*** We think it is important to get the discussion of socket and cores into the definition, as many in the industry view and define a single core on a multiple core chip as a processor.

Page 9; Line 359-62: Active State Definition: IBM finds the proposed "Active State" definition to be acceptable.

Page 10 L398: Processor Utilization: The definition of processor utilization should be changed to read: The percentage estimate of the server's compute activities relative to the full operational voltage and frequency of the processor(s). Processor utilization can be reported at the processor or CPU core level.

Page 10; L391-2: Reword Availability Features as follows: Features that support a server's ability to continue to perform it intended function at full capacity after the occurrence of one or more component failures.

Page 10; L393-4: Reword Serviceability Features to Read: Thos features that affect the duration and skill required to repair a server. These features allow a server to be serviced in the least amount of time through means such as automatic detection, isolation, and reporting of actual or potential failures, tool less parts removal, color coded touch points, and keyed connectors.

Page 10-11; L410-437: Product Family Definition

IBM strongly encourages EPA to broaden the product family definition to allow the qualification of a machine type or model (the highest designation for a product). IBM

has a machine type, model number, and feature code hierarchy and proposes that a product family be defined either as a single machine type or as a selected subset of machine type and model numbers.

Under the Tier 1 product family definitions, a four processor system is required to delineate a product family, and provide a full QPI/product data sheet for each level of populated processor socket and number of processor cores. Instead, we propose that for a product family, we define the power use and performance of the minimum and maximum configurations based on the highest power processer that can meet the specification requirements, with the minimum configuration defined as a machine type with one processor socket populated, a typical configuration as a system with 2 processor sockets populated, and a maximum configuration as a system with 4 processor sockets populated. This approach represents the full range of power and performance for the machine type (or with a defined subset of model numbers) and provides the information that our customers are typically interested in when they evaluate our products.

In order for this to work, the processor line of table 1 needs to be changed as follows:

Same Part Number Required in All Product Family Configurations: Should read no – multiple processor part numbers should be allowed based on both power sorts and speed sorts.

Change the processor "Note" to read:

- Processors must all be from the same manufacturer and model line.
- Processor core count and power requirements can vary. If so, the minimum, typical, and maximum configurations should be tested with the highest power/core count processor.
- Processor speed may vary within a product family.
- Power/Performance Datasheet should indicate the range of processor power and core count available in the product family.

For the table, remove the requirement that Part Numbers remain the same for power supplies, motherboard, and processor type. For motherboards, part numbers change as engineering changes (ECs) are processed for the system in order to differentiate different generations of the motherboard. Power supplies will have different part numbers depending on the manufacturer and may also change part number with ECs. Processors will have different part numbers for different speed sorts (same processor socket power) and different processor socket power levels.

IBM agrees with EPA's proposal for a range of capacities for storage devices and memory systems, as this accurately reflects the fact that customers buy a range of memory configurations depending on the requirements of their specific workloads. As we proposed for processors, it would be appropriate to include a minimum and maximum memory capacity and number of hard drives within a range of power use associated with the specific component at the minimum and maximum capacity of those components.

IBM would like to propose a face to face meeting with the ENERGY STAR® team sometime in the near future to discuss the product family concept and our proposed changes in detail. Specifically, IBM would like to discuss the concerns raised by EPA under "Power and Metrics" (Lines 429-434). Specific areas of discussion:

- The rationale for IBM's proposal when combined with the data that can be provided on power use by the Power and System x power calculator.
- How component power use is determined and characterized and the variability which exists between the same component sourced from different manufacturers.
- Data that EPA would be looking for to support IBM's proposal for product families.

With the changes being proposed for the testing and verification process, properly defining the product family and allowing a broad range of product configurations will enable companies to establish product families and minimize testing requirements while providing the information needed to characterize the systems, and provide purchasers the information they need to differentiate energy use in a system.

Page 12, L438-447:

IBM strongly opposes the changes to the definitions for the minimum and maximum configurations. The minimum and maximum configuration definitions should remain as defined in the Tier 1 requirements. Associating these configurations to the minimum and maximum possible active energy efficiency creates an indeterminate definition. In order to ascertain which configuration represents the minimum and maximum achievable active energy efficiency a detailed active energy versus configuration versus power use matrix would need to be tested to allow statistical determination of the minimum and maximum active energy state for that product family. The extent of the testing required to make this determination would negate any benefit of the product family categorization of products. The only logical course of action for a manufacturer would then be to qualify one or two configurations per machine type as ENERGY STAR® and then sell component adders to enable a purchaser to configure the ENERGY STAR® model to their specific needs.

Page 22, 23 Line 840-857: Product Family Qualification: If EPA accepts IBM's comments proposals on the product family discussion regarding pages 10 to 12, the Product Family Qualification requirements as stated here are acceptable. If EPA does not make changes to the table 1 product family requirements, IBM intends to provide additional comments on this section, as it minimizes the flexibility of the product family qualification and drives requirements for additional, extensive testing over and above what is needed to characterize the power use and performance of a product family.

Page 12, 13 L467 to 489: IBM supports the inclusion of resilient servers, multi-node servers and blade servers with up to four processors as products which can be qualified as ENERGY STAR® under the Tier 2 requirements.

Resilient Servers: IBM has currently qualified Computer Servers which are covered by the resilient server definition. This approach is particularly relevant if EPA chooses to

remove the current idle limits for the 1 and 2 processor socket servers as suggested by the Tier 2, Draft 1 document. If the 1 and 2 processor socket idle limits were maintained, then it may be appropriate to exempt resilient servers from those product groups.

Multi-node Servers: IBM agrees that multi-node servers should be included in the list of products eligible for qualification under the Computer Server requirements and that they can be addressed through the simple approach of dividing the power use of the server by the number of nodes in the system. IBM has evaluated this proposal and finds that it makes technical sense. In addition, IBM would like to recommend that EPA expand the multi-node server definition to include both 2 node and 4 node systems.

Blade Systems: IBM agrees that Blade servers should be included in Tier 2, however we do have specific concerns with and recommendations for the criteria proposed by EPA.

Page 13; L496-501, 504-15: These lines should be removed and should be replaced with the following statement: "All computer systems eligible for qualification under these requirements must have PSUs which meet the efficiency and power factor criteria detailed in tables 2 and 3."

Page 14; L517-22: IBM believes EPA has acted appropriately by tabling discussion on the use of Net Power Loss.

Page 13, 14: IBM supports the use of the EPRI 80+ Gold power supply standard for Tier 2.

Page 14, 15: Server Power Management Criteria:

IBM supports EPA's proposal to modify the power management requirements as follows:

1. Eliminate the criteria for one and two processor socket systems.
2. Require processor level power management implementation in all shipped ENERGY STAR qualified systems.
3. Require reporting of the system power use at system idle and maximum system work load for the active energy (SERT™) metric.
4. Require reporting of the active energy metric score for the overall server system and its SERT metric components.
5. Use the reported data, and SERT or active energy metric data generated from non-ENERGY STAR systems, to generate an active energy management criterion and determine what, if any, server categories are necessary to properly characterize and compare server system performance under the active energy metric. We think that the metric can eliminate or reduce the need for categorization of servers, though we believe that will have to be proven out by the data collected as the metric is run on different servers. As noted earlier, we think its important to keep the resilient server definition in the requirements as we believe that the overhead inherent in a server with a lot of redundancy could drive different

performance/power attributes than a non-redundant server which may justify a separate category.

6. This criterion would be evaluated through and incorporated into the Tier 3 Computer Server requirements.

Blade Servers should be evaluated according to the criteria listed above.

It is the opinion of IBM that the SERT metric is the right tool to integrate, performance, idle power, and power use over the operating range into a single metric.
There are several methodologies which could enable incorporation of idle into the SERT metric results:
1. Weight the score for the idle state or the idle state and 10% utilization at 20 or 30% of the total score to give more weight to idle power draw.
2. Set a criterion for the percentage reduction in power use between full power and idle power as measured by the SERT test and incorporate it into the SERT metric.
3. Depending on the power management functions of a computer system, the SERT metric over the range of utilization of a server will vary between a straight line and a concave curve. So you are looking for higher power use at higher utilizations and a lower power use (a pretty flat change in power use) at lower utilization. Considering this, you could use the slope of the curve from 0-20 (or 30%) utilization and 80% (or 70%) to 100% utilization as an indicator for of the power efficiency of the unit. You are interested in a flatter curve at low utilization and a steeper curve at higher utilizations which may enable you to consider some form of metric using the slope values.

These 3 options for incorporating idle power into the active energy metric are provided to stimulate thought and demonstrate that there are viable options available to incorporate consideration of the idle requirement into the active energy metric. However, the final methodology chosen will depend on the generation of data for the range of computer servers available in the marketplace and analysis of the data to determine which of these ideas, or other ideas not offered here, offers the best means to incorporate consideration of idle power into the active energy metric.

For this reason, we think that EPA's stated proposal to requiring reporting of the active energy metric on the PPD for all ENERGY STAR qualified models is sound, as it will begin the process of data generation and collection that will be needed to determine the best algorithm/calculation to generate a SERT score.

Page 15, 16: Blade System Criteria

As discussed above, IBM support covering blade servers under the Computer Server Power management requirements proposed in the Tier 2, Draft 1 requirements. If EPA chooses to continue setting idle power criteria for one and two processor socket servers, then IBM recommends that blade servers be evaluated under the server power management criteria for three and four processor systems because of the differences in overhead distribution between blade and rack server systems and the resulting impact on the server idle power calculation.

With regards to EPA's proposal for qualifying a blade chassis, IBM strongly recommends 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:

A complete blade system of minimally configured blades and a full chassis should be tested using the active energy metric. A manufacturer may choose to test a partially populated blade chassis, as long as a given power domain (the maximum number of slots supported by a single power supply or a pair of redundant power supplies) is fully populated with blades. It should be noted that the EPA proposal for partially populating a blade chassis does not make sense, as it does not assure that a single power domain is fully populated. The total power data can be divided by the number of blades in the chassis to calculate the fully burdened, per blade power usage in a fully populated chassis. The minimum blade server configuration should be used for the full chassis power test, as it provides the best means to allow meaningful comparisons between manufacturer's systems. It is likely that typical and maximum configurations will vary sufficiently in power use to make it difficult to derive meaningful comparisons between systems.

The power associated with a single, minimally configured blade server can be calculated by removing one blade server from the chassis and re-measuring the power use at idle and maximum power. The difference in power measurements represents the unburdened, single blade power requirements. The chassis power can be calculated by subtracting the individual blade power times the number of blades in the chassis from the total chassis power.

Similarly, the idle and maximum power for a typical and maximum blade configuration can be determined by inserting and removing a single blade of each configuration into the chassis populated with minimally configured blades.

This procedure, working off a single, fully populated blade chassis using minimally configured servers will provide the power measurements needed to provide the information needed to populate the Power Performance datasheet.

Page 16, L577-82: EPA should remove the requirement that blade servers ship with a list of blade chassis that allow the system to qualify as an ENERGY STAR system. In enterprise data centers, the person who installs the computer may be a contractor or system technician that is not directly involved in the data center operation.

Page 17-19: Active Mode Efficiency Criteria

IBM has provided separate comments on the SERT design document.

Page 17; L625-630: IBM supports the requirements for manufacturers to disclose information on the results of the SERT test on the Performance/Power Datasheet, including both the final score and the intermediate results under the Computer Server Tier 2 requirements. As we discuss in our SERT comments, EPA should not settle on a methodology for the final score until system results are collected and the most meaningful score aggregation method can be determined.

IBM supports the EPA core standards for the SERT metric and is working with SPEC to assure that the metric conforms with the core standards.

See the comments on page 14, 15 (page 6 of this document) regarding IBM's comments on reporting of idle and maximum power and performance data.

Page 19; L696-705: Additional Systems Requirements: IBM agrees with EPA's decision to remove the requirements for Energy Efficient Ethernet from the current draft specification, as the relevant requirements have not been finalized.

Page 20, L729-43: At a high level, if EPA is going to require the running and reporting of the active energy metric for a product or 3 different server configurations for a product family that should provide adequate data and information about the servers power and performance attributes. There should not be an additional requirement for reporting another performance benchmark.

If EPA is going to require reporting of a specified list of server performance or power/performance benchmarks, the manufacturers should be allowed to report the benchmark results for the configuration of the manufacturer's choice for the machine type. The Power Performance datasheet should provide an entry for both the benchmark type and score as well as tested configuration.

There are significant costs and time associated with running a benchmark test. As the manufacturers are already being required to run the SERT metric for the minimum, maximum, and typical configuration, it is unreasonable to require that the benchmark be run on these same configurations. Instead, they should be allowed to report the benchmark for the optimal configuration for executing that benchmark.

Page 21; L786-89: Input Power: The requirement for 5% measurement accuracy at the system level is not consistent with a power supply accuracy that allows +-10W below 100W input. Accuracy will not make a jump from 10% to 5% when the power supply crosses a given power threshold. In addition, systems with multiple power supplies will keep increasing (by 100 W for each power supply added) the minimum error levels. IBM recommends that the measurement accuracy be measured at the power supply level and that it be maintained at 10% accuracy below 100 W to be consistent with the power supply requirements and the capabilities of the sensor systems.

Page 22; L793: Inlet temperature probes are currently capable of +- 2^{o}C, but the validation process creates a measurement/validation accuracy of +-3 ^{o}C. It is our recommendation that we keep the specification requirement at +-3 ^{o}C, as that is the consistent with the level of accuracy that can be verified by the validation test.

Page 22; L811-17: IBM objects to the requirement to report a 30 second rolling average ever second. This quantity of data, when considering that a typical data center has hundred's or thousand's of servers, is completely unmanageable and requires a significant amount of computing infrastructure to just collect and file the data. Cumulating and storing all this data will have several minutes of latency – so the data availability and usefulness will be delayed by several minutes. The data should be collected at a minimum of 30 seconds for power and processor utilization and one to 5 minutes for temperature. There is no value to collecting data on a more frequent basis.

Page 28; L1036-1048: The blade test procedure provided here is not appropriate for blade testing. It is important that the test protocol require that a single power domain, supported by redundant power supplies where applicable, be fully populated with the blades to be tested where the manufacturer chooses not to test a full blade chassis. IBM supports making the all blades identical and testing with minimally configured blades. Allowing heterogeneous selection of blades for testing will enable different manufacturers to load their power supplies to different levels changing the efficiency levels measured by the different blade configurations. IBM's proposal for blade testing is presented on pages 8 & 9.

Page 30; Section 6.2: This section should be eliminated. Blade chassis should not be power tested. See IBM's comments on page 8.

IBM intends to provide comments on the Power and Performance Data sheet once the requirements of the specification are better defined.

The IBM team is available to discuss its technical concerns in more detail. Jay Dietrich (jdietric@us.ibm.com) is the IBM interface to the ENERGY STAR® program and would be happy to answer any questions you have or schedule a meeting with our technical team.

Thank you for considering our comments.