IBM Comments to the ENERGY STAR ® Computer Server Requirements
Version 2, Draft 3

IBM appreciates the opportunity to continue to work with the EPA ENERGY STAR® program to develop Version 2 of the computer server requirements. EPA has made substantial improvements in the requirements from Draft 2 to Draft 3. There are a few key areas that need additional refinement to improve the workability of the requirements and the inclusion of the Resilient Server category.

IBM appreciates EPA's incorporation of the Resilient Server definition, developed by IBM and other interested Green Grid members, in Draft 3. IBM believes that the Resilient Server category defines a class of server with a power signature that is distinctly different from that of a Managed Server. The higher power use is driven by the additional circuitry components required to support the functionality listed in the Resilient Server definition. IBM will discuss the reasons for the higher power requirements in more detail below. Based on the evaluation of the currently qualified 2 processor socket systems, IBM believes it is appropriate to establish a separate idle power criteria for 2 processor socket Resilient Servers, based on the higher average idle power measurements for currently qualified systems that meet the Resilient Server definition.

IBM continues to have concerns with the blade testing procedure and the proposal to focus on the use of half chassis data to set qualification criteria for Version 3 of the requirements. The nature of the blade chassis and the blade server, and the many different approaches to distributing system overhead in a chassis, will lead to "apple to oranges" comparisons between half populated chassis from different manufacturers which will introduce bias into any criteria developed from the submitted data. IBM believes that it is appropriate to allow manufacturers to only have to test a full chassis to provide the required power use and performance data for blade systems.

IBM also urges EPA to post a blinded SERT dataset for submitted products for the first 18 months of Version 2. Because the SERT metric is new, the relative and absolute value of the worklets have not been determined. The stakeholder community is interested in evaluating and assessing the SERT worklets for a variety of purposes. IBM believes that it is best to blind the initial data set to prevent attempts to compare manufacturers systems and/or to establish and assess identified product performance against a single metric, before the data set has been adequately vetted and evaluated by EPA, SPEC, and other interested and informed stakeholders.

We suggest that this can be accomplished by submission of two data sheets to EPA; with and without the SERT data section completed. The PPDS without the SERT data would be used to complete the public information on the ENERGY STAR website, the complete PPDS would be used to compile the blinded datasheet with identifiers removed to prevent matching of the data. Server manufacturers should be required to have the complete PPDS with the SERT data available to their customers.

IBM offers the following specific comments and recommendations with regard to the ENERGY STAR® Program Requirements Product Specification for Computer Servers: Eligibility Criteria Draft 2 Version 2.0.

Lines 5-19, Computer Server Definition: IBM recommends that EPA consider replacing (A) "Product Type" with (A) "Computer Server" and then make the different server types sub-bullets of Computer Server beginning with (1) "Managed Server" continuing through (9) "Large Server." As currently written, it appears that all of the items under (A) "Product Type" represent ten different product types of which a Computer Server is one of ten product types. In fact, the Product Type requirements cover Computer Servers, of which EPA has identified 9 types relevant to the requirements. This distinction, while subtle, is important for the proper application of the definitions and requirements in the Specification that follow.

Lines 20-21: IBM supports the removal of the EEC exemption for systems that are larger than 50 nodes which share the same chassis.

Line 31-48, Blade server definition: IBM recommends that EPA add a fifth type of blade server type: "(5) Multi-node Blade Server: A blade server which has multiple nodes. The blade server will be hot swappable, the individual nodes will not." At least one manufacturer currently has a multi-node blade product and others are expected within the Version 2 timeframe.

Lines 67 to 78 and Appendix B, Resilient Server Definition: IBM appreciates EPA's inclusion of the Resilient Server definition in the Specification. In discussions both within IBM and in industry groups such as ITI and Green Grid, there is general agreement that the Managed Server and the Resilient Server each have their own distinct power signature. These distinct power signatures result from the different levels of functionality and RAS delivered by managed servers which use Intel EN and EP and AMD 3xxx through 5xxx processors and Resilient Servers which use Intel EX and Itanium, AMD 6xxx, and IBM Power™ processors. While the processor type currently provides the first criterion to segregate the two server types, the additional server infrastructure and more complex firmware capabilities drive the higher power requirements for resilient systems. The table below details the functionality/circuitry differences that drive the higher power for a resilient server and provides an estimated range of the higher power requirements for some of the attributes.

EMBEDDED FEATURE OR CAPABILITY	ADDITIONAL POWER RANGE (W)
Use of memory buffers: DIMM cards have an additional chip between the memory chips and the processor to cache specific data pieces to improve system performance. Typically, one memory buffer chip will support two memory chips.	4.0 W per DDR port
Resilient systems include either 1. a dedicated RAID 5 slot or on-board system support for the RAID 5 capability; or 2. dedicated slots for additional RAID cards. The additional functionality requires additional infrastructure overhead and power demand.	1.12-18 W (measured and estimated) 2. 2-5 W (measured and estimated).
Built in chip I/O support for dedicated, high speed connectivity to external I/O and disk expansion drawers requires additional circuitry and power in the EX, Itanium and Power processors.	Included in higher processor power demand.
Increased I/O capability through a larger number of dedicated I/O slots in the chassis and/or the use of a dedicated, high performance I/O controller (either a separate processor or through inclusion in the processor socket).	Additional power from the infrastructure. Depends on the type of slots and controller.
Higher power demands require larger power supplies with higher losses.	2-5 W (measured and estimated)
Higher capacity fans based on the higher power demands and cooling requirements associated with the server.	15-28 W (est.)
RAS-ECC functionality that is detailed in the resilient server characteristics in Appendix B will drive more power consumption through the OS and firmware activity.	Additional Power is embedded in the system and cannot be estimated.
Higher Processor Frequency: Resilient Servers tend to run with higher frequency processors, which use more power both at full load and at idle.	1 W for each 0.1 GHz additional frequency (est.)

The table above provides estimates, based on a minimal number of measurements, of the different distinguishing components and functions that drive additional power use for Resilient Servers as compared to Managed Servers.

Comparison of Power Use for IBM Managed and Resilient Servers:

IBM also prepared a comparison (managed resilient comparison.xls) for three of its products each for both 2 processor socket and 4 processor socket systems – x3750 M4 (managed, EP based server, 2 and 4 sockets); x3690 X5 (resilient, EX 2 socket server); 8205-E6 Power 740 system (resilient, 2 socket power system); x3850 X5 (resilient, EX based 4 socket system); 8233-E8B Power system (resilient, 4 socket power system). The individual product data sheets, with the exception of those for the x3690 X5, are available on the IBM ENERGY STAR website (http://www-

03.ibm.com/systems/hardware/energy_star/). The data clearly shows the different levels of power required managed versus resilient systems and the significant differences in minimum configuration idle power between an EP, EX, and Power based system.

Comparison of Idle Power Criteria for one and two socket Managed and Un-Managed Servers

Another proof point that demonstrates the impact of additional functionality and capability in systems with higher RAS capability is the difference in idle power criterion for un-managed and Managed Servers. For a one socket system the difference is 10 W and for a two socket system it is 50 W. This difference cannot be attributed to the addition of the just the server processor, which will draw between 5 W and 15 W depending on its complexity. It is clear that other infrastructure requirements – more circuitry, more firmware operations, etc. – are driving additional power use in the server. This same infrastructure and associated power increases are seen, and expected, when you move from a Managed Server to the greater functionality and RAS capability of a Resilient Server.

Analysis of ENERGY STAR Qualified Servers to Identify Managed and Resilient Servers and Propose 2 Socket Resilient Server Idle Criteria

IBM analyzed the idle power data for one, two and four processor socket servers that are qualified as either individual products or product families to the ENERGY STAR requirements (Copy_of_enterprise_servers_family_product_list combined and defined – base_idle final.xls). We identified the processor and server type based on the use of EN or EP x86 processor(s) (Managed Server) or EX x86, Power, Itanium, or equivalent processor(s) (Resilient Server). The data for the IBM x3690 X5 server, a 2 socket resilient/scalable server was also included. We had done the qualification work prior to the V1.1 initiation, but we were not able to submit the data in time to have it consider under the V1.0 qualification process. There are several observations that can be drawn from the analysis of the ENERGY STAR qualified products.

1. There are no qualified one socket resilient/scalable servers in the database. We believe this is because the processor power penalty precludes qualification of a resilient server to the one socket idle minimum. IBM does not manufacturer any one processor socket x86 EX servers, but does manufacture two system p one socket servers that would meet the resilient/scalable definition. We did not attempt to qualify the server because of the low, managed server derived idle criteria.

To provide an example of an x86 EX 1 socket system, the table below provides idle data for a minimum configuration (4 GB memory, 1 HDD, and one PCI card) one socket populated (in a two socket system) x3690 X5 server. The idle limit for all but the first entry is 73 W; for the first entry it is 78 W. As you can see from the idle data, even discounting for the extra infrastructure to support the second processor socket, a one socket EX system could not meet the idle threshold. A one socket power system would have a similar power profile.

processor	qty	x	voltage (V)	freq (Hz)	idle (W)
130W, 8C, 2.26GHz, 24MB, X7560	1		229.63	59.995	128
130W, 6C, 2.66GHz, 12MB, X7542	1		228.47	59.999	122
105W, 6C, 2.00GHz, 18MB, E6540	1		228.63	59.996	130
105W, 4C, 1.73GHz, 12MB, E6510	1		228.32	59.888	127
95W, 8C, 1.86GHz, 24MB, L7555	1		230.42	60.008	128
95W, 6C, 1.86GHz, 18MB, L7545	1		230.2	59.997	130
95W, 4C, 1.86GHz, 18MB, E7520	1		228.23	59.997	122

2. There are 8 resilient/scalable server product families from three manufacturers qualified as two processor socket systems. I also added data from 7 product families for the IBM x3690 X5 systems to bring the total number of product families to 15 (lines 16 to 30 of the spreadsheet). Based on idle information data for the minimum configuration, adjusted using the adders to create consistent configurations for comparison, we estimated the idle power threshold for resilient two socket systems using two criteria:

a. Idle power at the 25th percentile: 180 W
b. Idle power at the 50th percentile: 186 W

The calculations for the data minimum, maximum and 25th and 50th percentile values can be found from cells BE256 to BG260 of the spreadsheet. As a note, we think that the data set is a bit small and the idle criterion is potentially skewed by the fact that half of the data is for resilient scalable x86 systems. While it would be beneficial to try to get a broader data set to establish the idle criteria for the two socket resilient/scalable servers, we recognize the time constraints in the current EPA schedule and believe it is appropriate to use the available data set to establish a two processor socket resilient/scalable server idle criteria.

We recommend that you use the idle power at the 50th percentile to set the idle criteria for 2 processor socket resilient/scalable servers in Version 2, as this recognizes that only those products that can meet the idle criteria (which is biased to EP servers, as they make up the majority of the 2 processor socket systems on the market) are reported. Because of the low idle criteria, we believe that most companies did not attempt to qualify resilient/scalable 2 processor socket systems, so using a 50 percentile cutoff provides representation of the current market.

3. The majority of qualified 4 processor socket systems are resilient scalable servers. Until the release of the Sandy Bridge x86 processors, only EX processors could be configured for 4 socket systems. The release of the Sandy Bridge processors enable EP based 4 socket systems, which accentuates the importance of establishing and collecting data on the resilient/scalable category in Version 2. The IBM x3750 M4 EP based 4 socket server discussed above are the first EP based 4 socket servers offered by IBM.

The analysis above details the case for expanding the types of servers to include a Resilient Server category with a separate Version 2 Idle Criterion for two socket systems and separate performance/power criterion in Version 3. Resilient Servers have a distinct

power signature when compared to managed servers and make up a significant percentage of IBM server sales. IDC data that shows that High End servers, which would be classified as resilient servers, accounted for 14% of total global server revenue (all manufacturers) in the first quarter of 2012.

Lines 81 and 82: Changed the sentence "Server nodes in a multi-node server are not designed to be hot-swappable." to "Server nodes in a multi-node server **NEED NOT** be hot-swappable". We do offer one or two systems with hot-swappable nodes that we would want to qualify to ENERGY STAR.

Lines 92-105 and 310, High Performance Computer (HPC) System Definition and Exclusion: IBM agrees with the proposed HPC definition and its inclusion in the requirements. An HPC system can be based on a standard product offering, such as the IBM dx360 M4, or it can be a purpose built system, such as the IBM Blue Gene series of HPC systems. In either case, the HPC system is much more than the processor or base server: it is an optimized, highly integrated cluster of server, storage, and interconnect systems which operates as a single data processing/management system.

HPC Systems based on Standard Products: IBM offers the "Intelligent Cluster" systems which combine industry standard server, storage, and networking solutions with proprietary software to create HPC systems. The datasheet for these systems is posted at: http://public.dhe.ibm.com/common/ssi/ecm/en/cld00221usen/CLD00221USEN.PDF
Page 4 lists the IBM system x servers that can be configured into an intelligent cluster. IBM has qualified a group of those servers; the x3850 X5, dx360 M4, x3550 M4, x3650 M4, and x3755 M3, to the ENERGY STAR requirements. Power systems offers the same capabilities: the Watson system that competed on Jeopardy was based on IBM Power 750 servers, which IBM has qualified to the ENERGY STAR requirements. It should be noted that the HPC product is significantly more than just the server systems – its capabilities are the product of a well designed and integrated combination of hardware; server, storage, and network systems, and software.

Purpose Built HPC Products: IBM also designs and sells purpose built HPC systems such as the Blue Gene line of products. Details on Blue Gene can be found at: http://www.research.ibm.com/bluegene/hardware.html
These systems are designed and architected from the processors up to tightly integrated processors, GPUs, memory, storage and I/O to efficiently execute computationally intensive workloads. Because they utilize an extremely dense configuration, they also utilize specially designed, highbred water/air or high efficiency air cooling systems. Typically, their performance/power scores both for input power and for total power will be very good.

While individual components of the HPC system may be qualified to the appropriate ENERGY STAR requirements, the complexity and size of the integrated HPC systems, either based on standard products or purpose built, do not lend themselves to an ENERGY STAR qualification and IBM agrees with their exclusion in section 2.2 (lines 303-315).

Lines 108 to 112 and 316, Large Server Definition and Exclusion: IBM agrees with the addition of the large server definition to the Computer Server product types. IBM mainframe system z servers and comparable products made by other computer server manufacturers may be offered in 4 processor socket configurations, but are substantially different from a managed or resilient server. While the resilient/scalable server definition underpins the definition of the large server, the system is ultimately differentiated by its increased I/O connectivity. A resilient/scalable calls for a minimum of 8 dedicated I/O slots and typically cannot support more than a maximum of 16 I/O slots whereas a Large Server must have a minimum of 32 I/O slots. The power signature/characteristics of a Large Server are materially different from the power signature/characteristics of a Managed or Resilient server and it is appropriate to exclude Large Servers from the ENERGY STAR requirements.

Line 124, 1.D Computer Server Components: EPA needs to add a definition for an Auxiliary Processing Accelerators (APAs) and GPGDUs. While EPA references APAs in Section 3.9 and provides GPGPUs installed in a PCI slot as an example, the intent is not clear. But there is a significant and growing cadre of components which can be plugged into server PCIe slots to provide additional function, some of which may include a CPU. Examples include:

1. A graphics processing card with independent processor and cache.
2. I/O expansion adapter that allows for channel attachment to racks of external storage and includes read and write caches as well as processing capabilities to accelerate I/O processing.
3. A card which contains 4-8 SSD daughter components and acts as a small storage subsystem within the card cage.
4. A highly specialized I/O adapter that includes a full function x86 processor and operating system within a Power System server.
5. A security card, again with an independent processor or processing capability, to provide encryption capabilities.

EPA needs to determine which of the listed types of APA cards EPA wishes to include in its data collection process and specify the specific types in the definition. At this point in time, it may be appropriate to request data only for the graphics cards. From a qualification standpoint, APA cards should not be included in the configurations tested for the SERT metrics, as SERT is not designed to exercise APA cards. Inclusion of the APA cards in a test configuration will only serve to add power without improving performance. IBM is willing to work with EPA and other industry stakeholders to develop a workable definition for APA cards.

Lines 243-248: IBM recommends that EPA add an item 1.H.2.d which states: “All tested configurations shall fully populate the available processor sockets. Qualified products will include all fully populated and partially populated processor socket systems within the defined product family.” The definition needs to fully define the range of the product family.

In addition, the "socket" in many 2 Socket systems can accept a CPU that is designed only for 1 Socket operation. These are typically lower cost, low to mid-wattage CPU options. Use of these 1Socket CPU SKUs in a 2S system should also be covered under the product family, as they have the same general system characteristics as systems using only one dual socket capable CPU SKU. Customers choose to place these CPU SKUs in a 2S system to gain features not found in the low end / lowest cost 1S systems. In order to qualify, the manufacturer needs to include the SKU for this processor type on the PPDS.

Lines 253 to 275, Product Family Configurations: With the addition of the requirement that only fully populated systems be tested for ENERGY STAR qualification, IBM recommends that EPA revise the definition for the Lower and Higher Cost and Performance Configurations. When IBM proposed the 5 corner product family approach, we believed the best means to allow companies to define the product family was to use the lowest socket power, lowest core count processor for the low and high configurations on one side of the 5 corners and the highest qualifiable socket power and core count on the other side of the product family. As a rule, these should make the low power and core count processors the low cost/low performance option for the minimum and maximum configurations and the higher power/core count processors will provide a higher cost/performance option. We believe that this approach better clarifies the requirements for the four corners of the product family and minimizes the problems with qualified products being outside of the 4 corners as discussed in lines 276 to 280. If EPA wishes to keep its current definitions, it would be beneficial if EPA could identify the use of socket power and core count to set the 4 corners as an acceptable approach to defining the product family.

Line 276, Product Family Adjustment for 1 processor socket systems: IBM recommends that EPA add an item 1.H.2.e which states: "A product family for a 1 processor socket system is defined by 3 configurations: Minimum Power, Maximum Power, and Typical Configurations. Testing of all 5 defined configurations in 1.H.2.a to c must be completed to qualify a two or four processor socket server." Given the limited range of configurations for a one processor socket system, the Minimum Power and Low-end Performance and Maximum Power and High-end Performance configurations are indistinguishable. There is no benefit to requiring the testing of five configurations for a one processor socket family. It appears appropriate to include this distinction in the definition. Adding the requirements for the two and four socket systems prevents potential confusion that could be created by the three configuration one socket product family.

Lines 334 to 342, Power Supply Efficiency Criteria: IBM agrees with the EPA decision to accept power supply qualifications performed against the 6.4.2, 6.4.3, and 6.5 revisions of the *Generalized Internal Power Supply Efficiency Test Protocol,* as well as Rev 6.6, as there were no material changes in the test procedure which would change the reported data.

Line 353, Table 1, DC Power Supplies: As DOE has removed DC power supplies from the Server Test procedure and SERT does not support systems with DC to DC power

supplies, EPA should remove the references to Dc-Dc power supplies from the Power Supply Type. If EPA is interested in qualifying DC based systems under ENERGY STAR, it could consider qualifying a Dc powered server where the AC powered version of the server has been qualified to ENERGY STAR, as the Dc-Dc power supply where the Dc-Dc efficiency is equivalent or better than the Ac-Dc conversion at all load points.

Lines 387-389, Power Management Disclosure: Change the sentence to read "…all power management techniques that are enabled by default **and specifically listed in the power management section of the PPDS** must be identified on the PPDS. **Manufacturers should also identify additional user enabled/managed power management functionality available on the server system**." It is not reasonable to require that all power management techniques, even those which are proprietary and embedded in the systems operating functionality, firmware, or component capability and cannot be altered, be disclosed. It is likely that server manufacturers have developed or will develop methods for quiescing individual components, slowing others, etc. in a way that does not affect performance, which cannot be manipulated by the user, and which a server manufacturer does not want to reveal to its competitors.

Lines 391-394, Blade Chassis Thermal Management: Change the requirement to read "..must provide real-time chassis or **blade inlet** temperature monitoring…" This change is important, as some chassis systems collect their temperature readings based on the blade inlet temperature to better match fan speed to the temperatures at the blade server. It does not change the intent of this requirement, which is that the fan speed is controlled by and varied according to the temperature of the blade server.

Lines 395-403 and 581-587, Documentation Delivery: IBM appreciates EPA's decision to allow companies to deliver required documentation electronically to the purchaser of the server.

Line 400, EPA approved format for Document Delivery: IBM finds the statement to be confusing "A list of qualifying blade chassis and ordering information must also be provided as part of product collateral provided with the blade in either a printed format or an alternative format approved by EPA. These requirements may be met via either printed materials, electronic documentation provided with the Blade Server, or information publically available on the Partner's website where information about the Blade Server is found." As written, the phrase "…or an alternative format approved by EPA." seems to suggest that if a company chooses not to use a printed format, their alternative choice has to be approved by EPA. The next sentence then goes on to define three acceptable, alternate formats. It appears appropriate that EPA remove the "…or an alternative format approved by EPA." phrase from this section, put a period after …provided with the blade.", and then let the final sentence set the three methods by which the information can be supplied. IBM strongly encourages EPA to remove the requirement for their approval on the documentation presentation, as this will create a whole host of unnecessary work.

Line 453, Table 3: As the Low-end Performance and Minimum Power Configurations do not specify a minimum quantity of memory, EPA should consider reducing the idle criterion by 3 W and have the manufacturer add the memory adder based on the total quantity of memory in the system, similar to the change made for the hard drives. It would serve to simplify the calculations.

Line 458, Table 4: EPA needs to add Power Allowances for memory buffers and RAID cards, as these are components that drive extra power use in Resilient Servers. We offer the following data for EPA's consideration:

Additional Component Power Measurements: We performed power use testing of two component adders.

1. Raid Card f/c 5631; IBM Part #74Y3288: SAS controller which splits the six drives bays of Storage Backplane #5618 into a pair of 3 drive bay groups and enables RAID-10 Capability. No PCI slot is used for the controller.

Measured Power draw = 8 W
2. We tested the power use of memory buffer chips on a memory riser card which has 2 memory buffer chips with 2 DDR ports each. We took a system with 4 memory riser cards and removed two of the cards. The DIMMS were distributed to the remaining 2 cards to maintain the same number of DIMMs on the server.

2 Socket Power system, 2 memory riser cards, 16x8 GB DIMMS = 331.84 W
2 Socket Power system, 4 memory riser cards, 16x8 GB DIMMS = 363.73 W

After the removal of the two memory riser cards and reallocation of the DIMMS, the power measurement was reduced 31.89 W. This indicates that each memory buffer chip contributes 7.97 watts to the server system power draw, which is distributed over 2 DDR ports. It is possible that in the future memory buffer chips may support 4 ports, but that the adder will be roughly proportional to the number of DDR ports supported. Therefore, we would recommend establishing the adder per DDR port rather than per memory buffer chip.

We would propose an adder for memory buffer = 4 W per DDR port.

Line 458, Table 4: EPA needs to add the allowance of 46 W for Auxiliary Processing Accelerators (APA), and its associated conditions.

Line 478 to 499, Blade Server Testing: IBM continues to have serious concerns about the proposal to test half blade chassis, as we know there are situations where this will not properly distribute the chassis overhead or penalize particular chassis configurations. It will not result in more comparable data between vendors (line 497) because the test procedure allows the vendor to decide what the required number of options (I/O, fans, and/or power supplies) will be placed in the chassis for the ½ chassis test and this allows

for some amount of customization/optimization to the test versus the configuration that would typically be utilized by the customer.

As an example where the half chassis approach does not provide a representative measurement of power use, IBM has a blade system where "half" of the chassis from a "power consumption" perspective includes the entire power of the infrastructure. The design of our BladeCenter E product is for two power domains: one domain with 6 blades and all infrastructure (fans, I/O, chassis management); and the 2nd domain with only blades, 8 of them for a total of 14 blades. The ½ chassis test procedure requires ½ the bays (7 blades in this case) to be populated, instead of populating only one "power domain" (with it's 6 blades). This product would be penalized by the "½ of the bays" requirement in the test procedure (the per blade average contains the entire infrastructure in the first power domain). Yet this is our most power efficient chassis when viewed from the perspective of 14 blades divided by the total chassis power.

The ½ chassis test is not possible for all IBM Blade Chassis where the chassis is populated with Double-wide blades. A double wide design allows for only 7 blades to be populated, requiring the half chassis test to be performed in 8 slots and two power domains. Note that most non-IBM chassis have 16 blade slots and a double wide product could still be tested in a half chassis configuration. This difference in chassis configuration creates an unequal comparison between IBM and other IT vendors' products which can only be remedied by allowing IBM to test only a full chassis.

IBM's position is that the only test procedure that allows customer (and the EPA) to truly compare per blade results from different vendors is to require the full chassis testing. The data that IBM provided on blade system measurements demonstrated that there were differences in the per blade measurement for a full and half filled chassis. IBM accepts the additional "testing burden" as being part of normal development operation. Each vendor must at some point test a fully loaded/maximum power chassis in order to qualify their product as ready for customer shipment. Making this configuration available for Energy Star testing can be planned so as to minimize development costs and impact, especially with the fewer tests require by adoption of the 5 point test procedure.

At a minimum, EPA needs to allow companies to only have to provide test data on a full chassis if they so choose. This does not represent any risk to the integrity of the ENERGY STAR measurement and reporting process and it provides data that will be more comparable where there are different approaches to overhead distribution between blade systems.

Lines 501 to 503: IBM has concerns regarding EPA’s plan to attempt to include APAs in the measurements and reporting for an ENERGY STAR server.
1. As noted in the definitions section, EPA has not provided an adequate definition of APA in the requirements.
2. The range and diversity of APA type cards is expanding rapidly as software and processing capabilities enable stand alone PCIe cards to perform specialty functions for security, graphics, computationally intensive calculations, and I/O capabilities to extend

server system functionality. IBM believes the rate of change and development in this space will outstrip any effort by EPA to define and set criteria for APAs.

While there may be sense in using Version 2 requirements to survey the universe of APAs and determine the range of power signatures, IBM submits that EPA is likely to be better served by focusing the requirements on the core server system and accept that APAs are too varied to be effectively incorporated into the requirements. If EPA persists in its efforts to incorporate APAs in the requirements, then IBM recommends that EPA work with industry stakeholders to clearly define sub-categories of APAs for which specific criteria can be set and provide an exclusion for any APA type card which is not covered by a category so it can be purchased by customers as a system add-on.

Also, APAs should be specifically excluded as a consideration for both the High End and Maximum Cost configurations as the presence of APAs could distort the performance and power results, biasing comparisons against a future performance/power criteria.

Line 557: IBM continues to believe that the field "Delta Temperature at Exhaust at Peak Temp" either should be removed or modified. The power dissipation is reported from the benchmark run where the temperature is 18-27 C. The peak temperature at 35 C is never tested for ENERGY STAR and power use would be somewhat higher than the 18-27 C due to leakage and fans. Without testing at the 35 C point, causing significant additional time and expense, the "Delta Temperature at Exhaust at Peak Temp" cannot be accurately calculated. We also question its value. We propose that EPA change the requirement to nominal delta temperature, which can be calculated rather easily with power and nominal airflow. If the inlet temperature is known, the exhaust temperature can be calculated. I suggest either removing the row or changing it to "Delta Temperature at Exhaust at Nominal Temp."

Line 568, 5.1.1 and 601, 5.3.2: IBM recommends changing the "…utilization of all logical CPUs…" to "**AVERAGE** utilization of all logical CPUs". In the case where there are many hardware threads running on many cores on several processors, it would not be practical nor would it provide value to report the utilization values for every thread on the system.

Lines 606 to 609, 5.4.1: The reporting frequency should be set at a 60 second period for reporting data that is not time stamped. As we have discussed previously, collecting data from hundreds or thousands of servers on a 10 second frequency will consume a significant portion of the data center network infrastructure with no attendant benefit in clarity of thermal conditions or operational response time.

Lines 612 to 613, 5.4.3: The intent of the time-stamping capability is to reduce the communication burden associated with the collection of power and thermal information on the data center infrastructure. EPA should set the reporting interval for time stamped data at 10 minute intervals. As we discussed above, reporting intervals of 10 to 30 second impact data center infrastructure with no attendant data center operating benefit.

A collection time of 10 minutes of 30 second averages, 20 data points, provides adequate information and generates sufficient response times to identify and react to non-catastrophic thermal excursions in the data center.

TEST METHOD DOCUMENT:

Line 15, Table 1: DOE should include the option for 208 V, 3 phase AC Power.

Line 21 to 22, 4.B: The ambient range should be made consistent with the SERT test requirements (20-28 C).

Lines 44-45, 4.D.6: Complete the sentence with "…including any external shunts**, SHALL MEET THE FOLLOWING CRITERIA.**"

Lines 83 to 88, 5.1.F: This section needs to be reworded to properly reflect the fact that EPA is now allowing qualification of systems with no installed HDD or SSD. IBM recommends that it be revised to read: "Products **SOLD WITH INTERNAL STORAGE SYSTEMS** shall be tested…". The rest of the discussion then follows logically.

Line 136-138, 6.1.I.3.c.iii: Add the following at the end of this section "…or redundancy for a populated power domain where redundant power is a standard offering for the product". Some blade servers may ship with a single power supply, even where there are slots for redundant supply. As with rack servers, the manufacturer should be allowed to specify the standard configuration for the product and test accordingly.

Line 156: IBM recommends changing "manufacturer specified workload software" with "the SERT software".

Line 172, Memory Scrubbing:

For a Power Systems server, the management module manages memory scrub performing the error correction/detection function over a three minute interval in every 23 minute period. The effect of memory scrubbing differs by configuration but we have observed the memory scrubbing cycle to increase idle power by 19 to 46 W, depending on memory size, as a result of moving the server out of idle into an active state. The file "Memory scrub power 3min_23min 09-24-2012.ppt" details a memory scrubbing cycle for two systems with 128 and 1024 GB of memory respectively. The energy use of the scrubbing cycle materially affects the idle power energy reading if it falls within the idle measurement period because of its higher intensity over the short scrub cycle.

Memory scrubbing is handled differently on x86 systems, being initiated once per day and running as long as necessary to complete the memory scrub. We do not have measured data on the memory scrub cycle power use for x86 based systems, but the

product team indicates that it increases the energy use by roughly 3 to 5 W depending on the quantity of memory in the system.

As IBM had mentioned in previous communications to EPA, an implementation of memory scrubbing could cause an intermittent increase in an idle power measure. It is possible that current or future systems may have other functions and self-checks that are needed for server stability that might also be performed on an intermittent basis.

We suggest allowing the idle test to take up to three 5 minute measurements over a 15 minute period, with the lowest of the three measurements reported as the idle power use for the SUT. This will enable identification and removal of power impacts driven by intermittent maintenance activities and assure the idle measurement is representative of what the consumer will see for the majority of time that their system is standing idle. This maps well to the test method for idle from Version 1 of the specification, which allowed a delay of 5-15 minutes prior to taking a 5 minute idle measurement. By making the use of three intervals optional, a Partner could choose to shorten their measurement time to a single 5 minute interval if they are confident that they will not require the other two intervals. Therefore, we recommend adding the following wording to the Test Method document, at or near line 172.

"Some server solutions have periodic asynchronous background processes that are required for system reliability. The overall impact of these processes does not greatly impact the idle power requirements of a computer server, but the processes may create an anomaly in the idle measurement. To accommodate this, a partner is allowed to make three consecutive 5-minutes measurements of idle power and select the best value."

PPDS:

Line 45: Change “24x7x365” to “24x365”.