ITI

the green grid®
get connected to efficient IT

September 9, 2016

Steven Hanson
Manager, ENERGY STAR for Data Center Products
United States Environmental Protection Agency
Washington, D.C. 20460
email: servers@energystar.gov

Re: Draft 1 Version 3.0 ENERGY STAR specification for Computer Servers

The Information Technology Industry Council (ITI) and the Green Grid (TGG) have prepared a joint set of comments in response to EPA's request for comments on the ENERGY STAR® Specification for Computer Servers Version 3 Draft 1.

ITI and TGG have prepared a set of detailed, document specific comments that address questions or comments raised by EPA in the Draft 1 document and changes that the industry considers necessary in light of recent or planned product changes or technology developments. These comments are provided in the document "Power Supplies and misc comments on ENERGY STAR V3 Draft 1".

ITI and TGG have collaborated with the SPECpower Committee on the development of a server active efficiency metric. We have collectively drafted a joint statement on the key aspects of an active efficiency metric and the product configurations which should be tested with the SERT® tool, report SERT data in the xml format and be measured against an active efficiency threshold. The joint position is provided in the document "Creating a Single Efficiency Metric from SERT results final – SPEC TGG".

In addition, we have prepared a detailed proposal for an active energy efficiency metric using SERT™ data. ITI and TGG have collaborated with member companies and EPA to collect and aggregate all of the available SERT data for ENERGY STAR certified servers. We have also collected the SERT results detail sheets with power and performance interval data from member companies to broaden the detail available in the dataset. We have done extensive analysis on the dataset, evaluated over 20 different algorithms for creating a single efficiency metric and chosen a specific algorithm to recommend to EPA for use in assessing server energy efficiency under the ENERGY STAR program. We have also made a recommendation regarding which configurations should be tested, data reported, and thresholds established. The details of the analysis and recommendations and the supporting analysis work that was done are provided in the following documents:

"Recommendation for energy efficiency metric for ENERGY STAR V3 Draft 1" provides the key reasons for the recommendation, observations made during the data analysis work, an explanation of the criteria used to compare the different assessed metric options, and presentation of the summarized data analysis.

"ENERGY STAR V3 Draft 1 recommendation Executive Summary Final" provides a text description of the information provided in "Recommendation for energy efficiency metric for ENERGY STAR V3 Draft 1".

"Deployed Power Analysis Methodology Description Final" which explain the Deployed Power validation method and its utility as a means to assess and validate the efficacy of a given active efficiency metric.

"Draft One Proposed Idle Limits Final" details our position that an idle power limit should not be set under the ENERGY STAR program and provides the data analysis to support this position.

"Creating a Single Efficiency Metric from SERT Results Final" details the equations used to aggregate the worklet level performance and power interval data into a single active efficiency metric. The SPECpower committee has produced a similar paper that begins with the aggregation of the worklet interval efficiency scores. The two methods are equivalent.

"Storage Servers Final" provides an assessment of how to best handle storage and I/O based servers which will be used in software defined storage and network systems in a data center.

"Mean Calculation Methodology Analysis" and "SERTMetric080116" provide an analysis of the geomean function and an explanation of why the geomean function should be used to aggregate the performance and power interval data into the single active efficiency metric.

Questions regarding this submittal should be directed to the TGG SERT Analysis Working Group, chaired by Jay Dietrich (jdietric@us.ibm.com). The Working Group is available to meet with EPA to review and discuss all of the submitted documents.

Sincerely,

Alexandria McBride
Director, Environment & Sustainability
ITI

Micro-servers: Section 1.A Product Types

The TGG SERT Analysis Working Group (WG) recommends that two definitions be added to product types.

1. Micro-servers: Micro-servers use processors with low power and low performance characteristics, with a unique performance/power profile as compared to managed or resilient servers. They are typically configured with multiple nodes in a 2U to 4U configuration.
2. Storage Server: The WG has provided a separate paper “storage server final review draft 1.pdf” which discusses the need to create a storage server definition and develop a separate, single value active efficiency metric using different weightings than proposed for the general server active efficiency metric. This recommendation is a partial response to the EPA request for comment on active efficiency options, addressing lines 409 to 413.

The TGG WG will work with EPA during the draft 2 phase of the process to define these two server product types and determine how best to address them in V3.

Removal of managed server definition: Lines 23-31:

While unmanaged servers have not been certified to ENERGY STAR® to date, the WG is aware of that some servers are offered without management controllers. The majority of these are pedestal servers or dual form factor servers where the rack server can be mounted on a pedestal in an office environment.

Removal of the managed server definition and its two sub-clauses eliminates a management controller and the capability to support dual power supplies as characteristics of a server. While this appears to be appropriate, the WG believes it is important to understand the product types which do not have a management controller, verify that they are included in the computer server definition 1.A.1 and validate that the exclusion of a management controller from the computer server definition does not have unintended negative consequences. The WG proposed to work with the EPA during Draft 2 development to complete the proposed analysis.

If the managed server definition is retained, it is recommend that the 1.2.A be removed, as some computer servers with a management controller are configured to run with only a single power supply.

Memory definition: Lines 155-156:

The TGG WG recommends that an additional sentence be added to the memory definition: “Memory devices include Dynamic Random Access Memory and Non-volatile memory devices.”

Server manufacturers are beginning to offer alternative memory types as new volatile and non-volatile memory and storage device types are developed. It may also be necessary to provide different adders for flash or other non-volatile memory devices. This should be investigated during draft 2 development.

Storage Device definition: Lines 157 to 166:

The TGG WG agrees that the Storage Device definition in the Server Requirements should be consistent with the definition in the Data Center Storage Version 1.0 Specification.

Buffered DDR channel: Lines 230-234

The definition needs to include the requirement for the presence of memory buffer hardware. As worded any DDR channel would qualify. It is recommended that the working be modified to read "Channel or Memory Port connecting a memory controller through a piece of memory buffer hardware to a defined number of memory devices..." Add at the end of this section "In this instance buffered DDR is not referring to DIMM Type."

Low-end Performance Configuration: Lines 255 to 257

The definition needs to specify the lowest performance processor, rather than the lowest socket power. The sentence should be changed to read "The combination of *lowest performance processor as measured by the product of core count and nominal frequency*, PSUs, memory…"

High-end Performance Configuration: Line 260

Include the word nominal in front of frequency to clearly specify the processor frequency that should be used for the calculation.

Minimum Power Configuration: Lines 267 to 281

The TGG WG intends to analyze the minimum power and low-end performance configurations, as was done with the maximum power and high-end performance configurations to demonstrate that a single low-end configuration will satisfactorily define the low-end configuration of a given server product.

Power Supplies: Lines 325 to 238

The TGG WG agrees that the minimum level of power supply efficiency for single output power supplies should be set at 80plus Platinum or the equivalent.

For multi-output power supplies, the TGG WG recommends that EPA set the requirements at 80plus gold or the equivalent. The members of the working group investigated where multi-output power supplies were being used; they are largely and possibly only used in tower servers. There are two key reasons to set the multi-output supply requirements at 80plus gold.

1. In general, multi-output supply efficiency has lagged single output supply efficiency because power losses from additional switches and voltage regulators within the supply make it harder to reach the required efficiency levels.
2. It is more difficult to engineer higher efficiency supplies at lower power outputs. Because the multi-output supplies are used in tower servers, the power output values are lower increasing the engineering challenges and cost to achieve platinum efficiency.

3.3.2 Supervisor Power Management: Lines 358 to 360

EPA needs to clarify the intended minimum power management capabilities required to meet this requirement. TGG WG to provide recommended language during the V3 draft 2 development process.

Require SERT xml output to be submitted to EPA: (lines 387 to 390)

The TGG WG strongly supports the requirement that manufacturers submit to EPA the SERT xml output file. The SPECpower committee offers an automated extract macro which enables easy extraction of the full SERT data set so that the data can be easily accessed for analysis and assessment activities.

The TGG WG recommends that EPA work with the SPECpower Committee to assure that field definition and entry requirements are clear and are crafted to insure consistency across the datasheet submittals. During our work with the SERT datasheets we have found that some fields have very different content from lab to lab.

Section 3.5.3; Active State Efficiency Requirements: Lines 394 to 413

The TGG WG has provided EPA three documents detailing a proposed approach for setting and validating active state efficiency thresholds using data collected by the SERT tool.

1. Recommendation of Energy Efficiency Metric for ENERGY STAR V3 Draft 1.pdf
2. ENERGY STAR V3 Draft 1 metric recommendation Executive Summary.pdf
3. Deployed Power Analysis Methodology Description.pdf
4. Storage servers final review draft.pdf

Question regarding idle and active being separate or combined: Lines 403-404

Should EPA choose to retain an idle power limit for servers, the idle power limit and the chosen active efficiency threshold should be managed as two separate metrics. EPA needs to assess the interaction of the idle power and active efficiency thresholds and insure that the idle power limit does not disqualify server products with high active efficiency scores. As discussed in the paper “Draft 1 proposed idle power limits final review draft.doc” which the WG has provided to EPA as part of its Draft 1 comments, an idle power limit is not a measure of server efficiency and the idle limit and associated adders, as proposed in Draft 1, will result in the exclusion of servers with high efficiency as rated by the active efficiency metric. Idle power is unrelated to energy consumption in customer’s hands for a deployed set of servers and inclusion of idle into the active efficiency metric will have the tendency to increase the energy consumption of servers in the data center.

Idle Limit and Adder Changes: (line 468)

From v2 to v3 EPA is proposing ~40% reduction in idle base power along with reductions in idle power adders for power supplies, memory devices and storage devices. While there is justification for making reductions in the base idle power and the idle power adders, the TGG WG believes that the overall result is overly aggressive and will result in servers with high active efficiency scores being excluded from ENERGY STAR certification. The WG has provided a detailed analysis of the idle limits based on an assessment of the servers and their associated configurations in ITI database in a separate comment titled “Draft 1 proposed idle power limits final review draft.doc”. There are several key points made in the analysis:

1. Using the recommended geomean efficiency active efficiency metric and identifying the threshold score that would certify 25% of the current database of ENERGY STAR certified servers, 22 of 66 systems that would pass the active efficiency metric threshold would not pass the idle power limit.
2. Low-end configurations (minimum power and low-end performance) have a much higher pass rate for the idle power limit than the typical and high-end configurations.
3. With the reduction of the base idle power limit and the component adders, an adder needs to be provided for the higher performance processors. Processor TDP ranges from 13 W to 130 W in the current database – the higher TDP processors will not be able to meet the idle power limits.

CPU socket power or TDP adder: The WG recommends that an idle power adder be given for specific TDP levels or for compute capacity as measured by the product of core count and frequency. The WG is willing to work with EPA to determine the best indicator and the associated idle adder values to address this concern.

Memory adder: The TGG working group supports the reduction of the memory adder to 0.25 W/GB. The WG also recommends that emerging memory technologies such as the use of Flash or other Non-volatile memories be analyzed. Current technical and power measurement information should be evaluated to determine the idle power characteristics of the emerging memory technologies.

Integrated I/O adder:

Current allowance is up to 10 Gbps which will have to be extended to north of 250 Gbps. Idle power allowance for highest speed need to be more than double of current 10G allowance. Possible range: >10G <=25G, >25G <=50G, >50G <=100G, >100G <=200G, >200G

Four Socket server idle power limits: Lines 388 to 390

As a general note, the TGG WG believes that EPA should not set idle power limits as idle power is a poor measure by which to assess server energy efficiency.

The TGG WG agrees with the EPAs observation that there is insufficient data to set idle power limits for four socket servers. The complexity of the 4 socket server configurations and the limited number of 4 socket server products, for both managed and resilient servers, makes it difficult to set idle power limits that will not be exclusionary and limit product offerings. The limited product range is also likely to make it difficult to set active efficiency thresholds as well.

The TGG WG recommends that no idle limits be established for 4 socket servers, but that idle power data continue to be reported.

GPU and component idle and active power adders: Lines 529 to 546:

APA idle power reduction: For certain applications, significant energy efficiency increases can be gained by use of an APA vs. traditional computation methods. Draft 1 3.10.1 outlines a method to accommodate removable APA cards. However, recent advancement in server designs include an attached APA where the APA die is direct attached to the mother board or integrated into the CPU package (i.e. non-removable from the system). Some of these direct attach solutions have notable additional energy efficiency and idle power improvements versus removable APAs. Servers with direct attached APAs either should be excluded, handled as a separate class of servers with specific idle and active efficiency requirements or they will need to receive:

1. An idle adder to account for the presence of an APA in the system.
2. An active power reduction based on a performance related characteristic of the APA, to account for the effect of the APA power debt with no attendant performance improvement, on the active efficiency metrics. This will get complicated, as companies will have to make computational adjustments to the measured power demand in the SERT test in order to adjust the worklet efficiency scores for the APA power debt.

These adjustments will be necessary to avoid exclusion of these products from Energy Star certification.

Based on the cost and complexity of adding attached APAs to a server, TGG is not concerned that this allocation or exclusion would be utilized as a loop hole.

FPGA devices: We expect to see products that may have FPGA die included in system, either on cards which will attach to a PCI slot or direct attached to the motherboard. For PCI cards, the FPGA device should be addressed in accordance with the 3.10.1 requirements. Where the FPGA die is direct attached, they will need to be handled with the same requirements as a direct attached APA.

Executive Summary: Recommendation for a Single, Combined Server Efficiency Metric

The Green Grid SERT™ Analysis Working Group (WG) has assessed the options for combining the performance and power data generated by the SERT™ tool. Based on the analysis of an extensive data set consisting of SERT data from ENERGY STAR® certified systems and the SPECpower Committee's dataset, the TGG WG and ITI are proposing a recommended single, combined metric for assessing the energy efficiency of server products. In addition, the TGG WG has developed a methodology, the deployed power assessment, which enables validation that the a higher server active efficiency score can result in a lower deployed power when a higher efficiency server or servers is/are installed in the data center. An effective efficiency metric will have a good correlation between the efficiency score and the deployed power; an ineffective metric will not.

The slide deck "Recommendation for energy efficiency metric for ENERGY STAR V3 Draft 1" provides the key reasons for the recommendation, observations made during the data analysis work, an explanation of the criteria used to compare the different assessed metric options, and presentation of the summarized data analysis. Specific data items presented include percent correlation of the rankings of deployed power and efficiency score metrics, the graphs of deployed power for seven power/utilization/workload combinations versus efficiency scores for the selected weightings for the Weighted Geomean metric, the average efficiency scores for the 5 different configuration types reported to ENERGY STAR, and graphical comparison of the deployed power versus weighted geomean efficiency scores for three power/utilization/workload combinations . This paper provides a summary text to support the slide deck.

Recommended Metric:

ITI/TGG is proposing that the ENERGY STAR® adopt a single, combined efficiency metric which uses the geomean to combine the SERT normalized performance and power interval data into individual worklet performance/power efficiency scores, combines the worklet efficiency scores by geomean by component/workload type (CPU, memory, and storage) and then uses a weighted geomean to calculate the combine the component/workload scores into a single metric.

The recommended weighting for the component/workload scores is 60% CPU, 35% memory, and 5% storage. The analysis of server workloads which led to the selection of the CPU to memory weighting of 60% CPU to 40% memory is described in the paper "SERT Metric Weighting Proposal". It was determined, in consultation with the SPECpower committee that it was important to include a 5% weighting of the combined storage worklets. The memory weighting was reduced to 35%, recognizing that the CPU rating needed to be maintained at 60% given the importance of the CPU functionality to server operation. The description of the calculation method and equations to combine the SERT tool performance and power data into a single combined efficiency, power and performance value for a server configuration is provided in the paper "Creating a Single metric from SERT results".

Justification for the Choice of Metric

The WG evaluated over 15 different data combination approaches, reducing those to the 7 options presented in page 6 of the slide deck. The TGG WG completed a final evaluation on the top two of the seven options presented on page 6: Weighted Geomean Efficiency with Geomean intervals (60%, 35%, 5%) and the Weighted Geomean Efficiency with Geomean intervals (65%, 30%, 5%). The evaluation criteria used to assess the efficiency metrics are listed and defined in the table on page 6. The metrics are graded on page 7 of the slide deck with either yes (green check) or no (red x) indicating a pass or fail or high (green battery), medium (yellow battery) or low (red value) rankings indicating relative adherence to the criteria. The criteria assessment illustrate that the Weighted Geomean (60%, 35%, 5%) metric with geomeaned intervals best met the assessment criteria.

Based on the data available to the WG today, we believe the 60/35/5 weighting is the correct weighting for a combined metric. The SPECpower committee is doing additional evaluation on the balance between different CPU workload types across the 12 worklets and an assessment of how best to assess and aggregate memory bandwidth and capacity worklets. The WG will continue to collaborate with the SPECpower committee on these activities and perform additional analysis as the SPECpower committee refines their analysis and proposed potential refinements to the metric. Ultimately, the correlation of the deployed power and active efficiency metric ranking will be the key criteria by which the TGG WG will assess additional metric options and adjust/update its current recommendation for the weighting for the single active efficiency metric.

The TGG WG is available to review the full analysis work that we have done to arrive at our final selection and recommendation of the combined metric. We have done a detailed look at different methods to combine the interval data, aggregation methods and other activities

Deployed Power Assessment Methodology

As the WG worked on its assessment of the different methods to combine the SERT tool data to create a single server energy efficiency metric, it sought to find a means to validate which metric would best represent the ability of a server to do more work with less energy use in a data center or office environment. In order to evaluate potential metrics, the WG determined that a graphical and ranking assessment of the power consumption of a deployment of a group of servers required to execute a defined workload versus the aggregated efficiency score for that server, plotted and ranked for all the servers in the available dataset, offered the most effective means to assess the efficacy of each combined efficiency metric. The assessment is based on determining the ability of a server to deliver a given workload for the minimum expenditure of energy and validating that the better combined efficiency score will result in a lower deployed server power demand to execute that work . The use of a deployed power calculation enables differentiation between the effectiveness of a low performance, low power server and a high performance, high power server, as it enables an assessment of the number of servers and their associated energy use required to deliver a given workload in a datacenter or office environment.

To calculate the number of servers needed to perform a workload and their associated deployed power, it is necessary to select a workload level against which to calculate the number of deployed servers, a workload performance value for the server which represents the ability of the tested configuration to meet the performance goal, and the associated server power use. The detailed explanation of the deployed power evaluation process is provided in the paper “Deployed Power Analysis Methodology Description”.

Executive Summary: Recommendation for a Single, Combined Server Efficiency Metric

1. It was determined that the workload performance level needed to be large enough to avoid quantization effects, so a value of 100 * the maximum performance value of the tested servers was selected.
2. Because servers execute a diverse range of workloads, three assessment workloads were selected: a CPU intensive workload, a memory intensive workload and a weighted workload which mimics the workload weighting of the combined metric being assessed. The maximum weighted geomean performance of a given server configuration is used in determining the number of deployed servers for this assessment, as that is the workload a data center operator would use when selecting a server. The weighted geomean server performance is divided into the workload performance level to determine the number of servers required to execute the work.
3. The deployed power is calculated by multiplying the number of servers required to execute the workload times the server power calculated for 7 different combinations of power levels and workload types - idle, 25%, 50%, and 100% utilization, the weighted power calculated for the workload, the CPU intensive workload and the memory intensive workload. A range of server powers were calculated to validate that the combined metric is balanced and representative of efficiency across the range of workloads that servers are expected to perform. The seven utilization/power/workload combinations are described in detail in the paper “Deployed Power Analysis Methodology Description”.

<u>Deployed Power Assessment and Ranking Analysis:</u>

Page 8 and 9 of the slide deck display tables that provides the correlation between the product rankings, the correlations between the deployed power and server efficiency rankings and the average rank mismatch for each of the deployed power/server efficiency metric plots. The correlation values indicate that the efficiency metric does an excellent job of predicting a lower power consumption in the data center. Higher correlations between the deployed power and efficiency metric rankings indicate that a metric will properly represent server efficiency as measured by a larger workload delivered per unit of energy consumed.

The graphics on pages 10 to 17 show the plots of the various deployed power rankings against the server efficiency ratings and provide the curve fit factors for each of the 4 metric choices.

<u>Idle workload</u>: Deployed idle power has the lowest correlation of all the deployed power analyses. There is a very poor correlation between idle power and the proposed active efficiency metric. The reasons for the poor correlations are discussed in the paper “Draft 1 proposed idle power limits”. The reader is referred to that document; we will not repeat the text here.

As has been discussed in other submittals, lower idle power is captured in the difference between the minimum and maximum power under the SERT test and servers with a smaller dynamic range (min power/max power) will receive a better efficiency score. A smaller dynamic range value indicates a lower power consumption at low utilization or idle and lower power consumption in the data center or office environment. This point reinforces the importance of using the efficiency metric, not idle power, to assess server efficiency.

Evaluation of Low-end and High-end configurations:

The power profile and functional capabilities of a server will be dictated by the quantity and functional characteristics of the CPU, memory and storage components. There is a wide diversity of functional capabilities and power use both within the choice of configurations for a single server product and across the range of server products available in the market.

In order to assess how the configuration type affected the efficiency score, we calculated the geomean, minimum, maximum and standard deviation of the Weighted Geomean efficiency scores for each configuration type (page 18 of the chartset). The data shows that the maximum power and high-end performance (high-end) configurations have significantly higher efficiency scores as compared to the low-end performance and minimum power (low-end) configurations. This is not surprising, as the low-end configurations largely used processors with lower power and performance characteristics and had a lower memory capacity, while the high-end configurations used high performance, high power processors. This point is graphically illustrated in pages 19-21 of the charts set, where the geomean scores are plotted against the weighted deployed power, the CPU intensive workload power and the 25% utilization power, and the high-end configurations and low-end configurations are plotted separately. These graphics illustrate that there are very few high-end configurations with efficiency scores below 20 and very few low-end configurations with efficiency scores above 60.

This assessment of the data illustrates the importance of setting separate thresholds for the low-end and high-end configuration groups. It will also be important to set clear requirements for the component selections and quantities for the low-end and high-end test configurations as described below.

Low-end configuration: Lowest available performance processor, as characterized by lowest product of the core count and CPU nominal frequency, at least the minimum memory capacity required by SERT (SERT Users Guide 1.3.3, page 6), and one HDD drive of the highest rotation speed and largest form factor available for the server. Accommodation will need to be made for servers that do not support HDDs or require a drive for operation.

High-end configuration: Highest performance processor as characterized by the highest product of the core count and CPU nominal frequency, available for the server, three to four times minimum SERT memory capacity requirement, and two SSD drives.

In addition to the two configurations that would be evaluated for conformance to an efficiency level, the typical configuration should also be tested under SERT and the SERT results reported via the .xml report. Because the typical configurations will not be consistently configured, an active efficiency threshold should not be established for this configuration. The current definition of the typical configuration should be maintained.

Trademark reference:

SPEC and its product names SERT and SPECpower are registered trademarks of the Standard Performance Evaluation Corporation (SPEC), see spec.org.

Deployed Power Workload Type Descriptions

The Green Grid SERT™ analysis work group (TGG WG) is recommending that a single server energy efficiency metric can be created using the data collected from the SERT tool. To create the efficiency score, the individual worklet power and normalized performance interval data must be combined using the geomean function and dividing the combined normalized performance by the combined power data to create a worklet efficiency score. The worklet efficiency scores are then combined by component/workload (CPU, memory and storage) group using the geomean function. Finally, the three workload/component values are combined into a single performance/power efficiency score using a weighted geomean. The recommended weighting of the three workload/component groups is 60% for CPU, 35% for memory, and 5% for storage. The basis for this weighting was discussed above. For simplicity, we are recommending that the SERT tool calculate the efficiency metrics using the geomean so that the worklet efficiency scores can be combined for the single efficiency metric.

Overview of the Deployed Power Analysis Methodology for validation of the proposed server efficiency metric

In order to evaluate potential metrics, we determined that a graphical and ranking assessment of the power consumption of a deployed group of servers required to execute a defined workload versus the aggregated efficiency score for that server, plotted and ranked for all the servers in the available dataset, offered the most effective means to assess the efficacy of the combined efficiency metric. This assessment process will be referred to as the "Deployed Power Analysis". The TGG WG selected the Deployed Power approach to demonstrate that the most efficient server as measured by Weighted Geomean Efficiency metric delivers the most workload per unit of power consumed and minimizes the power consumption in the data center and to compare the ranking of different metric options.

The TGG WG strongly believes that the use of the Deployed Power Assessment is an essential step in validating the efficacy of any server energy efficiency metric. Ultimately, the measure of efficiency is the fact that the chosen server results in a lower power consumption in the data center.

The Deployed Power assessment is designed to validate that a better efficiency score will result in a lower energy demand from a set of servers deployed to execute a targeted workload in a data center. The use of a deployed power calculation enables differentiation between the effectiveness of a low performance, low power server and a high performance, high power server, as it enables an assessment of the number of servers and their associated energy use required to deliver a given workload in a datacenter or office environment.

To calculate the number of servers needed to perform a workload and their associated deployed power, it is necessary to select a workload level to use to calculate the number of deployed servers.

1. It was determined that the workload performance level needed to be large enough to avoid quantization effects, so a value of 100 * the maximum performance value of the group of tested servers was selected.
2. Because servers execute a diverse range of workloads, three assessment workloads were selected: a CPU intensive workload, a memory intensive workload and a weighted workload which mimics the workload weighting of the combined metric being assessed. For simplicity, we are using the weighted geomean of each workload type to calculate the number of servers required to deliver the selected work level. Given SERT™ normalization methods and consistency across the worklets, the weighted, normalized performance is directly proportional to the maximum performance level by a factor of 2.5. Therefore, the geomean of the interval levels can be used to create a comparison that is proportional to

a maximum performance assessment. It should be noted that a data center operator would use the maximum server performance to determine the number of deployed servers that would be used in a data center.

3. TGG decided after much consideration on the following power/utilization/workload types to be used in our comparison analysis. The intent is to assess a combination of workload types and power consumption levels that a set of servers will experience in an operating environment.
    1. Idle power as measured by the SERT tool;
    2. Geomean of the power for all workloads at the 25% utilization (light workload)
    3. Geomean of the power for all workloads at the 50% utilization (medium workload)
    4. Geomean of the power for all workloads at the 100% utilization (heavy workload)
    5. Weighted deployed power using the metric weighting
    6. Weighted deployed power using the CPU intensive workload weightings (85% CPU, 15% memory)
    7. Weighted deployed power using the Memory intensive workload weightings (40% CPU, 60% memory)

    Multiple workloads and power use scenarios were assessed to validate that the combined metric is balanced and representative of efficiency across the range of workloads that servers are expected to perform and to avoid the assertion that use of a single power/workload/utilization might be biased to a particular outcome.

Details on creation of these workload types are provided in section on deployed power calculations below.

Determining the number of deployed servers

In order to determine the number of servers required for any given server model we must determine both a performance target for the dataset and a performance capability for each individual server. We combine and use the individual worklet performance values reported in the SERT tool in order to determine the number of servers required to meet the performance target.

Any attempt at a single efficiency metric based upon the SERT tool must use some method to combine the individual worklet efficiency scores to create a single efficiency metric. For any given efficiency metric being evaluated we will use the geomean combinatory method and the designated component (CPU, memory and storage) weightings to calculate the number of servers required and their associated deployed power for the workload scenarios indicated above.

We will use the recommended weighted geomean (CPU 60%/Memory 35%/5% Storage) metric, designated as the weighted geomean metric, as an example to illustrate the process in the following sections.

Establish Target Performance

We first begin with defining a target performance level for the set of deployed servers.

Deployed Power Workload Type Descriptions

We use 100 times the maximum performance of the highest performance server in the data set for the performance target, in order to minimize quantization issues since we will do all our calculations on an integral number of servers.

The number of servers required to meet a desired performance level is calculated according to Equation 1.

Equation 1

$$Deployed\_QTY_n = Roundup(\frac{100*Max(Perf_{All\ Servers})}{Perf_n})$$

Where:

$Perf_{AllServers}$ = the performance values for all servers in the data set

$Perf_n$ = the performance of server n.

We need a performance value for each server, $Perf_{n,}$ which is related to the way the efficiency score is calculated in order to determine the number of deployed servers. For the weighted geomean efficiency method we combined and weighted the performance data as follows:

1. For each worklet, we used the geomean to combine the performance and power interval data for each worklet, normalized the performance scores to a reference system and created individual worklet efficiency scores by dividing the normalized performance and power values.
2. We used the geomean to combine the individual worklets by workload types (CPU, memory, storage). Because we are only using Flood for the memory workload, there was no need to use the geomean.
3. We then combined the geomeaned scores using the weighted geomean, a log function with a multiplier indicating the weighting of the workload types to create the single weighted geomean efficiency score.

See the “mean combination methodology analysis” and the “SERT Metric Analysis” papers for detailed explanation of why the geomean was chosen to combine the data.

We used Equation 2 to combine the geomean of the normalized interval performance scores in order to obtain the performance of each individual server. In addition, peak performance and the combination of the interval performance values are related by a defined ratio (see pages 4 to 6 of the “SERT Metric Analysis” paper), so the analysis will be proportional to normalized maximum performance by a factor of 2.5.

Equation 2

$$Perf_{Wght} = EXP(0.60 * \ln\Big(Geomean\big(Perf_{Compress\,,} Perf_{Crypto}, Perf_{LU}, Perf_{SOR}, Perf_{SORT}, Perf_{SHA256}, Perf_{SSJ}\big)\Big) \\ + 0.35 * \ln\big(Perf_{memory}\big) + 0.05 * \ln\Big(Geomean\big(Perf_{Sequential}, Perf_{Random}\big)\Big))$$

$Perf_{worklet}$ represents the geomean of the normalized interval performance scores from the SERT data base.

$Perf_{Wght}$ represents the newly calculated weighted performance score.

Calculate the Number of Deployed Servers

The weighted performance for each server is divided into the value of 100 * maximum server performance in the dataset to calculate the number of servers required to deliver the designated workload. We have chosen to use the weighted performance to calculate the number of servers for each configuration type because the value is readily available from the data aggregation calculations that are performed to create the efficiency scores and because the nature of the SERT test results in a the geomeaned performance data being a factor of 2.5 larger than the maximum performance. A data center operator will use the maximum performance value to size the number of servers required to support a workload in the data center.

Now that we have a deployed quantity of servers it is necessary to determine what the deployed power of these sets of servers will be while running the different workload types we described above.

We do this by calculating an estimated power for each server while performing the workload types under evaluation.

Deployed Power Calculations

We calculate a power value for each server based upon the workload type under evaluation. We calculate the power demand by combining the interval power data by worklet, where necessary, and then by component type and combining the component values using the weighting of the efficiency metric under evaluation. The following sections describe how this is done for each of the workload types under evaluation. Where we use interval power data, the interval data, the combined power data at the component level, and the component level data will be combined using the geomean. The component level data will be combined with the appropriate weightings.

Deployed Idle

We calculate a deployed idle power for each server by multiplying the number of deployed servers times the idle power of the server as reported by the SERT tool. The idle power value is taken directly from the SERT tool report.

Equation 2

$$DeployedPwr\_idle_n = Deployed\_QTY_n * Pwr\_idle_n$$

Where:

$Deployed\_QTY_n$ = Qty of systems calculated by 100*maximum performance of all systems divided by the performance of server n.

XX% Utilization [25%, 50%, 100%]

For our three load scenarios with 25%, 50%, and 100% utilization levels we combine the individual power values at the utilization level using the geomean for each component types (CPU, Memory, storage) and the combine the component types according to the workload weightings of the efficiency metric under evaluation. Some of the worklets do not have interval data for the desired utilization level and for these we use a linear approximation of the power for the intended utilization using the provided utilization power values. For the weighted geomean metric, we calculate the deployed power according to Equation 3.

Deployed Power Workload Type Descriptions

Equation 3

$$DeployedPwr(tgt\%)_n = 0.6 * Geomean({}_1^{j}CPUworklet(Ut\%)_n) + 0.35 * Geomean({}_1^{j}Memoryworklet(Ut\%)_n) + 0.05 * Geomean({}_1^{j}Storageworklet(Ut\%)_n)$$

Where:

$Geomean({}_1^{j}CPUworklet(Ut\%)_n) =$ Geomean of the j CPU worklet power values at Ut% utilization for server n.

Weighted Deployed Power

For our weighted deployed power scenario we need to calculate a power value that is more general in nature than the idle power or our three utilization scenario power values. We use the worklet type weightings methodologies of the specific metric in question. For each worklet we created a power value that takes into account the variation in power across different utilization levels to reflect the power scaling capabilities of the individual servers. This is done by taking the geomean of the worklet interval power values as the representative power use for the worklet. For the weighted geomean metric we use Equation 4 to calculate the individual server power.

Equation 4

$$Pwr_{Wght} = 0.6 * Geomean\left(Geomean\left({}_1^{j}Pwr_{CPU}\right)\right) + 0.35 * Geomean(Geomean({}_1^{j}Pwr_{Memory})) + 0.05 * Geomean(Geomean({}_1^{j}Pwr_{Storage}))$$

Where:

Pwr_{Wght} is the new weighted power value

$Geomean({}_1^{j}Pwr_{worklet})$ is the geomean of the j worklet interval power values

CPU Intensive Workload

The CPU Intensive workload is used to emulate a very compute intensive server workload type. In this scenario we change the worklet type weightings to 85% CPU, 15% memory and 0% storage.

Equation 5

$$Pwr_{CPUIntensive} = 0.85 * Geomean(Geomean({}_1^{j}Pwr_{CPU})) + 0.15 * Geomean(Geomean({}_1^{j}Pwr_{Memory})) + 0 * Geomean(Geomean({}_1^{j}Pwr_{Storage}))$$

Where:

$Pwr_{CPUIntensive}$ is the weighted power value for the CPU intensive workload

$Geomean({}_1^{j}Pwr_{worklet})$ is the geomean of the j worklet interval power values

Memory Intensive Workload

The memory Intensive workload is used to emulate a workload whose performance is generally limited by the size and or performance of memory in the system. In this scenario we change the worklet type weightings to 40% CPU, 60% memory and 0% storage.

Equation 6

$$Pwr_{Wght} = 0.4 * Geomean\left(Geomean\left({}_{1}^{j}Pwr_{CPU}\right)\right) + 0.6 * Geomean(Geomean({}_{1}^{j}Pwr_{Memory})) + 0 * Geomean(Geomean({}_{1}^{j}Pwr_{Storage}))$$

Where:

Pwr_{Wght} is the new weighted power value

$Geomean({}_{1}^{j}Pwr_{worklet})$ is the geomean of the j worklet interval power values

Conclusion Regarding the Deployed Power Methodology

The assessment of deployed power, as represented by the power required by the number of servers of a given product type and configuration needed to execute a defined workload, offers an effective means to validate server energy efficiency as measured by the energy required to deliver a given workload. The power profile and functional capabilities of a server will be dictated by the quantity and functional characteristics of the CPU, memory and storage components. There is a wide diversity of functional capabilities and power use both within the choice of configurations for a single server product and across the range of server products available in the market. Assessing server efficiency using the deployed power method addresses the fact that a lower power, lower performance server may have a better efficiency or desirable, low power profile but will require more total power to deliver a given workload as compared to a higher performance, higher power server because of the number of units that need to be deployed.

The TGG WG strongly believes that the use of the Deployed Power Assessment is an essential step in validating the efficacy of any server energy efficiency metric. Ultimately, the measure of efficiency is the fact that the chosen server results in a lower power consumption in the data center.

Draft One Proposed Idle power Limits

TGG analyzed the pass rate of systems to the proposed idle power limits using the ITI/TGG data base of 265 ENERGY STAR certified server configurations consisting of 1 and 2 socket managed rack servers. The measured SERT idle power was compared to the allowable idle limit for each configuration as calculated by adding the proposed version 3 base idle limit to the allowed adders for each configuration's redundant power supplies and additional memory capacity, storage devices, and I/O devices as recorded from the SERT data sheets. Table 1 shows the results of that analysis for the entire data base split out by Energy Star 2.0 defined configurations. The idle limit pass rate column shows the pass rate, where the measured idle power was less than the calculated, allowable idle power, as a percentage of the total systems of that configuration type. The Average Weighted Geomean Efficiency column shows the average efficiency score of the same subset of servers (both passing and failing) using the Weighted Geomean Efficiency Metric with Geomean intervals (60% CPU, 35% memory, 5% storage).

Table 1b

	Qty Pass Idle Limit	Qty Fail Idle Limit	Idle Limit Pass Rate %	Average Weighted Geomean Efficiency Score by configuration
High-End Performance	24	22	52.2%	43.7
Low End Performance	30	16	65.2%	20.3
Maximum Power	22	30	42.3%	34.5
Minimum Power	43	11	79.6%	27.5
Typical	39	27	59.1%	36.0
All Configs	158	107	59.6%	

Note: The weighted geomean efficiency and deployed power calculations were done excluding the capacity data. The addition of capacity changes the numbers slightly, but not in a way that changes the conclusions.

Of particular note is that the configuration with the highest pass rate using the idle power limit is the minimum power configuration followed by the low end configuration. The lowest pass rate configuration for the proposed idle power limit is the maximum power configuration followed by the high end configuration. When we compare the pass rates for idle power to the average efficiencies using the TGG metric we see a significant contrast.

The average weighted geomean efficiency score is the lowest for the low end configurations followed by the minimum power configuration. The highest average weighted geomean efficiency occurs with the high end configurations. The analysis of the idle data and the efficiency metric data illustrates the point that systems with higher power processors and larger component counts can have a higher overall efficiency, as measured by the workload delivered per unit of energy consumed, because of the higher performance delivered by the more capable, higher powered processors and components. The improvement in performance greatly exceeds the increase in power demand, resulting in a better efficiency rating than a server with lower performance, lower power processors and components. The TGG proposed weighted geomean efficiency metric shows a high correlation between more efficient servers and lower energy use in the data center according to our deployed power analysis. Dependence on the proposed idle power limit to identify more efficient servers will, in many cases, favor systems that will have a higher energy use when deployed in the data center.

Investigation of the TGG data set shows that 22 of the 66 servers in the top quartile (highest efficiency) of the proposed weighted geomean efficiency score values fail the proposed idle power limit in draft 1. This indicates that the idle limit is set too low, as it is excluding 1/3 of the high efficiency systems as assessed by the weighted efficiency score.

Configuration differences vs idle pass rate

An evaluation of server platforms that have some configurations passing and some configurations failing against the idle limits shows a tendency for servers with the high TDP processors, large memory footprints and high counts of 10 or 15K HDDs in the configurations that fail idle power limits. Servers with a large numbers of SSDs, as compared to HDDs, have a tendency to pass. Servers with a medium number (4 to 6) of 10 K or 15 K HDD drives are more likely to fail than a comparably configured system with lower (1-2) HDD counts.

A comparison was also done between the idle power pass rates for high-end performance and maximum power configurations and the low-end performance and minimum power configurations the pass rate against the top 25th percentile of the Weighted Geomean Efficiency score for high-end and low-end configurations. Provide data analysis.

This evaluation highlights several key issues with the proposed adder values/methodology proposed in Draft 1.

1. There needs to be idle power adders for processors based on the compute capacity as defined by the product of the core count and the nominal CPU frequency, CPU performance, or the processor TDP. Based on the data in the ITI dataset, processor power consumption in managed servers varies from 13 watts for some lowest performance processors to 160 W for high performance processors with high core counts and nominal frequency. Under the version 2 base idle power and component adder allowances, the differences in processor power were covered; this is no longer true with the more restrictive base limits and adder values. In order to fairly compensate for the higher idle power in higher performing process, a processor adder must be defined in ENERGY STAR V3.
2. An evaluation of DC power consumption specifications for Enterprise HDD’s from 2 primary manufacturers’ shows only 2 of the 28 drives evaluated have DC idle power specifications below the 4 watt AC power adder proposed in draft 1. This means that virtually no data center specific 3.5 in. HDD’s can be added in any significant number without the system failing the idle power limit. EPA should increase Idle power allocation should be provided for 3.5 in, xxK+ RPM drives. This addition is particularly important where high capacity drives are needed for an application.

Correlations of system parameters to idle power

Correlations were run on system configuration elements to the idle power of the managed 1 & 2 socket systems in the ITI data set. Table 2 shows the correlations to CPU thermal design power, CPU Capacity Memory amount and rotating media disk drive amount. All these show correlations above 50%, indicating that higher performance processors, as assessed by core count and nominal frequency, and higher component counts often drive higher idle power.

Table 2: Correlation of server component characteristics and idle power

CPU TDP	CPU Capacity	Memory Amount (MB)	Disk Drive Count
0.5755	0.4306	0.62045	0.5159

Notes:

1. Disk drive count includes 2.5in and 3.5in of all platter speeds
2. CPU TDP not available on 9 configurations in the data base.
3. CPU Capacity calculated by number of sockets times the number of cores times the core base frequency

Correlation of Idle Power to Performance and Efficiency Scores:

Analysis of the data also shows that a higher idle power has a similar level of correlation to performance scores. In previous studies correlations were evaluated between idle power and efficiency scores and the correlation was not seen to be significant.[1] The data is depicted in figure 1 below. Of particular note is that for CPU worklets (Compress, Crypto AES and LU) we see significant negative correlation of worklet efficiency scores (the efficiency scores discussed in this section are the efficiency scores calculated and reported by SERT V1.0 and V1.1) to idle power. This would indicate that for these worklets, lower idle power corresponds to decreased active energy efficiency. This is also true of the hybrid SSJ worklet which is the intended overall system performance worklet within SERT. There is significant positive correlation to idle power only for the memory efficiency scores which include a memory capacity scaling factor. Higher memory capacity often drives higher idle power for the systems due to a significant increase in the number of memory DIMMs.

The performance scores, on the other hand, have higher correlation to idle power than the efficiency scores in all cases. This indicates that idle power is a better predictor of performance, as opposed to efficiency, and that lower idle power generally follows lower performance. The correlation and trend is consistent with technology and computer architecture expectations, as higher power processors within a family have higher levels of performance which outweighs the additional power consumption, improving efficiency.

This analysis indicates that idle power is a reasonable predictor of performance, as opposed to efficiency, and that lower idle power generally follows lower performance. The correlation and trend is consistent with technology and computer architecture expectations, as higher power processors within a family have higher levels of performance and functionality which outweighs the impact of the additional power consumption improving efficiency.

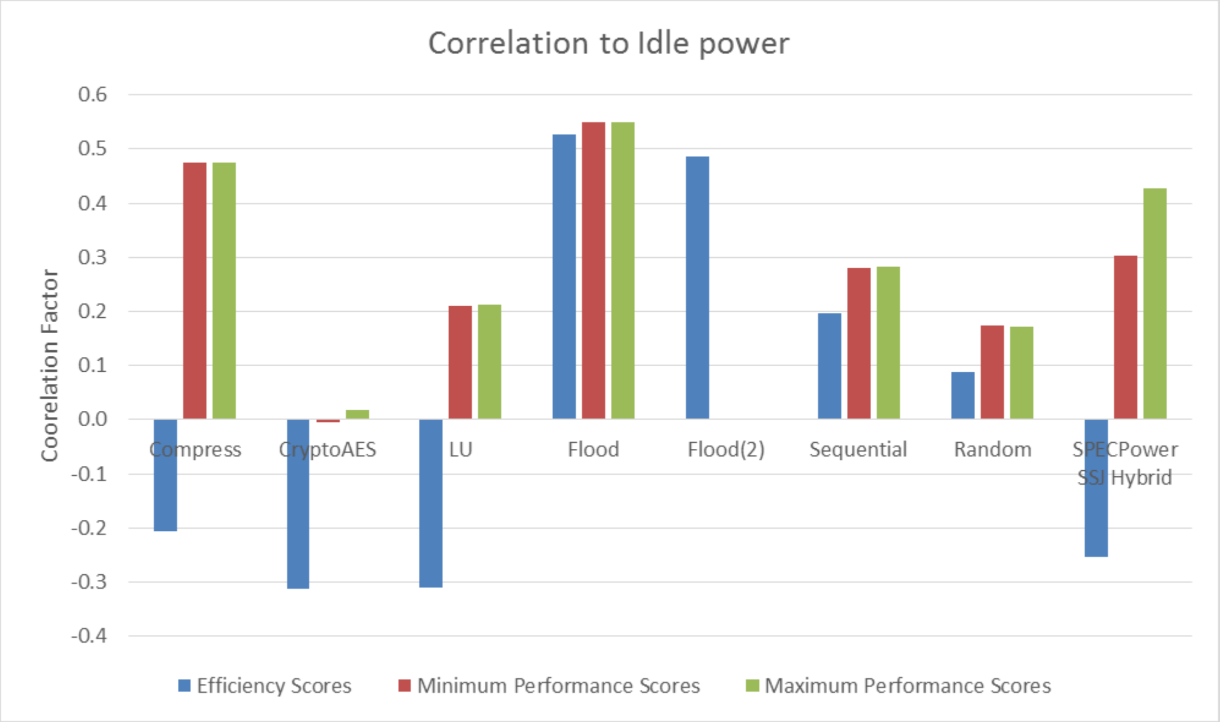

Figure 1: Correlation of Idle Power to Worklet Efficiency and Performance Scores

Given the above correlations and evaluation of the ITI data set, the setting of aggressive idle power limits will need to factor in CPU TDP and/or performance capacity in order not to favor lower power, lower performance processors/systems. Low power, low performance processors generally result in lower efficiency servers and higher energy use in the data center. Further analysis will be needed to determine if the memory adder is sufficient after adjustments are made for HDD's and CPU's.

Reduction of idle power on a given server can be a useful tool to reduce energy consumption of the server when utilization is not high during actual use in the data center. Fixed Idle power limits are not an effective method of comparing one server to another in order to determine which will result in lower energy use in the data center.

TGG has developed a deployed power analysis methodology which highlights the ineffectiveness of idle power as a predictor of energy use in the data center. The deployed power methodology is discussed in detail in the document "Deployed Power Analysis Methodology Description". Using this analysis, a correlation of idle power to deployed power was run on the ITI data set of 1 and 2 processor socket rack servers. This correlation was then compared to the correlation of TGG proposed active efficiency metric to deployed power. The correlation of idle power to the deployed power is 7.76% and the correlation between TGG Efficiency metric and deployed power is -58.2%. An efficiency metric should have a strong negative correlation to deployed power as higher efficiencies should result in lower energy use in the data center. The very low (to positive) correlation of the deployed power to server idle power demonstrates that idle power is a poor indicator of power consumption when those servers are deployed in the data center, and may actually correspond to increased energy consumption in the data center. The analysis shows that the TGG proposed active efficiency metric is significantly better at predicting energy use in the data center than idle power. The analysis also shows a fixed idle power limit would conflict with energy efficiency goals in the data center. TGG continues to recommend removal of the fixed idle power limits or at minimum scale idle power to the (measured and non-measured) performance capabilities of the server.

Table 3: Examples of two systems: one that passes the idle limits with a low Efficiency score and one that fails idle power but has a high Efficiency Score

	Family	Configuration	Idle Limit Pass/Fail	Weighted Geomean Efficiency	Deployed Power (W)
Lowest Quartile TGG Eff. With majority passing idle limit	a	High-End Performance	Pass	11.62	951,638.3
	a	Low End Performance	Pass	14.60	757,336.3
	a	Maximum Power	Fail	11.24	983,746.6
	a	Minimum Power	Pass	11.75	941,221.6
	a	Typical	Fail	12.15	910,296.3
Highest quartile TGG Efficiency all Fail idle limit	cef	High-End Performance	Fail	96.48	114,629.9
	cef	Low End Performance	Fail	82.90	133,916.8
	cef	Maximum Power	Fail	81.66	135,388.3
	cef	Minimum Power	Fail	32.00	345,673.5
	cef	Typical	Fail	64.93	170,796.2

Note: The weighted geomean efficiency and deployed power calculations were done excluding the capacity data. The addition of capacity changes the numbers slightly, but not in a way that changes the conclusions.

To further illustrate the issues with the currently proposed idle power limits an analysis was run on two systems at opposite ends of the TGG efficiency spectrum as shown in Table 3. Family "a" is in the lowest quartile of TGG efficiency and has a majority of configurations passing the proposed idle limit while family "cef" is in the highest quartile of TGG efficiency metric and has all configurations failing against the proposed idle limit. Of particular note is that the average deployed power of the failing configurations in family "cef" is ~ 180,000 watts while the average deployed power of the passing systems in family "a" is ~ 883,000 watts. Systems passing the idle limit have about 4.9 times the deployed power of systems failing when comparing these two families.

Storage Servers

The ENERGY STAR® Program was developed to measure and compare the energy efficiency of hardware appliances. While this was initially focused on household and commercial appliances, the philosophy was extended to office and data center equipment. The products under test were expected to be capable of performing their primary functions, as configured and shipped, without requiring additional software that did not ship with the unit.

Data center IT equipment has, historically, followed the appliance model. Servers, storage systems and switches were shipped to customers in a "ready to run" configuration, just as PCs, laptops, monitors, etc. were for earlier specs. While you might need a discrete second product to drive a workload, the identity of that product was irrelevant to the testing in question.

The last few years have seen the emergence of "software-defined" offerings that address what were previously managed as appliance functions. For example, a software-defined storage (SDS) product allows a data center to purchase software that runs on a server from any server vendor and achieve similar features and functions to those they would obtain from a storage appliance. The concept is also being used to address switching needs. This phenomenon is what has given rise to the "storage server" product offering that has been discussed in in analyses of the ENERGY STAR data, and is also driving a growing "I/O server" segment where a purchaser may configure unusually large numbers of Ethernet connections to provide switching functions (software-defined networking or SDN) or storage network connectivity.

Storage servers and the more general case of servers configured to provide shared I/O functions in place of dedicated appliances are really a special case of server usage, and should not be addressed in either appliance specifications or within the requirements for traditional servers that run application workloads. The software-defined storage or network functionality runs as just another workload on a general purpose data center compute appliance (aka server), and rarely runs as the only function executing on that server. These environments do have an unusually heavy or skewed I/O or storage device configuration and as such, a radically different performance/power profile than one finds on a server configured to run a defined software application/workload. Overall, they are still functioning as servers running additional non- storage or non-switching applications, but they have a different configuration as compared to a typical server with balanced CPU, memory, storage and I/O components. Many I/O server configurations most closely resemble blade server configurations, rather than other I/O appliances.

This means that there is no I/O appliance to test. Any given SDx (software defined network or storage) software product can run on a wide range of servers and configurations, independently of the efficiency of the hardware components hosting it. The only appliance that can be tested is the server. However, the server is configured in a distinct way with processors and memory size to best manage I/O centric activities and at least 12 storage devices or 12 I/O ports and/or 4 or more 10 Gb I/O ports. In addition, the server will typically not ship with any defined software to manage the storage or network functionality. Our recommendation is that I/O servers be addressed in the server program as a distinct server category with a defined set of configuration requirements and a category specific weighting of the SERT component/worklet group's efficiency scores. Additional analysis work needs to be done to determine the best weighting of the component/worklet scores, and whether the existing storage worklets are appropriate for this subcategory. The TGG SERT analysis and Emerald Analysis working groups believe that the assessment of the appropriate metric for storage/I/O servers can be completed within the timeline EPA envisions for the completion of the ENERGY STAR V3 server requirements development.

Mean Calculation Methodology Analysis

Based on comments received from our discussions with Intertek regarding the choice of the geomean to combine worklet efficiency scores, we investigated the method of mean calculation that should use when combining interval data for determining a weighted performance and/or power value for a server to use in our weighted power analysis scenarios. We evaluated three means calculation methods: average, geomean and harmonic mean. The average tends to favor larger numbers in a data set and Harmonic means tend to favor smaller values in a data set. The Geomean always generates an aggregated value between the average and harmonic means when calculating a mean of a set of values.

We took the set of rack servers in the overall combined ITI and SPECpower SERT™ database and used each of the above methods to calculate the combined metrics and compared the results.

For each server we took the power and performance interval data and calculated an interval efficiency. We then used each of the three mean calculation methodologies to calculate an efficiency from the interval efficiency values per Equation 1 (EFF_{base}) and a second efficiency value based upon the means of the interval performance and power values according to Equation 2 (Eff_2).

Equation 1

$$Eff_{base} = Mean(Eff_{interval})$$

$$\text{Where } Eff_{interval} = \frac{Perf_{interval}}{Pwr_{interval}}$$

Equation 2

$$Eff_2 = \frac{Mean(Perf_{interval})}{Mean(Pwr_{interval})}$$

We then compared the two efficiency scores to see if there was a difference. We calculated an error for each mean method by Equation 3.

Equation 3

$$\%Error = \frac{Eff_2 - Eff_{base}}{Eff_{base}}$$

For all cases evaluated using the geomean as the mean type in Equation 2 yielded the exact same value as that achieved using Equation 1. Using either an average or harmonic mean as the mean method resulted in different values using Equation 1 and Equation 2. The averaging method calculates a larger value using the ratio of performance and power means than it does for the mean of the interval efficiencies. The harmonic mean calculates a lower value using the ratio of performance and power means (Equation 2) than it does for the mean of the interval efficiencies (Equation 1). The size of the error between the two methods varied and seemed to be proportional to the magnitude of the difference between minimum and maximum power measurement (dynamic range) values for the worklets. We chose to use calculate the dynamic range from the maximum to the minimum active power, as many servers have a deep idle state that results in an idle power lower than would be expected if you extended the active power curve to an "active" idle measurement. Using the range of the active power measurements provides a better representation of the active range of the servers.

Mean Calculation Methodology Analysis

Based on the observation that the size of the difference between the combined efficiencies, the calculated %Error was plotted versus the dynamic range as calculated in Equation 44 below.

Equation 4

$$Dynamic\ Range = \frac{(Pwr_{100\%} - Pwr_{\min active})}{Pwr_{100\%}}$$

Figure 1 shows the plot of error in the harmonic means calculation versus the dynamic range and Figure 2 shows the plot of average means calculation versus the dynamic range.

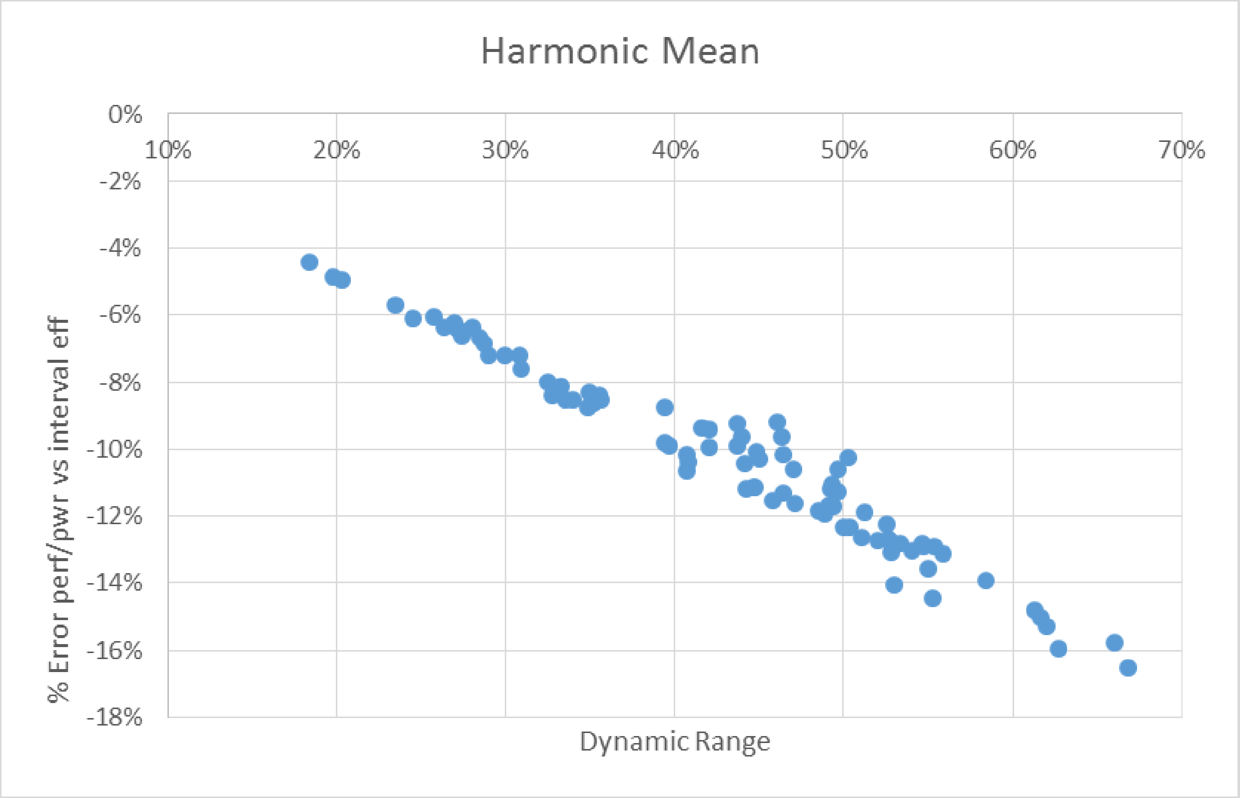

Figure 1

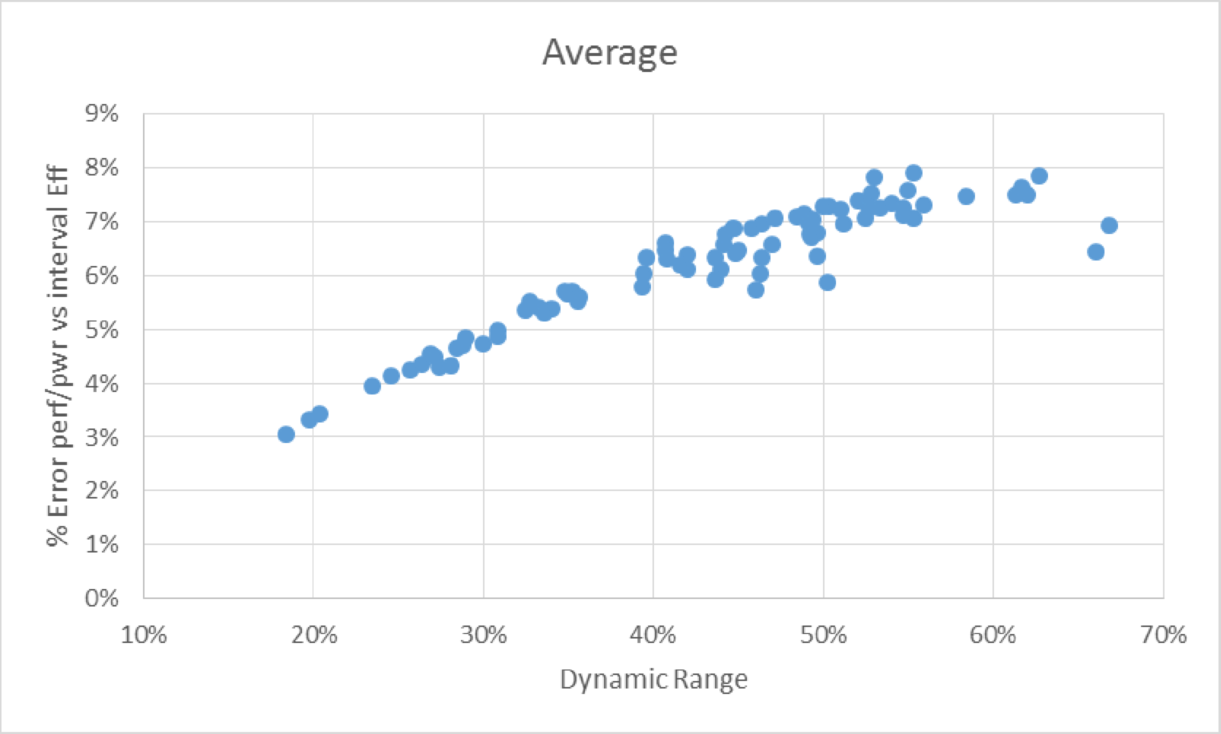

Figure 2

Conclusion

We have chosen to use the geomean function to combine the performance and power interval data. The use of the geomean function to combine performance data, which is the standard SPEC approach, is also supported by the document "[How not to lie with statistics: the correct way to summarize benchmark results]" which indicates that the geomean is the best method for combining performance measurements. Using the geomean prevents any single performance score from unduly influencing the combined metric. It is also appropriate to use the Geomean to combine either SERT worklet efficiency scores or interval power and performance measurement data, as it delivers the same overall efficiency results regarding the relative values of the final combined efficiency score.

The above analysis highlights mathematical properties of means calculation methods. Of the three means methods considered only the geomean provides the same answer when dealing with ratios (i.e. the mean of a ratio = the ratio of means). As the server efficiency metric will inherently be a ratio of performance to power, the geomean is the only method that will provide a consistent efficiency value when combining either ratio metric efficiency values or performance and power values.

<u>Trademark References</u>:

SPEC and its product names SERT and SPECpower are registered trademarks of the Standard Performance Evaluation Corporation (SPEC), see spec.org.