Comments on Energy Star for Computer Servers Draft 2

General comments:
Overall the draft specification seems to be an effective approach given the current, rapidly developing state of understanding, development of metrics and technology.

Simple Labelling;
We need to be quite careful with a program that proposes simple labelling of the form, 'compliant' / 'not compliant' for items as complex and variable as compute servers. The 'efficiency' of a compute server depends upon multiple factors, not all of which are characteristics of the server under consideration. External factors can greatly influence the achieved efficiency of the server, primarily these are the workload and the environment in which it is housed (data centre or comms room). This creates a, potentially significant difference between the device's design efficiency and the achieved efficiency within the system.

Workload;
Considered as a simple value of 0%-100% the workload that the server operates under needs to be considered in conjunction with the server's workload to power draw characteristics. A server with very high 'performance' per watt at 100% load might be very good in an HPC environment but if it has a relatively flat load to power function and thus a relatively high power draw at low workloads it may well be less efficient at lower workloads than a server with a lower 'headline' efficiency but more pronounced load to power relationship. Therefore we need to consider the target workload before selecting the server. This is why the multiple workload points in the SPECPower graphs are so useful.

Environment;
We also need to consider the environment in which the server is operated before we can understand the achieved efficiency. Data centres have some common characteristics, primarily they all have some fixed overhead which is independent of the IT electrical load and some variable overhead which responds to the IT electrical load. Other factors such as external temperature are peripheral to this discussion.
The fixed overhead presents the same type of issue at the facility level that the idle power presents at the server level. As servers (and other devices) are installed into a data centre power and cooling is allocated for these devices and thus cannot be used elsewhere. This used to be based on the 'nameplate' or power supply rating but the industry is now moving toward the 'as configured' power rating of the device. It is important to recognise that each device that has power and cooling provisioned to it should also be allocated a share of the fixed overhead energy use of the facility.
The variable overheads of the facility are applied to the power actually drawn by the server, both idle and workload power. This can be approximated for many facilities with a (facility specific) multiplier value.

Configuration;

Each model of server from a manufacturer has many thousands of possible combinations in which it can be ordered. The relative energy performance of two comparable devices from two manufacturers can vary substantially as these configuration options are changed and within the lifetime of the product. The DRAM is a known area of substantial variation.

Also, there are several references in the document to the 'Maximum, Minimum and Typical Configurations', how are these to be determined and how will we ensure that these are fair representations of each product and between products?

Power Benchmarks;
The SPECPower benchmark, whilst being an excellent start, uses only the SPEC Java workload whilst the industry has a much broader range of benchmarks (SPEC have integer, floating point, matrix etc). The reason for multiple benchmarks is that we do not consider the performance or capacity of a compute server as a single value. There are a number of characteristics that define the performance under a given workload each of which impacts the quantity of installed hardware, such as processors or memory and the power drawn by the server under load.

Extending Benchmark Power Reporting
To allow operators to make effective choices we would like to see a power reporting protocol such as that in SPECPower applied to a range of other benchmarks that are commonly used to assess and scale systems. This would allow each manufacturer to continue to select an appropriate set of internal components to achieve their best score on each benchmark but with the addition of energy reporting to the broader range of benchmarks in a common format.

An adoption of a power measurement and reporting protocol by a body such as the EPA may assist in driving this change.

Extended Labelling;
Given these issues we believe that there are a set of base requirement that can be usefully set for energy star compliance in compute servers but that it is not appropriate to try and rank servers for their overall energy efficiency as this is too dependent on specific configuration options and the context in which they are deployed.

To assist non specialist operators in selection of appropriate, energy efficient servers it may be appropriate to develop a set of 'standard' workloads and environments chosen to represent common deployment scenarios and provide some sort of ranking or scoring system.

For specialist operators with significant estates of ICT equipment who have the available skills we believe that it would be more effective to ensure that the available information on the servers energy characteristics was made available in a common format. This would include a machine readable format to allow asset management software to automatically obtain this information for large and existing estates.

Comments on the draft document;

1) Definitions, no comment, these seem like a reasonable compromise in what is a complex area with soft and moving boundaries.

2) Qualifying products
The elimination of server appliances also seems reasonable.

3) Efficiency requirements.
I am not sure that I understand the proposal to use net power loss. Presumably this would be net power loss at a rated load point?

10% Loading;
We think the important part of the discussion here is that we understand the efficiency of the equipment at any load point. This is where the first issue arises in the 'compliant' / 'not compliant' labelling of servers against variable workloads. If some servers are exempted from the 10% loading this will be evident in the supplied data and they can be marked as unsuitable for low workload conditions.

Multi Voltage;
We in the EU would, of course, be happy with 230V testing. As Energy Star has global impact we think it would be useful for all Energy Star tests to consider the 230V used in much of the rest of the world. There is also consideration of running data centres in the US at 230V to avoid certain inherent inefficiencies in the way 208V is derived and distributed.

Fan Power;
For the tier 2 specification it is worth considering testing the systems under a range of inlet air temperatures as fan power can vary substantially.

Exemptions for redundant supplies;
Where systems with redundant supplies show lower efficiency it is important that this is clearly communicated to the purchaser. Where redundant supplies are an option both efficiencies should be clearly shown on specification data to inform purchasers. Redundant power supplies are frequently used where they are simply not necessary due to resilience elsewhere in the system, such as at the network level.

B Idle Power;
Available Power Management Features;
We agree that these are highly specific, they will also create a terminology problem and cloud the issue for purchasers, they are not a replacement for overall power consumption measurement. Further, WE would suggest that there should be no difference between the enabled power management features for testing and those that are enabled by default in the shipped product and accompanying management software. Efficiency is about performance against energy use and some of the power management features can substantially impact performance.

Categorisation;
We believe that attempting to categorise systems based on their configuration will result in two significant issues. The first is that the amount of installed memory rises quickly and these categories will have to change frequently. The second is that it opens the door to further extensions and subdivisions of the categorisation scheme by, individually, reasonable and fair objections.
Reporting idle power as a fraction of maximum seems to be the most practical choice at the moment. We agree that a suitable stress test will need to be selected as the SPECPower Java benchmark only stresses a subset of components.
There are inevitable compromises whichever way is chosen to achieve this goal but part of the program is to obtain performance per watt results for servers and these will clearly identify those whose overall performance is not poor. This information will lead purchasers to other systems unless there is a specific reason for this higher power use.

Standardised Data Measurement;
Most systems currently available report instantaneous power draw when polled which is of low value. We suggest that systems which report their energy use as by using a counter instead of a gauge approach (miles instead of MPH) be awarded some benefit in assessment.
Equally credit should only be given to systems that can report this data in an open and royalty free standard.

Appendix A
We would like to see a more effective and standardised approach in the short term to labelling the ‘as configured’ power for a server. The practice of provisioning to the ‘nameplate’ rating has driven substantial cost and energy inefficiencies but many operators do not fully understand how to provision to the power the server can draw in the configuration purchased.

Links to vendor supplied power calculators should be superseded in tier 2 by a standard system of both human and machine readable data ‘ labels’. Many of the vendors have worked to provide significant information to their customers but this is hampered by the lack of consistency and difficulty of automated access.

Only machine readable data formats will allow operators to populate their asset and management databases with the information about their estate and this is an essential step towards real energy management in the data centre. Power and cooling provisioning is frequently managed through physical asset management software and this will require such data under ‘as configured’ provisioning as operators move away from ‘nameplate’ as specified in the EU Code of Conduct.