Server metrics

Anson Wu

Intertek PLC

Ecodesign Technical Assistance Study on Standards for Enterprise Servers and Data Storage (Lot 9)

Goals for server metric development

Metric objective: to provide an indicator of the energy efficiency and energy consumption of a particular model and configuration under 'normal' use conditions.

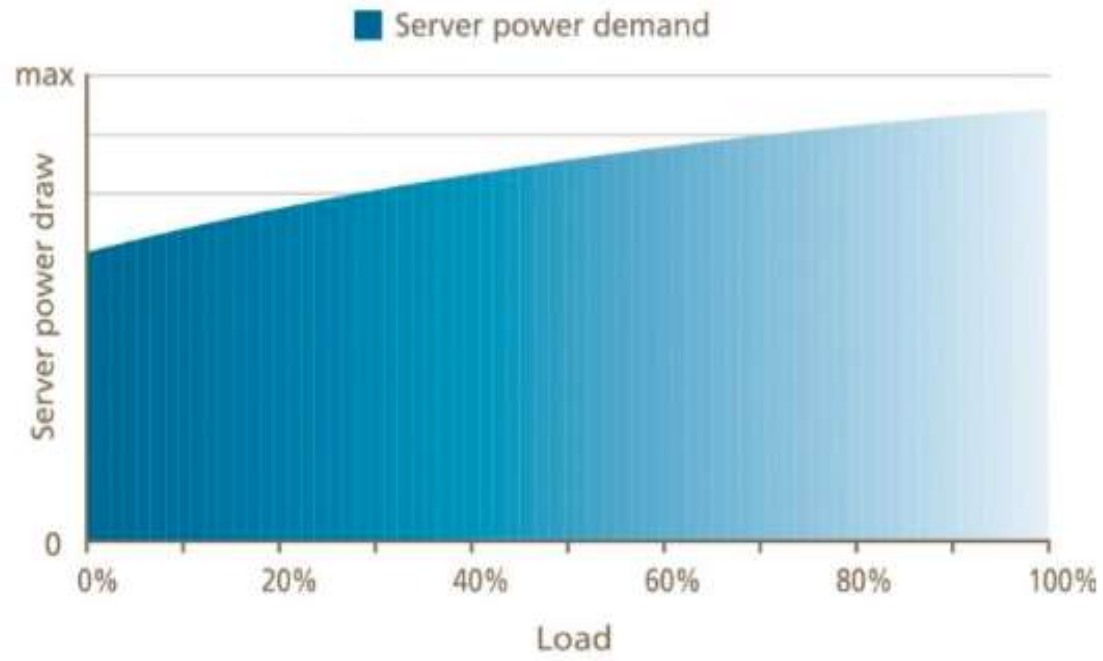
- Focus upon the maximum potential for savings.
  - Avoidance of weighting towards max utilization
  - Means to take into account low utilization
  - Means to take into account idle power overhead
- Technology neutrality
- Interoperability
- Scalability (expandability, redundancy, system level scalability)
- Avoidance of negative market influence
- Defining product to test / product families

Importance of SME market

- Still represents very large market sector ~50%
- Least efficient deployment in terms of PUE
- Least efficient in terms of utilization
- Purchase lower performance servers which are less efficient
- Least capacity to improve efficiency
- Hardest to engage – too many individual stakeholders compared to large organisations
- Still many obstacles to migrating to cloud

Simple metric which can be easily interpreted has most potential

Importance of idle and low loads in metric



Idle and low load power consumption is high proportion of max power

- Long idle periods possible at night, including geolocated cloud services
- 100% (CPU) utilisation is mainly for bursts
- Utilisation 25-75% level

SERT Design & Operation

Intertek
Valued Quality. Delivered.

Server efficiency rating tool (SERT) is a tool to measure server energy efficiency.

Developed and licensed by SPEC for use

Uses a set of **worklets** to test different components – CPU, Memory and storage.

- Worklets = transaction based software simulations.
- Performance = throughput, (i.e. number transactions per second)
- Worklets scale automatically with the server configuration/components.
- Power and performance tested at different load levels

Worklets are grouped into **workloads** based on subsystem/components tested

Summary of worklets

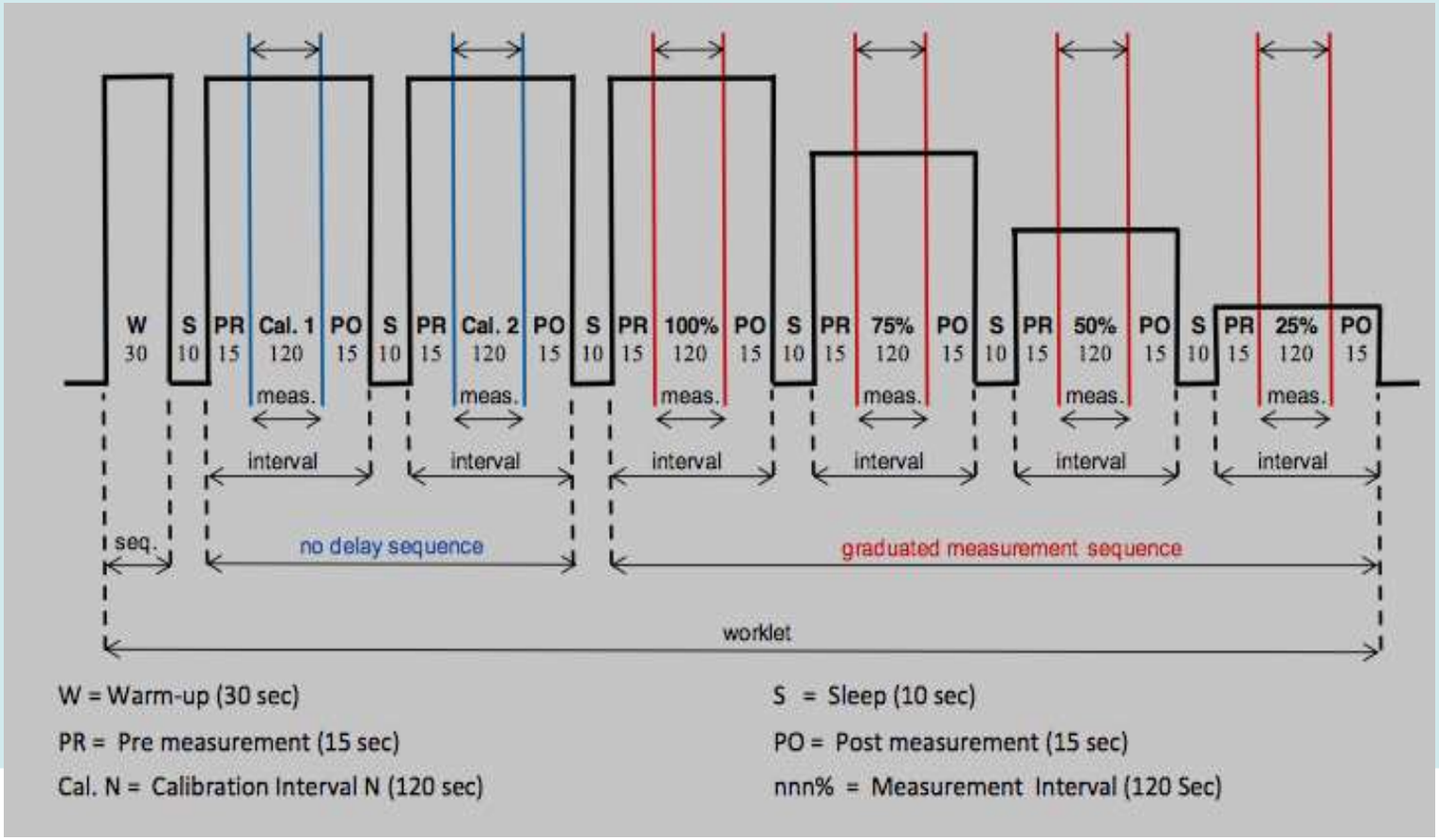
Intertek
Valued Quality. Delivered.

Workload	Load Level	Worklet Name
CPU	100%, 75%, 50%, 25%	Compress
		CryptoAES
		LU
		SHA256
		SOR
		SORT
		XMLValidate
Memory	Flood: Full, Half	Flood
	Capacity: 4GB, 8GB, 16GB, 128GB, 256GB, 512GB, 1024GB	Capacity
Storage	100%, 50%	Random
		Sequential
Hybrid	100%, 87.5%, 75%, 62.5%, 50%, 37.5%,25%, 12.5%	SSJ
Idle	idle	Idle

Summary of worklets

Intertek
Valued Quality. Delivered.

Worklets tested at different load levels



Worklet testing and reporting

Intertek
Valued Quality. Delivered.

Reports performance (throughput) and power consumption at **each load level** for **each worklet.**

Normalises the performance score against baseline server.

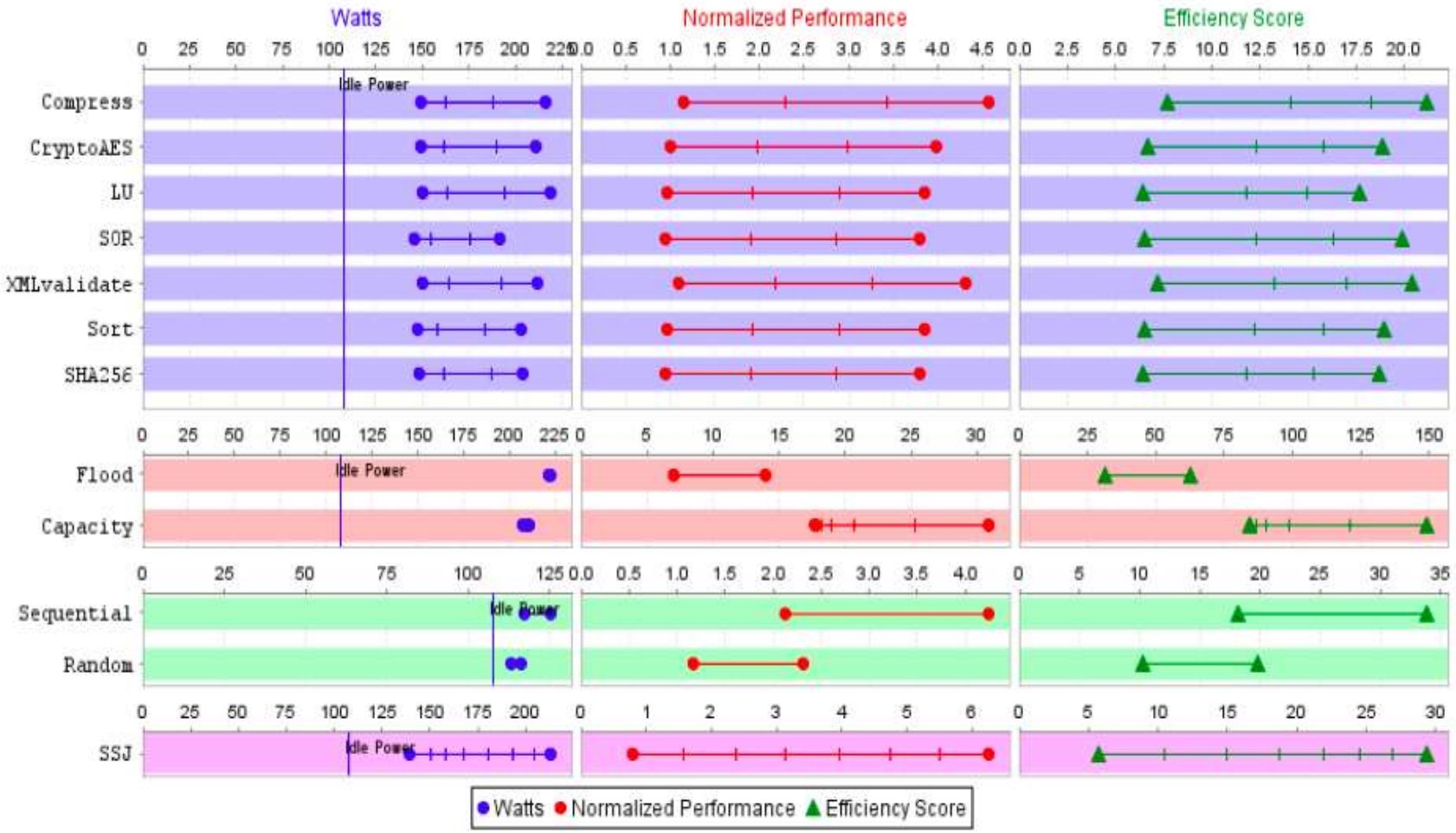
Calculates efficiency score for each worklet:

$$\frac{\sum performance\_score_{load\_level}}{\sum power\ consumption_{load\_level}}$$

The efficiency score is measured in transactions/Joule.

Report output

Intertek
Valued Quality. Delivered.



Scope and scalability

Intertek
Valued Quality. Delivered.

System performance should scale in proportion to the system configuration.

- Sockets: 1-4
- CPU architectures: x86, ARM, POWER
- Form factors: Rack 1U-4U, Tower, blade
- Resilient servers

Not applicable to:

- Mainframes
- Compute clusters (multiple servers)
- Server appliances
- Some expansion cards GPUs, networking, FPGAs,

Scalability

System and network

- SERT not designed to scale to multiple servers
- However, key issue is the comparison between one large high performance server compared to many low performance servers

Ambient temperature

- SERT is designed to run through at one controlled temperature. It is expected that manufacturers will be choosing the lowest temperature
- However, operators are interested how power varies with ambient temperature to optimise data centre power consumption and PUE.

Environmental testing conditions

The SERT tool must be run within constrained environmental conditions:

- Ambient temperature lower limit: 20°C
- Ambient temperature upper limit: within operating specification of the SUT.
- Elevation and Humidity: within documented operating specification of the SUT.
- No overt direction of air flow inconsistent with normal data centre practices.
- AC power supply (single or 3-phase).
- Not compatible with low voltage and 48V DC power supply servers).

Inlet temperature affects fan speed and energy consumption. Servers will be tested as close to the lower limit as possible as this is where they perform most efficiently.

Worklet Power and performance analysis

Intertek
Valued Quality. Delivered.

CPU

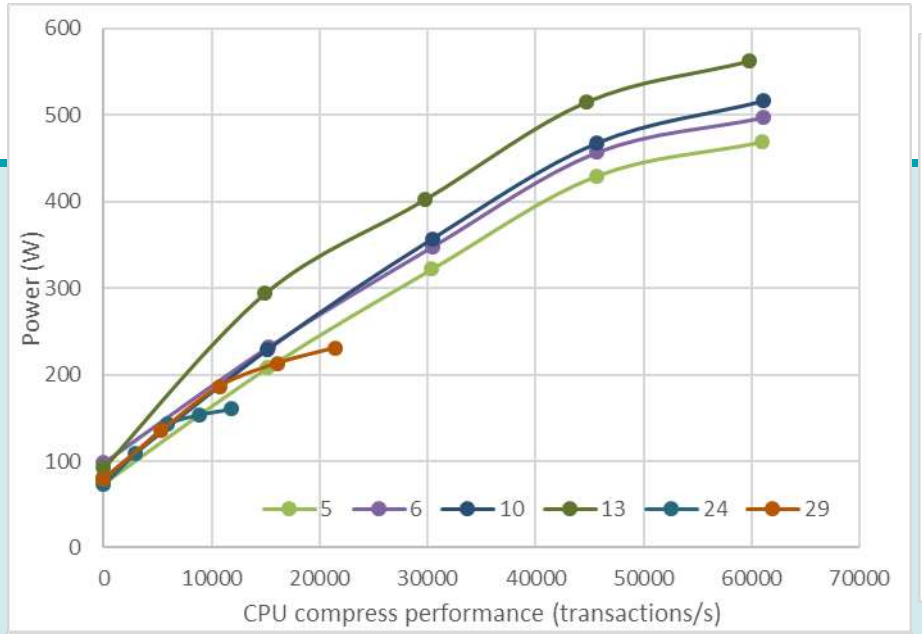
- Performance and power scale, minor impacts of RAM and storage due to power consumption
- Smaller CPUs more efficient at lower performance, larger CPUs at high performance
- Power consumed per unit performance is high relative to RAM, storage

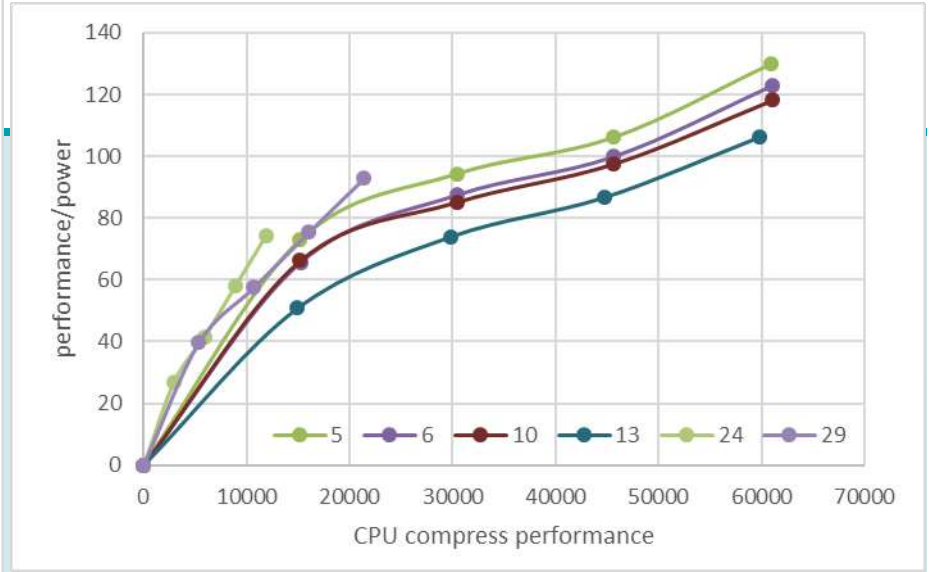
Hybrid

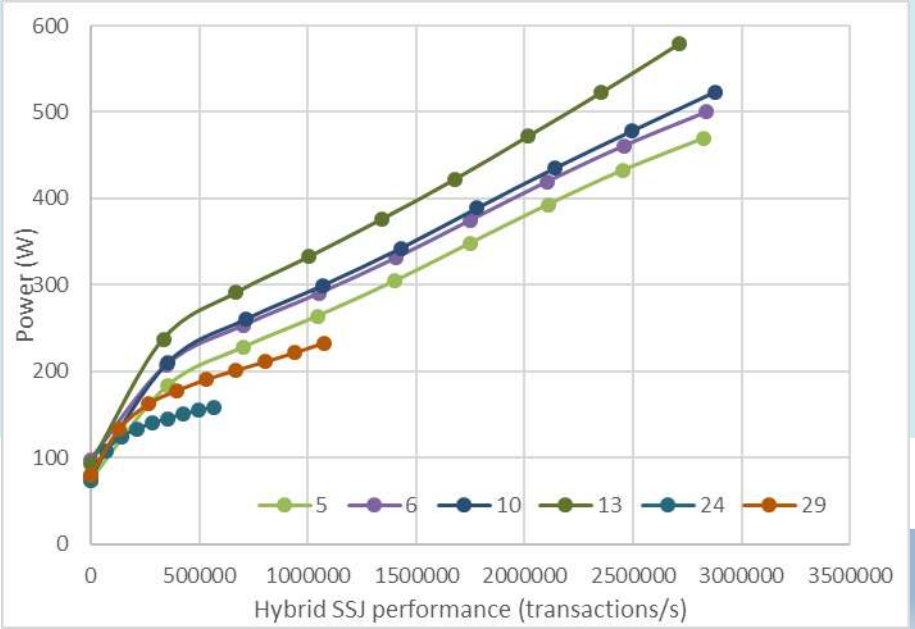
- Performance and power scale
- Very little performance influence by RAM.
- Smaller CPUs more efficient at lower performance, larger CPUs at high performance - more variation in power between CPUs

Compress and Hybrid SSJ

Intertek
Valued Quality. Delivered.







Sample number	CPU		Memory		Storage
	Number of CPUs x cores	Frequency (GHz)	DDR4 Modules / dimms (MB)	RAM (GB)	Number o HDDs
5	2 x 18	2300	8	64	1
6	2 x 18	2300	8	64	8
10	2 x 18	2300	16	256	1
13	2 x 18	2300	16	1024	1
24	2 x 6	1600	16	256	1
29	2 x 6	2400	16	256	1

Worklet Power and performance analysis (2)

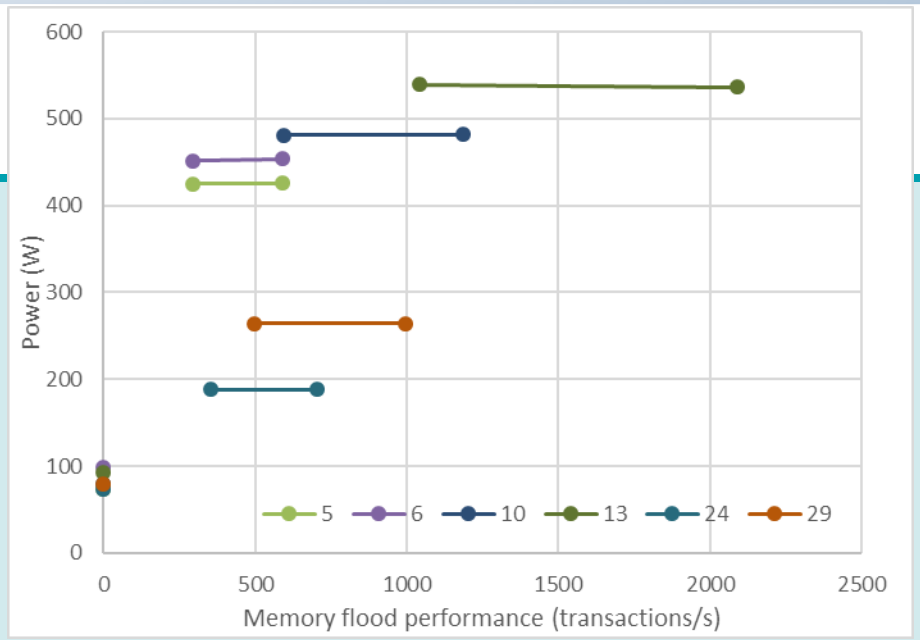
RAM

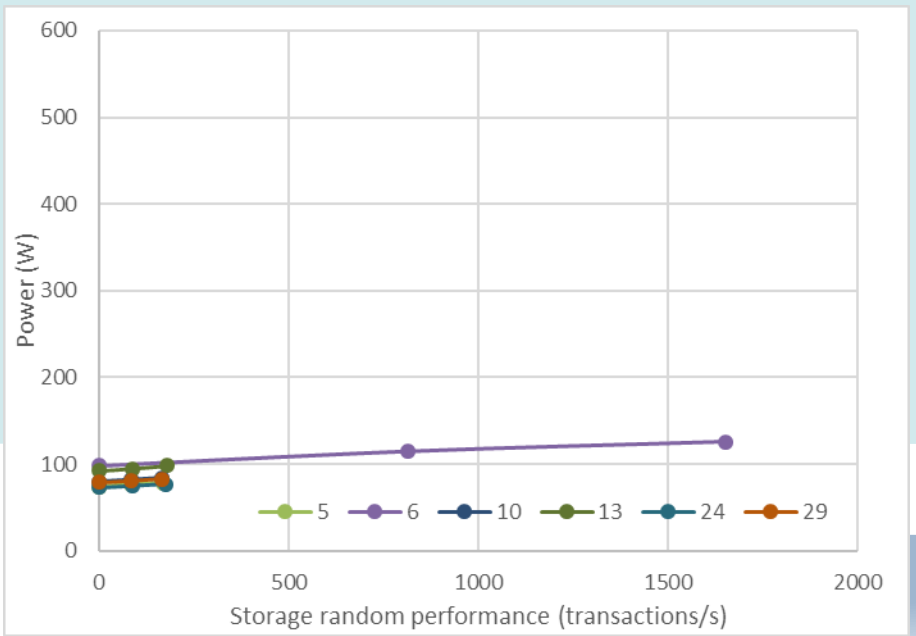
- Power does not vary, power and performance influenced by CPU
- Doubling RAM doubles performance, but power increase relatively small

Storage

- Largely independent of RAM and CPU (except idle power)
- Doubling number of drives doubles performance
- SSD score much higher than HDD

Memory flood and Storage random



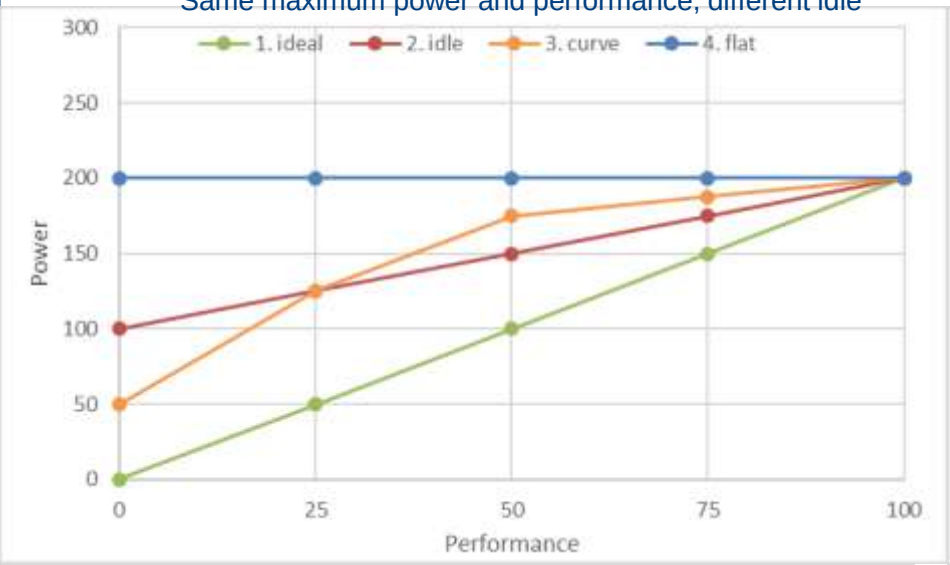


Sample number	CPU		Memory		Storage
	Number of CPUs x cores	Frequency (GHz)	DDR4 Modules / dimms (MB)	RAM (GB)	Number of HDDs
5	2 x 18	2300	8	64	1
6	2 x 18	2300	8	64	8
10	2 x 18	2300	16	256	1
13	2 x 18	2300	16	1024	1
24	2 x 6	1600	16	256	1
29	2 x 6	2400	16	256	1

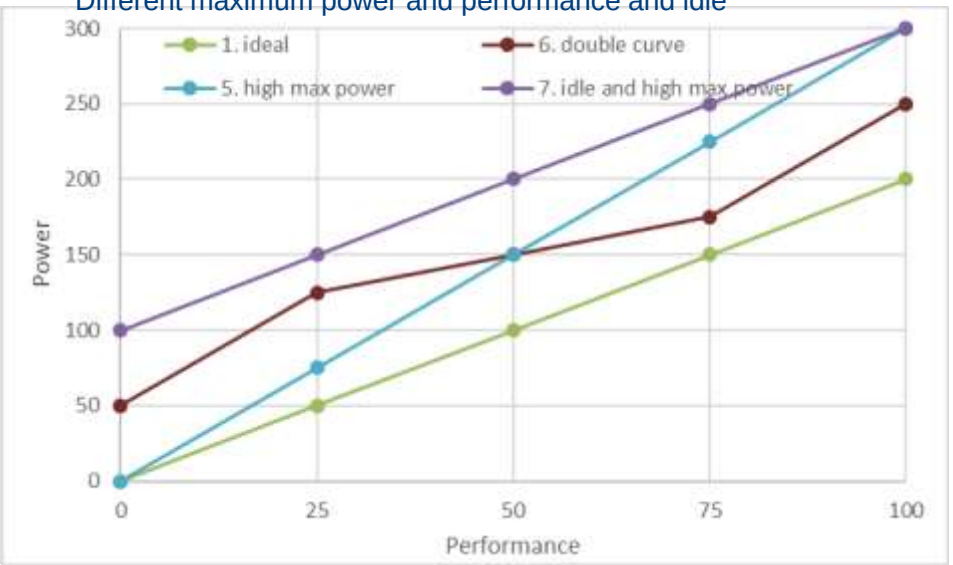
Metric Development – Hypothetical Curves

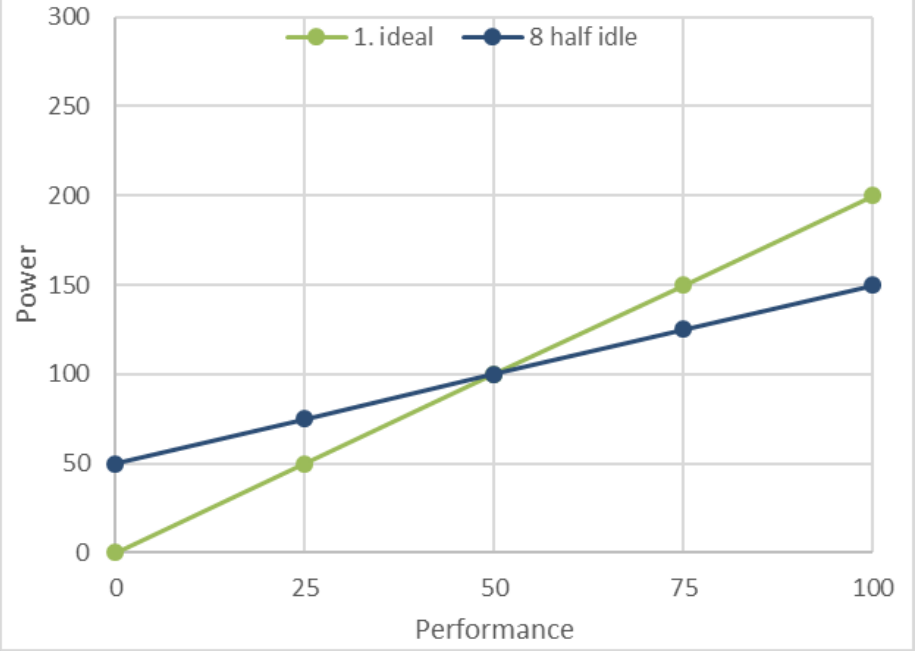
Intertek
Valued Quality. Delivered.

Same maximum power and performance, different idle



Different maximum power and performance and idle



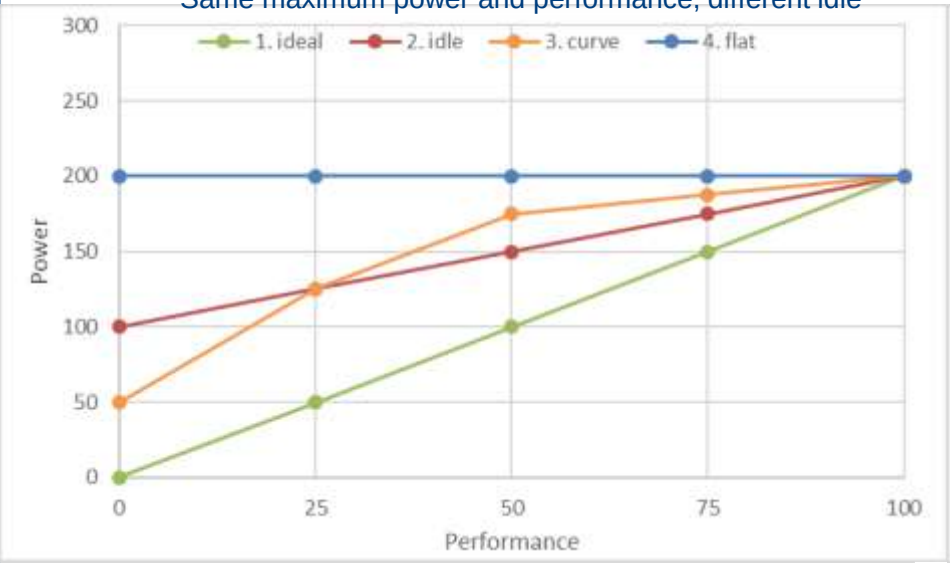


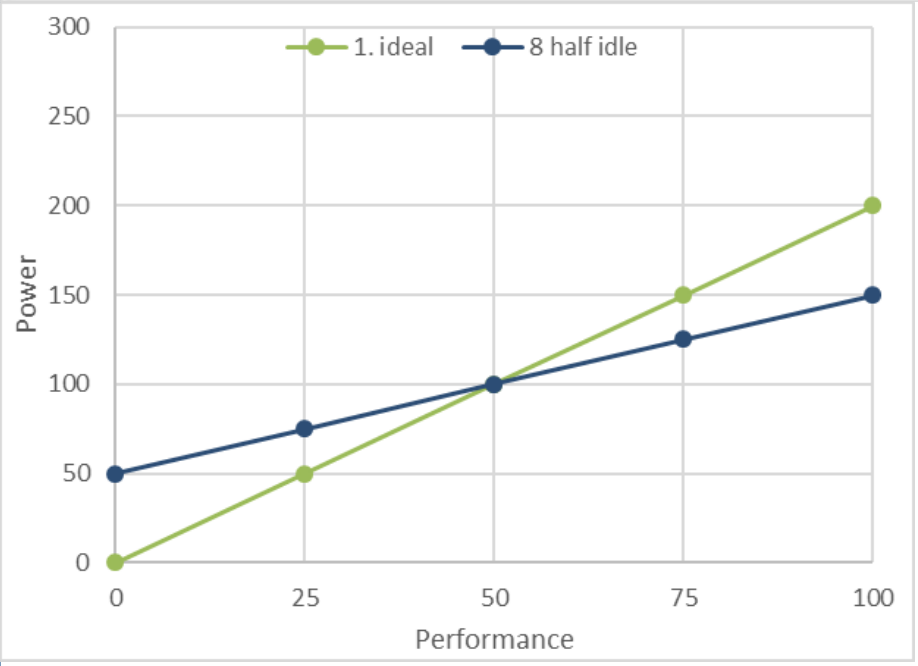
Hypothetical curve	Line type	Idle power	Max power
1. Ideal	Linear	0W	200W
2. Idle	Linear with non zero idle	100W	200W
3. Curve	log curve	50W	200W
4. Flat	Linear (flat),	200W	200W
5. High max power	Linear (similar to "Ideal" but higher max power)	0W	300W
6. Double curve	Inverse S curve	50W	250W
7. Idle and high max power	Linear with non zero idle and high max power	100W	300W
8. Half idle	Linear with non zero idle and low max power	50W	150W

Different metric approaches

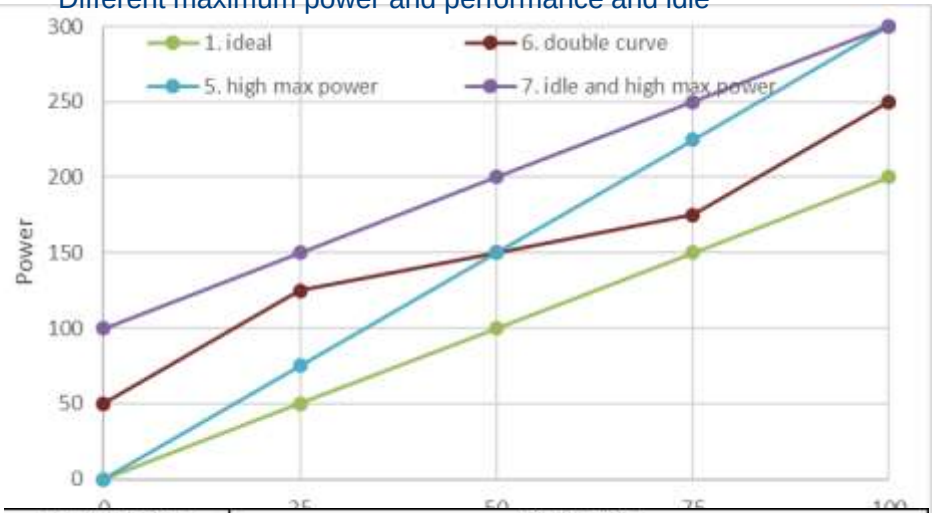
Intertek
Valued Quality. Delivered.

Same maximum power and performance, different idle





Different maximum power and performance and idle

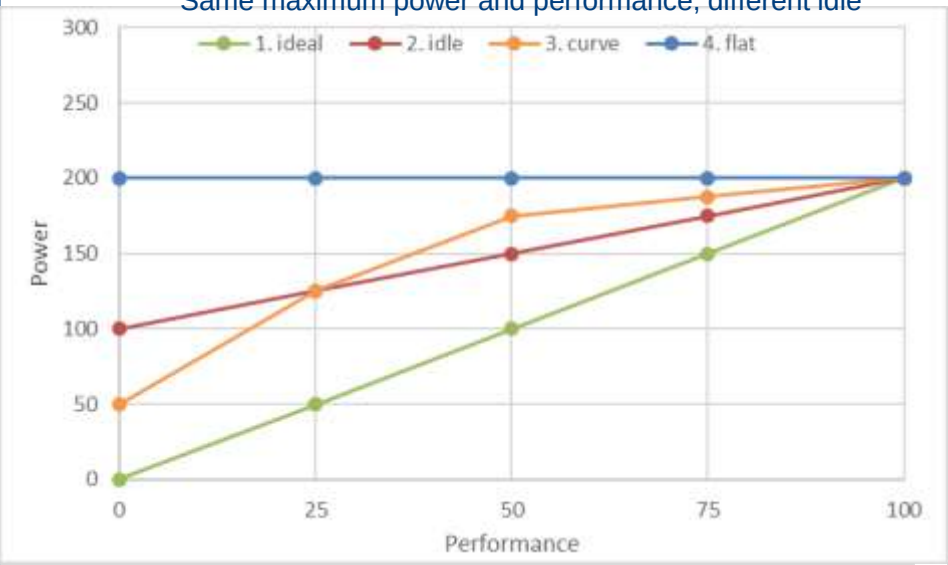


Metric Approach	Explanation
SERT	Sum of performance/sum of power, as explained previously. Weighted towards maximum performance efficiency
Mean	Mean of performance/power at each load level This is the simplest metric which gives equal weighting to the efficiency score at each load level
Weighted	Same as mean but lower load levels weighted higher Increases the weighting of lower load levels efficiency score to represent reality because servers operate at low utilisation level.
Heavy weight	Same as weighted but stronger weighting of low load levels
Peak:Idle	Ratio of peak performance/idle power Approach used by Japan Top Runner program. This encourages maximum performance as well as minimising idle power.
Peak:(Idle/Max)	Ratio of peak performance/ (1+proportion idle power/max power) Variation of Peak:idle. Using the Idle/max power gives an indication of how much the power scales instead of just idle.
SERT:(Idle/Max)	SERT efficiency score / (1+proportion idle power/max power) Variation of SERT to rebalance the high performance weighting of SERT

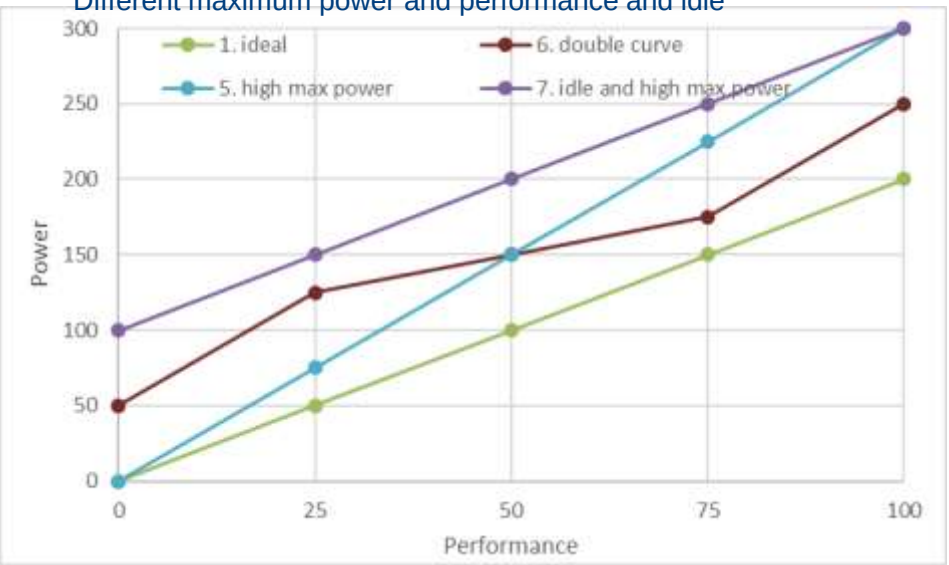
Performance of metrics against curves

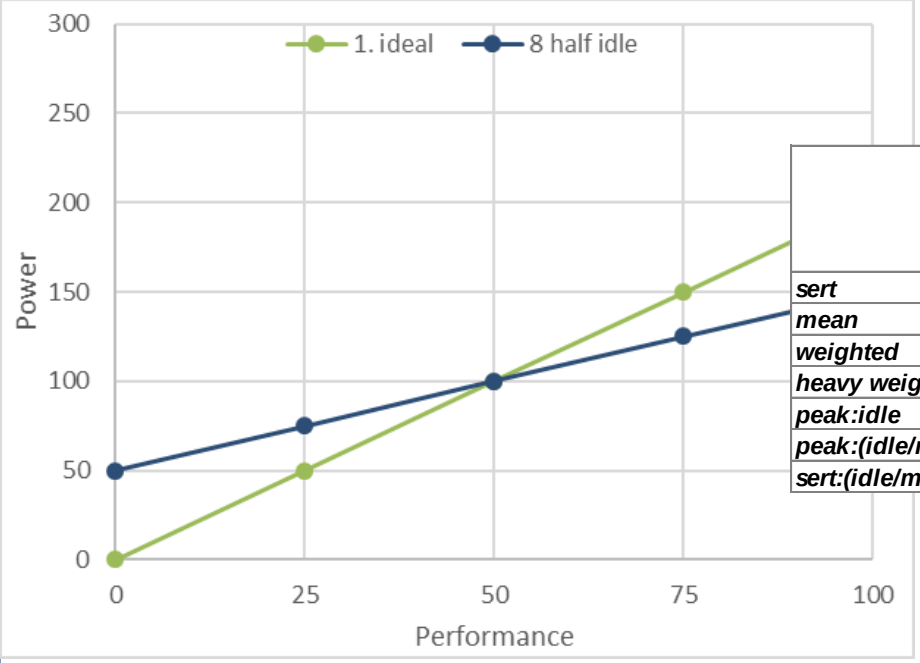
Intertek
Valued Quality. Delivered.

Same maximum power and performance, different idle



Different maximum power and performance and idle





	1. ideal	2. idle	3. curve	4. flat	5. high max power	6. double curve	7. idle and high max power	8 half idle
sert	100%	77%	73%	63%	67%	71%	56%	111%
mean	100%	73%	69%	63%	67%	68%	53%	105%
weighted	100%	68%	63%	55%	67%	66%	50%	99%
heavy weight	100%	63%	59%	50%	67%	61%	47%	94%
peak:idle	100%	1%	2%	0%	100%	2%	1%	2%
peak:(idle/max)	100%	67%	80%	50%	100%	83%	75%	75%
sert:(idle/max)	100%	51%	58%	31%	67%	60%	42%	83%

Conclusion: the **SERT approach, combined with an idle/max factor** provides the most representative differentiation between the various efficiency curves.

Proposed efficiency metric

Intertek
Valued Quality. Delivered.

DRAFT

CPU and hybrid workloads key components

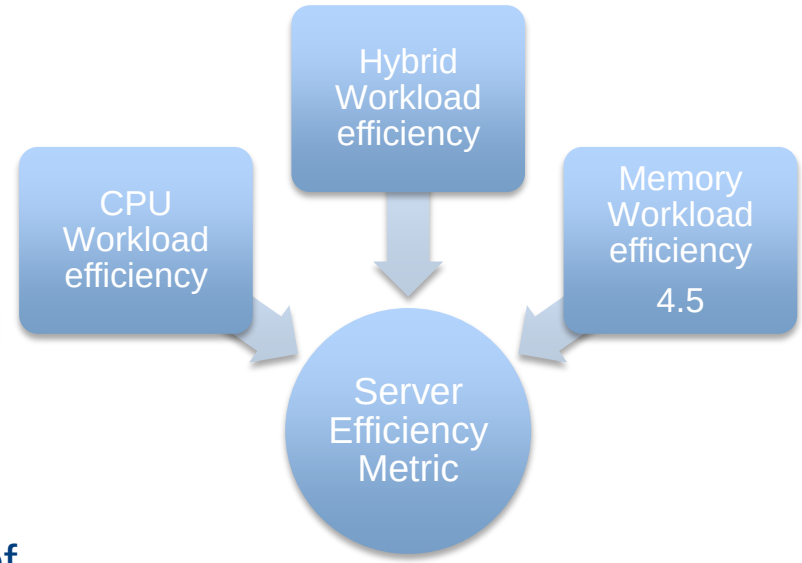
- Large influence on power consumption and performance across CPU & memory worklets.

Memory workload as a secondary component

- Divided by a factor (4.5) (through experimental tuning).

Storage workload not included

- Performance almost completely independent of the RAM and CPU.
- Distortion: 100-fold difference between SSD and HDD performance and efficiency scores.



$$\textbf{system efficiency metric} = CPU\ workload\ efficiency + hybrid\ workload\ efficiency + \frac{memory\ efficiency}{4.5}$$

Proposed efficiency metric

Intertek
Valued Quality. Delivered.

DRAFT

$$workload\ efficiency = \left(\prod_i^n worklet\ efficiency_i\right)^{\frac{1}{n}} \quad \text{(geometric mean)}$$

$$worklet\ efficiency = \frac{normalised\ SERT\ efficiency\ score}{\frac{idle\ power}{\max\ power}+1}$$

Dynamic range

- Workload efficiency is the geometric mean of the worklets – used by SERT and is more suited to values of very different magnitudes
- Worklet efficiency is adjusted by dynamic range

Dynamic range adjustment

Intertek
Valued Quality. Delivered.

DRAFT

$$Dynamic\ range = \frac{idle\_power}{maximum\_power} + 1$$

Varies from:

- 1 (zero idle power, perfect dynamic range) to
- 2 (idle and max power identical, no dynamic range)

Takes into account **relative, not just absolute idle power**

Rewards servers improving efficiency at **real world utilisation** levels.

Addresses **network scalability** issue and enables better comparison between high performance and many low performance servers

Proposed efficiency metric

Intertek
Valued Quality. Delivered.

DRAFT

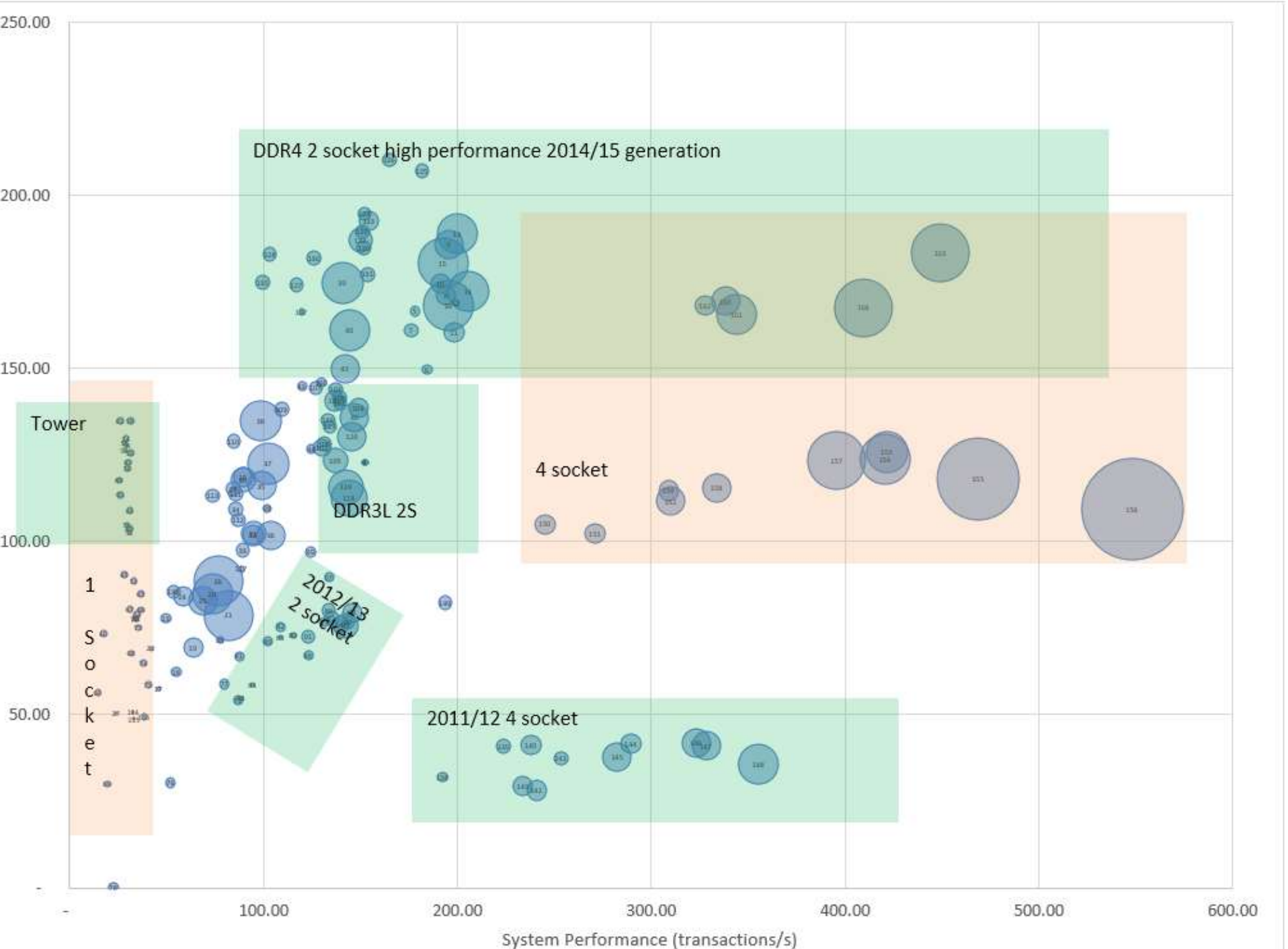
$$\textbf{system efficiency metric} = CPU\ workload\ efficiency + hybrid\ workload\ efficiency + \frac{memory\ efficiency}{4.5}$$

Where,

$$workload\ efficiency = \left(\prod_i^n worklet\ efficiency_i\right)^{\frac{1}{n}} \quad \text{(geometric mean)}$$

$$worklet\ efficiency = \frac{normalised\ SERT\ efficiency\ score}{\frac{idle\ power}{\max\ power}+1}$$

Dynamic range


DDR4 2 socket high performance 2014/15 generation
Tower
DDR3L 2S
4 socket
1 Socket
2012/13 2 socket
2011/12 4 socket
250.00
200.00
150.00
100.00
50.00
-
100.00
200.00
300.00
400.00
500.00
600.00
System Performance (transactions/s)

Observations on data

Intertek
Valued Quality. Delivered.

- Dataset of 180 different server configurations
- Shows difference between generation of servers
- Shows difference between different number of sockets
- Shows difference between efficiency of 2 sockets servers from low to high performance

Evaluation: proposed efficiency metric

Intertek
Valued Quality. Delivered.

Goal	
Focus upon the maximum potential for savings. Avoidance of weighting towards max utilization Means to take into account low utilization Means to take into account idle power overhead.	✔
Technology neutrality	✔
Interoperability	✔
Scalability (expandability, redundancy, system level scalability)	✔
Avoidance of negative market influence	✔
Defining product to test / product families	✔

Possible metric improvements

Intertek
Valued Quality. Delivered.

DRAFT

- **Increasing impact of dynamic range** (by decreasing the +1),
- Calculating dynamic range **from lowest utilisation point** instead of idle power,
- Simplifying application of dynamic range in metric by **applying to system efficiency** rather than worklet

Status of other metric developments

Development of metric is ongoing by industry, and regions including China, Korea, Germany,

Current industry proposal for CPU intensive metric:

- Correlates with Lot 9 draft proposed metric.
- More analysis and comparison required.

$$SPEC\ CPU$$

$$= e^{0.8 \ln(CPU\ workload + Hybrid\ SSJ\ workload) + 0.15 \ln(memory\ workload) + 0.05 \ln(storage\ workload)}$$

$$= CPU\ workload^{0.8} * hybrid\ workload^{0.8} * memory\ workload^{0.15} * storage\ workload^{0.05}$$

Defining products to test: the problem

Testing every possible configuration would be extremely time consuming, if not unfeasible: server models = very many of permutations of CPU, RAM, HDDs, I/O devices PSUs.

Other policies have created performance bands or different “representative” configurations for different performance levels.

Goals:

- Allow buyer to compare efficiency between models and configurations.
- Enable monitoring and verification activities – configuration used to declare efficiency often cannot be purchased for market surveillance.
- Not entail excessive costs or resources for the manufacturer.

Defining products to test: the solution

Intertek
Valued Quality. Delivered.

DRAFT

Manufacturers test and declare **at least three** configurations, representing **low, typical and high performance.**

Configurations align with those **advertised** on the manufacturers website (to assist consumer insight and enforcement activities).

Include two HDD/SSD (since the storage worklets are not included in the metric, standardising this improves comparability).

Don't include additional I/O cards (as the metric does not measure performance of these devices).

Potential **model to estimate** the efficiency of other configurations:

- Configuration boundaries – what exotic and atypical configurations can be excluded?
- What additional data is required/available, such as component level power data?
- What accuracy range is possible/desirable?

Intertek

THANK YOU!

Catriona McAlister
Consultant to Intertek
catriona.mcalister@seagreentree.com

Anson Wu
Consultant to Intertek
anson.wu@hansheng.co.uk

Davy Avenue, Knowlhill,
Milton Keynes
Bucks, MK5 8NL
T: +44 1908 857777.
http://www.intertek.com