Maximising Server IT Energy Efficiency through optimal Hardware and Software configuration.

Kat Burdett
Techbuyer
Harrogate, United Kingdom
k.burdett@techbuyer.com

Nour Rteil
Interact DC Ltd
Harrogate, United Kingdom
nour@interactdc.com

Rich Kenny
Interact DC Ltd
Harrogate, United Kingdom
rich@interactdc.com

***Abstract*—This paper analyses the impact of memory configuration and BIOS power profile on the energy efficiency of a Data Centre server undertaking a range of benchmarks that mimic real world application. It highlights some easy "wins" for achieving these goals, as well as some more specialized improvements that can be made through server profile customization.**

Keywords—Data Centre, Energy Efficiency, Green IT, BIOS, Memory Optimisation

I. Introduction

Digitialisation is a key driver in achieving our route to net zero emissions globally and digital solutions underpin each of the 17 United Nations sustainable Development goals (SDGs) to help build a better future but increased access to digital services come at a cost. Data Centres (DCs) are the facilities that house the server infrastructure underlying the digital world. These facilities are heavily energy intensive, with the DC sector accounting for between 1-2% of the worlds electrical power consumption [1] and contributing as much carbon emissions as the aviation industry (1.5-2.5%) [2], as such there is a great deal of value looking into improving the performance and efficiency of DC IT hardware.

These improvements can be done at a hardware level, a bios level, and a software level; this paper will consider the first two of these elements using a collection of servers from 2014 to 2021 and covering both Intel and AMD CPUs. Firstly, we will address the role that memory configuration plays in efficiency then we will consider the impact of BIOS settings on energy efficiency.

II. Methodology

In this paper we will consider the impact of changes in memory and bios settings on the performance and efficiency using two industry standard benchmarks, namely SPECpower_ssj2008 [3] and SERT[4]. These tests are undertaken in an attempt to capture some 'easy wins' in terms of improving performance for data centre servers, while still being realistic about the wider impact some of these changes may have.

SPECpower_ssj2008 is the older of the two benchmarks and undertakes a single Java workload, SSJ, which mimics applications that may be seen in enterprise or eCommerce scenarios. SERT is built upon SPECpower_ssj2008 with a further ten worklets which stress the CPU, memory and storage for a more rounded perspective on behaviour. SERT Score is calculated through a series of nested geometric means measured based on a reference server. In both benchmarks power consumption of the server is recorded for the duration of these tests to supply an energy efficiency metric upon completion – SSJ Ops per Watt for SPECpower_ssj2008 and SERT 2 Score for SERT.

The benchmarks are tested on a PowerEdge R7525 and a PowerEdge R6525 with AMD EPYC 7552 and 7352 processors respectively, using Windows Server 2019 OS and OpenJDK 11 with SERT-supported JVM client configurations.

The first half of the paper looks at proper memory configuration as a means to improving hardware configuration with minimal expenditure and the significant impact that will have on performance, while the second half of the paper looks

TABLE I. SERT and SPECpower_ssj2008 Results for The Tested scenarios

SERT results for PowerEdge R6525 with x2 EPYC 7352							
scenario #	DIMMs	channels populated	capacity (GB)	full power	SSJ performance	SSJ efficiency	SERT 2 score
1	4 x 64GB	25%	256	384	4,211,715	10,967	42.8
2	8 x 64GB	50%	512	486	6,576,007	13,540	53.3
3	16 x 8GB	100%	128	489	6,938,602	14,179	52.1
4	16 x 16GB	100%	256	500	7,003,540	14,010	55.7
SERT results for PowerEdge R7525 with x2 EPYC 7552							
scenario #	DIMMs	channels populated	capacity (GB)	full power	SSJ performance	SSJ efficiency	SERT 2 score
5	4 x 64GB	25%	256	379	2,676,375	7,063	53.9
6	8 x 16GB	50%	128	498	7,700,415	15,474	59.5
7	8 x 64GB	50%	512	524	7,962,754	15,183	76
8	16 x 16GB	100%	256	521	9,580,400	18,402	71.2
SPECpower_ssj 2008 results for PowerEdge R7525 with x2 EPYC 7552							
scenario #	DIMMs	channels populated	capacity (GB)	idle power	power at full load	performance at full load	energy efficiency
9	4 x 64GB	25%	256	82	199	1,542,574	7,501
10	4 x 16GB	25%	64	95	319	2,730,221	8,561
11	8 x 16GB	50%	128	101	464	6,646,970	14,236
12	16 x 16GB	100%	256	104	516	8,988,510	17,419

at the change in performance from switching and customizing BIOS system profiles to suit a workload.

III. MEMORY CONFIGURATION

According to SERT, memory configuration is one of the primary influences on the SERT 2 score [5]. Memory selection and configuration comes down to two primary choices: number of DIMMS and DIMM capacity. Previously, we explored the impact of these 2 options on power consumption, SSJ efficiency (CPU-intensive workload) and the overall SERT 2 efficiency score for Gen 9 HPE servers with Intel CPUs [6]. In this paper, we plan to analyse the impact of memory configuration on newer 1U and 2U servers and CPU generations (release year 2020) to see whether the same behaviour still applies.

A. *Results and Analysis*

We varied the memory configuration for the PowerEdge R6525 and R7525 by changing the number and capacity of the DIMMs populated. From the obtained results we observed no direct correlation between total RAM capacity and efficiency or power consumption. Yet, there is a noticeable link between the percentage of memory channels populated and energy efficiency, as demonstrated in TABLE I.

As the number of memory channels populated increases, SSJ performance and SSJ efficiency increase. The SERT 2 efficiency score generally follows that trend, except for scenarios #4 and #8. This is explained by looking at how the SERT 2 efficiency score is calculated. SERT 2 metric is the aggregate of the geometric means of the CPU workloads (SSJ is part of that), memory workloads and storage workloads, with respective weights of 65%, 30%, and 5%. So, for scenarios #2 and #7, the high capacity of 512 GB increases the memory scores (due to the way the Capacity workload operates), which in turn increases the SERT 2 efficiency score. It is noteworthy to mention that power increases as the number of DIMMs increases, but the power increase is compounded at higher utilizations as the number of memory channels increases.

B. *Capacity vs. Channel Population*

Unlike channel population, total RAM capacity doesn't have a direct impact on power consumption or energy efficiency unless the workload is memory intensive (such as the Capacity3 worklet of SERT).

To further demonstrate this, consider the same dual-core PowerEdge R6525 server, with a total capacity of 256GB (scenarios #1 and #4), and the only difference being the # of

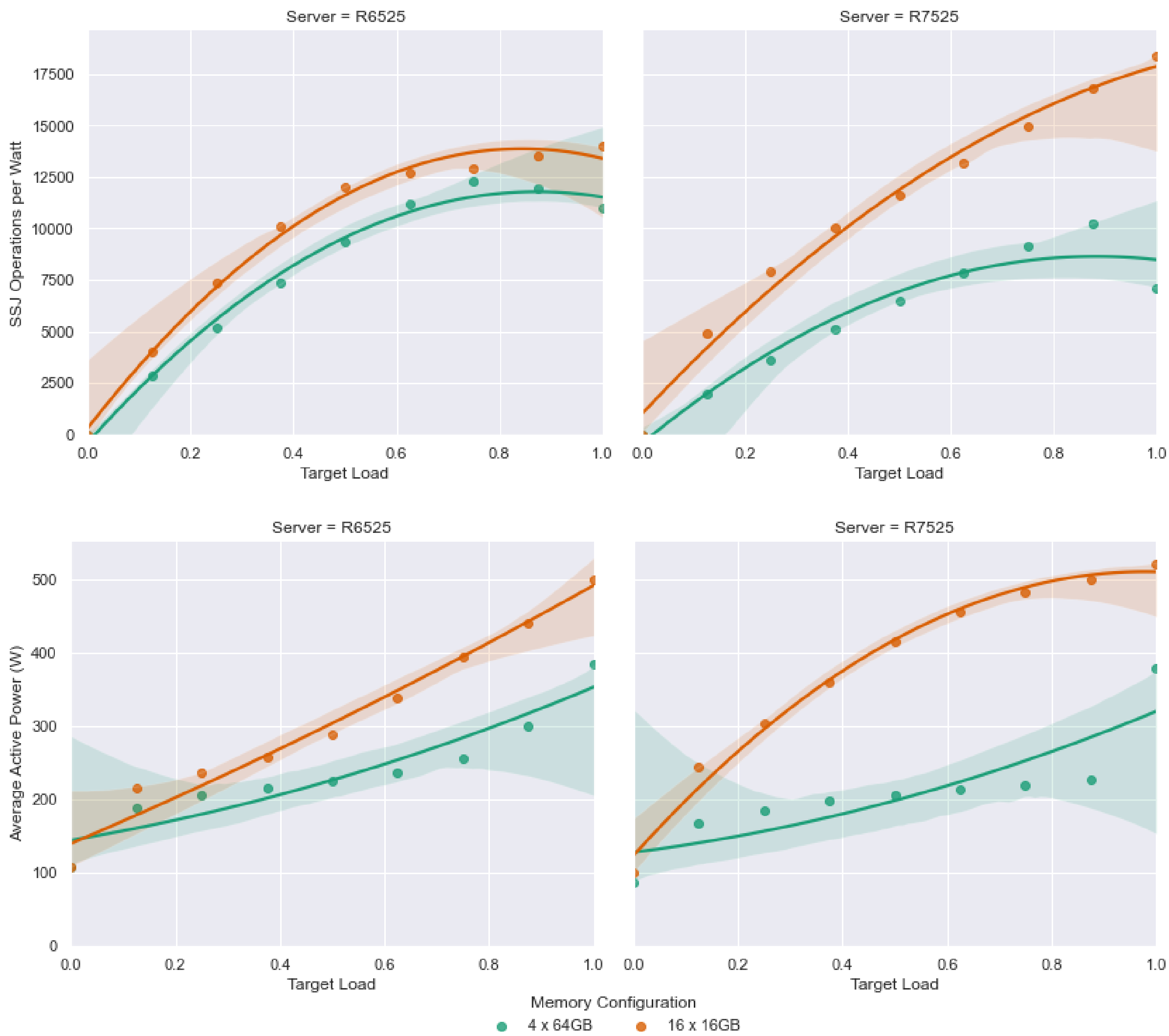

Fig. 1. Efficiency (top) and Power Consumption (bottom) across the range of target load levels for R6525 (left) and R7525 (right) with varying memory configurations and a total of 256GB total memory capacity during SSJ workload.

DIMMs installed/number of channels populated as seen in Fig. 1 (left side). The configuration with 16 DIMMs consumed 116 more watts than the model with 4 DIMMs (30% increase) but was 27.8% more efficient at full load. This increase in efficiency is attributed to the high increase in performance (66% increase) which outweighs the increase in power consumption.

The same can be observed for the R7525 with 256GB capacity using SERT (scenarios #5 and #8) and SPECpower_ssj2008 (scenarios #9 and #12). As seen in Fig. 1 (right side), for the R7525, the 16 DIMM configuration had 37% higher power, 258% higher performance, and 160% higher efficiency than the 4 DIMM configuration. A much higher increase in SSJ efficiency than seen on the R6525.

It is noteworthy to mention that power increases as the number of DIMMs increases, but the power increase is compounded at higher utilizations. For the scenarios we have tested, the power increase at idle state is low compared to the power increase at full load as the number of channels are increased.

In this section we were able to show that the key determinant to efficiency, in terms of RAM configuration, is channel population over total capacity. Optimal efficiency can be achieved by maximising the number of channels populated, where possible. By doing so, we were able to increase the server's SSJ efficiency up to 28% for the R6525 and 160% for the R7525, and the overall SERT efficiency up to 30% for the R6525 and 32% for the R7525. The next section explores the impact of BIOS configuration on efficiency and how optimal configuration can be achieved by tuning the BIOS settings.

IV. BIOS CONFIGURATION

The BIOS is firmware that sits on the divide between hardware and software, providing the translation layer for the operating system. The BIOS can be configured to modify hardware behaviour, and on modern servers these settings can be grouped into system profiles. There is often a profile tailored toward maximizing efficiency and a profile tailored toward maximizing performance, although others also exist for more specialized scenarios like power-capped Dense Configuration Optimisation in Dell servers before 14th Generation.

A. BIOS Server Profiles and their effect on Efficiency

Previous work exploring the relationship between these server profiles and the performance and efficiency of these benchmarks has highlighted the importance of server profile awareness, demonstrating a reduction of between 20-50% power consumption for negligible change in performance when shifting from balanced mode to the performance mode [7].

Building on that, this paper will consider two of the system profiles found in the BIOS of the PowerEdge R7525 and R6525 – Performance Mode and Balanced Power and

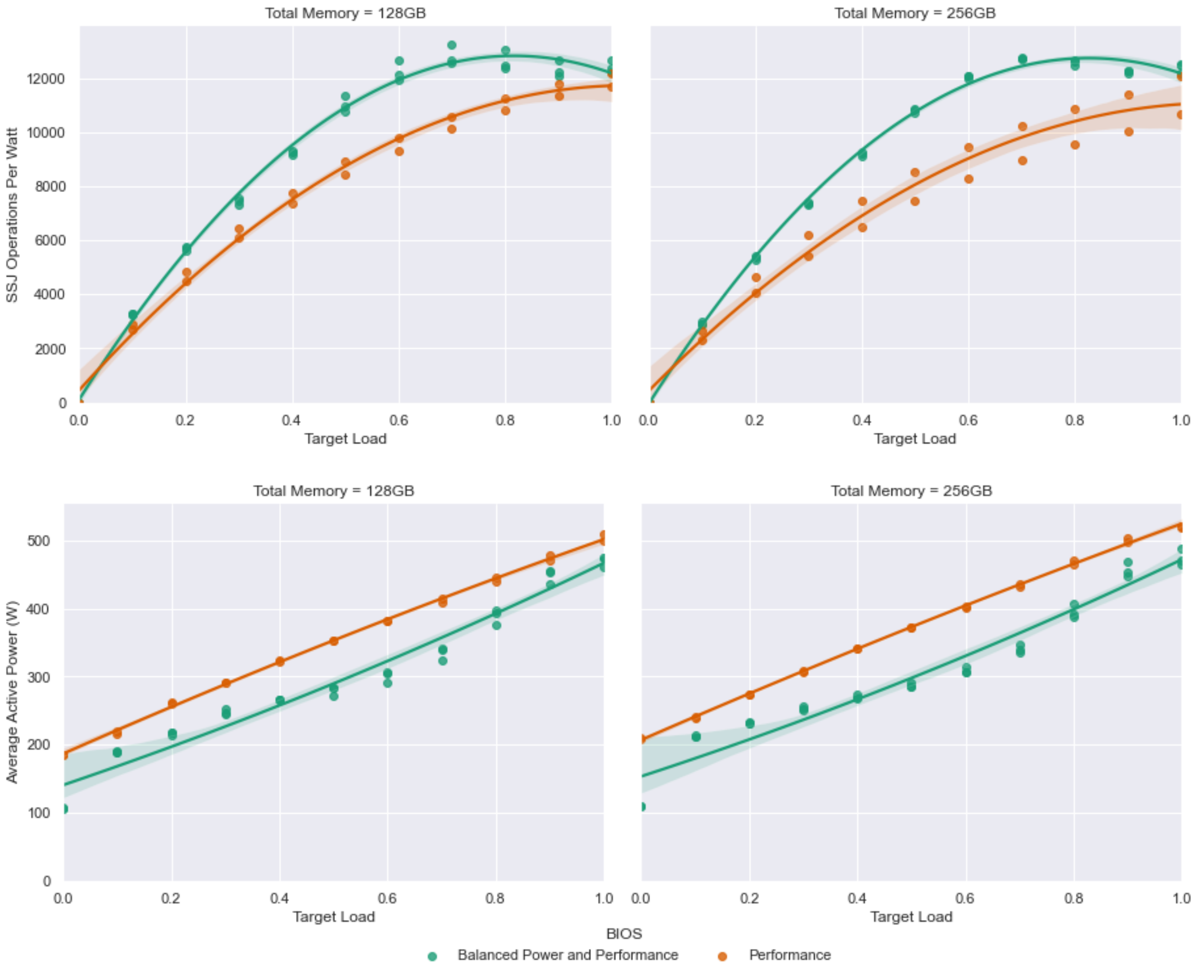

Fig. 2. Demonstrating the impact on efficiency and power consumption for the R6525 with balanced memory configurations of 128GB and 256GB when switching BIOS server profile from Performance Mode to Balanced Power and Performance.

Performance Mode – before looking at some tuning options for specific applications.

This BIOS system profile variation can be see seen in Fig. 2 for changing the R6525 server profile from Balanced Power and Performance to Performance mode while performing the SSJ worklet as part of a SPECpower_ssj2008 benchmark run. Tests were repeated multiple times for each configuration, maintaining all other variables as equal and ensuring the memory configuration was properly balanced for a total of nearly 400 tests performed. The results seen in Fig. 2 show an increase in overall power consumption at all load levels and both total memory amounts and negligible or no change in performance, resulting in a significant change in efficiency (measured in SSJ Operations per Watt for SPECpower_ssj2008).

B. BIOS Server Profile Custom Tuning

BIOS settings can also be manually modified by setting a system profile to Custom. This can allow for modification of individual parameters for attributes such as NUMA Nodes per Socket or whether Turbo Boost is enabled. This is a common practice among the published results for the SPECpower_ssj2008 benchmark, with server manufacturers commonly heavily modifying the BIOS settings of the SUT (Server under Test) to optimise the performance of the SSJ worklet. Furthermore, the modification of the attributes underlying the Java application that hosts the benchmark also have a significant impact on the overall performance. This falls outside the scope of this paper and have been held constant for the sake of this work. This practice of heavily customising the firmware (and software) environment for running the benchmark is not only allowed but also encouraged by SPEC, with the Quick Start Guide recommending reviewing published results with configurations similar to the SUT you look to test to determine how best to customise before benchmarking [8].

In this section, we emulate this behaviour by reviewing the Custom Bios configurations of two published results, a Dell Inc. PowerEdge R7525 (AMD EPYC 7702, 2.00 GHz) [9] and a Dell Inc. PowerEdge R6515 (AMD EPYC 7702P, 2.00 GHz) [10]. While these servers do not have the same hardware configuration as the SUTs tested in this paper they are the closest of the published results available at the time of testing. Both made similar modifications to BIOS settings as seen in TABLE II.

Each of these attributes were incrementally changed to determine the sensitivity performance may have before considering the fully modified Custom Server Profile against the Balanced Power and Performance Server Profile as a whole. The final Custom comparison will be presented here for simplicity, but if the results are to be replicated it is best to change the variables incrementally in case problems are encountered as this will allow easy isolation of the variable at fault.

The only significant deviations from the BIOS settings seen in TABLE II were the NUMA nodes per socket was set to 2 as opposed to 4 and Fmax may have varied from 2200 based on the architecture of the AMD processors tested.

Fig. 3 shows the results of this Custom Server Profile on worklet score for the R7525 with 2 AMD EPYC 7552 processors. While this test was conducted on both SPECpower_ssj2008 and SERT benchmarks, the improvement seen by the SSJ worklet behaviour was comparable between both, seeing on average 12% improvement in efficiency across load levels. This worklet score is calculated by taking the geometric mean of the efficiencies at each load level and comparing it to a baseline reference server.

The Custom profile improved worklet efficiency scores for 9 out of the 11 worklets tested, with the greatest improvements occurring in Compress, CryptoAES, and SSJ, all of which are benchmarks mimicking similar kinds of applications. There was a small drop in efficiency of the SOR worklet of 0.76% and a more significant drop in efficiency of the Capacity3 worklet of 7%.

TABLE II. CUSTOM BIOS SETTING MODIFICATIONS BY ORDER OF TUNING

Modifications to BIOS
System Profile set to custom
CPU Power Management set to OS DBPM
Memory Frequency set to 2666MHz
Turbo Boost enabled
Memory Patrol Scrub disabled
Memory Refresh Rate set to 1x
PCI ASPM L1 Link Power Management enabled
Determinism Slider set to Power Determinism
Efficiency Optimized Mode enabled
Logical Processor enabled
L1 Stream HW Prefetcher disabled
L2 Stream HW Prefetcher disabled
NUMA nodes per socket set to 4
Cstates set to Autonomous
MADT Core Enumeration set to Linear
L3 Cache as NUMA Domain set to Enabled
Memory Interleaving disabled
Opportunistic Self-Refresh enabled
ApbDis enabled
ApbDis Fixed Socket P-state set to P3
Boost FMax set to Manual
Manual Boost FMax set to 2200

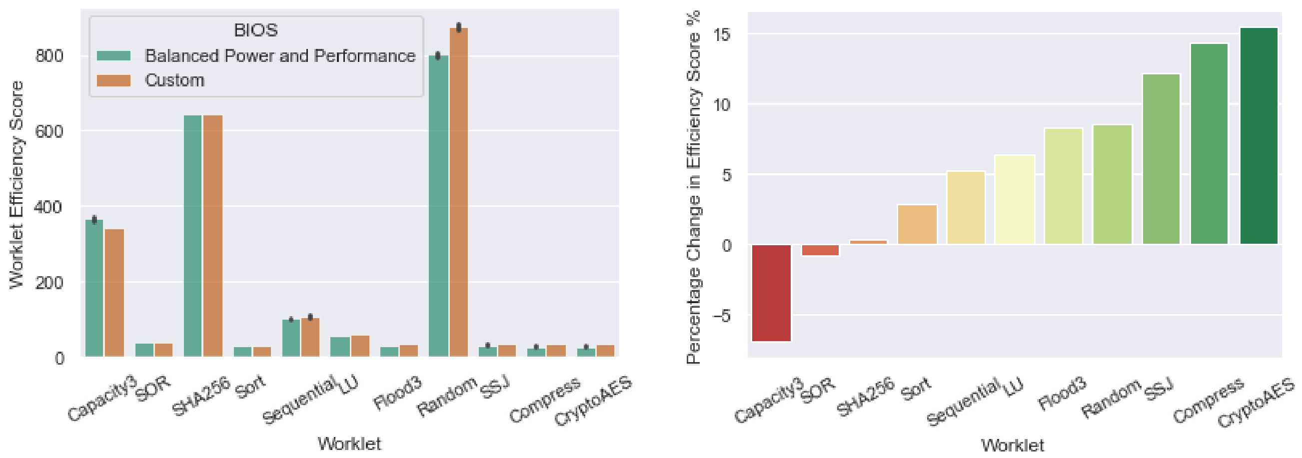

Fig. 3. SERT Worklet Efficiency Scores for changing BIOS settings from Balance Power and Performance to Custom for the R7525 configuiration, annotated with percentage change in score for each worklet. While the majority of worklet applications became more efficient, both SOR and Capacity3 became less efficient due to the BIOS customised toward SSJ optimisation.

It is also worth noting that while the change to Custom profile provided improvements to the majority of SERT worklets when considering each load level equally, it did increase idle and low utilisation power consumption up slightly. Given the high propensity for idle or "zombie" servers in Data Centres [11], such a change should only be made not only if there is a solid understanding of the application a server is going to undertake, but also the average utilisation with which that server performs that application. If the server is mostly idle, any efficiency savings from this specific BIOS tuning will be negated by the increased power consumption while idling.

C. *Memory Configuration and BIOS Customisation*

While the tests seen in the previous subsection assume that memory configuration has been properly accounted for, as highlighted by Section III, this may not always be possible or desired – for example, if a hardware owner has chosen that configuration for a specific reason. In this case, BIOS customisation for a specific application can still provide some improvements in efficiency – although any efficiency "wins" will likely be relatively small compared to the savings created by proper hardware memory allocation.

V. CONCLUSION

The memory configuration of a server can have a significant impact on its performance and efficiency. Populating all available memory channels provides the optimal configuration for energy efficiency. This holds true for older and newer servers and CPU generations as shown in this paper. Subsequently, memory selection needs to be planned as such. For example, for a server with 16 available memory channels which requires 128GB RAM capacity, opting for x16 8GB DIMMs instead of x4 32GB or x2 64GB etc. will offer the highest performance per watt across all the load levels.

Another relatively easy way to improve the server's efficiency, with minimal to no loss in performance, is by changing the BIOS power profile from Performance mode to the Balanced mode. This is especially suitable for general purpose workloads. Further efficiency optimisations can be achieved by customising and tuning the individual BIOS settings to fit the workload by using the Custom BIOS mode. This will rely on a deeper understanding of the applications a server will be expected to run in order to utilise effectively.

ACKNOWLEDGMENT

This work was supported by a UKRI Future Leaders Fellowship [grant number MR/T04389X/1].

REFERENCES

[1] IEA. (2021) Data centres and data transmission networks. [Online]. Available: https://www.iea.org/reports/data-centres-and-datatransmission-networks

[2] M. Avgerinou, P. Bertoldi, and L. Castellazzi, "Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency," Energies, vol. 10, p. 1470, 9 2017.

[3] Standard Performance Evaluation Corporation (SPEC), SPECpower_ssj 2008. https://www.spec.org/power_ssj2008/ (accessed May 18, 2023).

[4] Standard Performance Evaluation Corporation (SPEC), SERT Suite. https://www.spec.org/sert/ (accessed May 18, 2023).

[5] SPEC RG Working Group, The SERT 2 Metric and the Impact of Server Configuration, Sep. 2018. URL: https://www.spec.org/sert2/SERTmetric.pdf

[6] R. Bashroush, N. Rteil, R. Kenny, and A. Wynne, "Optimizing server refresh cycles: The case for circular economy with an aging moore's law" IEEE Transactions on Sustainable Computing, vol. 7, no. 01, pp. 189–200, jan 2022.

[7] N. Rteil, K. Burdett, S. Clement, A. Wynne and R. Kenny, "Balancing Power and Performance: A Multi-Generational Analysis of Enterprise Server BIOS Profiles," 2022 International Conference on Green Energy, Computing and Sustainable Technology (GECOST), Miri Sarawak, Malaysia, 2022, pp. 81-85, doi: 10.1109/GECOST55694.2022.10010599.

[8] Standard Performance Evaluation Corporation (SPEC), SPECpower_ssj2008 Quickstart Guide v1.12 https://www.spec.org/power/docs/SPECpower_ssj2008-Quick_Start_Guide.pdf

[9] Dell Inc. PowerEdge R7525 (AMD EPYC 7702, 2.00 GHz) published SPECpower_ssj2008 result, https://www.spec.org/power_ssj2008/results/res2020q2/power_ssj2008-20200324-01021.html (accessed May 18, 2023).

[10] Dell Inc. PowerEdge R6515 (AMD EPYC 7702P, 2.00 GHz) published SPECpower_ssj2008 result, https://www.spec.org/power_ssj2008/results/res2020q3/power_ssj2008-20200728-01041.html (accessed May 18, 2023).

[11] J. G. Koomey, E. R. Masanet, N. Horner, I. L. Azevedo, and W. Lintner, "United States Data Center Energy Usage Report," Berkeley Lab, no. June, p. 65, 2016. [Online]. Available: https://eta.lbl.gov/publications/united-states-data-center-energy