Optimizing server refresh cycles: The case for circular economy with an aging Moore's Law

Rabih Bashroush, *Member, IEEE*, Nour Rteil, Rich Kenny, and Astrid Wynne

***Abstract*— Demand for digital services is increasing significantly. Addressing energy efficiency at the data center mechanical and electrical infrastructure level is starting to suffer from the law of diminishing returns. IT equipment, specifically servers, account for a significant part of the overall facility energy consumption and environmental impact, and thus, present a major opportunity, not the least from a circular economy perspective. To reduce the environmental impact of servers, it is important to realize the effect of manufacturing, operating, and disposing of servers on the environment. This work presents new insights into the effect of refreshing servers with remanufactured and refurbished servers on energy efficiency and the environment. The research takes into consideration the latest changes in CPU design trends and Moore's law. The study measures and analyzes the use phase energy consumption of remanufactured servers vs new servers with various hardware configurations. Case studies are used to evaluate the potential impact of refurbished server refresh from an economic as well as environmental perspectives.**

***Index Terms*—balanced server configuration, circular economy, data centers, energy efficiency, environmental impact, green computing, memory configuration, power consumption, processors, refurbished servers, SERT, sustainable computing.**

I. INTRODUCTION

PEOPLE and organizations rely heavily on digital services, to the extent that the annual global Internet traffic in 2017 was estimated to be 1.5 ZB, and is expected to reach 4.8 ZB by 2022 [1]. This equates to an increase of 220% over five years or an average of 26% annually. Yet, internet penetration is only at 58.8% worldwide (as of June 2019 [2]), with numerous upcoming digitalization projects and services fueled by emerging trends and technologies such as smart cities, Internet of Things, and 5G amongst others. The explosive growth in demand is matched by an increase in demand for equipment and energy to power the rapidly expanding infrastructure. The infrastructure encompasses the servers, storage, and networking devices as well as the supporting mechanical and electrical plants. This amplified demand for computing power creates a global environmental challenge. Energy consumption for data centers was estimated to be 103 TWh in Europe in 2014 [3] and rose to 130 TWh in 2017 [4], an average increase of approximately 25% over 3 years. This equated to 5% of the electricity consumption in Europe. In China, the energy consumption of data centers was estimated to be 160 TWh in 2018 [5]. The greenhouse gas emissions of the ICT sector have increased by half since 2013, rising from 2.5% to 3.7% of global emissions [3].

Many metrics have been created to help measure the energy efficiency of data centers. One of these is Power Usage Effectiveness (PUE) [6], although PUE technically is not an energy efficiency metric because it does not cover IT. According to the Uptime Institute's Global Survey 2019, the industry average PUE was 1.67 in 2019 [7]. This means for every 1 kWh consumed by IT, 0.67 kWh is consumed by the mechanical and electrical infrastructure needed to maintain an appropriate operating environment. Accordingly, the biggest part of energy consumption is attributed to IT, and particularly to servers (65%) [8].

The environmental impact of servers can be attributed to three phases: 1) the embodied energy associated with the manufacturing of servers; 2) use phase energy consumption; and 3) end of life impact (e.g. waste to landfill).

In previous work (using server data up until 2016), we studied the optimal server refresh cycles along with the environmental impact associated with procuring new servers vs prolonging the life of the existing kit [8]. The work served as the basis for several international best practices and standards (e.g. EN50600-99-2) as well as policies (e.g. the European EcoDesign Legislation for servers and storage devices).

However, since then, technological developments, or lack thereof, warranted a revisit of the work. Current server processor technology has not witnessed any significant changes in performance per watt during the past few years, something that was historically driven by major gains, as originally prescribed by Moore's Law [9].

This paper is submitted for review on June 2020. This research is funded in part of the Knowledge Transfer Partnership program by InnovateUK under partnership number 11485.

R.B., is with Uptime Institute, London E1 7LS, UK (e-mail: r.bashroush@qub.ac.uk) and University of East London, London E16 2RD, UK.

N.R., is with the University of East London, London E16 2RD, UK and Techbuyer, Harrogate HG2 8PB, UK (e-mail: n.rteil@uel.ac.uk).

R.K., is with Techbuyer, Harrogate HG2 8PB, UK (e-mail: r.kenny@techbuyer.com).

A.W., is with Techbuyer, Harrogate HG2 8PB, UK (e-mail: a.wynne@techbuyer.com).

This paper studies the environmental and performance impact of the slowdown in technological advances on server refresh, and its implication on employing circular economy practices (e.g. use of refurbished/remanufactured servers vs new servers).

Within this work, we define the term *refurbished servers* to indicate used servers that are tested and cleaned, and *remanufactured servers* those rebuilt with component upgrades on the core product (i.e. motherboard, processor and main memory). The reliability of refurbished servers is found to be close to that of new servers as their peripheral components, such as HDDs and power supplies that tend to fail the most, are often replaced during the refurbishment cycle. Conversely, remanufacturing servers is constrained by the incompatibility between new CPU models with older motherboards so is limited to compatible component upgrades for the server model.

Section 2 describes the rationale behind this work, the emerging technology trends, and research questions. Section 3 provides an overview of the benchmarking methodology used. Section 4 covers the experiments conducted. Section 5 discusses the results, followed by a case study presented in section 6 demonstrating the potential impact of this work. Finally, section 7 highlights the study limitations, with conclusions and future research directions discussed under section 8.

II. RATIONALE AND TECHNOLOGY TRENDS ANALYSIS

A previous study into the impact of hardware refresh [8] demonstrated refresh gains from the use phase were dominant, thanks to major sustained performance advancements from improvements in processor technology, dubbed Moore's law. However, this has substantially slowed down over the past three to four years.

Moore's Law observed that transistor count on microchips would double every two years due to advancements in lithography. This led to a reduction in transistor size, meaning less energy and faster switching. However, the continued validity of this trend is under question. To better understand this, data from SPECpower [10] containing energy performance of volume servers was studied, against advancement in lithography [11]. The results are shown in Fig. 1 below.

Fig. 1 shows how server performance per watt (the orange line) was maintained over the past decade. The move from one lithography to another (e.g. from 65nm to 45nm, 45nm to 32nm, etc.) presented major performance gains, as well as a substantial reduction in idle power (the blue line). This is by large due to efficiency gains from reducing transistor size in microchips.

However, another interesting observation in Fig. 1 is the way idle power increased with the introduction of more cores to sustain performance gains. This can be seen with the introduction of 8 cores during the 45nm lithography, and then more evidently with the stagnation at 14nm over the past few years. This increase in idle power to maintain performance gains per watt has major implications on overall server energy consumption in production environments. According to various studies [4][12], the average server utilization level is 25%. As such, servers spend most of their life idling. Thus, increasing performance per watt at the expense of increasing idle power does not necessarily make the server more efficient in production environments.

This is further reflected in the dynamic range of servers, which is the ratio between full load server energy consumption and idle power energy consumption. The higher the dynamic range, the more energy proportionate the server is, and the more efficient it is in a real-life deployment with a fluctuating workload.

Fig. 2 shows how the dynamic range changed over the past decade, with a clear turn of events since the last study, which covered performance data until 2016 [8]. Accordingly, this work examines what this development means to refresh cycles, in particular, the case for refreshing with new servers rather than refurbished or remanufactured servers.

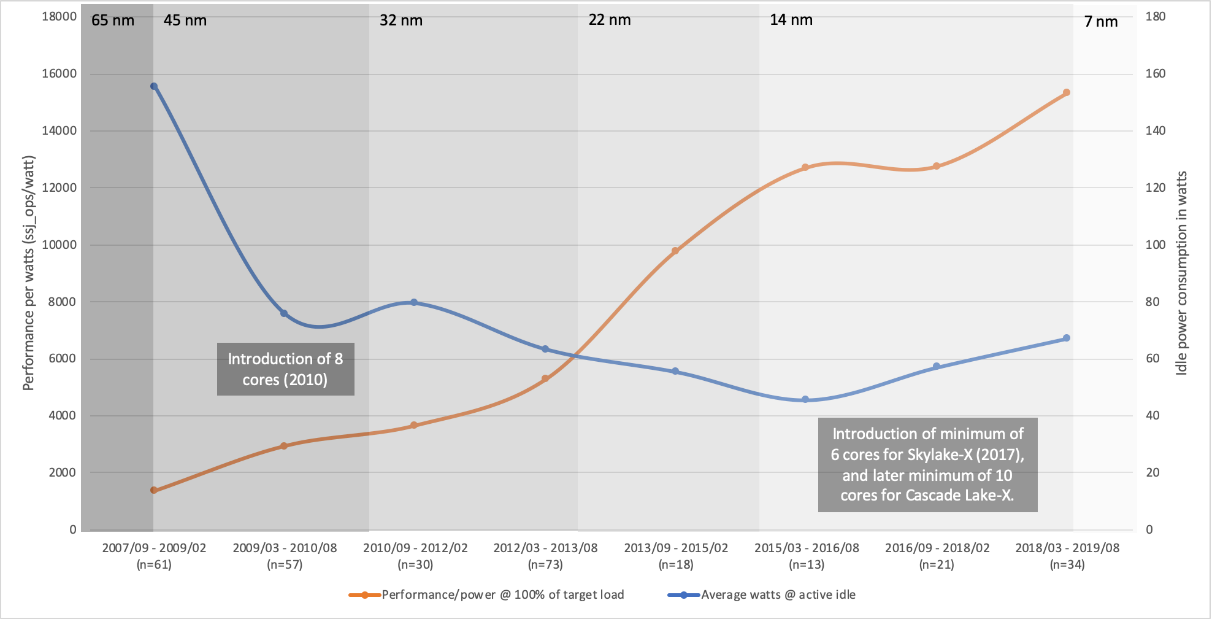


Fig. 1. Impact of CPU lithography on performance and idle power (n represents the number of servers sampled in each time frame)

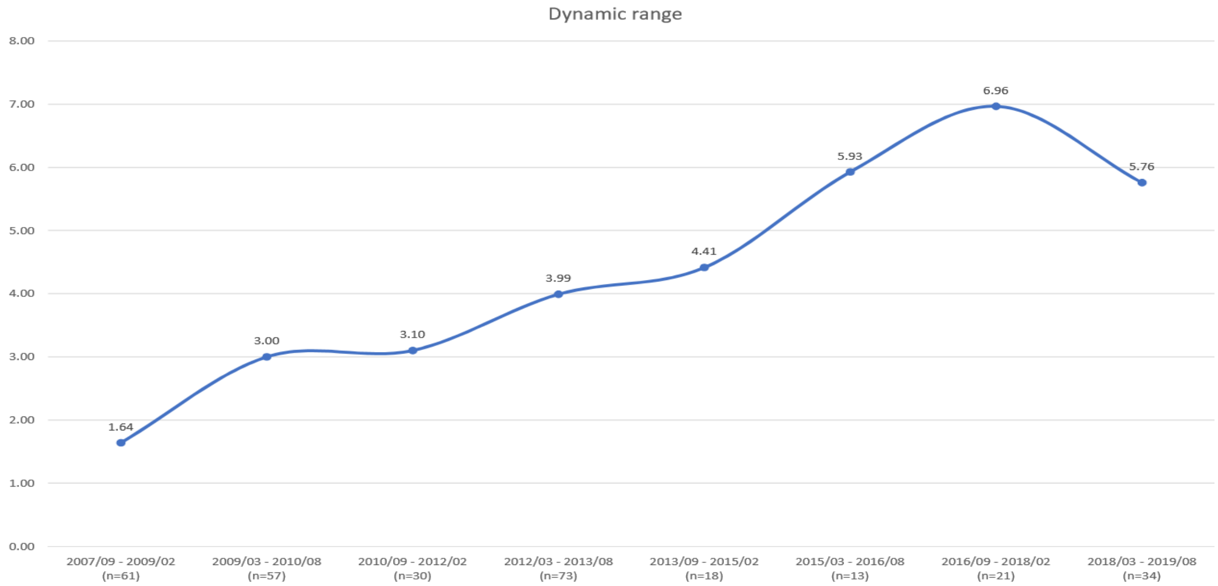


Fig. 2. Dynamic range of sampled servers (n represents the number of servers sampled in each time frame)

According to the EURECA study [4], which examined over 300 European data centers, 40% of deployed servers are older than 5 years, yet, they consume 66% of the overall facility energy and contribute 7% of the compute capacity. This presents a significant opportunity to eliminate waste through a server refresh.

However, do we still have a viable case for refreshing older servers with new ones? For what age range? And, what opportunities exist to increase the efficiency of younger servers? These questions will be addressed in the subsequent sections. Namely, the gains that can be achieved through remanufacturing and component level upgrades are discussed in sections 3, 4, and 5. The viability of refreshing older servers with new vs refurbished servers is then discussed in section 6.

III. METHODOLOGY

The first step towards determining the impact of the aging Moore's Law on energy efficiency was to measure the efficiency of new vs remanufactured servers, using different hardware configurations and component level upgrades (section 4).

The measurements were carried out using the Server Efficiency Rating Tool SERT. The SERT environment setup [14] comprises of:

1. System under test (SUT): the server for which the measurements are being recorded.
2. Controller System (CS): the system that executes workloads on SUT and records generated values.
3. Power analyzer: used to measure the power consumption of the SUT.
4. Temperature sensor: used to measure the ambient temperature where the SUT is located.
5. SERT: Software that runs on the CS and SUT and contains subcomponents (PTDaemon, Chauffeur, Director, Reporter, and GUI) responsible for configuring, measuring, gathering, and reporting environmental, power, and performance data after a run is complete.

The components are interlinked as per the setup shown in Fig. 3 [15].

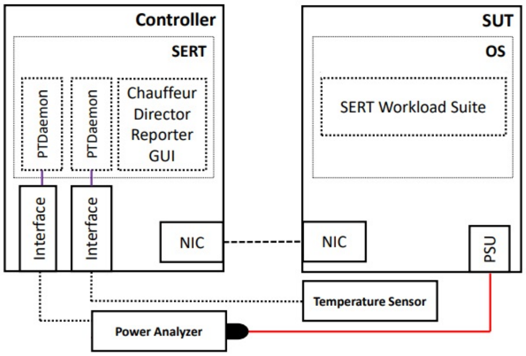


Fig. 3. SERT Environment Setup [15]

The SERT suite is composed of 4 workloads: CPU, memory, storage, and idle. Each workload (apart from idle) is composed of a group of worklets designed to stress a particular aspect of the SUT. Table I summarizes the main functionality of each worklet [16].

For each worklet, data are reported for a set of load intervals and the total values are calculated as the geometric means across all loads and worklets, resulting in a balanced overall server efficiency score.

The results from the SERT experiments were then analyzed, particularly around how remanufactured servers perform against new servers (section 5).

Finally, representative use case scenarios were evaluated for different refresh cycles, using refurbished and new servers, based on cost as well as environmental impact (section 6).

TABLE I
Description of SERT Worklets

Workload	Worklet	Description
CPU	Compress	Compresses and decompresses data using Lmepel-Ziv-Welch (LZW).
	CryptoAES	Encrypts and decrypts data using the AES block cipher algorithm.
	LU	Computes the LU factorization of a dense matrix using partial pivoting.
	SOR	Performs access patterns in finite-difference applications.
	Sort	Sorts a randomized 64-bit integer array.
	SHA256	Performs SHA-256 hashing transformations on a byte array.
	SSJ	Performs multiple different transactions that simulate an enterprise application.
Storage	Sequential	Reads and writes data to/from random file locations.
	Random	Reads and writes data to/from file locations, picked sequentially.
Memory	Flood3	Performs a sequential memory bandwidth test that uses arithmetic operations and copies instructions.
	Capacity3	Performs a memory capacity test that uses XML operations on a data set.

IV. EXPERIMENTS

This section describes the experiments conducted with several server configurations and their corresponding generated SERT scores.

A. Server Under Test

The HPE ProLiant DL380 Gen 9 server was used as the basis for the experiments (being a representative volume server), with the base configuration shown in Table II. BIOS configuration is as per the default manufacturer settings with the power saving feature enabled. An asterisk designates the variable parts across the different test scenarios.

TABLE II
HPE ProLiant DL380 Gen 9 Base Configuration

Hardware configuration	
CPU Name*	2 x E5-2690v3
CPU Characteristics*	2.60GHz, 12-core, 30MB
Power Supply	2 x 800W
Memory*	2 x 16GB DIMMs
Disk Drive*	2 x 600 GB SAS HDDs
Disk Controller	HP smart array p440ar/2gb fbwc 12gb 2-ports SAS controller
NICs Installed	HP 1GB quad-port 331flr ethernet adaptor
Network Speed	1000 Mbit/s
Software configuration	
Operating system	Microsoft Windows Server 2012 R2
Java Virtual Machine	Oracle Hotspot 64-bit server JVM 1.7.0

* These are variable parts across the different test scenarios

B. Test Scenarios

Twenty-two test scenarios (TS1 to TS22) were conducted as per Fig. 4. The base configuration for all these scenarios is described in Table II. For the Refurbished vs new scenarios, they compare like for like (e.g. a new CPU E5-2690v3 is replaced with the same model but refurbished). The remaining test scenarios cover the impact of changing memory, processor, and storage configuration on the server's energy efficiency.

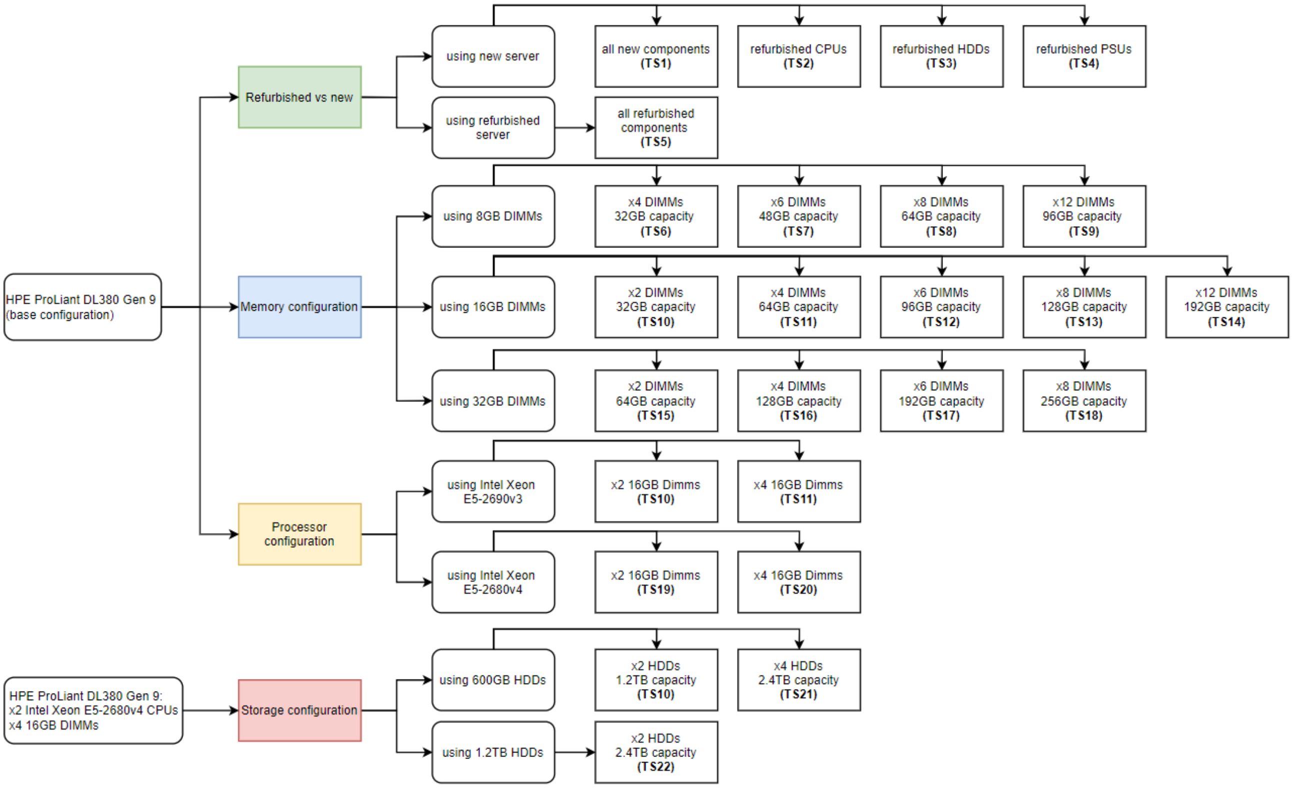


Fig. 4. SERT test scenarios

C. Efficiency Scores

The SERT benchmark produces an efficiency score representing the overall server efficiency, calculated as follows:

- First, the performance score for each worklet is calculated by dividing the transactions count by elapsed measurement time and taking the geometric mean across loads.
- Second, the worklet efficiency score is calculated by dividing the performance score by measured active power and taking the geometric mean across loads.
- Workload efficiency scores are then calculated by taking the geometric means of the worklet efficiency scores.
- Finally, SERT 2 efficiency score Es is computed as the geometric mean of workload efficiency scores (Equation 1), based on the following weightings: 65% for CPU, 30% for memory, and 5% for storage.

$$E_s = exp\,(0.65 \times ln(E_C) + 0.3 \times ln(E_M) + 0.05 \times ln\,(E_S)) \qquad (1)$$

Where E_C, E_M, and E_S are the energy efficiency scores for CPU, memory, and storage workloads respectively.

D. Results

Table III summarizes the workload efficiency scores and idle power values for all scenarios illustrated in Fig. 4.

TABLE III
SERT Workload Efficiency Scores

Test Scenario	Workload Efficiency Score			Idle Watts	SERT 2 Efficiency Score	Inlet Temp. (C)
	CPU	Storage	Memory			
TS1	13.2	29.0	13.6	73.7	13.8	21.4
TS2	13.1	25.0	13.6	75.3	13.7	22
TS3	13.1	52.9	13.6	78.9	14.2	24.7
TS4	13.3	24.5	13.6	77.8	13.8	21.2
TS5	12.9	32.0	13.5	76.2	13.7	21.9
TS6	15.3	37.3	19.7	73.7	17.2	29.4
TS7	15.2	37.4	22.1	74.4	17.8	27.9
TS8	16.3	37.1	30.7	75	20.5	28.7
TS9	16	35.3	33.3	77.4	20.7	28.8
TS10	13.5	37.5	13.9	73.6	14.3	26.4
TS11	15.7	37.4	24.7	73.7	18.8	24.1
TS12	15.5	36.9	28.1	74.4	19.3	25.1
TS13	16.3	36.4	37.2	75.1	21.7	24
TS14	15.7	34.8	38.4	76.1	21.4	24.5
TS15	13.3	37.8	16.5	72.9	15	22.6
TS16	15.3	36.9	28.4	74.4	19.3	22.4
TS17	15	35.5	32.4	76.7	19.8	20.8
TS18	15.8	34.8	42.4	76.2	22.1	24.2
TS19	15.4	36.3	16.7	78.8	16.4	30.3
TS20	18.2	35.1	30.5	79.6	21.9	30.1
TS21	17.8	41.2	29.8	88.3	21.7	24.1
TS22	18.6	24.5	30.8	80.1	21.9	27

V. OBSERVATIONS & ANALYSIS

A. New vs Refurbished

The comparison of SERT 2 efficiency server scores between an all-new server – TS1 (13.8) – and an all refurbished one – TS5 (13.7) – for the same server make and model, indicates no statistically significant variation. The slight difference in scores can be attributed to manufacturing discrepancies and minor temperature changes in the benchmarking environment. Moreover, swapping in new components with refurbished components (TS2 to TS4) resulted in the same efficiency scores, with a standard deviation of 0.26 (2% variation).

B. Memory Configuration & Energy Efficiency

The effect of memory channels on the overall energy efficiency score can be seen in Fig. 5. The graph shows that efficiency scores of servers with the same number of channels populated are similar (average variation of 0.8%), regardless of the total memory capacity. Taking the case when two memory channels are utilized, for example, the overall efficiency scores using 16GB (scored 18.8) and 32GB DIMMs (scored 19.3) were very close (2.6% variation), even though using 16GB DIMMs provided a total capacity of 32GB whereas using 32GB DIMMs amounted to a total capacity of 64GB.

The graph also shows that populating all 4 channels with 1 DIMM per channel provides the best efficiency. Going beyond 1 DIMM per channel will result in slight efficiency degradation because multiple DIMMs will have to share the same channels.

To stress the importance of memory channel utilization, as opposed to memory capacity, on energy efficiency, Fig. 6 shows the difference in efficiency scores using 16GB and 32GB DIMMs for the same total memory capacity of 64 GB (7.2% variation), 128 GB (2.9% variation), and 192 GB (1.3% variation).

C. Memory Configuration & Idle Power

Idle power increased by approximately 4% when memory capacity increased from 32GB to 192GB using 16GB DIMMs. However, increasing the number of DIMMs while keeping the memory capacity constant did not affect the baseline idle power significantly.

Fig. 7 shows the variation of idle power for different memory combinations. As shown, the higher the capacity, the higher the idling power of the server. Increasing active state efficiency at the expense of idle power will result in a net increase in energy consumption over the server's useful lifetime within a production environment given the disproportionate amount of time the server spends in idle state.

D. Workload & CPU Efficiency

Fig. 8 demonstrates the load-interval efficiency scores for an HPE Gen 9 server. SERT load-interval is defined as the percentage of transactions executed per second from the maximum number of transactions. As shown in the graph, energy efficiency for the CPU worklets was highest at 75% load-interval (orange line), followed by 50% load-interval (gray), 100% load-interval (blue), and lowest at 25% load-interval (yellow) in most cases.

Energy efficiency scores doubled moving from 25% to 75% load-interval. This jump stresses the importance of workload management on the overall energy efficiency by increasing utilization levels. While achieving high utilization levels above 50% requires significant planning (particularly considering the impact on performance, response time, redundancy, etc.), utilization levels of 35% to 40% should be achievable with minimal investment.

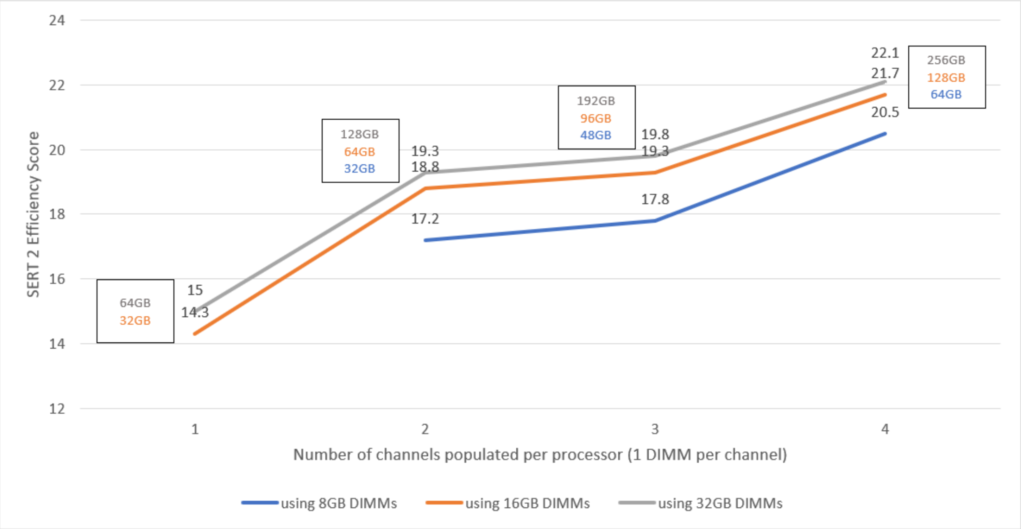


Fig. 5. Impact of the number of memory channels populated on SERT 2 energy efficiency score

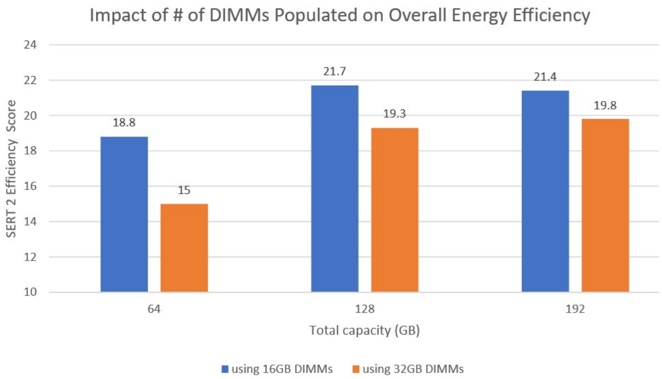
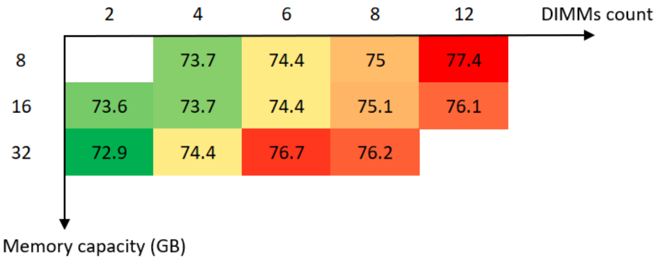


Fig. 6. Impact of # of DIMMs populated on SERT 2 energy efficiency score



Memory capacity (GB) \ DIMMs count	2	4	6	8	12
8		73.7	74.4	75	77.4
16	73.6	73.7	74.4	75.1	76.1
32	72.9	74.4	76.7	76.2	

Fig. 7. Idle power consumption (in watts) for different memory configurations (capacity vs DIMM count)

E. CPU & Energy Efficiency

Fig. 9 displays the energy efficiency scores of HPE Gen 9 and Gen 10 (latest at the time of conducting this research) servers with different processors and memory capacities. The processor specs are presented in Table IV.

The higher the number of cores (for the same CPU lithography), the higher the energy efficiency score of the server (as the score does not take into consideration idle performance – see figure 1), regardless of the processor make. For example, a Gen 9 with Intel Xeon E5 v4 processor (Q1 2016) outperformed a Gen 10 with Intel Xeon Skylake processor (Q3 2017).

Furthermore, adding 32GB RAM to a Gen 9 (total memory capacity of 64GB) with E5 v3 (22nm, Q3 2014) processor produced major efficiency improvements outperforming a baseline Gen 10 with a Skylake processor (14nm, Q3 2017) and 32GB total memory capacity. This observation emphasizes the opportunity discussed under rationale (section 2). Performance gains in the last three to five years were minimal. This meant a small change in memory configuration of an existing five- or less-year old server can lead to a performance boost, outperforming a newer server with less memory capacity and underutilized memory channels. This highlights the importance of evaluating existing server configurations as the first resort to increasing performance and efficiency as compared to a complete refresh with new servers. The economic and environmental opportunities are significant.

F. Storage Configuration & Energy Efficiency

The overall server efficiency score decreased from 21.9 (TS20) to 21.7 (TS21) when adding two additional 600GB HDDs. The additional hard drives caused the storage workload efficiency to increase, CPU workload efficiency to decrease, and idle power to increase significantly. Given the small weight of the storage workload efficiency on the server's overall efficiency score, this increase did not outweigh the decrease of the CPU workload, causing the server to be slightly less energy efficient with x4 600GB HDDs than x2 600GB HDDs. Moreover, using x2 1.2TB HDDs was as efficient as using x2 600GB HDDs.

However, this does not give the full picture. While CPU and memory configurations directly impact server performance per watt, increasing storage does not. It simply increases the server's ability to store more data. As such, storage should be optimally sized depending on the application needs to avoid energy waste.

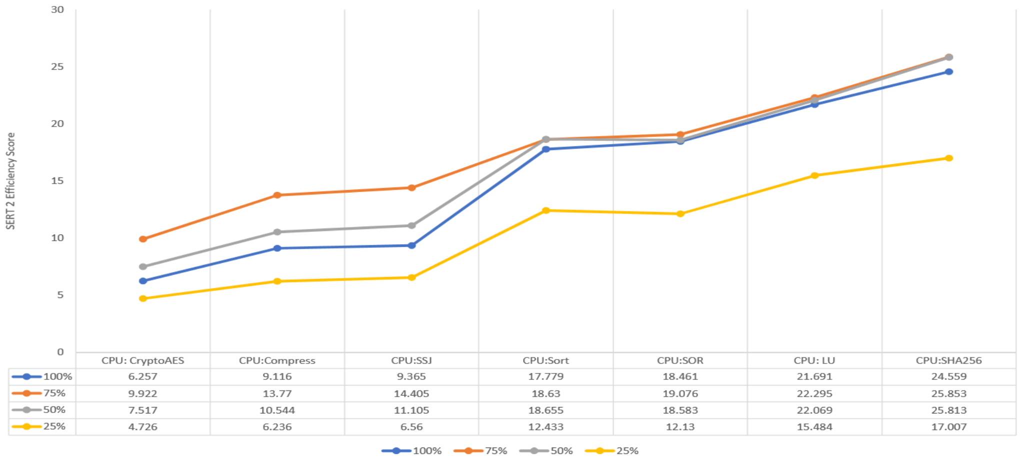


Fig. 8. Load level efficiency scores for CPU worklets

Table IV
CPU Specs

CPU model	Release date	Lithography (nm)	Cores	Frequency (GHz)
Intel Xeon E5-2690v3	Q3'14	22	12	2.6
Intel Xeon E5-2689v4	Q1'16	14	14	2.4
Intel Skylake Silver 4114	Q3'17	14	10	2.2
AMD EPYC 7401	Q3'17	14	24	2

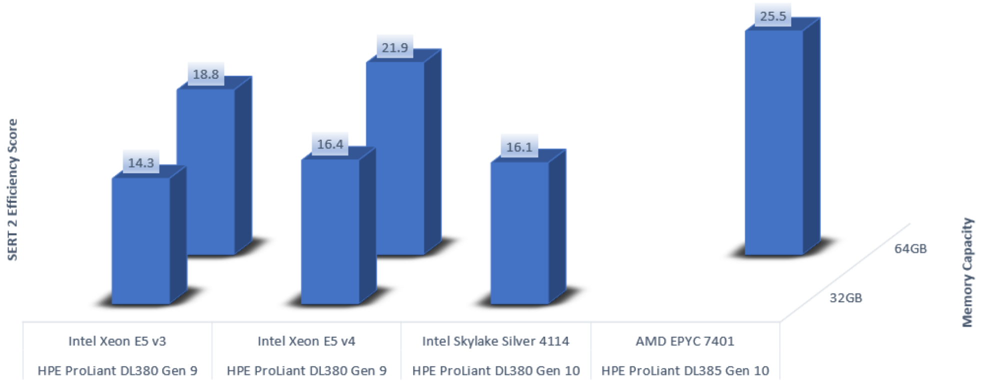


Fig. 9. Energy efficiency score of different servers/CPU and RAM configurations

VI. CASE STUDY

Following the observations above, there is a major opportunity for refreshing servers older than five years to provide boosts in performance for the same number of watts consumed. Yet, given the minor gains in CPU performance over the last three to four years, does it make sense to refresh servers that are released after 2015? The following case study was conducted to compare the impact of refreshing servers (new vs refurbished) on energy, cost, and the environment.

A. Analysis of Dual-socket Volume Servers (2010 - 2019)

To carry out this analysis, the SPECpower dataset containing server energy performance results based on the power_ssj2008 benchmark was analyzed. To track trends and eliminate outliers (e.g. high-end machines), only dual-socket servers were considered for trend consistency. These were then broken down into 18-month intervals (based on the published date of the server release date in SPECpower), roughly in line with Moore's law, and the performance averaged for each period. The results are shown in Table V.

TABLE V
Analysis of dual-socket volume servers (2010-2019)

Interval	Time intervals (1.5 years)	# of servers reported	Average watts @ 100% load	Average idle watts	Performance/power @100% load
1	2010/09 - 2012/02	30	247	80	3,648
2	2012/03 - 2013/08	73	253	63	5,277
3	2013/09 - 2015/02	18	245	55	9,791
4	2015/03 - 2016/08	13	270	46	12,710
5	2016/09 - 2018/02	21	398	57	12,754
6	2018/03 - 2019/08	34	387	67	15,335

B. *Workload Energy Consumption*

Based on the server models in Table V, the workload energy consumption for different scenarios was calculated, assuming a fixed workload ω of 200 million ssj_ops, under different deployment scenarios (varying PUE and utilization levels) according to Equation 2 [8].

$$E_\omega = \left(\sum_{m=1}^{\delta} \left(P_i^{sm}\alpha_{sm} + P_f^{sm}\beta_{sm}\right)\right) \times 8.76 \times PUE \qquad (2)$$

Where δ is the number of servers. β_{sm} is the average utilization rate of server and α_{sm} is the average active idle rate defined as 1- $\beta_{sm.}$ P_i^{sm} and P_f^{sm} are the average active idle and 100 percent capacity power (in Watts) respectively.

Equation 2 can be written as Equation 3 if an identical set of servers are used with a balanced load (same average utilization) and the same active idle P_i and 100 percent load power P_f.

$$E_w = \delta \times \left(P_i + \beta\left(P_f - P_i\right)\right) \times 8.76 \times PUE \qquad (3)$$

The results are shown in Table VI. Worst, Average and Best-case scenarios presented in Table VI are for illustration purposes as the PUE and utilization levels for such cases could vary depending on many other parameters (e.g. location, operating temperature, season, type of workload, etc.). Table VII shows the number of servers needed to run 200M ssj_ops for each scenario, calculated using Tables V and VI.

Running on 7.5-year-old servers (interval 1), the workload consumes 3,778 MWh/year of electricity in an on-premise environment, with an average scenario (10% utilization and PUE of 2). Running the same workload within the same environment using the latest servers (interval 6) drops the energy consumption to 580 MWh. This translates into a reduction of 85% in energy consumption. Similarly, running the same workload using servers that are 1.5 years old (interval 5), produced a reduction of 82% in energy consumption.

However, if existing servers are only 3 years old (interval 4), running this workload would consume 697 MWh within an average on-premise scenario. Yet, refreshing these machines to the latest servers (interval 6) would only lower the energy consumption by 16%, a significant reduction in gains compared to earlier scenarios. This is attributed to the slowdown in Moore's law and reduction in server dynamic range (section 2).

TABLE VI

Use phase annual energy consumption for 200M ssj_ops workload for various deployment scenarios (in MWh)

Environment	**Scenario**	**Utilization rate**	**PUE**	**Intervals**					
				1 (7.5Y old)	2 (6Y old)	3 (4.5Y old)	4 (3Y old)	5 (1.5Y old)	6 (Current)
On-premise (not virtualized)	Worst	5%	3	10,349	5,846	2,747	1,746	1,534	1,459
	Average	10%	2	3,778	2,202	1,049	697	629	580
	Best	25%	1.5	1,428	889	435	313	294	258
Colocation (not virtualized)	Worst	5%	2.5	8,624	4,872	2,289	1,455	1,279	1,216
	Average	10%	1.8	3,400	1,982	944	627	566	522
	Best	25%	1.3	1,238	770	377	271	255	224
On-premise (virtualized)	Worst	6%	3	8,788	4,998	2,356	1,512	1,337	1,263
	Average	30%	2	1,696	1,072	528	386	366	318
	Best	60%	1.5	882	592	298	231	226	189
Private cloud	Worst	7%	2.5	6,394	3,661	1,730	1,121	997	935
	Average	30%	1.8	1,527	965	475	347	330	286
	Best	60%	1.3	764	513	258	200	195	164
Public cloud	Worst	7%	2	5,115	2,929	1,384	897	798	748
	Average	40%	1.5	1,077	698	347	260	250	214
	Best	70%	1.1	606	412	208	163	160	134

TABLE VII

Number of servers needed to run 200M ssj_ops workload for various deployment scenarios

Environment	**Scenario**	**Utilization rate**	**PUE**	**Intervals**					
				1 (7.5Y old)	2 (6Y old)	3 (4.5Y old)	4 (3Y old)	5 (1.5Y old)	6 (Current)
On-premise (not virtualized)	Worst	5%	3	4,473	3,049	1,609	1,168	788	667
	Average	10%	2	2,237	1,524	804	584	394	333
	Best	25%	1.5	895	610	322	234	158	133
Colocation (not virtualized)	Worst	5%	2.5	4,473	3,049	1,609	1,168	788	667
	Average	10%	1.8	2,237	1,524	804	584	394	333
	Best	25%	1.3	895	610	322	234	158	133
On-premise (virtualized)	Worst	6%	3	3,728	2,541	1,341	974	657	556
	Average	30%	2	746	508	268	195	131	111
	Best	60%	1.5	373	254	134	97	66	56
Private cloud	Worst	7%	2.5	3,195	2,178	1,149	834	563	476
	Average	30%	1.8	746	508	268	195	131	111
	Best	60%	1.3	373	254	134	97	66	56
Public cloud	Worst	7%	2	3,195	2,178	1,149	834	563	476
	Average	40%	1.5	559	381	201	146	98	83
	Best	70%	1.1	320	218	115	83	56	48

C. Payback Point for refreshing to new vs refurbished

Payback time for refreshing servers (P_s) was calculated based on the cost of procuring new servers (the price of servers plus procurement costs) against energy-saving gains due to the reduction in energy consumption. This is captured in Equation 4.

$$P_S = \frac{(1+\theta)\times \sum_{m=1}^{\delta}(Q_m^s C_m^s)}{(\mathrm{E}_{\mathrm{old}} - \mathrm{E}_{\mathrm{new}})\times P_e} \tag{4}$$

Where δ is the total number of servers. Q_m^s is the quantity and C_m^s is the cost of servers to be procured. θ is the procurement overhead (assumed to be 15% in this case study). $\mathrm{E}_{\mathrm{old}}$ and $\mathrm{E}_{\mathrm{new}}$ are the annual energy consumption for running workload ω on the old and new (to be procured) servers, respectively. P_{e} is the price of energy per kWh (assumed to be USD 0.1 in this case study).

A representative server price (for volume dual-socket servers released in interval 6) is assumed to be $2,800 (by averaging the market price of a number of the servers covered in interval 6 at the time of conducting this analysis), while refurbished server price (for servers released during Interval 5) is approximated to $1,200 based on refurbished market data [17].

Tables VIII and IX show the payback time for refreshing servers of different ages with current (Interval 6) and refurbished servers (Interval 5), respectively. The result demonstrates a good economic case for refreshing servers with new ones if servers are older than 7.5 years (Interval 1). And an even stronger case for refreshing servers with refurbished ones for servers older than 6 years (Intervals 1 & 2), with return on investment as low as 1 year. Yet, payback for refreshing servers newer than 5 years old (Intervals 3 and 4) is very high for all cases, regardless of whether refreshed with new or refurbished servers.

Given that there is no financial benefit for refreshing servers that are newer than 5 years (Intervals 3 and 4), optimal hardware reconfiguration and proper utilization management become a significant option. These servers can be reconfigured to produce higher efficiency as shown in the experiments described in the previous section. Improving server's efficiency can be done by populating all memory channels, upgrading existing processors to more efficient ones (higher number of cores), and/or optimizing storage capacity.

TABLE VIII
Payback point in years after refreshing to new Interval 6 generation servers

Environment	**Scenario**	**Utilization rate**	**PUE**	**Intervals**			
				1 (7.5Y old)	2 (6Y old)	3 (4.5Y old)	4 (3Y old)
On-premise (not virtualized)	Worst	5%	3	2 🟢	5 🟡	17 🔴	75 🔴
	Average	10%	2	3 🟢	7 🔴	23 🔴	92 🔴
	Best	25%	1.5	4 🟡	7 🔴	24 🔴	79 🔴
Colocation (not virtualized)	Worst	5%	2.5	3 🟢	6 🔴	20 🔴	90 🔴
	Average	10%	1.8	4 🟡	7 🔴	25 🔴	102 🔴
	Best	25%	1.3	4 🟡	8 🔴	28 🔴	91 🔴
On-premise (virtualized)	Worst	6%	3	2 🟢	5 🟡	16 🔴	72 🔴
	Average	30%	2	3 🟢	5 🟡	17 🔴	53 🔴
	Best	60%	1.5	3 🟢	4 🟡	16 🔴	43 🔴
Private cloud	Worst	7%	2.5	3 🟢	6 🔴	19 🔴	82 🔴
	Average	30%	1.8	3 🟢	5 🟡	19 🔴	59 🔴
	Best	60%	1.3	3 🟢	5 🟡	19 🔴	50 🔴
Public cloud	Worst	7%	2	4 🟡	7 🔴	24 🔴	103 🔴
	Average	40%	1.5	3 🟢	6 🔴	20 🔴	58 🔴
	Best	70%	1.1	3 🟢	6 🔴	21 🔴	52 🔴

TABLE IX
Payback point in years after refreshing to refurbished Interval 5 generation servers

Environment	**Scenario**	**Utilization rate**	**PUE**	**Intervals**			
				1 (7.5Y old)	2 (6Y old)	3 (4.5Y old)	4 (3Y old)
On-premise (not virtualized)	Worst	5%	3	1 🟢	3 🟢	9 🔴	51 🔴
	Average	10%	2	2 🟢	3 🟢	13 🔴	80 🔴
	Best	25%	1.5	2 🟢	4 🟡	15 🔴	120 🔴
Colocation (not virtualized)	Worst	5%	2.5	1 🟢	3 🟢	11 🔴	62 🔴
	Average	10%	1.8	2 🟢	4 🟡	14 🔴	89 🔴
	Best	25%	1.3	2 🟢	4 🟡	18 🔴	138 🔴
On-premise (virtualized)	Worst	6%	3	1 🟢	2 🟢	9 🔴	52 🔴
	Average	30%	2	1 🟢	3 🟢	11 🔴	94 🔴
	Best	60%	1.5	1 🟢	2 🟢	13 🔴	169 🔴
Private cloud	Worst	7%	2.5	1 🟢	3 🟢	11 🔴	63 🔴
	Average	30%	1.8	2 🟢	3 🟢	12 🔴	104 🔴
	Best	60%	1.3	2 🟢	3 🟢	14 🔴	195 🔴
Public cloud	Worst	7%	2	2 🟢	4 🟡	13 🔴	78 🔴
	Average	40%	1.5	2 🟢	3 🟢	14 🔴	137 🔴
	Best	70%	1.1	2 🟢	3 🟢	16 🔴	260 🔴

D. Environmental Impact of Manufacturing and Recycling a Server

Table X shows the production impact of a typical rack server on 15 environmental impact parameters, classified under three main categories according to a study published in July 2015 [18].

While this is a reliable and widely used study, other studies reported higher figures due to variations in manufacturing technologies and the fact that the carbon and materials cost for ICT is on an upward trend. Dell published a series of carbon footprint reports on its servers in early 2019 [19]. Analysis of these would put the carbon cost of manufacture significantly higher than [18]. The Dell carbon footprint data allocates between 9.5% and 22.5% to the manufacturing of rack servers. CO_2e set against manufacture ranges from 1140.56 kg CO_2e to 1782.2 kg CO2e, with a mean average of 1333.4kg CO_2e. While much work is still being carried out in this area, studies like this do suggest that the benefits of avoiding manufacturing are more significant than originally thought.

As for the material breakdown, servers contain a high proportion of steel, aluminum, and plastic; three of the top materials for industrial greenhouse gas emissions worldwide. Even though current legislations impose proper disposal and recycling of e-waste, some materials are non-recyclable and recyclable ones will not be fully recoverable. Table XI shows the number of recoverable materials following End-of-Life (EoL) scenarios 2 and 3, where scenario 2 represents servers that are recycled after some parts are manually separated and treated (like batteries, hard drives, etc..) and scenario 3 represents servers that are recycled without any previous treatment [20].

TABLE X
Lifecycle impact (per unit) of Rack Server during the production phase

Resources Use and Emissions	Unit	Material	Manufa.	Total
Other Resources & Waste				
Total Energy	MJ	8451	552	9002
Of which, electricity	MJ	5809	245	6053
Water (process)	ltr	1730	16	1746
Water (cooling)	ltr	560	142	702
Waste, non-haz./landfill	g	36016	2011	38027
Waste, hazardous/ incinerated	g	2214	4	2218
Emissions (Air)				
Greenhouse Gases in GWP100	Kg CO2 eq.	475	33	508
Acidification	g SO2 eq.	3747	154	3901
Volatile Organic Compounds (VOC)	g	19	2	22
Persistent Organic Pollutants (POP)	ng i-Teq	442	45	487
Heavy Metals	mg Ni eq.	1435	105	1540
PAHs	mg Ni eq.	493	3	497
Particulate Matter (PM, dust)	g	2506	31	2538
Emissions (Water)				
Heavy Metals	mg Hg/20	744	3	748
Eutrophication	g P04	12	1	13

There is a real need to look at the limitations of electronic recycling as it is currently not the answer to the growing crisis of e-waste. Servers typically have a short refresh cycle, lasting on average 3-5 years but sometimes as little as one year, due to various factors including maintenance and lease contracts. Thus, it is essential to have effective server lifecycle management in place and use them for as long as possible to minimize irrecoverable damage to the environment.

Table XI
Description of the materials and their quantities in servers, and the recycled amounts under EoL scenarios 2 and 3 [20]

Component/Material	Amount in server (g)	Recycling EoL 2 (g)	Recycling EoL 3 (g)
Aluminium	1,263	1,185.1	1,149.07
Brass	7	6.65	4.9
Copper	806.56	747.98	483.72
Steel	14,861	13,996.18	13,970.21
Ferrous metals	216	151.11	151.11
Zinc	96	67.2	57.6
ABS	360	266	266.4
EVA	75	0	0
HDPE	210	97.76	0
PBT	240	0	0
PC	289	0	0
PCABS	324.28	0	0
PCFR40	51	0	0
PCGF	52	0	0
PUR	2	0	0
PVC	145	0	0
Styrofoam	1026	0	0
Synthetic rubber	35	0	0
Other materials			
Cables	31	7.4	7.4
Electronics	3,966	596.12	444.32
Paper	3,629	0	0
Others (solder)	2	0	0
Neodymium magnets	68	0	0
Batteries	44.6	20	20
Total	27,799	17,142	16,555

VII. STUDY LIMITATIONS

The energy efficiency scores used in this study are based on the SERT benchmark which is tailored more to measure the performance of servers with a transactional type of workloads. This score might not be equally representative of the performance of other types of workloads. For example, a server deployed mainly for memory-intensive tasks, will not do much computational work. However, given that the CPU represents the largest proportion of the energy consumption of the server, and that interactive transactional workloads are dominant in the market, this should not influence the findings of this study.

Additionally, SERT results for refurbished/remanufactured kits are sensitive to the quality of refurbishment/remanufacturing. For example, running SERT on a server with over-tightened heatsinks, causing improper heat dissipation, can result in very low SERT scores. As such, care was taken throughout the experiments to ensure that refurbished and remanufactured servers used for benchmarking were built following relevant best guidelines and procedures to ensure consistency.

The data used in the experiments from the SPECpower database contains benchmarking results from high-end as well as low-end servers. To ensure consistency, only dual-socket servers were selected, which eliminated any potential outliers influencing trend analysis.

Another parameter to consider is the impact of the ambient temperature of the different SPECpower experiments reported in the database on server energy performance [21]. This has been addressed by using averages across multiple servers (ranging between 18 and 61 servers per data point) in the analysis, which is representative of performance across the SPECpower allowable temperature range.

Furthermore, the slowdown in Moore's Law influenced the findings of this study and strengthened the case for the use of refurbished servers to eliminate inefficient equipment older than 5 years. However, the microprocessor landscape is an ever-moving picture with 7nm lithography just introduced by AMD. This could give Moore's Law a new lease of life for the next few years. However, the physical, as well as economic limits are being reached for the current approach to processor design.

Another potentially viable way to increase energy efficiency would be to replace general-purpose CPUs with more domain-specific ones (e.g. TPUs, GPUs, ASICs, etc.) when applicable (e.g. AI applications, hashing, rendering, etc.). However, this is outside the scope of this work.

VIII. Conclusion and Future Work

Concerns over carbon footprint and climate change have led to a rise in awareness and demand for increasing the energy efficiency of computing [22], whilst at the same time reducing its environmental impact. Recent studies show that the past decade witnessed significant improvements in data center efficiency, which is attributed to improved cooling designs and optimized power management, among other factors. Many data centers are following best practices to increase the energy efficiency of their facilities; this can be seen in the substantial drop in power usage effectiveness (PUE) between 2007 and 2014 [7]. However, the same cannot be said about IT equipment, where 66% of facility energy is consumed by servers producing 7% of the compute capacity [11].

This work has highlighted the savings that could be achieved by utilizing refurbished servers for refresh cycles and remanufacturing practices to enhance the performance and efficiency of younger servers. Particularly, the work demonstrated how the use of refurbished equipment could create a viable economic case to refresh servers that are five to six-year-old. As demonstrated in Tables VIII and IX, refreshing such servers would only make an economic sense when refurbished equipment is used.

Furthermore, this study reveals the significant impact of server remanufacturing and reconfiguration on younger servers through the use of balanced memory configurations, upgraded processor technology, and storage reconfiguration on the overall efficiency score. This presents a strong opportunity to increase the efficiency of younger servers that do not make a business case for a full refresh (where we found the threshold to be servers younger than five to six-year-old).

Not only does reusing servers save a significant amount of landfill and reduced waste and toxic emissions produced during the manufacturing phase of servers, employing professionally remanufactured servers can be more energy efficient than using the latest generation of servers if configured properly. This performance gain can be attributed to the slowdown in Moore's law and the fact that newer servers are not maintaining the same efficiency improvements seen in the past.

Finally, going forward, this work will be expanded to check the impact of other server features (e.g. optimal BIOS energy settings) and ambient parameters (e.g. inlet temperature vs leakage current vs fan speed) on overall server energy consumption in production environments. Additionally, the cost analysis will be expanded to include other parameters such as maintenance and support costs.

Acknowledgment

This research is co-funded by InnovateUK as part of the Knowledge Transfer Partnership program under partnership number 11485.

References

[1] Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper, Feb. 2019 URL: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[2] World Internet Usage and Population Statistics, 2019 mid-year estimates. Internet World Stats. URL: http://internetworldstats.com/stats.htm [last accessed: January 2020]

[3] European Commission FP7 PEDCA Pan-European Data Centre Academy – Final Project Report. January 2015.

[4] European Commission H2020 DC EURECA Project – Final Project Report. April 2018.

[5] Greenpeace. Powering the cloud: How China's internet industry can shift to renewable energy. 2018

[6] PUE: A Comprehensive Examination of the Metric. The Green Grid. 2012. Accessed 25 March 2015. http://www.thegreengrid.org/~/media/WhitePapers/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf?lang=en

[7] Andy Lawrence, "Uptime Institute global data center survey 2019." Uptime Institute Intelligence, Report 26, April 2019. URL: https://uptimeinstitute.com/2019-data-center-industry-survey-results

[8] R. Bashroush, "A Comprehensive Reasoning Framework for Hardware Refresh in Data Centers." IEEE Transactions on Sustainable Computing, 2018. DOI: https://doi.org/10.1109/TSUSC.2018.2795465

[9] Jonathan Koomey, Stephen Berard, Maria Sanchez and Henry Wong. 2010. "Implications of historical trends in the electrical efficiency of computing." IEEE Annals of the History of Computing vol. 33(3): 46-54. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.323.9505&rep=rep1&type=pdf

[10] K. D. Lange, "Identifying shades of green: The SPECpower benchmarks," Comput., vol. 42, no. 3, pp. 95–97, 2009.

[11] Rabih Bashroush, Andy Lawrence, "Beyond PUE: Tackling IT's wasted terawatts.", Uptime Institute Intelligence, Report 34, January 2020. URL: http://go.the451group.com/K00PhDY00e9313obt0Fg0Y0

[12] Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, and William Lintner. 2016. United States data center energy

usage report. https://eta.lbl.gov/publications/united-states-data-center-energy

[13] Standard Performance Evaluation Corporation, SERT 2, 2020. URL: https://www.spec.org/sert2/

[14] SPECpower Committee, Server Efficiency Rating Tool (SERT) Design Document 2.0.2, March 2019. URL: https://www.spec.org/sert2/SERT-designdocument.pdf

[15] SPEC SPECpower_ssj2008 Committee, Server Efficiency Rating Tool (SERT) User Guide 2.0.2, March 2019. URL: https://www.spec.org/sert2/SERT-userguide.pdf

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

[17] Quote from Techbuyer as of February 2020. URL: https://www.techbuyer.com/

[18] Bio by Deloitte, Preparatory study for implementing measures of the Ecodesign Directive 2009/125/EC DG ENTR Lot 9 - Enterprise servers and data equipment - Task 5: Environment & Economics, Jul.

[19] Dell Carbon Footprint Assessments. URL: https://www.dell.com/learn/us/en/uscorp1/corp-comm/environment_carbon_footprint_products

[20] Laura Talens Peiro, Fulvio Ardente September 2015 Environmental Footprint and Material Efficiency Support for Product Policy. URL: https://publications.jrc.ec.europa.eu/repository/bitstream/JRC96944/lb-na-27467-en-n%20.pdf

[21] J. Sarkinen, R. Brännvall, J. Gustafsson, and J. Summers, ‘Experimental Analysis of Server Fan Control Strategies for Improved Data Center Air-based Thermal Management.’ In proceedings of the Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2020.

[22] R. Bashroush, E. Woods, A. Noureddine. ICT Energy Demand: what got us here won’t get us there! IEEE Software, Institute of Electrical and Electronics Engineers, 2016, 33 (2), pp.18 - 21. ff10.1109/MS.2016.53ff. ffhal-01897887

Rabih Bashroush received a BEng degree in computer and communications engineering from AUB, in 2001 and a PhD degree in systems engineering from Queen’s University Belfast, in 2005.

He is the Director of Research at Uptime Institute and holds a faculty position with UEL. Before joining Uptime, he spent two decades holding various positions in academia (e.g. QUB and CMU) as well as industry (e.g. Philips Research and Danfoss).

Dr. Bashroush was named in the UK Universities top 100 Best Breakthrough list for his work on energy conservation in ICT and was the recipient for the Industry Initiative of the Year DCD Global Awards 2018.

He advises various governments (e.g. EU Commission, UK, etc.) and influenced several international policies and legislation (e.g. EcoDesign, EMAS, and GPP) on ICT energy conservation. He serves on the main standardization committees for CEN/CENELEC/ETSI as well as ISO/IEC. He serves as an editorial board member on various journals, and on the steering committee of the IEEE Transactions on Sustainable Computing journal. He has been a member of the IEEE and IEEE Computer Society since 1998.

Nour Rteil received a BEng degree in computer and communications engineering from AUB in 2015.

She worked for 4 years as a Solution Developer at Dar Al Handasah, primarily developing web solutions using ASP.net. Before working as a full-time employee, she did an internship at the Technical University of Dresden in 2014, and an internship at Ericsson in 2015. In 2019, she joined the University of East London and Techbuyer as a Knowledge Transfer Partnership (KTP) associate working on a tool to model the energy efficiency of servers using optimal hardware configuration.

Richard Kenny received a BA in International Development, specializing in Economics, from UEA in 2009 and postgraduate qualifications in Law specializing in Data Protection and Information Governance from Northumbria University in 2019.

He has worked as a Technical Project Manager since 2009 primarily in software development and information security. Since 2017 he has been working for Techbuyer where he is the Group IT Director. He introduced various ICT refurbishment programs and solutions at Techbuyer addressing data center as well as end-user devices.

Astrid Wynne is the sustainability lead at Techbuyer. She works on the Interreg funded CEDaCI (Circular Economy in the Data Centre Industry) project. She has submitted evidence to the Circular 4.0: Data-driven Intelligence for a Circular Economy lead by the University of Exeter, the Climate Protection Potentials of Digital Transformation (CliDiTrans) at the Bordersteps Institute in Germany, and the Environmental Audit Committee’s inquiry into electronic waste and the circular economy. She writes on digital sustainability for the industry press with a particular focus on materials usage and the embodied energy in ICT.