A Comprehensive Reasoning Framework for Hardware Refresh in Data Centers

Rabih Bashroush, *Member, IEEE*

Abstract—With the increased trend of moving towards the consolidation of computing in larger facilities, and with the rise of paradigms such as Cloud, Smart Cities, and IoT, data centers have been highlighted as a major energy consumer. Yet, data centers exist purely to host IT services, which on average tend to account for the largest part of the overall facility energy consumption. Frequent refresh of IT hardware has emerged as a trend in hyper-scale data centers. However, the wider environmental impact of hardware refresh has become a concern. This work provides a comprehensive framework that helps identify the energy saving opportunities, while demonstrating the overall environmental impact related to hardware refresh. The work sheds new light on key relationships such as the one between hardware utilization and Power Usage Effectiveness (PUE) to drive efficiency. Various data center deployment scenarios are used as case studies (based on real-life datasets) to validate the proposed concepts.

Index Terms—Data centers, energy efficiency, hardware refresh rate, environmental impact

1 Introduction

Perhaps many of the world's population are unaware how much our everyday economic and social wellbeing depends on reliable, secure and efficient data centers and the services they provide. Data centers are facilities that provide the connectivity hubs, power distribution, operational environment and physical security for some of the critical equipment needed to support our digital age.

Indeed, we are entering a new era of digital activity. An explosive growth in data fuelled by the proliferation of paradigms such as the Internet of Things, Cloud, and Smart Cities, paired with improved connectivity (with 5G on the horizon) is making the digitalisation of our economy and social interaction move at an unprecedented pace.

Yet, this has created a new challenge. The level of energy consumption and the environmental impact of data centers have been highlighted as major concerns [1]. For example, the energy consumption of data centers in Western Europe was estimated to be 86 TWh in 2013 (3 percent of annual electricity consumption of Western Europe) and was projected to increase to 104 TWh by 2020 [2]. Nonetheless, this trend was recently predicted to slow down in some parts of the world [3].

Over the last decade, and with the emergence of metrics such as the Power Usage Effectiveness (PUE – the ratio of the total facility load divided by the IT load) [4], tangible progress has been made in increasing the effectiveness of the data center cooling and power infrastructure, with PUE of near 1.01 reported in some cases [5]. However, PUE does not factor in the efficiency of IT, which is ultimately the purpose of the data center. Less attention has been paid in the industry to the interplay between the useful workload and facility to drive true efficiency [6]. In IT, work has focused largely on consolidation and resource throttling [7] in order to increase utilization [8], though recently addressing energy efficiency in software design [9], [10] has started to attract more attention.

With the current increased interest in adopting Life Cycle Assessment (LCA) approaches [11] (which factor in the environmental impact of the manufacturing, transport, usage and end-of-life of equipment), there is a lack of understanding in industry as to how this would apply to data centers, and particularly to IT hardware refresh, a rising trend in hyper-scale data centers [12]. The question becomes; Is it better for the environment to keep an old existing IT kit, or refresh? When is the right time to refresh? Do we need to follow a full LCA to identify ideal refresh intervals? To conduct an LCA exercise within a data center environment, one would need to obtain the details of the various existing equipment elements so that analysis can be conducted. This is a major exercise and could be a challenge, or even a barrier, to many large data centers (comprising thousands of servers), that have a range of equipment, coming from different vendors and with varying age. In many cases, such information is not realistically available.

This work addresses these questions and provides a comprehensive reasoning framework to help guide hardware refresh decision making, presenting a major energy and environmental conservation opportunities. Previous work has highlighted the importance of IT lifecycle management [13], where for example [14] used exergy analysis for lifecycle assessment of a computer server. However, no work has been done to provide in-depth analysis to guide decision making and help identify the environmental saving opportunities, without the need for a full LCA. More recent

• *The author is with the School of Architecture, Computing, & Engineering, University of East London, London E16 2RD, United Kingdom. E-mail: R.Bashroush@qub.ac.uk.*

Manuscript received 19 June 2017; revised 13 Jan. 2018; accepted 16 Jan. 2018. Date of publication 18 Jan. 2018; date of current version 6 Dec. 2018. (Corresponding author: Rabih Bashroush.)
Recommended for acceptance by R. G. Melhem.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TSUSC.2018.2795465

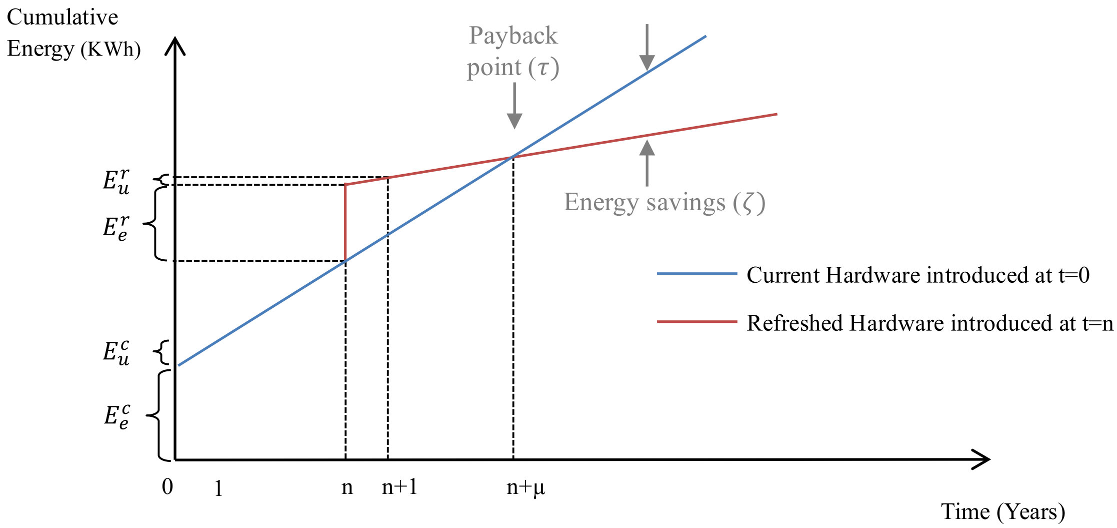

Fig. 1. Cumulative energy consumption of existing hardware versus refreshed hardware.

work has considered a similar breakeven approach [15]. In this work, three case studies were discussed, varying from changes in the technology node and area of the CPU and caches, to guiding main memory refresh, while taking into consideration environmental savings.

The next section describes the methodology used to calculate potential use phase energy savings due to hardware refresh, factoring in the server embodied energy (as the other major source of energy consumption in addition to use phase energy), and excluding wider LCA energy elements to preserve simplicity. In Section 3, these equations are used to develop the reasoning framework for enterprise servers, where also the relationship between hardware utilisation and PUE is formulated. Then, several case studies representing various data center deployment environments are discussed in Section 4. In Section 5, the wider LCA environmental analysis is conducted, demonstrating the validity of the reasoning framework in not only saving energy, but also reducing the overall environmental impact. Section 6 then highlights the study limitations, with Section 7 rounding off with a conclusion and future work.

2 Methodology

To identify the optimal time interval (if it exists) for hardware refresh rates, consider Fig. 1 below. The diagram shows the cumulative energy consumption of two scenarios. The first scenario (blue line) shows the energy consumption of existing hardware. The second scenario (orange line) shows the cumulative energy consumption of refreshed hardware at a point n in time.

In this scenario, the embodied energy (energy consumed to produce the hardware) of current hardware (in KWh), introduced at time $t = 0$, is designated as E_e^c; and the annual use phase energy (in KWh) as E_u^c. Similarly, for the refreshed hardware, introduced at time $t = n$, E_e^r and E_u^r represent the embodied and annual use phase energy, respectively. Hardware disposal energy (end of life) is very small compared to embodied and use phase energy (see Section 5) and was not included in the model. Additionally, when the hardware is recycled or redeployed, disposal energy is considered net energy saving.

The point of intersection between the two scenarios (represented as $n + \mu$ in Fig. 1) is calculated. The intersection point, if exists, represents the point in time where the refreshed hardware and existing hardware would have the same cumulative energy consumption. Beyond that point, energy savings will be accrued (compared to no hardware refresh). To find the intersection point, the equations of the two scenario lines need to be calculated.

For current hardware, the cumulative energy line (blue) passes through two points with coordinates (0, E_e^c) and (1, $E_e^c + E_u^c$) as shown in Fig. 1. Accordingly, the equation of the line can be calculated as:

$$E^c(t) = E_u^c\, t + E_e^c, \tag{1}$$

where $E^c(t)$ is the cumulative energy consumption (in KWh) for the current hardware at a point in time $t \geq 0$.

Similarly, we calculate the equation of the refreshed hardware cumulative energy line (orange), which passes through the two points (n, $E_e^c + E_e^r + nE_u^c$) and (n + 1, $E_e^c + E_e^r + nE_u^c + E_u^r$), to be:

$$E^r(t) = E_u^r\, t + n\left(E_u^c - E_u^r\right) + E_e^c + E_e^r, \tag{2}$$

where $E^r(t)$ is the cumulative energy consumption for the refreshed hardware (in KWh) at a point in time $t \geq n$, and n is the refreshed hardware introduction time (in years).

Given the above two Equations (1) and (2), the intersection point of the two lines can be calculated by setting $E^c(t) = E^r(t)$, which gives

$$E_u^r\, t + n\left(E_u^c - E_u^r\right) + E_e^c + E_e^r = E_u^c\, t + E_e^c, \tag{3}$$

solving for t, the intersection payback time, τ, is found to be

$$\tau = n + \frac{E_e^r}{E_u^c - E_u^r}. \tag{4}$$

Expressing the intersection time τ in terms of n (refreshed hardware introduction time)

$$\tau = n + \mu, \tag{5}$$

where μ is the time interval after which efficiency is achieved by the newly refreshed hardware (payback time). Then, replacing (5) in (4) produces

$$\mu = \frac{E_e^r}{E_u^c - E_u^r}. \tag{6}$$

Equation (6) shows the amount of time needed to start saving energy after a hardware refresh (factoring in embodied energy). The first observation in (6) is that it makes no reference to the embodied energy of existing hardware, E_e^c. Thus, deciding on an optimal hardware refresh rate does not require knowledge of the embodied energy of existing hardware as currently widely believed. Indeed, this is referred to as the sunk cost fallacy in behavioural economics [16].

Similarly, the amount of energy saved, ζ, due to hardware refresh can be calculated based on the difference between $E^c(t)$ and $E^r(t)$ over a time period $\sigma \geq \tau$, from Equations (1) and (2), as

$$\zeta_\sigma = E^r(t) - E^c(t) = \left(E_u^c - E_u^r\right)(\sigma - n) - E_e^r. \tag{7}$$

As with Equation (6), Equation (7) shows that energy savings are also independent of the embodied energy of current hardware E_e^c.

3 Application – Enterprise Servers

This section builds on the equations defined in the previous section to develop a reasoning framework for identifying optimal enterprise server refresh rates given specific run-time parameters. For this, server embodied energy and server use phase energy are calculated. Then, the refresh rate is analysed based on current server technology trends. Finally, a more accurate way to calculate efficiency (compared to using PUE [17], [18]) is discussed.

3.1 Server Embodied Energy

Over the past few years, there have been various estimates released about the embodied energy of servers in different blogs, reports and white papers [19], [28]. Yet, most of these figures are estimates, and without data from vendors, it is difficult to produce a reliable figure. However, generating such figures by vendors is also not a precise science and requires a lot of estimations about the manufacturing process.

For the purpose of this work, the study by InnovateUK [20] is considered. The report first estimates the embodied energy of servers based on an earlier work on Personal Computers (PC) [21], but takes into consideration the differences between servers and PCs (e.g., extra disk and more RAM, but no additional cards like graphics and audio, and no monitors in servers). Based on this, a figure of 1,300 KWh (for 2003 manufacturing processes) is calculated. Additionally, the report cites a figure by Fujitsu of 730 KWh (in 2008), and concludes by adopting an average figure of 1,000 KWh.

Hence, a fixed figure for server embodied energy of 1,000 KWh is adopted (after further consulting with HP, DELL, and IBM, as the major server manufacturers). This is considered high enough to be representative of most servers, especially considering the rise of unbranded servers or so-called ‘White Label’ servers, which are built to order in large volumes (less packaging, paint, etc.) and vanity-free servers such as Open Compute servers [22]. Thus,

$$E_e^s = 1{,}000\ KWh \tag{8a}$$

$$E_e^r = \eta \times E_e^s \tag{8b}$$

where,

E_e^s is the embodied server energy, and

η is the number of new servers to be procured (the refresh hardware).

3.2 Use Phase Energy

A server use phase energy, including the facility footprint, is calculated using the following equation:

$$E_u^s = AE_u^s \times PUE \tag{9}$$

where,

E_u^s: Annual energy consumption of a server in KWh

AE_u^s: Annual energy consumption of server while running at the average utilization level in KWh

PUE: Facility Power Usage Effectiveness [4], which indicates the facility overhead for running the server (to cover cooling, power infrastructure, etc.).

AE_u^s can either be physically measured at the facility (if this is existing hardware) or estimated using the following equation

$$AE_u^s = (P_i^s\,\alpha + P_f^s\beta) \times 8.76 \tag{10}$$

where, P_i^s and P_f^s are the average power consumption (in watts) of a server running active idle and at 100 percent capacity, respectively. These figures can mostly be found in published benchmarks (SPECpower [23]) or measured for a particular model (e.g., using SERT [24]). It is important to ensure that the comparison (between current and new hardware) is made using comparable workload. For example, if SPEC benchmarks are used, it is operations/watt that should be normalised (see Section 4). β is the annual average utilization of a server, and α the annual average active idle $(\alpha + \beta = 1)$. 8.76 is used to convert the figure from watts to KWh (multiplying by 8760 hrs/year and dividing by 1000). Equation (10) assumes linear power proportionality, which is generally the case as per SPECpower data.

PUE is calculated by measuring the average annual total facility energy consumption, divided by the IT load energy consumption [4], as

$$PUE = \frac{Total\ facility\ load}{IT\ load}. \tag{11}$$

3.3 Refresh Rate Analysis

As shown in Equation (6), the payback time interval, in terms of energy savings due to hardware refresh, is only dependent on the embodied energy of the new hardware (8), and the difference between the use phase energy in the current and to be installed hardware. Thus, to calculate the optimal refresh rate interval, the relationship between the current and to be installed hardware use phase energy needs to be identified.

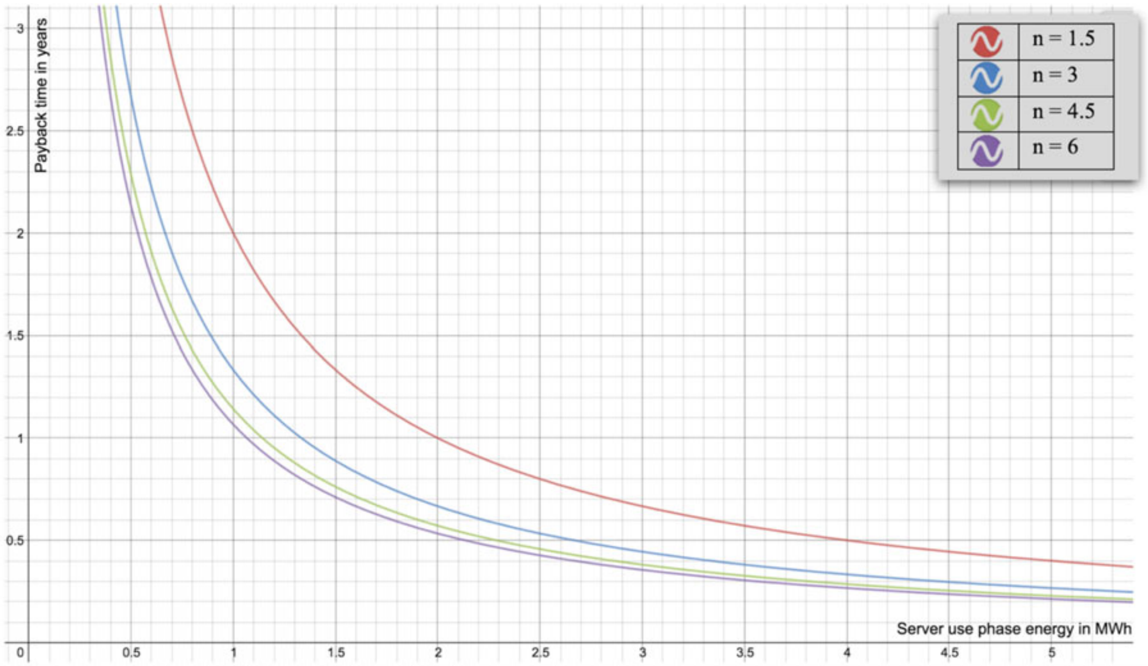

Fig. 2. Payback time after hardware refresh (one server) introduced at point n in time (in intervals of 1.5 years) vs. current annual server use phase energy in MWh, for $\mathrm{n} = 1.5$, 3, 4.5, and 6.

Currently, there are two main models used when predicting performance evolution of servers. The first referred to as Moore's Law [25], which is interpreted as the doubling in performance (based on the number of transistors built onto the processor chip), every 18 months. The second, Koomey's Law [26], indicates that the performance per watt doubles every 1.57 years (using mathematical regression on historical power and performance data). Given the similarity between the two models, and for simplicity, a doubling in performance per watt every 1.5 years will be assumed. Thus, for a fixed workload, the energy consumption will halve every 1.5 years. Accordingly, E_u^r can be expressed in terms of E_u^c as

$$E_u^r = \frac{E_u^c}{2^{\left(\frac{n}{1.5}\right)}}. \tag{12}$$

where n is the time in increments of 1.5 years.

Substituting (12) in (6),

$$\mu = \frac{E_e^r}{E_u^c}\left(1 - \frac{1}{2^{\left(\frac{n}{1.5}\right)}}\right)^{-1} \tag{13a}$$

For analysing one server, (8) can be substituted in (13a) to reduce the number of variables to two: E_u^c and n, producing:

$$\mu = \frac{1000}{E_u^c}\left(1 - \frac{1}{2^{\left(\frac{n}{1.5}\right)}}\right)^{-1} \tag{13b}$$

Equations (13a) and (13b) can help identify the time it takes to recoup the embodied energy incurred by the introduction of new hardware. Fig. 2 below shows Equation (13b) plotted for various values of n (1.5, 3, 4.5, and 6 years).

As can be seen from Fig. 2, the payback time on hardware refresh is below one year for servers with annual use phase energy consumption of 1 MWh or more.

To better understand the energy saving opportunities, Equation (12) is substituted in (7) to produce Equation (14a) below, followed by a one server version (14b).

$$\zeta_\sigma = E_u^c\left(1 - \frac{1}{2^{\left(\frac{n}{1.5}\right)}}\right)(\sigma - n) - E_e^r \tag{14a}$$

$$\zeta_\sigma = E_u^c\left(1 - \frac{1}{2^{\left(\frac{n}{1.5}\right)}}\right)(\sigma - n) - 1000 \tag{14b}$$

Equations (14a) and (14b) can calculate the energy savings over a period of time σ given hardware refresh rate n, and current energy consumption E_u^c.

Fig. 3 shows Equation (14b) plotted for various values of n (1.5, 3, and 4.5 years) looking at potential savings over a 5-year window ($\sigma = 5$).

For example, Fig. 3 shows that for a server with E_u^c of 1 MWh, delaying hardware refresh from 1.5 years to 3 years could yield an overall energy waste of 0.25MWh over 5 years. On the other hand, there would be no energy benefit in refreshing a server with E_u^c of less than 0.5MWh over 5 years (given the three refresh intervals in Fig. 3). More details are discussed in Section 4 for various data center types and hardware utilization scenarios.

3.4 Efficiency Calculation

To be able to address efficiency, the facility energy consumption needs to be related to the useful IT load performed. Otherwise, we risk referencing PUE as an efficiency metric, which is evidently not [17], [18]. For this, consider a workload, ω, which runs over ϑ servers, where each server s_m has an average utilization of β_{s_m} and active idle and 100 percent capacity power as $P_i^{s_m}$ and $P_f^{s_m}$, respectively. Based on Equations (9) and (10), the use phase energy consumption, E_u^ω, of workload ω, can be calculated as

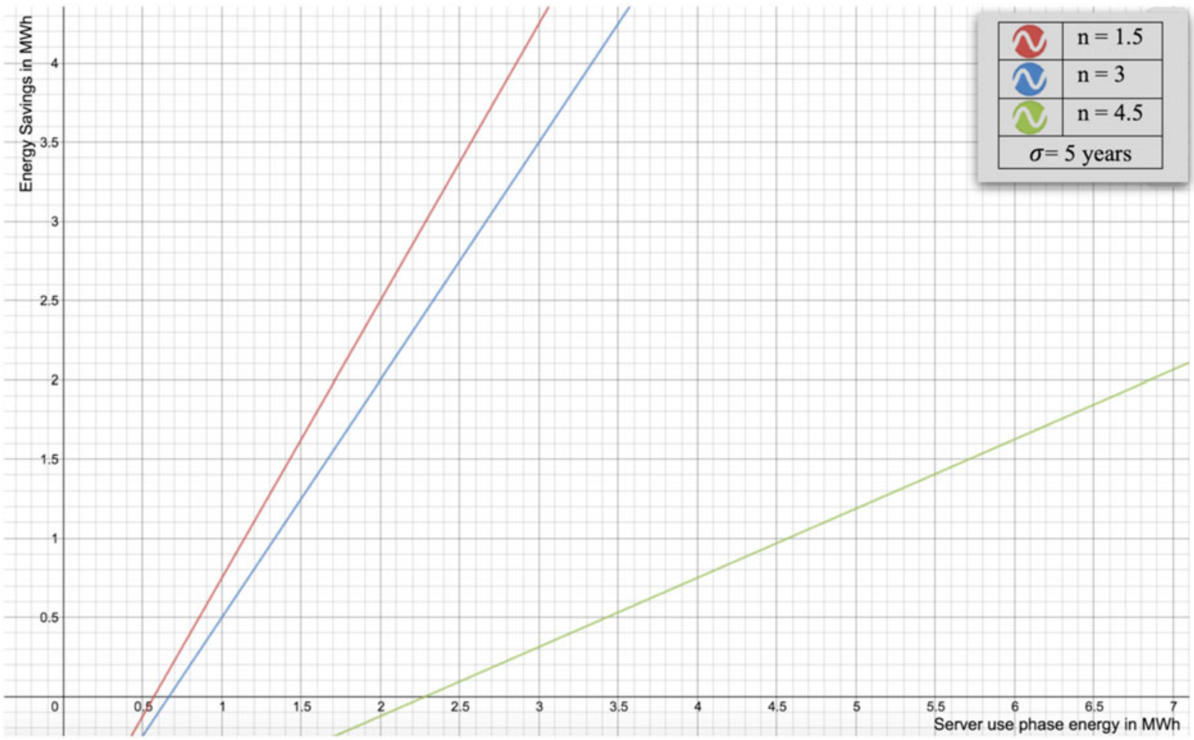

Fig. 3. Potential energy savings (ζ_5) over $\sigma = 5$ years vs. current annual server use phase energy, for hardware refresh rates of $n = 1.5$, 3, and 4.5 year.

$$E_u^{\omega} = \left(\sum_{m=1}^{\vartheta} (P_i^{s_m} \alpha_{s_m} + P_f^{s_m} \beta_{s_m}) \right) \times 8.76 \times PUE \tag{15a}$$

Where an identical pool of servers is used (same active idle and 100 percent capacity power), with balanced load (similar average utilization), Equation (15a) can be written as:

$$E_u^{\omega} = \vartheta \times (P_i^s + \beta_s (P_f^s - P_i^s)) \times 8.76 \times PUE \tag{15b}$$

Equations (15a) and (15b) can be used to calculate true efficiency by comparing the energy savings running a workload within different deployment scenarios (given specific values of PUE, utilization and hardware profile). For example, the current average P_i^s and P_f^s figures for a server are 33 and 171 watts respectively, producing 1,986,066 Server Side JavaScript Operations (ssj_ops using SPECpower benchmark) at 100 percent capacity (see Section 4 for calculation details). Assuming a hardware profile consistent with Equation (15b), an average workload of $\omega = 1{,}986{,}066$ (a workload equivalent to the average full capacity of 1 recent server), we can calculate the use phase energy of running the workload on 1 server (100 percent capacity), 2 servers (50 percent capacity), 3 servers (33 percent capacity), 4 servers (25 percent capacity), etc. for different PUE values. This is shown in Fig. 4.

One of the observations that can be made from Fig. 4 is that running workload ω in a facility with a PUE of 3 and utilization level of 50 percent, is more energy efficient than running the workload in a data center of PUE 1.1, but with utilization level of 5 percent. This allows us to make a direct comparison between the impact of server utilization level and PUE on the overall energy consumption for a given workload to better understand use phase energy efficiency.

4 Case Studies

In this section, various case studies representing real-life data center scenarios are evaluated. First, a workload model is calculated based on benchmark data and used across all experiments. Then, the running of the workload is assessed in various data center deployment types, including virtualised vs non-virtualised, assuming worst, average and best case scenarios.

4.1 Workload and Hardware Profile

For running the experiments, a model workload, ω, is needed as a benchmark. Additionally, average server watts @ 100 percent capacity (P_f^{ω}), and average server watts @ active idle (P_i^{ω}) for the workload are also required for the various time intervals.

For this, server performance data published by SPECpower is used [23]. SPECpower_ssj2008 is an industry-standard benchmark that evaluates the power and performance characteristics of volume server class computers and spans server models from November 2007 until April 2016. The data provided in the benchmark include system type, technical specification (cores, memory, etc.), and system performance in terms of server-side java operations (ssj_ops) @ 100 percent capacity, average watts at 100 percent capacity, average watts @ active idle, and overall ssj_ops at 100 percent capacity.

First, in order to calculate a representative power performance for various server vendors, the SPECpower data is broken into six, 18-month brackets, in line with the analysis window (see Section 3.3). Table 1 below shows the time intervals (starting with the first available SPECpower measurement) along with the number of samples in every interval. To ensure consistency of trends (and minimise bias

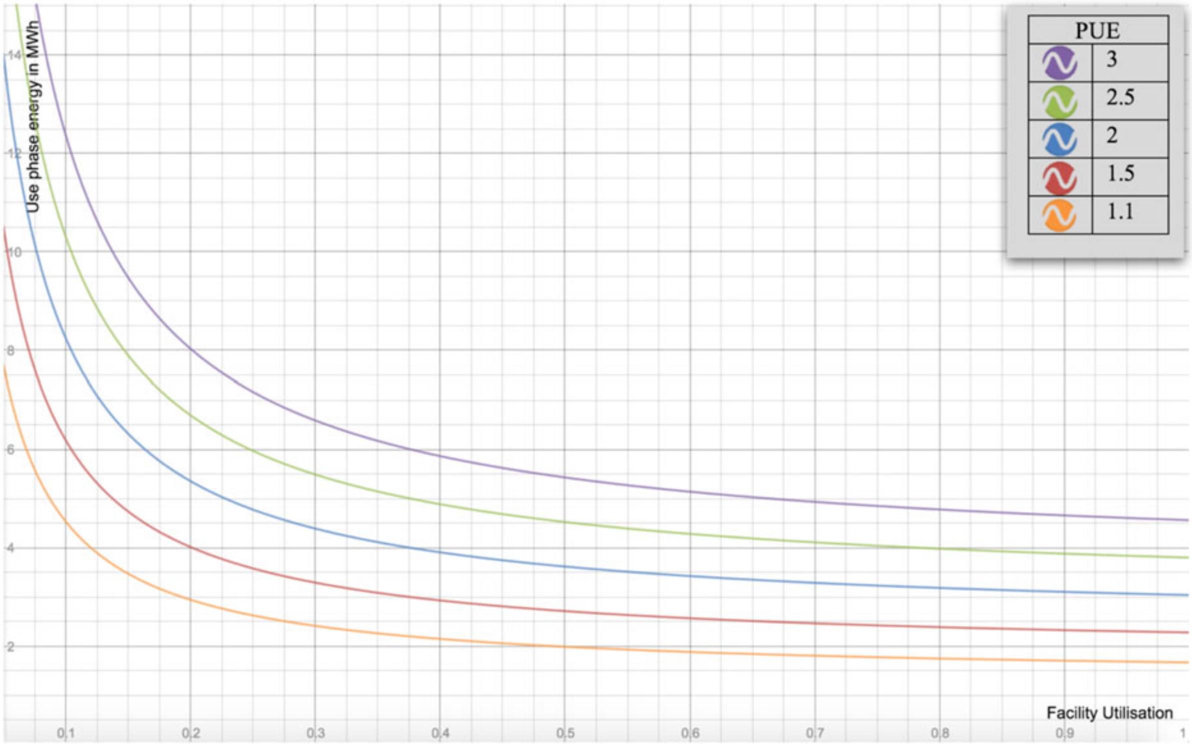

Fig. 4. Annual use phase energy consumption (in MWh) for running a workload $\omega = 1,986,066$ ssj_ops (workload equivalent to the average full capacity of 1 current server) versus facility utilization level given PUE's of 1.1, 1.5, 2, 2.5, and 3.

from high-end servers), only single node server measurements are used (see Section 6).

Then, for every interval, the average for all reported measurements in the time frame is calculated for: ssj_ops@100 percent, watts @100 percent, watts @ active idle, and overall ssj_ops/watt. Additionally, the *Dynamic Range* ratio for every interval (Avg watts @100 percent / Avg watts @active idle) is also calculated. These calculations are shown in Table 2.

The data reported in Table 2 provides a snapshot of how server power performance has been evolving over the past decade, in 18-month windows. Most notable is the major drop in active idle power.

Using the data from Table 2, the power performance characteristics of a model workload ω over the six intervals can be calculated. For the purpose of this experiment, a workload equivalent to the average of 100 servers running at full capacity using today's performance characteristics (interval 6) is used. Accordingly, ω equals 200M ssj_ops.

TABLE 1
18-Month Time Intervals for Calculating Average Server Performance

Interval	From (inclusive)	To	# of Servers reported
1	01/11/2007	01/05/2009	81
2	01/05/2009	01/11/2010	82
3	01/11/2010	01/05/2012	84
4	01/05/2012	01/11/2013	112
5	01/11/2013	01/05/2015	15
6	01/05/2015	01/11/2016*	13

* *This interval includes measurements up until April 2016, the time the analytical work was concluded.*

Looking at the fifth column (Avg Overall ssj_ops/watt) in Table 2, it is interesting to see how the technology trend has roughly followed Koomey's law so far (note that interval 6 only has partial data as it covered 10 months rather than 18 – see Table 1).

Table 3 shows the normalised average watts @ 100 percent and active idle for running ω over the six time intervals.

The next section studies how this workload performs in various typical data center deployments, namely: On-Premise (virtualised and non-virtualised), Colocation, Public Cloud, and Private Cloud. For every case, the worst, average and best case scenarios are analysed. The parameters used for the different scenarios (PUE and utilization values) are based on published data [27] and can be adjusted to suit other deployment scenarios.

4.2 Deployment Scenarios

In this section, five main deployment options are discussed, with worst, average, and best case scenarios for each deployment. The PUE and utilization (β) values of the different scenarios are based on [27].

TABLE 2
Average Server Power Performance over the Time Intervals Defined in Table 1

Interval	Avg ssj_ops @100%	Avg watts @100%	Avg watts @ active idle	Avg Overall ssj_ops/watt	Dynamic Range
1	294,105	237	139	823	1.71
2	638,826	230	86	2,066	2.67
3	939,688	287	106	2,788	2.71
4	1,402,689	290	73	4,402	3.99
5	2,349,339	226	46	8,213	4.93
6	1,986,066	171	33	9,139	5.26

TABLE 3
Normalised Average Watts @100 percent and Average Watts @ Active Idle for $\omega = 200M\ ssj\_ops$ Scenario

Interval	Avg watts @100%	Avg watts @active idle
1	161,131	94,405
2	72,161	27,072
3	61,012	22,504
4	41,298	10,358
5	19,244	3,901
6	17,268	3,284

Using the normalised workload data in Table 3, the values of PUE and β for the various deployment scenarios, and Equation (15b), the annual use phase energy for running the workload using the 6 different hardware profiles (representing the six time intervals) can be calculated. These calculations are presented in Table 4.

As can be seen from Table 4, the annual use phase energy consumption for workload ω varies substantially between the different deployment scenarios, and across the 6 different hardware profiles over time. It is interesting to note that, for example, a worst case scenario for on-premise non-virtualised data center, adopting best practices (decreasing PUE and increasing utilization), and refreshing its hardware (from 1 to 6) can achieve annual energy savings of 99 percent. Yet, by just refreshing the hardware from interval 1 to interval 6, energy savings of 94 percent can be achieved for the same scenario.

Additionally, using the information in Table 2 and Equations (9) and (10), the average annual use phase energy of a given server can be calculated by specifying its approximate age interval, for the various deployment scenarios as shown in Table 5.

While various figures tend to be referenced to estimate the average annual server use phase energy consumption, Table 5 demonstrates that without considering the specific deployment scenario, such estimations could be substantially imprecise.

Using the data from Tables 4 and 5, the number of servers (rounded to the nearest server) needed to run the workload can be calculated for each deployment scenario. This is presented in Table 6.

Table 6 shows the substantial reduction of servers needed to run the same workload. While this sounds common sense given Moore's Law, yet in practice, it is common to see servers being refreshed with the same number of new servers to run the same workload (based on existing infrastructure and the way budgets are allocated) [12]. Unless there is an increased workload scalability requirement, without carefully considering the number of new replacement servers genuinely needed, the overall server utilisation could go down after a refresh leading to decreased efficiency (due to the excess capacity introduced by the newly installed servers).

5 Life Cycle Impact Analysis

In this section, the overall environmental impact of hardware refresh is analysed, taking into consideration the age of existing hardware, deployment environment, and the evaluation time interval (number of years). The aim is to assess whether decisions made based on the reasoning framework discussed earlier would also result in reduced environmental impact, following a full LCA approach.

5.1 The LCA Model

The environmental impact dataset from [28] as shown in Table 7 is used as the basis for the LCA analysis.

Table 7 shows the production (embodied), distribution, usage and end-of-life impact for a typical rack server on 15 environmental impact parameters, classified under three main categories [28]. While the use phase impact is directly dependent on the energy consumption of a server (depending

TABLE 4
Annual Use Phase Energy Consumption of Workload ω for Various Deployment Options Using Worst, Average, and Best Case Scenarios for the Six Different Hardware Profiles

	Scenario	PUE	β	Annual Use Phase Energy in GWh (for running workload ω)					
				Hardware 1 (7.5Y old)	Hardware 2 (6Y old)	Hardware 3 (4.5Y old)	Hardware 4 (3Y old)	Hardware 5 (1.5Y old)	Hardware 6 (Current)*
On-Premise (non-virtualised)	Worst	3	5%	51.37	15.41	12.84	6.26	2.45	2.09
	Average	2	10%	17.71	5.53	4.62	2.36	0.95	0.82
	Best	1.5	25%	5.84	2.02	1.69	0.95	0.41	0.36
Colocation (non-virtualised)	Worst	2.5	5%	42.81	12.85	10.7	5.21	2.04	1.74
	Average	1.8	10%	15.94	4.98	4.16	2.12	0.86	0.74
	Best	1.3	25%	5.06	1.75	1.46	0.82	0.35	0.31
On-Premise (virtualised)	Worst	3	6%	43.1	13.04	10.87	5.35	2.11	1.81
	Average	2	30%	6.68	2.37	1.99	1.15	0.5	0.44
	Best	1.5	60%	2.94	1.19	1	0.63	0.29	0.26
Private Cloud	Worst	2.5	7%	31	9.46	7.88	3.92	1.56	1.33
	Average	1.8	30%	6.01	2.13	1.79	1.03	0.45	0.39
	Best	1.3	60%	2.55	1.03	0.87	0.55	0.25	0.22
Public Cloud	Worst	2	7%	24.8	7.57	6.31	3.13	1.25	1.07
	Average	1.5	40%	3.98	1.48	1.25	0.75	0.33	0.29
	Best	1.1	70%	1.94	0.81	0.68	0.44	0.2	0.18

* *As of 2016*

TABLE 5
Annual Use Phase Energy Consumption of a Server Given Various Deployment Options Using Worst, Average, and Best Case Scenarios for the Six Different Hardware Profiles

	Scenario	PUE	β	Annual Use Phase Energy in KWh (for a server)					
				Hardware 1 (7.5Y old)	Hardware 2 (6Y old)	Hardware 3 (4.5Y old)	Hardware 4 (3Y old)	Hardware 5 (1.5Y old)	Hardware 6 (Current)
On-Premise (non-virtualised)	Worst	3	5%	3,777	2,462	3,016	2,194	1,441	1,040
	Average	2	10%	2,604	1,767	2,169	1,653	1,119	815
	Best	1.5	25%	2,146	1,609	1,984	1,667	1,194	885
Colocation (non-virtualised)	Worst	2.5	5%	3,148	2,051	2,514	1,829	1,201	866
	Average	1.8	10%	2,344	1,591	1,953	1,488	1,007	733
	Best	1.3	25%	1,860	1,395	1,719	1,445	1,035	767
On-Premise (virtualised)	Worst	3	6%	3,803	2,500	3,064	2,251	1,489	1,076
	Average	2	30%	2,948	2,272	2,803	2,413	1,750	1,301
	Best	1.5	60%	2,598	2,272	2,816	2,665	2,023	1,523
Private Cloud	Worst	2.5	7%	3,191	2,115	2,593	1,924	1,280	927
	Average	1.8	30%	2,653	2,045	2,523	2,172	1,575	1,171
	Best	1.3	60%	2,251	1,969	2,440	2,310	1,753	1,320
Public Cloud	Worst	2	7%	2,553	1,692	2,074	1,539	1,024	742
	Average	1.5	40%	2,340	1,893	2,340	2,095	1,549	1,158
	Best	1.1	70%	2,000	1,805	2,239	2,164	1,657	1,251

on age, utilization and deployment environment as explained earlier), it can be assumed that the embodied, distribution and end-of-life impact is comparable for most server types.

In order to analyse the environmental impact of hardware refresh for a particular scenario, the net environmental impact over a chosen time duration T, NEI_T, is defined to be the difference between the Hardware Refresh scenario environmental impact (HWR) and the Business As Usual scenario environmental impact (BAU) over the same time period, as shown in Equation (16).

$$NEI_T = HWR_T - BAU_T \tag{16}$$

BAU_T is calculated by adding the total of the environmental impact of Production, Distribution, Use and End-of-life for current hardware. The Production impact (PI), Distribution impact (DI), and End-of-life impact (EI) are calculated based on the values in Table 7, multiplied by the number of servers currently deployed (η). To calculate the Use phase impact (UI), first, the energy consumption of deployed servers need to be calculated over the comparison duration ($E^c_{u(T)}$). For example, this can be done by using annual server energy consumption values from Table 5 (after converting from end-use energy to primary energy), multiplied by the number of years. Once the Use phase energy is calculated, the environmental impact can be extrapolated from the values in Table 7 for every server ($\mathrm{UI}^{s}_{\mathrm{E}^{c}_{u(T)}}$), then the total is calculated for all servers. This is summarised under

$$BAU_T = \eta \times (PI + DI + EI) + \sum\nolimits_{s=1}^{\eta} \left(UI^{s}_{E^{c}_{u(T)}} \right) \tag{17}$$

HWR_T is calculated in a similar way; However, for HWR_T, the PI, DI, and EI of existing hardware are also included in addition to those of the new refreshed

TABLE 6
Number of Servers Needed to Run Workload ω Given Various Deployment Scenarios and Hardware Profiles

	Scenario	PUE	β	Number of servers needed to run workload ω					
				Hardware 1 (7.5Y old)	Hardware 2 (6Y old)	Hardware 3 (4.5Y old)	Hardware 4 (3Y old)	Hardware 5 (1.5Y old)	Hardware 6 (Current)
On-Premise (non-virtualised)	Worst	3	5%	13,601	6,261	4,257	2,852	1,703	2,014
	Average	2	10%	6,800	3,131	2,128	1,426	851	1,007
	Best	1.5	25%	2,720	1,252	851	570	341	403
Colocation (non-virtualised)	Worst	2.5	5%	13,601	6,261	4,257	2,852	1,703	2,014
	Average	1.8	10%	6,800	3,131	2,128	1,426	851	1,007
	Best	1.3	25%	2,720	1,252	851	570	341	403
On-Premise (virtualised)	Worst	3	6%	11,334	5,218	3,547	2,376	1,419	1,678
	Average	2	30%	2,267	1,044	709	475	284	336
	Best	1.5	60%	1,133	522	355	238	142	168
Private Cloud	Worst	2.5	7%	9,715	4,472	3,041	2,037	1,216	1,439
	Average	1.8	30%	2,267	1,044	709	475	284	336
	Best	1.3	60%	1,133	522	355	238	142	168
Public Cloud	Worst	2	7%	9,715	4,472	3,041	2,037	1,216	1,439
	Average	1.5	40%	1,700	783	532	356	213	252
	Best	1.1	70%	971	447	304	204	122	144

TABLE 7
Life Cycle Impacts (per unit) of Rack Server (Source: [28])

Life Cycle phases / Resources Use and Emissions	Unit	PRODUCTION Material	PRODUCTION Manuf.	PRODUCTION Total	Distribution	Use	End-of-life Disposal	End-of-life Recycl.	End-of-life Total	Total
Other Resources & Waste							debet	credit	balance	
Total Energy (GER)	MJ	8 451	552	9 002	92	119 659	9	-4 654	-4 646	124 107
of which, electricity (in primary MJ)	MJ	5 809	245	6 053	0	119 632	0	-3 239	-3 239	122 446
Water (process)	ltr	1 730	16	1 746	0	17	0	-961	-961	802
Water (cooling)	ltr	560	142	702	0	5 320	0	-212	-212	5811
Waste, non-haz./ landfill	g	36 016	2 011	38 027	70	61 981	298	-19 990	-19 693	80 385
Waste, hazardous/ incinerated	g	2 214	4	2 218	1	1 909	0	-1 233	-1 233	2 895
Emissions (Air)										
Greenhouse Gases in GWP100	kg CO2 eq.	475	33	508	7	5 109	0	-263	-263	5 361
Acidification, emissions	g SO2 eq.	3 747	154	3 901	20	22 624	0	-2 080	-2 080	24 465
Volatile Organic Compounds (VOC)	g	19	2	22	1	2 671	0	-11	-11	2 683
Persistent Organic Pollutants (POP)	ng i-Teq	442	45	487	0	283	0	-245	-245	525
Heavy Metals	mg Ni eq.	1 435	105	1 540	4	1 223	0	-801	-801	1 966
PAHs	mg Ni eq.	493	3	497	4	284	0	-231	-231	555
Particulate Matter (PM, dust)	g	2 506	31	2 538	137	503	2	-1 398	-1 396	1 782
Emissions (Water)										
Heavy Metals	mg Hg/20	744	3	748	0	522	0	-415	-415	855
Eutrophication	g PO4	12	1	13	0	23	0	-6	-6	29

hardware. Thus, considering the refreshed hardware will encompass κ servers, HWR_T can then be calculated as per

$$HWR_T = (\eta + k) \times (PI + DI + EI) + \sum_{s=1}^{k} \left(UI^s_{E^r_{u(T)}} \right) \tag{18}$$

By substituting Equations (17) and (18) in (16) the net environmental impact NEI_T can then be calculated for the refresh scenario as per

$$NEI_T = k \times (PI + DI + EI) + \sum_{s=1}^{\kappa} \left(UI^s_{E^r_{u(T)}} \right) - \sum_{s=1}^{\eta} \left(UI^s_{E^c_{u(T)}} \right). \tag{19a}$$

TABLE 8
Life Cycle Impact (per unit) of Rack Server, Factored Under Use PHASE IMPACT (UI) and Other Impact (OI) – which Designates the Total of All Remaining Production, Distribution, and End-of-Life Impact

Other Resources & Waste	UI Reference	OI
Total Energy (GER) – in MJ	119659	4448
of which, electricity – in primary MJ	119632	2814
Total Water (process + cooling) – in ltr	5337	1276
Waste, non-haz./ landfill – in g	61981	18404
Waste, hazardous/ incinerated – in g	1909	986
Emissions (Air)		
Greenhouse Gases in GWP100 – in Kg CO_2 eq.	5109	252
Acidification, emissions – in g CO_2 eq.	22624	1841
Volatile Organic Compounds (VOC) – in g	2671	12
Persistent Organic Pollutants (POP) – in ng i-Teq.	283	242
Heavy Metals – in mg Ni eq.	1223	743
PAHs – in mg Ni eq.	284	271
Particulate Matter (PM, dust) – in g	503	1279
Emissions (Water)		
Heavy Metals – in mg Hg/20	522	333
Eutrophication – in g PO_4	23	6

Equation (19a) can be further simplified by adding all the fixed environmental impact, that is PI, DI, *and* EI, under one impact, OI, which would represent all the environmental impacts, with the exception of use phase impact (the variable part).

$$NEI_T = k \times OI + \sum_{s=1}^{k} \left(UI^s_{E^s_{u(T)}} \right) - \sum_{s=1}^{\eta} \left(UI^s_{E^s_{u(T)}} \right) \tag{19b}$$

Table 8 shows a simplified version of Table 7, with overall environmental impact factored under OI and UI, and water usage combined under 1 entry.

Using the data from Table 8, and Equation (19b), the environmental impact of different hardware refresh scenarios can then be evaluated and analysed.

5.2 Case Study

Consider the case study from Section 4 where we had a specific workload ω along with its energy consumption under various deployment environments (Table 4). Assume we have an On-premise, virtualised, average scenario (PUE of 2 & 30 percent utilization) with hardware average age of 4.5 years (Hardware 3 case). According to the previous section, this workload required 709 servers (Table 6), each consuming 2,803 KWh per year (Table 5). To upgrade the hardware to current hardware (Hardware 6), this will require 336 new servers consuming 1,301 KWh each annually (assuming PUE and utilization levels remain the same). To analyse the environmental impact of the hardware refresh for this scenario, the environmental impact of the Use phase ($\mathrm{UI}^{\mathrm{s}}_{\mathrm{E^s_{u(T)}}}$) need to be calculated for both Business as usual and the Refresh scenario, over a fixed period of time, say $\mathrm{T} = 5$ years. The Use phase impact can be calculated by extrapolating data from Table 8, then using Equation (19b) to compare the overall impact. Note that the end-use energy calculated for servers in the previous section need to be converted into primary energy to be used with Table 8 data. For

TABLE 9
Life Cycle Impact Assessment for a Hardware Refresh Scenario of 4.5-Year-Old Kit Deployed within a PUE of 2 and 30% Utilization Data Center, with $k = 336$ and $\eta = 709$ and Five Year Assessment Period

Other Resources & Waste	UI Reference	OI	UI per server BAU	UI per server HWR	NEI_5
Total Energy (GER)	119659	4448	126135.00	58545.00	-68264067
of which, electricity (in primary MJ)	119632	2814	126106.54	58531.79	-68797350.59
Total Water (process + cooling)	5337	1276	5625.84	2611.21	-2682618.946
Waste, non-haz./ landfill	61981	18404	65335.44	30325.15	-29949831.22
Waste, hazardous/ incinerated	1909	986	2012.32	934.01	-781609.4886
Emissions (Air)					
Greenhouse Gases in GWP100	5109	252	5385.50	2499.66	-2893763.904
Acidification, emissions	22624	1841	23848.42	11069.14	-12570724.04
Volatile Organic Compounds (VOC)	2671	12	2815.56	1306.83	-1553102.919
Persistent Organic Pollutants (POP)	283	242	298.32	138.46	-83670.84613
Heavy Metals	1223	743	1289.19	598.37	-463334.4057
PAHs	284	271	299.37	138.95	-74509.82438
Particulate Matter (PM, dust)	503	1279	530.22	246.10	136505.9378
Emissions (Water)					
Heavy Metals	522	333	550.25	255.40	-192426.649
Eutrophication	23	6	24.24	11.25	-11392.49986

this, 40 percent grid efficiency factor is used to convert end-use energy to primary energy (which is the EU average, meaning 40 percent of primary energy produced make it as end-use energy, factoring in various losses such as production, transmission, etc.).

A negative NEI value indicates an environmental impact reduction, while a positive NEI value indicates an increased environmental impact over the assessment period, if the hardware is refreshed. As can be seen from Table 9, all environmental impact indicators show a significant negative value (improvement) except for Particulate Matter. Fig. 5 shows these values as percentage change compared to business as usual scenario over the assessment period of 5 years.

This demonstrates the validity of the refresh reasoning framework in not only reducing energy consumption but also the wider environmental footprint, without conducting a full LCA. This can be attributed to the fact that in a data center environment, the majority of the environmental impact is associated with the use phase (the focus of the reasoning framework).

The data and equations presented in this section can be similarly applied to the remaining case studies (discussed under Section 4) or any deployment scenario for validation.

6 Study Limitations

As with any research, this work exhibits a number of limitations, which are discussed in this section along with the mitigation measures.

First, the case studies discussed under Section 4 depend on the quality of the SPECpower benchmark dataset. One of the main limitations of the dataset is that the servers reported are not randomly selected (e.g., servers reported in the first few years were higher end products rather than volume servers). In order to address this limitation, only single

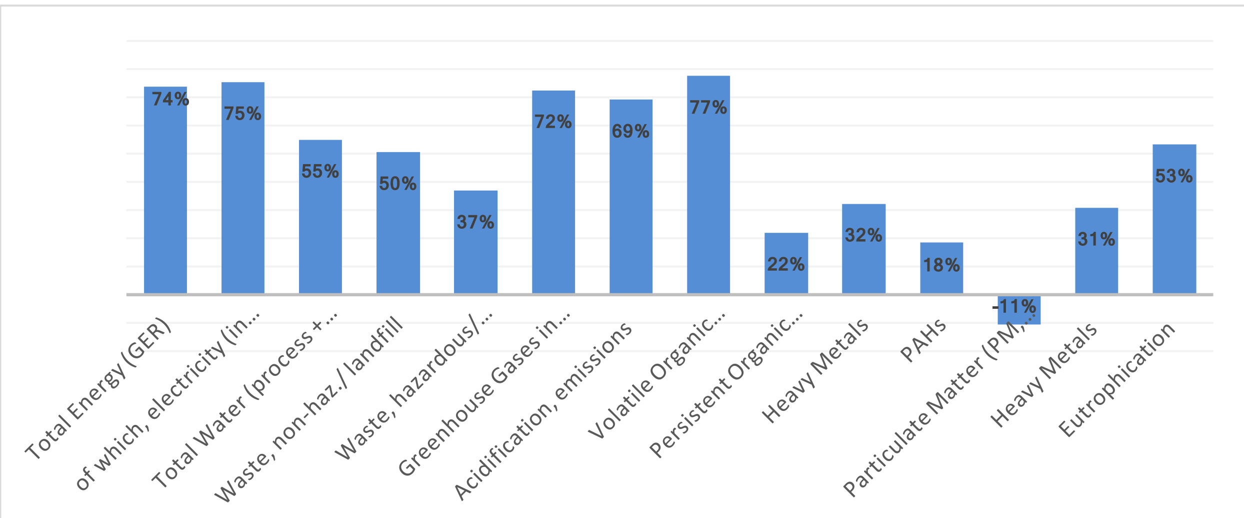

Fig. 5. Percentage change in environmental impact for the given hardware refresh scenario showing significant improvement in all indicators, except for Particulate Matter which is 11 percent worse off.

node servers were selected in the analysis which ensured trend consistency. Additionally, the dataset has a substantial number of records to be representative enough of the wider server population. Furthermore, dates reported in the benchmark are measurement dates rather than production dates (age). Thus, a manual check on various models was conducted to ensure reliability. This confirmed that measurement dates were in line with production dates for the selected measurements.

Second, the quality of the environmental impact analysis section depends on the quality of the dataset used. However, this is the only such dataset available and is the one officially acknowledged by the EU Commission. Additionally, extrapolating the dataset could introduce a margin of error in the analysis provided under Section 5. Yet, studying the methodology used in creating the dataset in [28], the potential error due to extrapolation seems to be marginal.

The third limitation is the lack of published data by OEMs about the thermal influence on server energy performance. It is known that server performance is dependent on the operating temperature and it is generally surmised that increases in operating temperature reduce energy efficiency; However, this could not be captured in this work due to the lack of such thermal performance data. It is anticipated that legislative work in the EU under EcoDesign will soon require OEMs to publish such data, particularly in relation to the increase in server fan power consumption at elevated temperatures. This is also defined as a requirement in the EU Code of Conduct for Data Center Energy Efficiency published by the EU Commission Joint Research Center (JRC).

Finally, the case studies assumed a fixed PUE and utilisation levels for the various hardware refresh scenarios. While this is true when the hardware being refreshed represents a small percentage of the overall data center IT load (which tends to be the case in most large data centers due to business continuity concerns), when refreshing a significant proportion of IT equipment (with a smaller and more efficient kit), the PUE will tend to increase due to the baseline overhead of the facility (this is one of the limitations of PUE when considered as an efficiency metric). Nonetheless, the overall facility energy consumption would still decrease, which suffices for this work.

7 Conclusion and Future Work

This work sheds new light on the relationship between IT footprint and the overall data center energy consumption. The work shows that optimising IT equipment (e.g., hardware refresh or increased utilization) can yield more energy savings in a data center than decreasing the PUE. Additionally, the work demonstrates that the embodied energy of existing equipment is not a deciding factor when it comes to calculating the best data center hardware refresh rate from an environmental impact perspective. Equations and models are provided to help guide refresh strategies that would reduce the environmental impact and running cost of facilities. These can be adapted to the various types of data centers (e.g., enterprise, co-location, wholesale, etc.) to help identify the optimal refresh rate for each scenario. Most importantly, the work emphasises the importance of linking energy consumption to IT workload (rather than just IT equipment) to better understand and optimise for energy performance.

Going forward, some of the underlying technology evolvement assumptions made in this work will have to be revisited and redefined. For example, computing efficiency cannot continue to evolve per Koomey's law. It is expected by 2048, Koomey's law will come to an end due to Landauer's principle [29]. Yet, Moore's law is expected to seize much earlier, potentially within the next decade [25].

Finally, this work will next consider the creation of complementary business case analyses models to support the uptake of the methodologies discussed. From a cost perspective, the work will capture elements related to hardware refresh such as procurement costs, project management, installation, decommissioning and energy consumption. On the other hand, additional benefits related to the reduction of IT equipment footprint (e.g., less maintenance and support requirements, reduced risk of incidents/outages, etc.) will also be considered, including scalability which could extend the life of existing facilities (minimising the environmental impact of building new ones).

Acknowledgments

This work was funded as part of the EURECA project which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 649972. The authors would like to thank Paolo Canfora, European Commission - JRC, for his feedback on the initial version of this work in relation to LCA. They would also like to thank Mark Acton, Chair of the EU Code of Conduct on Data Centers Best Practice Committee, for his valuable input and reviewing initial drafts of this work.

References

[1] R. Bashroush, E. Woods, and A. Noureddine, "Data center energy demand: What got us here won't get us there," *IEEE Softw.*, vol. 33, no. 2, pp. 18–21, Mar./Apr. 2016.
[2] P. Bertoldi, "The european programme for energy efficiency in data centres: The code of conduct, EU DG JRC," *Inst. Energy Transport*, 2016.
[3] A. Shehabi, et al., "United States data center energy usage report," Lawrence Berkeley Nat. Laboratory, Berkeley, CA, Tech. Rep. LBNL-1005775, 2016.
[4] ISO/IEC 30134-2, Information technology – Data centres – Key performance indicators – Part 2: Power usage effectiveness, 2016.
[5] DataTank Immersion Cooling Data Centre winning the Best Green ICT Award 2014 for Most Efficient Data Center – PUE 1.01, [Online]. Available: : https://www.allied-control.com/publications/DataTank_Product_Info.pdf, Last accessed: October 2017.
[6] R. Bashroush and E. Woods, "Architectural principles for energy-aware internet-scale applications," *IEEE Softw.*, vol. 34, no. 3, pp. 14–17, May/Jun. 2017.
[7] A. Beloglazov, R. Buyya, Y. C. Lee, and A. Zomaya, "A taxonomy and survey of energy-efficient data centers and cloud computing systems," *Adv. Comput.*, vol. 82, no. 2, pp. 47–111, 2011.
[8] Y. C. Lee and A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," *J. Supercomputing*, vol. 60, no. 2, pp. 268–280, 2012.
[9] N. Jebraeili, A. Noureddine, J. Doyle, S. Islam, and R. Bashroush, "gUML: Reasoning about energy at design time by extending UML deployment diagrams with data centre contextual information," in *Proc. 13th IEEE World Congress Services*, Jun. 2017, pp. 59–66.
[10] S. Islam, A. Noureddine, and R. Bashroush, "Measuring energy footprint of software features," in *Proc. 24th IEEE Int. Conf. Program Comprehension*, May 2016, pp. 1–4.
[11] A. K. Jones, et al., "Green computing: A life cycle perspective," in *Proc. Int. Green Comput. Conf.*, Jun. 2013, pp. 1–6.
[12] Interview with Mark Acton, Chair of the EU Code of Conduct on Data Centres Best Practice Committee, May 2016.

[13] E. Masanet, A. Shehabi, and J. Koomey, "Characteristics of low-carbon data centres," *Nature Climate Change*, vol. 3, no. 7, pp. 627–630, 2013.

[14] C. R. Hannemann, V. P. Carey, A. J. Shah, and C. D. Patel, "Lifetime exergy consumption of enterprise servers," *Int. J. Exergy*, vol. 7, no. 4, pp. 439–453, 2010.

[15] D. Kline, et al., "Holistically evaluating the environmental impacts in modern computing systems," in *Proc. 7th IEEE Int. Green Sustainable Comput. Conf.*, 2016, pp. 1–8.

[16] H. R. Arkes and C. Blumer, "The psychology of sunk costs," *Organizational Behavior Human Decision Processes*, vol. 35, pp. 124–140, 1985.

[17] N. Horner and I. Azevedo, "Power usage effectiveness in data centers: overloaded and underachieving," *The Electricity J.*, vol. 29, no. 4, pp. 61–69, 2016.

[18] G. A. Brady, N. Kapur, J. L. Summers, and H. M. Thompson, "A case study and critical assessment in calculating power usage effectiveness for a data centre," *Energy Convers. Manag.*, vol. 76, pp. 155–161, 2013.

[19] C. R. Hannemann, V. P. Carey, A. J. Shah, and C. D. Patel, "Lifetime exergy consumption of enterprise servers," *Int. J. Exergy*, vol. 7, no. 4, pp. 439–453, 2010.

[20] InnovateUK, Energy Efficient Computing Special Interest Group, "Towards the estimation of the energy cost of Internet mediated transactions," UK, Sep. 2013.

[21] E. Williams, "Energy intensity of computer manufacturing," *Environ. Sci. Technol.*, vol. 38, pp. 6166–6174, 2004.

[22] The Open Compute Project, [Online]. Available: http://www.opencompute.org/, Last Accessed: June 2017.

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

[24] K. D. Lange and M. G. Tricker, "The design and development of the server efficiency rating tool (SERT)," *ACM SIGSOFT Softw. Eng. Notes*, vol. 36, no. 5, pp. 145–150, 2011.

[25] Interview with Gordon Moore, Mar. 30, 2015, *IEEE Spectrum*.

[26] J. Koomey, S. Berard, M. Sanchez, and H. Wong, "Implications of historical trends in the electrical efficiency of computing," *IEEE Ann. History Comput.*, vol. 33, no. 3, pp. 46−54, Mar. 29, 2010, doi: 10.1109/MAHC.2010.28.

[27] Report by WSP Environment & Energy and Natural Resources Defense Council, "The Carbon Emissions of Server Computing for Small- to Medium-Sized Organizations," Oct. 2012.

[28] 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. 2015.

[29] C. H. Bennett, "Notes on landauer's principle, reversible computation, and Maxwell's Demon," *Studies History Philosophy Sci. Part B: Studies History Philosophy Modern Phys.*, vol. 34,. no. 3, pp. 501–510, 2003.

Rabih Bashroush received the BEng degree in computer and communications engineering from AUB, in 2001 and the PhD degree in systems engineering from Queen's University Belfast, in 2005. He is a reader in computer science and the director of the Enterprise Computing Research Group, University of East London (UEL). Before joining UEL, he spent 10 years with Queen's University Belfast. He held visiting scientist positions with Carnegie Mellon University, Philips Research Labs, and Danfoss Power Electronics. He is the author of more than 70 publications and received various research grants in the field of energy efficiency, including EURECA (EU H2020 project) and PEDCA (EU FP7 project). In 2016, he was appointed as the coordinator of the EU Commission DG CONNECT Smart Cities Research Cluster on Energy Efficiency in Data Centers. He serves on the CEN/CENELEC/ETSI Coordination Group on Green Data Centers, the BSI TCT/7/3 Telecommunications, Installation requirements: Facilities and infrastructures - CEN/CENELEC TC215, and the BSI ST/46 ISO SC 39 Sustainability For and By IT standardisation committees. He served on the EU Commission Expert Working Group for Best Environmental Practice in the Telecommunication and ICT services sector and the EU Consultation Forum for the EcoDesign Legislation for Energy Efficient Servers. He is a fellow of the United Kingdom Higher Education Academy and has been a member of the IEEE and IEEE Computer Society since 1998.