Optimizing the Impact of Upgrading Computer Equipment

Gabriel Breuil,
R&D division Impakt, Constellation
Paris, France
gabriel.breuil@dlr.de

Olivier Hermant,
Mines Paris, PSL University
Fontainebleau, France
olivier.hermant@minesparis.psl.eu

Renaud Pawlak,
R&D Cincheo
Paris, France
renaud.pawlak@cincheo.com

***Abstract*—Upgrading hardware platforms and infrastructure, such as personal computers, servers, or supercomputers, suffer from a tension when it comes to carbon neutrality. Increasing performances are an incentive for fast renewal, while manufacturing environmental costs urge us to extend as much as possible the lifetime of devices. Thus, carbon neutrality seems to be difficult to reach.**

In this article, we propose a theoretical model with an implementation available online. We study a refreshment of only the CPU and we try to answer quantitatively the question "what is the optimal point for which we may tend to carbon neutrality"? Our approach tackles a model of the greenhouse gas emissions, the primary energy and the monetary cost of two parameters that vary country by country: the use of a microprocessor, and its fabrication. We investigated three different renewal scenarios comparing two and three CPU versions: (1) maintaining a constant workload capacity, (2) maintaining the amount of processor cores fixed, and (3) keeping the amount of microprocessors fixed.

Our results suggest that the most efficient solution (*i.e* with the lowest impacts) is to replace the entire batch of CPUs with the newer batch as soon as possible, which backs up the idea that modular replacement of hardware components is worth studying to help optimizing the impacts of computer equipment in general.

***Index Terms*—Data centers, hardware upgrading rate, carbon-neutral system, impact models, life cycle assessment, resource management.**

I. INTRODUCTION

In 2019, 4.1 billions users of Information and Communication Technologies (ICT) have been identified, and the demand is still growing. This increase is partly driven by the desire of having more efficient ICT infrastructures. The energy cost of equipment manufacture (end-user devices, Data Centers and network) corresponds to 44% of the total ICT energy consumption [1]. Nicola Jones explains in [2] that by 2030 the ICT will increase up to 21% of the total electricity demand. These power consumption expectations split into a constant factor for the production of ICT and consumer devices, and an exponential increase for networks and Data Centers. The demand of Data Center services is indeed expected to expand within the next few years, driving an effort towards more efficiency. To do so, one has to enhance the workload capacity, which can be done for example by virtualization and increasing the performance of the electronic components used in Data Centers. As a consequence, each year, new generation of processors, hard disks, RAM, optical fibre, and so on are released on the market [3]–[8] and replace older equipment. Moreover, even if a Data Center can sustain a given workload without upgrading electronic components, hardware replacement still may occur. Indeed, components have generally a three-year warranty that coerces customers to a three-year life-cycle [9], [10], that drives people to renew equipment.

Carbon-neutral systems and infrastructure are a key issue for the green transition. However, neutrality in the ICT is no easy task and may be difficult to reach. There is only few investigations on carbon-neutral data-centers, which are quite recent [11]–[13]. These works suggest new policy instruments, technological methodology, and investigation on the overall energy consumption of Data Centers.

The huge amount of data transferred through the network, the periodic upgrade of electronic components, and the energy consumption of the infrastructure are the main reasons for the increase of energy consumption in Data Centers. The energy consumption of infrastructure can be divided in two almost equal parts: computing equipment (processor, service power supply, storage, communication equipment, and so on) and support systems (cooling, building switchgear, PDU, and UPS). Despite its numerous components, cooling is predominant in support systems and represents the majority of the energy consumption [14], [15]. The energy consumption of the computing equipment is balanced among various categories [2], [15]–[18].

One of the challenges of the digital 21st century is to reach carbon-neutral systems, infrastructures, architectures etc. Thus one needs to evaluate and reduce the energy consumption and the environmental impact of Data Centers. Over the last decades, metrics such as the Power Usage Effectiveness (PUE), Thermal Design Power (TDP), Power Consumption (PC), Power Effectiveness (PE), and so on have been used. Scientists have worked on the enhancement of Data Centers performance by optimizing these metrics [16], [19]–[21]. Among those, TDP and PUE are the most commonly used. TDP gives the maximal power value of a CPU when it is used at 100% and PUE gives the ratio of the total facility energy load (the support system and the computing equipment) divided by the ICT energy load (the computing equipment). During the past decade, the main focus has been to bring the PUE as close as possible to 1. Indeed, all the energy of the facility would be then dedicated to the ICT equipment. It has led to optimizing the cooling and power infrastructure [22], [23]. Nevertheless, tuning the PUE has no direct impact on the efficiency of ICT nor on the actual energy consumption and environmental impact of Data Centers and of their electronic components.

To go further on the improvement of the understanding of Data Center energy consumption, one must take into account the Life Cycle Assessment (LCA) of its components in order to describe the global impact on the environment [24]–[26]. The LCA considers the environmental impacts of the manufacturing, transport, usage and end-of-life of the components of Data Centers.

Rabih Bashroush and co-authors have focused their work on the upgrade of hardware in Data Centers [27]–[30]. In the study published in 2018 [27], Bashroush has put the spotlight on energy saving and the impact of the hardware upgrade. The model is based on the replacement of one server by another one, with performances assessed by Koomey's law. The payback point is defined as the time at which the upgrade starts showing energy benefits. Moreover, the model is validated by using real-life dataset, in particular a lifecycle impact analysis [31] of servers shows that the use phase of a Data Center has the strongest impact on the total consumption. In addition, Doyle *et al.* [30] have studied the economical impact – and so the money saving – of the hardware renewal with Bashroush's model. They established that it is possible to obtain a return on investment in the case of reducing the server population. Then, in [29], Bashroush *et al.* suggested that a renewal with re-manufactured servers is an efficient alternative for 5-6 years old servers.

It is essential to detail the hardware renewal policy and choose to replace CPUs, drive bay, etc. only when it has a global positive impact: immediate or amortized money, energy and CO_2 emissions gains, while reducing useless waste. The review of Jin et al. [32] analyses the servers' energy consumption and classifies the different components of Data Centers. The ranking for energy consumption in a server is as follows: the central processing unit (CPU) is the higher energy consumer with 32%, power supply represents 15% and memory 14%. At the exception of cooling, the dominating factor is therefore CPU. Moreover, the energy consumption of CPUs has been widely studied both in terms of measure and reduction [33]–[36].

Nowadays, the refresh in datacenter involves a full refresh of servers as if they were a unique entity. Since all components may not have the same optimal refresh period, it is worth questioning this strategy and study the theoretical renew of only one component in a server in order to reduce the overall server manufacturing energy. Since the CPU is the highest energy consuming component in a server (cooling put apart), we focus in this paper only on the refresh of CPUs. We describe a two-CPU renewal model and attempt to determine the ideal frequency and proportions to replace an old-generation CPU by a new one. We study several criteria with different scenarios in order to tend to a carbon-neutral system. We then discuss the extension to three generations of CPUs, allowing to target more realistic scenarios. Our models use a precise quantitative assessment of the manufacturing costs.

Section II describes our model, built on a comparison of performance, manufacturing energy, and energy use of the cores of two different CPUs. Our model has three degrees of liberty: the proportion of CPUs to replace, the replacement frequency, and the date of the first replacement. Three renewal policies are proposed: a CPU-to-CPU match, a core-to-core match, and a workload capacity-to-workload capacity match. Then, we extend this model to take into account different levels of hardware Mean Time Between Failure (MTBF) for three successive generations of CPUs.

Section III discusses the limits of our model. It highlights why the global energy is an approximation of the exact one and focuses on the sensitivity of the results to the MTBF. Section IV gives the results applying the two-CPU renewal model and determine the best CPU-replacement scenarios in a Data Center, according to the CPU respective specifications. Section V is a discussion focusing on a three-CPU renewal scenario, that compares the costs of upgrading a CPU twice in a row, or only once. For all scenarios, the model takes into account the market availability constraint of the newer-generation CPUs. The impact of these renewal scenarios is assessed on economical, environmental and energetic criteria, and we compare the optimal points of each of them. Sections VI and VII close the article by discussing limitations, improvements and uses of our model.

II. METHODOLOGY

In our model, we assume that the environmental impact of using a processor is proportional to its electric energy consumption. In other words, we consider that the power consumption of a processor and of the infrastructure (server, cooling and global Data Center power supply) is directly related to their environmental impacts. Similarly, we assess the manufacturing environmental costs in terms of energy [37]. The model involving the refreshment between two CPUs is available on [1]

A. Comparison between two different versions

Let P_1 and P_2 be two processors with respective release dates d_1 and d_2. We consider in this study that they belong to the same family, so that their characteristics are meaningful to compare, as it can be the case for the Intel® Xeon® processor family for instance. It is nevertheless unfair to compare global characteristics and performances, as P_1 and P_2 may have different amount of cores. Therefore, we employ two basic measures, the manufacturing and usage energy *per core*:

E_i^{m}	manufacturing energy for P_i	(Wh/core)
E_i^{u}	annual usage energy for P_i	(Wh/core)

The manufacturing energy is the fraction per core of the total manufacturing energy cost of the processor. The latter has been studied in-depth by the working group Boavizta[2], and it depends on the lithography technology, the amount of cores, and the die size.

The annual usage energy is, similarly, the fraction of the Thermal Design Power (TDP) of P_i, as specified by the manufacturer, multiplied by the annual use time – we consider a continuous round-year usage of $8\,760\mathrm{h}$. This overestimates the actual consumption, but high-availability

---

[1] https://github.com/Constellation-Group/OptImUp.git

[2] https://boavizta.org/blog/empreinte-de-la-fabrication-d-un-serveur

data centers approach this bound, and we want to evaluate the renewal of CPU at their maximal load. The Data Center starts operating at $t_0 \geq d_1$, with only P_1 processors, and stops at t_{end}. To evaluate the impact of different renewal policies, we introduce two parameters: t_{S}, the starting date of the renewal scheme (with $t_{\mathrm{S}} \geq d_2$), and Δ_{I}, a fixed time period that separates two renewals. Each renewal replaces part of P_1 processors by P_2 ones, and all periods and dates are counted in years.

We distinguish four stages in the lifespan of the Data Center, as depicted in figure 1. At the *initial stage*, the Data Center contains only P_1, then at each of the *replacement stages*, a certain amount of P_1's is replaced by some P_2's. Otherwise, the mix of P_1's and P_2's remains stable, and either a replacement will occur later (*interval stages*), or no further replacement is scheduled (*final stage*).

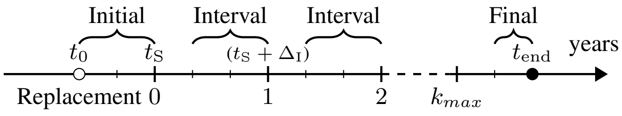

Figure 1: The different stages and times of the model

We track the quantity of cores through four sequences indexed by the replacement stages $k \in [\![0, k_{\max}]\!]$, with i the processor type ($i = 1, 2$):

$(N^r_{i_k})_{k\in[\![0,k_{\max}]\!]}$	removed P_1/added P_2 cores at stage k
$(N_{i_k})_{k\in[\![0,k_{\max}]\!]}$	amount of P_1/P_2 cores after stage k

In order to uniformly present equations and computations, we extend the last sequences: $N_{i_{-1}}$ represent the amounts of P_i cores *before* the first replacement, that is to say, $N_{1_{-1}}$ is the starting amount, and $N_{2_{-1}}$ is 0.

1) Initial stage: This stage spans from t_0 to t_{S}. We start with $N_{1_{-1}}$ cores and the total energy that is spent is:

$$E_{\mathrm{init.}}(\mathrm{P}_1) = N_{1_{-1}} \cdot \left(E_1^{\mathrm{m}} + E_1^{\mathrm{u}} \cdot (t_{\mathrm{S}} - t_0)\right) \tag{1}$$

The value of $E_{\mathrm{init.}}$ depends only on the characteristics of P_1 (E_1^{u} and E_1^{m}) and is indeed a function of P_1.

2) Replacement stages: The renewal dates t_k occur periodically $k_{\max}+1$ times, with intervals Δ_{I}. All renewal dates precede the end date: $t_k \leq t_{\mathrm{end}}$, which allows to derive

$$k \leq \frac{t_{\mathrm{end}} - t_{\mathrm{S}}}{\Delta_{\mathrm{I}}} \tag{2}$$

As $k_{\max}$ is the last stage, it is actually the integral part of the right-hand side of equation 2, and it is a function of the parameters t_{S}, Δ_{I}, and t_{end} of our model.

By convention, replacements happen at the first calendar day of the years. We add the annual usage energy consumption of the new P_2 and P_1 core numbers with the manufacturing cost of the new P_2 cores.

$$E_k(\mathrm{P}_1, \mathrm{P}_2) = N_{1_k} E_1^{\mathrm{u}} + N_{2_k} E_2^{\mathrm{u}} + N^r_{2_k} E_2^{\mathrm{m}} \tag{3}$$

The energy consumption is now a function of P_1 and P_2. We investigate degressive renewal strategies. Therefore, the sequence $(N^r_{1_k})_{k\in[\![0,k_{\max}]\!]}$, that determines the number of removed P_1 cores, is defined by

$$N^r_{1_k} = \rho \cdot N_{1_{k-1}} \tag{4}$$

Here, ρ is the replacement fraction, a percentage that is a parameter of the model.
Equation 4 is an approximation, as we cannot replace a fraction of a core or of a processor. The exact equation is

$$N^r_{1_k} = n_1 \cdot \lceil \frac{N_{1_{k-1}}}{n_1} \cdot \rho \rceil \tag{5}$$

Discretization is negligible for a large number of processors. Section III-A addresses the question quantitatively. Here and below, we only presents the continuous model.

The number of added P_2 cores depends on the replacement policy, defined by a renewal ratio R:

$$N^r_{2_k} = R \cdot N^r_{1_k} \tag{6}$$

The quantity of cores, are the result of integrating the successive removals and additions. They read as follows,

$$\begin{aligned} N_{1_k} &= N_{1_{k-1}} - N^r_{1_k} \\ N_{2_k} &= N_{2_{k-1}} + N^r_{2_k} \end{aligned} \tag{7}$$

Note that equations 4 and 7 are mutually recursive.

We define three scenarios that induce different renewal ratios R, that we call the *fixed-CPU scenario*, the *fixed-core scenario* and the *fixed-workload scenario*.
The **fixed-CPU scenario** preserves the number of CPU. There is a one-to-one match between removed an added processors. The renewal ratio reads, therefore,

$$R = \frac{\mathrm{n}_2}{\mathrm{n}_1} \tag{8}$$

The **fixed-core scenario** preserves the number of cores. The renewal ratio R is 1, which immediately entails, for all stages k, $N^r_{2_k} = N^r_{1_k}$ and $N_{2_k} + N_{1_k} = N_{1_{-1}}$.
The **fixed-workload scenario** preserves the workload. The renewal ratio is then the ratio of the performances per core, $\frac{B_i}{n_i}$, where B_i are the CPU PassMark® Software Pty Ltd performance benchmark values[3]. The ratio reads as follows

$$R = \frac{\mathrm{B}_1 \cdot \mathrm{n}_2}{\mathrm{B}_2 \cdot \mathrm{n}_1} \tag{9}$$

3) Interval stages: They happen $k_{\max}$ times, between two successive replacements. The only energy consumption stems from the usage of the P_1 and P_2 blend that has been set at the replacement stage k. We multiply this energy by the length of the stage, $\Delta_{\mathrm{I}} - 1$ (see figure 1). During those stages, for all $k \in [\![0, k_{\max-1}]\!]$, the total energy consumption then reads,

$$E_{\Delta k}(\mathrm{P}_1, \mathrm{P}_2) = \left(N_{1_k} E_1^{\mathrm{u}} + N_{2_k} E_2^{\mathrm{u}}\right) \cdot (\Delta_{\mathrm{I}} - 1) \tag{10}$$

---

[3] https://www.passmark.com/

4) Final stage: This only happens when the replacement scheme has come to its end, ie., if $t_{k_{\max}} < t_{\mathrm{end}}$ (equation 2 becomes strict on $k_{\max}$). The length of this stage is $\Delta t_{k_{\max}} = t_{\mathrm{end}} - t_{k_{\max}}$, which may be 0 in the degenerate case, and total energy consumption then reads as follows,

$$E_{\mathrm{end}}(\mathrm{P}_1, \mathrm{P}_2) = \left(N_{1_{k_{\max}}} E_1^{\mathrm{u}} + N_{2_{k_{\max}}} E_2^{\mathrm{u}}\right) \cdot \Delta t_{k_{\max}} \quad (11)$$

There still might a blend of P_1 and P_2 remaining, typically with a degressive replacement strategy ($\rho < 100\%$).

5) Additional manufacturing energy consumption: The Mean Time Between Failure (MTBF) is the life expectancy between two failures. Equivalently, every year, a certain average proportion $x\%$ of the processors break down.

If a P_1 processor fails at a replacement stage, we consider it to be one of the replaced processors, at no extra cost (hence the -1 factor in equations 12 and 13). This requires x to be lower than the replacement fraction ρ, which is, by an order of magnitude, verified in practice. Otherwise, we consider that broken processors are replaced by the same model, and we add the manufacturing costs.

We neglect replacement during the initial stage, CPUs are new. For all $k \in [\![0, k_{\max} - 1]\!]$, the additional energy consumption for the k-th replacement and interval stages, considered together, is:

$$E_k^{\mathrm{a}}(\mathrm{P}_1, \mathrm{P}_2) = x \cdot N_{1_k} E_1^{\mathrm{m}} \cdot (\Delta_{\mathrm{I}} - 1) + x \cdot N_{2_k} E_2^{\mathrm{m}} \cdot \Delta_{\mathrm{I}} \quad (12)$$

For the last $k_{\max}$-th replacement and final stages, the annual additional energy consumption is:

$$\begin{aligned} E_{\mathrm{end}}^{\mathrm{a}} = \quad & x \cdot N_{1_{k_{\max}}} \cdot E_1^{\mathrm{m}} \cdot \Delta t_{k_{\max}} \\ & + x \cdot N_{2_{k_{\max}}} \cdot E_2^{\mathrm{m}} \cdot (\Delta t_{k_{\max}} + 1) \end{aligned} \quad (13)$$

Global energy consumption. The global energy function is defined as being a function of P_1 and P_2 and corresponds to the energy consumption during $t_{\mathrm{end}} - t_0$.

$$\begin{aligned} G(\mathrm{P}_1, \mathrm{P}_2) = E_{\mathrm{init.}}(\mathrm{P}_1) + \sum_{k=0}^{k_{\max}-1} \Big[& E_k(\mathrm{P}_1, \mathrm{P}_2) \\ & + E_{\Delta k}(\mathrm{P}_1, \mathrm{P}_2) + E_k^{\mathrm{a}}(\mathrm{P}_1, \mathrm{P}_2) \Big] \\ & + E_{k_{\max}}(\mathrm{P}_1, \mathrm{P}_2) + E_{\mathrm{end}}(\mathrm{P}_1, \mathrm{P}_2) + E_{\mathrm{end}}^{\mathrm{a}} \end{aligned} \quad (14)$$

Section III discusses the limitations of equation 14.

B. Comparison between three different versions

Following the same guidelines, the model extends to more than two processors. The global energy consumption expression of equation 14 remains valid, but the strategies suffer a combinatorial explosion: For three different processors of the same family P_1, P_2, and P_3, manufactured at t_{P_1}, t_{P_2} and t_{P_3}, there are many stage types, up to three ratios ρ_{12}, ρ_{13}, and ρ_{23}, etc. To ease the comparison during the case studies, we constrain the replacement ratio to be $\rho = 100\%$ and the replacements to occur at t_{P_2} and t_{P_3}.

III. ACCURACY OF THE MODEL

We evaluate the model presented in sections II-A and II-B with three processors: Xeon® Silver 4114[4,5], Xeon® Gold 5220[6,7] and Xeon® Platinum 8380[8,9] by evaluating the economic, energetic and environmental impacts of CPU renewal in a Data Center. Both of them have the Skylake ®micro-architecture. For simplicity, the three processors are abbreviated Silver, Gold and Platinium, respectively. Their specifications are shown in table I. Three impacts are defined as follows:

- **Economical impact**: the manufacturing cost is defined to be the launch price listed on Intel's website and the use cost is defined to be equal to the price of the electric energy consumed by the CPUs. We assume in this article that the fictional Data Center studied is localized in France. In this case, TotalEnergies© [39] suggests a price of €0.174 0 per 1 kWh.
- **Energetic impact**: we consider the primary energy (PE) required to manufacture the processors and to run them.
- **Environmental impact**: it is estimated by the global warming emissions (GWP) which is expressed in carbon dioxide equivalent ($\mathrm{CO_2}e.$). The processor manufacturing weight of $\mathrm{CO_2}e.$ is determined by using the weighted average of the electrical mix from the factories location [40]. The electrical mix of each country for 2021 is taken from the website *Our World In Data* [38]. The carbon intensity of electricity for the processor manufacturing is then $0.413\,\mathrm{kgCO_2}e.$ for $1\,\mathrm{kWh}$. The carbon intensity of electricity for the processor usage in France is $0.056\,93\,\mathrm{kgCO_2}e$ for $1\,\mathrm{kWh}$ [41].

Unless otherwise specified, we consider that the Data Center is initially composed of 100 Silver in 2017. It stops at the end of 2026. Given the specifications of the Silver and Gold CPUs, listed in table I, the respective replacement ratios are 1.8 for the fixed-CPU scenario, 1 for the fixed-core scenario and 0.92 for the fixed-workload scenario. When the replacement is total, i.e. when the replacement fraction ρ is 100%, this respectively leads to 100 Gold CPUs (1800 cores), 55.56 Gold (1000 cores), and 50.99 Gold CPUs (917.83 cores). At this point, it is essential to highlight that the three CPUs have been chosen such that it is attractive to refresh the CPUs in a primary energy and carbon footprint point of view. New generations of CPU might have a higher primary energy or carbon footprint emission per core than the previous generation. In this case it might be indeed pointless to refresh the CPUs.

The validation of our model is detailed in the next two following subsections.

---

[4] https://www.intel.fr/content/www/fr/fr/products/sku/123550/intel-xeon-silver-4114-processor-13-75m-cache-2-20-ghz/specifications.html

[5] https://www.cpubenchmark.net/cpu.php?cpu=Intel+Xeon+Silver+4114+%40+2.20GHz&id=3095

[6] https://ark.intel.com/content/www/fr/fr/ark/products/193388/intel-xeon-gold-5220-processor-24-75m-cache-2-20-ghz.html

[7] https://www.cpubenchmark.net/cpu.php?cpu=Intel+Xeon+Gold+5220+%40+2.20GHz&id=3534

[8] https://ark.intel.com/content/www/fr/fr/ark/products/212287/intel-xeon-platinum-8380-processor-60m-cache-2-30-ghz.html

[9] https://www.cpubenchmark.net/cpu.php?cpu=Intel+Xeon+Platinum+8380+%40+2.30GHz&id=4483

Table I: Specification of three processors Intel® Xeon® of micro-architecture Skylake®

	Unit	Xeon® Silver 4114		Xeon® Gold 5220		Xeon® Platinum 8380	
Release date		2017		2019		2021	
PDT	W	85		125		270	
Core amount		10		18		40	
		Total	Per core	Total	Per core	Total	Per core
Manuf. primary energy	kWh	67.26	6.73	79.03	4.39	126.62	3.17
Manuf. carbon weight [38]	$kgCO_2e$	27.77	2.78	32.62	1.81	52.27	1.31
Manuf. price	€	704	70.40	1 664	92.44	8 666	216.65
Run on primary energy	kWh	744.60	74.46	1 095.00	60.83	2 365.20	59.13
Run on carbon weight	$kgCO_2e$	42.37	4.24	62.31	3.46	134.58	3.36
Run on price	€	129.56	12.96	190.53	10.59	411.54	10.29
Performance		13 125 [?]	1 312	25 740 [?]	1 430	62 318 [?]	1 558

A. Why does equation 14 approximates the actual global energy and by which margin?

In order to confirm the replacement model that we suggest, one has to know for certain that the global energy we compute is in good agreement with reality. Equation 14 is an approximation since potentially fractional ratios (eq. 8,9) can generate a fractional amount of Silver and Gold cores (eq. 6, 4, 7). The same holds on CPUs since the amount of cores is not necessarily a multiple of the amount of cores of a single CPU. To assess the difference, we have investigated two cases: a high amount of Silver (100 CPUs – 1 000 cores) and a low amount of Silver (5 CPUs – 50 cores). For both initial amounts of Silver, we have determined two amounts of Gold (labelled “rational” and “integer”) for each of the three scenarios. A rational value corresponds to the computed values of cores and CPUs from equations 6,7,4. An integer value stand for upper rounded values of cores and CPUs. They are gathered in table II. For each scenario, we computed the standard deviation of the global energy. They have been determined for the three impacts – PE, GWP, and Price. The standard deviations are gathered in table II as well. We see that for a fixed-workload replacement, the calculated amounts of Gold are 50.99 (51 if rounded upper) and 2.55 (3 if rounded upper). The standard deviations of the three impacts are $0.01\,\%$ in the case of 100 Silver. However, the standard deviations in the case of low amount of Silver are higher, between $9.99\,\%$ and $11.61\,\%$. Such an increase of the standard deviation is expected intuitively since rounding errors are absolute and benefit from the dampening of larger numbers. A similar behavior of the global energy standard deviations for a fixed-core replacement is observed as well. The standard deviations are lower for 100 Silver while it is higher for 5 Silver. Since the fixed-CPU replacement implies a conservation of the amount of CPUs, the standard deviations are $0\,\%$ in this case.

Table II: Integer and rational amounts of Gold that are manufactured to replace Silver for the three scenarios and the related standard deviation of the global energy from the computed amount of CPUs.

	Integer	Rational	σ(PE)	σ(GWP)	σ(Price)
		Fixed-workload			
100 Silver	51	50.99	0.01 %	0.01 %	0.01 %
5 Silver	3	2.55	11.61 %	11.10 %	9.99 %
		Fixed-core			
100 Silver	56	55.56	0.60 %	0.58 %	0.52 %
5 Silver	3	2.78	5.73 %	5.47 %	4.93 %
		Fixed-CPU			
100 Silver	100		0 %	0 %	0 %
5 Silver	5				

The conclusion to this question is that the non-discrete replacement model that we suggest is valid in the case of large values of CPUs, 100 being already large enough. Moreover one has to keep in mind that we only take into account the manufacturing and the use of processors. We do not consider the processors end-of-life due to a significant lack of available information, which means that the global energy is underestimated.

B. Does the Mean Time Between Failure (MTBF) have a significant effect on the impacts of CPU replacement?

Table III: Global energy for a total renewal in 2019 for the three scenarios without MTBF $G_{0\%}$ and the standard variations of the global energy with various MTBF values.

Scenario	$G_{0\,\%}$	$\sigma_{3.00\cdot10^{-3}\,\%}$	$\sigma_{0.35\,\%}$	$\sigma_{2.90\,\%}$
		Primary energy (MWh)		
Fixed-CPU	1 039.55	$2.00\cdot10^{-4}\,\%$	0.02 %	0.17 %
Fixed-Core	646.70	$2.00\cdot10^{-4}\,\%$	0.02 %	0.17 %
Fixed-Workload	606.35	$2.00\cdot10^{-4}\,\%$	0.02 %	0.17 %
		Global warming potential (tCO_2e)		
Fixed-CPU	64.36	$1.00\cdot10^{-3}\%$	0.13%	1.15%
Fixed-Core	40.75	$1.00\cdot10^{-3}\%$	0.13%	1.10%
Fixed-Workload	38.33	$1.00\cdot10^{-3}\%$	0.14%	1.09%
		Price (K€)		
Fixed-CPU	415.14	$8.00\cdot10^{-3}\%$	0.90%	8.63%
Fixed-Core	273.44	$8.00\cdot10^{-3}\%$	0.92%	7.61%
Fixed-Workload	258.88	$9.00\cdot10^{-3}\%$	1.04%	7.44%

By definition, the higher the MTBF the longer a processor operates without failure. The MTBF of electronic devices is not always provided by manufacturers. However, B. Schroeder and G. A. Gibson [42] have analyzed failure data from a high-performance computing Data Center. 23 000 failures have been recorded between 1996 and 2005, $64\,\%$ of them are coming from hardware breakdowns and among them $42.8\,\%$ result from CPU breakdowns. Thus, annually, 700 failures stem from CPU, over 24 101 processors, which gives a MTBF value of $301\,607\,\mathrm{h}$ (around 34 years). In the end, it means that every year $2.90\,\%$ CPUs fail due to hardware reasons. Two theoretical studies on hardware and software failures in Data Centers tackle CPU MTBF [43], [44]. In [43], the MTBF of CPU under normal operation condition is estimated to be $260\,000\,000\,\mathrm{h}$, equivalently, each year $0.003\,\%$ of the CPUs break. In [44], the MTBF value chosen as input parameter for their study is $2\,500\,000\,\mathrm{h}$. It corresponds to $0.35\,\%$ of CPUs failing every year. In order to evaluate the effect of the MTBF on the costs, we consider that the 100 Silver CPUs are all replaced in 2019. The Gold are used until the end of 2026. We consider four MTBF values ($0\,\%$, $3.00 \cdot 10^{-3}\,\%$, $0.35\,\%$ and $2.90\,\%$) and we compute their corresponding global energies ($G_{0\,\%}$, $G_{3.00 \cdot 10^{-3}\,\%}$, $G_{0.35\,\%}$, $G_{2.90\,\%}$). In table III, we gather $G_{0\,\%}$ and the variation of it with the global energies computed with MTBF values: $\sigma_{3.00 \cdot 10^{-3}\,\%}$, $\sigma_{0.35\,\%}$, $\sigma_{2.90\,\%}$. The global energies are expressed in MWh, tCO$_2e$ and K€ in order to evaluate the energetic, environmental and economic impact.

The answer is yes, MTBF has a non-negligible impact on the global energy consumption considering the energetic, carbon and economic impacts. The lower MTBF value ($3.00 \cdot 10^{-3}\,\%$) triggers very few variations on the global energy. The percentage variations are respectively $2.00 \cdot 10^{-4}\,\%$, $1.00 \cdot 10^{-3}\,\%$ and $8.00 \cdot 10^{-3}\,\%$ for PE, GWP, and the price for each of the three scenarios. The intermediate MTBF value ($0.35\,\%$) implies less than one percent variations for 8 out of 9 combinations impact/scenario. Only the percentage deviation of the price in the fixed-workload replacement case shows an increase of $1.04\,\%$. However it remains still low in regards to the initial values of the global cost $G_{0\,\%}$. The higher MTBF value ($2.90\,\%$) has a non-negligible impact on the variation of the global energy. Even though it is lower or equal to $1.15\,\%$ for the three scenarios for primary energy and global warming potential, one cannot bypass the effect on the price. Indeed, a MTBF of $2.90\,\%$ leads to a price increase of around $8\,\%$ for the three scenarios. In the worse case of the fixed-CPU scenario, the increase corresponds to a global cost of $G_{2.90\%} =$€$450.96\,\mathrm{K}$. It is €$35.82\,\mathrm{K}$ higher than the global cost for the fixed-core scenario with no MTBF ($G_{0\%} =$ €$415.14\,\mathrm{K}$).

From now on, we consider an MTBF value of $2.90\,\%$ for Silver and Gold CPUs.

IV. RESULTS APPLYING OUR MODEL

In this section, we apply our model to a comparison between Silver and Gold CPUs and address the problem of how and when we should renew them in the Data Center in order to minimize the impacts. The result of our application is that, in the case of the fixed-workload scenario, replacing all Silver by Gold in 2019 has the lowest economical, energetic, and environmental impacts. Otherwise said, if we plan to eventually upgrade all our processors, it is best to do it as soon as the new generation of CPUs becomes available. However, this answer holds only in the two-CPUs case, and CPUs with different characteristics, or run/produced in different countries, could have given different answers. We analyse the results on three parameters of our model: the CPU replacement fraction ρ, the frequency of the replacements (in the case of a partial replacement policy) and the starting date of the replacement scheme. The code used for this application of our model is available online[10].

A. *The fractional and frequency approaches*

We assume that for all scenarios, the first replacement date is 2019, at the Gold release. In figure 2, the orange lines represent a total replacement in 2019 and the blue lines represent an annual $\rho = 50\,\%$ replacement rate of the remaining Silver CPUs. The long dashed lines correspond to the fixed-CPU scenario, the short spaced lines correspond to the fixed-core scenario, the dotted lines correspond to the fixed-workload scenario, while the black plain lines represent a no-renewal scenario, where the Silver CPUs are kept indefinitely. We gather the amount of cores in (a) and (b), the primary energy values in (b) and (f), the carbon footprint data in (c) and (g), and the price in (d) and (h).

All cases of the $\rho = 100\,\%$ fraction of figure 2 quickly reach constant annual primary energy consumption, annual carbon dioxide equivalent emissions and annual expenses. Moreover, the fixed-CPU scenario increases the consumption (primary energy – $109.73\,\mathrm{MWh}$/year, carbon dioxide equivalent – $6.33\,\mathrm{tCO}_2e$/year, and price – €$23.81\,\mathrm{K}$/year), compared to the no-renewal scenario. However, the two other scenarios yield lower annual consumption, for all three impacts. The only excessive consumption – in regard to the no-renewal scenario – is observed in 2019 and is due to the manufacturing costs.

The annual $\rho = 50\,\%$ replacement means that every year starting from 2019, half of the remaining Silver CPUs are replaced with Gold CPUs. We can observe, in figure 2e), the case of the fixed-core renewal: in 2018 the Data Center is composed of $1\,000$ Silver cores, in 2019 it is composed of 500 Silver cores and 500 Gold cores, in 2020 it is composed of 250 Silver cores and 750 Gold cores and so on. Such fractional replacement allows to pay off the replacement in 2019 compared to a total replacement. In the end, the limit is the same as the total replacement. The primary energy consumption in 2019 is equal to $95.93\,\mathrm{MWh}$ (for an annual $\rho = 50\,\%$ replacement) while it is equal to $117.40\,\mathrm{MWh}$ (for a total replacement). Similarly, the carbon dioxide equivalent emission is equal to $6.87\,\mathrm{tCO}_2e$ while it is equal to $9.49\,\mathrm{tCO}_2e$ for a total replacement in 2019. The annual expenses in 2019 is almost divided by two for the fractional replacement (€$99.20\,\mathrm{K}$ in comparison to €$185.45\,\mathrm{K}$ for the total replacement). In addition, we see that the limit of the evolution of the core amount per year, primary energy consumption, carbon dioxide equivalent and annual expenses are reached very quickly. The core amount of Gold CPUs in 2023 is $1\,743.75$ (fixed-CPU scenario), 968.75 (fixed-core scenario) and 889.15 (fixed-workload scenario), which corresponds to $97\,\%$ of the total amount of Gold CPUs. The replacements can be considered as finished in 2023.
We can identify that for the chosen CPU Gold and Silver (see figure I), for both $\rho = 100\%$ and $\rho = 50\%$, for the fixed-workload and core, we reach an energy, carbon footprint and expenses lower than the no renewal scenario. We would then expect that after several years from the first refreshment year (2019), the cumulative energy consumption and carbon footprint would be lower.

Figure 3 presents the cumulative primary energy (a and d), the cumulative carbon dioxide equivalent (b and e), and

---

[10] https://github.com/Constellation-Group

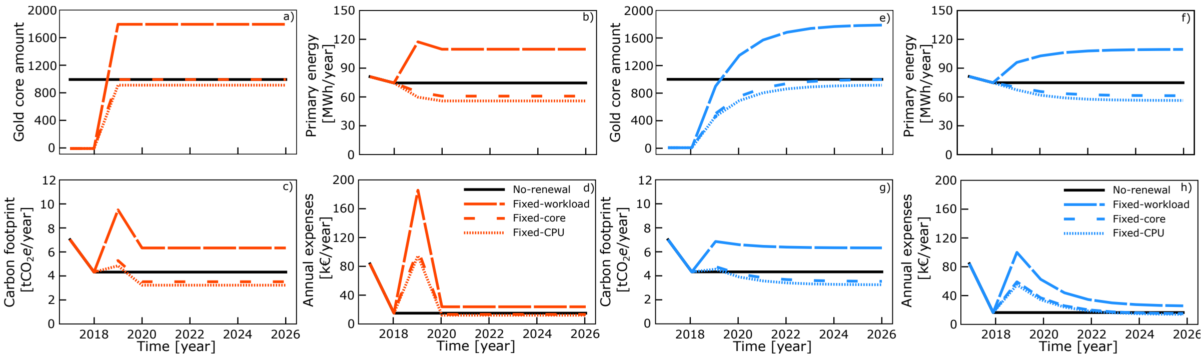

Figure 2: Multi-criteria evolution – PE [MWh] (b), GWP [tCO$_2e$] (c), and spendings [K€] (d)– for a total replacement in 2019 and (f–h) for a $\rho = 50\,\%$ annual replacement rate, considering four scenarios.

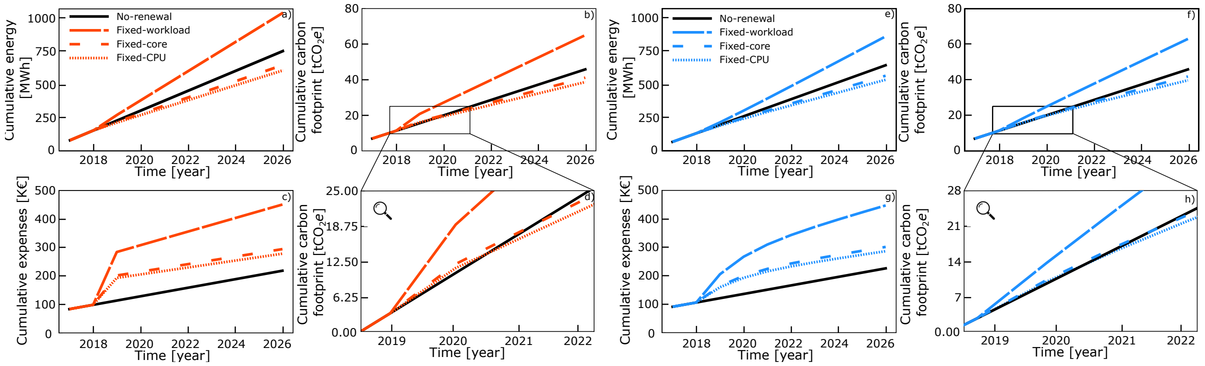

Figure 3: Cumulative multi-criteria evolution – PE [MWh] (a), GWP [tCO$_2e$] (b), and spendings [K€] (c)– for a total replacement in 2019 and (d–f) for a $\rho = 50\,\%$ annual replacement rate, considering four scenarios.

the cumulative expenses (c and f) over the years for the three scenarios in the case of a total replacement and of an annual $\rho = 50\,\%$ replacement, respectively. We can see that in the case of the fixed-CPU replacement – for both total and fractional replacement – the three cumulative impacts are always higher than the ones in the case of no replacement. It rules out the possibility that the fixed-CPU replacement may allow to reduce at least one of the three impacts. However, for a fixed-core and fixed-workload replacement, the cumulative primary energy are lower from 2019 than the one in the no-replacement case. Thus, the total and fractional replacement allow to reduce the primary energy consumption from the very first year of renewal dates. In addition to that, in the case of the fixed-core replacement, the carbon dioxide emission are higher in 2019 and 2020 in regards to the no-replacement case. It is higher only for 2019 in the fixed-workload replacement case. Then, one need to wait one (fixed-workload replacement) or two years (fixed-core replacement) to reduce the carbon dioxide emissions (see the zooms of the carbon footprints). Unlike primary energy and carbon dioxide equivalent, the cumulative expenses, for both total and fractional replacements and both scenarios, are higher than in the case of no replacement. If one expects to make savings, one has to wait until 2048 for the cumulative expenses of the replacements to become lower than the cumulative expenses in the case of no replacement. Such behavior of the cumulative expenses is due to the fact that the manufacture price per core for Gold is 1.31 higher than for the Silver and the run on price per core for Gold is 1.22 lower than for the Silver. One has to wait longer to soften the replacement expenses.

Then in our case, the manufacturing carbon weight per core is of 2.78 kgCO$_2e$. (CPU Silver) and 1.81 kgCO$_2e$. (CPU Gold), and the run on carbon weight per core is of 4.24 kgCO$_2e$. (CPU Silver) and 3.46 kgCO$_2e$. (CPU Gold). It is then obvious that since the manufacturing and run on per core is lower for CPU Gold, we reach a cumulative carbon footprint lower than the one in the no renewal case. It is then achieved in a maximum of two years. We need to highlight again that this is only due to the technical specification of the CPU we used. Another result would have been obtained with another type of CPU.

Figure 4 shows the evolution of the primary energy in MWh (in (a), (b) and (c)), the carbon dioxide equivalent in tCO$_2e$ (in (d), (e) and (f)), and the expenses in K€ (in (g), (h) and (i)), depending on the renewal frequency. For each impact, the three scenarios are considered. With a 10-years 2017–2026 lifespan and a 2019-market availability for the Gold, the frequency varies from 1 to 8 years. We consider five cases: $0\,\%$ (in black), $25\,\%$(in green), $50\,\%$ (in blue), $75\,\%$ (in yellow), and $100\,\%$ (in red) of the remaining Silver CPUs are replaced by Gold CPUs at each stage. The first and last strategies correspond to the no replacement and total replacement. The evolution of the three impacts depends on the frequency of the renewal stages, that start in 2019. For example, a frequency of 1 year means yearly replacements, while a frequency of 3 years yields replacements in 2019, 2022, and 2025. We see in the nine graphics that the costs for a total replacement and for no replacement do not depend on the frequency. Indeed, the total replacement policy leaves no Silver left after 2019 and the no-replacement policy runs the initial Silver from 2017 to 2026. Total fixed-workload and fixed-core replacements lead to 917.83 and 1 000 cores respectively. Both manufacture and use costs per core are lower for Gold than for Silver (see table I), and better performance per core entails a smaller amount of Gold cores compared to Silver. Those factors interfere constructively, so the total replacement is a lower bound for the primary energy and carbon dioxide equivalent, while the no-replacement case is the upper bound. Between these bounds, higher frequencies and fractions of replacements entail lower global primary energy and carbon dioxide equivalent. The situation is more complex for the economical impact, since the manufacturing cost of the Gold core is higher than for the Silver, but the use cost is lower. As it can be computed, in the case of a total and fixed-workload replacement, the replacement expenses are fully amortized only after 2048. This equilibrium point is out of reach of our 10-years 2017–2026 window. Consequently, the total replacement is an upper bound and the no replacement is a lower bound for global expenses. Because of the high manufacturing price, the lower the renewal frequency and the replacement fraction, the lower the global expenses are. We also observe a strong sensitivity to the frequency for fractional replacements, due to the 10-year window effect. Indeed, a 3-years period yields 3 replacements, while 4-7 years yield only two replacements. Note that for a frequency of 1, 2 and 3 years for a fraction of 75 % and for a frequency of 1 year for a fraction of 50 %, global expenses slightly exceed the total-replacement upper bound. We suppose that such a behavior is caused by the fact that due the fractional amount of replaced CPUs, the amount of Silver only tends to zero, but never totally reaches it. Finally, figures 4 c), f) and i) show that the fixed-CPU scenario, that preserves the amount of CPUs, is not efficient. It wastes extra computing power, primary energy, carbon dioxide equivalent, and money. Indeed, no matter the replacement fraction or frequency, a fixed-CPU strategy has always higher costs than no replacement.

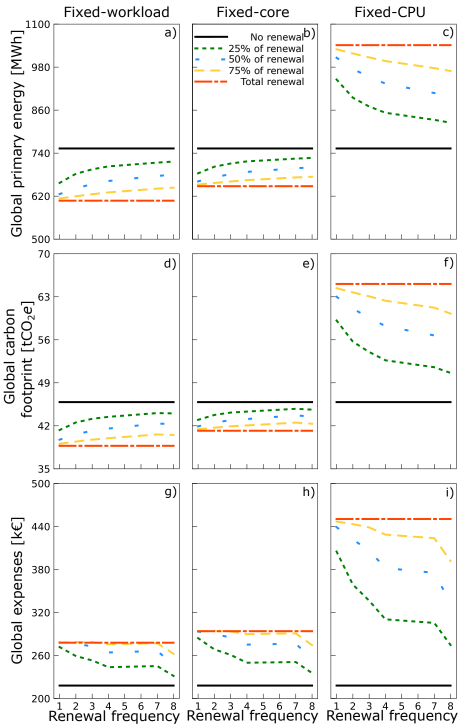

Figure 4: Evolution of PE [MWh] (a–c), GWP [tCO_2e] (d–f), and expenses [K€] (g–i) along the renewal frequencies, for the 25%, 50%, 75%, and 100% replacement rates.

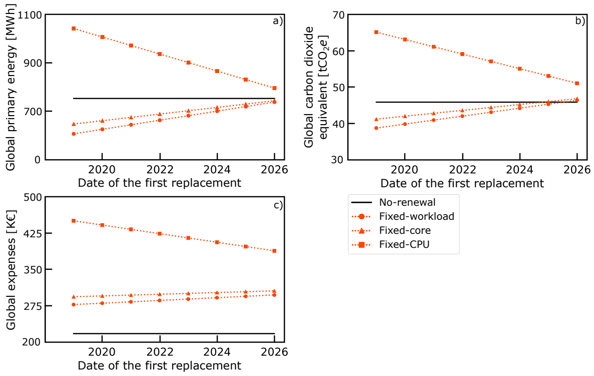

Figure 5: Global multi-criteria evolution – PE [MWh] (a), GWP [tCO_2e] (b), and expenses [K€] (c) – along the date of a total replacement.

B. The starting date approach

The evolution of the global primary energy, carbon dioxide equivalent, and expenses according to the first replacement date for a total replacement are shown in figure 5 (a), (b) and (c), respectively. We see that the fixed-CPU scenario leads to extra consumption of all three resources in regards to the no-renewal scenario.

In figure 5 (a), the fixed-workload and fixed-core scenarios allow to reduce the global primary energy. The lowest point is reached when the first replacement date comes as soon as possible. Additionally, the global primary energy of all scenarios converge as the first replacement date increases. One see, in figure 5 (b), a reduction of the carbon dioxide emissions only if the total replacement is performed in 2025 for the fixed-workload scenario and in 2024 at the latest for the fixed-core scenario. Indeed, we have seen previously that one has to wait one or two years – depending on the scenario – to soften the carbon dioxide emissions due to the CPU manufacturing. Thus, a fixed-workload total replacement in 2026 or a fixed-core replacement in 2025 are over consuming. Apart from not performing any replacement, which is the most economical scenario, it is better to replace (fixed-workload and fixed-core scenarios) on the very first year of the Gold market release.

Finally, the best replacement policy is a fixed-workload total replacement in 2019. For a Data Center composed of 100 Silver, it allows to save 145.70 MWh, 7.12 tCO_2e and costs 59 801 € in comparison to not performing any replacement at all. This specific conclusion is only valid in the case of replacing Silver with Gold, which specifications are shown in table I. Although the fixed-workload scenario is the most relevant one, we have to keep in mind that it is based on the conservation of the performance per core. Such criterion could be in high-performance computing systems best interest, since calculus depends a lot on the CPU's performances. In the case of Data Centers that host servers, cloud services, and virtual machines, it may better not to preserve the workload and to adopt a fixed-core policy. Such replacement would save $105.28\,\mathrm{MWh}$ and $4.66\,\mathrm{tCO}_2e$ and would cost €75 876.

V. DISCUSSION: EXTENDING OUR MODEL TO A THREE-CPU CASE

The previous case study has been assessed by simulating the two-CPUs model (see section II). However, our simulator is not able for the moment to support three or more CPUs models and suggest better replacement possibilities or even forecasts, by extrapolating future processors performances and price evolution. This section aims to discuss the model we presented earlier by adding a third CPU. Indeed, one would argue that the most ecological refreshment is to renew each CPU one after another, but this is very CPU-sensitive regarding their technical characteristics. We aim to broaden our model into a more general one. Nonetheless, we have used a spreadsheet to compare the Xeon®Silver 4114, Xeon®Gold 5220, and Xeon®Platinum 8380 CPUs (see table I). They are labelled Silver, Gold, and Platinum, respectively. We assume that the Data Center is filled with 100 Silver in 2017 and the Data Center is running on from 2017 to the end of 2026. Since Gold and Platinum do not appear on the market at the same time, one can enumerate thirteen cases to evaluate the replacement of Silver:

- Case 1: None of the processors P_1 are replaced;
- Case 2: 50% of Silver are replaced in 2019 by Gold;
- Case 3: All Silver are replaced in 2019 by Gold;
- Case 4: 50% of Silver are replaced in 2019 by Gold, the remaining are replaced in 2021 by Platinum;
- Case 5: 50% of Silver are replaced in 2019 by Gold, the remaining are replaced in 2021 by Gold;
- Case 6: 50% of Silver are replaced in 2019 by Gold, themselves replaced in 2021 by Platinum;
- Case 7: 50% of Silver are replaced in 2019 by Gold and, in 2021, everything is replaced by Platinum;
- Case 8: All Silver are replaced in 2019 by Gold and, in 2021, 50% of the Gold are replaced by Platinum;
- Case 9: All Silver are replaced in 2019 by Gold, themselves replaced in 2021 by Platinum;
- Case 10: 50% of Silver are replaced in 2021 by Gold;
- Case 11: All Silver are replaced in 2021 by Gold;
- Case 12: 50% of Silver are replaced in 2021 by Platinum;
- Case 13: All Silver are replaced in 2021 by Platinum.

Even though we have shown in IV that the fixed-workload scenario is the more suitable – and to a certain extent the fixed-core scenario – we evaluate the replacement of Silver by two different CPUs according to the three scenarios: fixed-workload, fixed-core, and fixed-CPU. The reader would have noticed that the fractional replacement is restricted to a 50% fraction. This choice has been made to optimize the number of cases, while still showcasing the effect of a fractional replacement. Table IV shows the global primary energy [MWh], the global carbon dioxide [tCO_2e], and the global expenses [K€] of the thirteen cases for a time window of 10 years (from 2017 to 2026). The fixed-workload scenario, case 9 – all of the Silver are replaced by Gold in 2019, then the Gold are replaced by Platinum in 2021 – is the one that allows the lowest global primary energy consumption (573.60 MWh) and the lowest global carbon dioxide equivalent (37.66 tCO_2e). However, it is the mot expensive one (466.08 K€).

It is quite different for the fixed-core scenario. Case 9 remains the one that allows the lowest global primary energy consumption for this scenario (640.43). Unlike the fixed-workload scenario, case 3 – all of the Silver are replaced in 2019 by Gold – is the case that allows the lowest global carbon dioxide equivalent (41.20 tCO_2e). It is due to the fact that replacing the Gold by Platinum does not allow to soften the first replacement of Silver by Gold (which takes more or less two years, see figure 3). It is then less environmentally impacting to keep Gold and not to replace them with Platinum. Once again, we can see that the fixed-CPU scenario is not feasible since case 1 – no P_1 processors are replaced – is the one that gives the lowest global primary energy (753.08 MWh), global carbon dioxide equivalent (45.87 tCO_2e), and global expenses (218.33 K€).

Both of the fixed-workload and fixed-core scenarios for cases 9 and 3 show significant costs. It is better not to replace Silver if one wants to save money. However, in the replacement cases for the fixed-workload scenario, case 2 – 50% of Silver are replaced in 2019 by Gold – is the one that is the least expensive (247.22 K€), but it leads to extreme global primary energy (680.13 MWh) and global carbon dioxide equivalent (43.80 tCO_2e) values. For the fixed-core scenario, case 2 is the least expensive as well (255.67 K€).

Case 3 for the fixed-core scenario seems to be the best replacement to suggest. Indeed, the global primary energy is not the lowest but among them (647.79 MWh), we have seen previously that the global carbon dioxide is the lowest for this case and the global expenses are 294.24 K€, which is the same order of magnitude as for case 1.

VI. LIMITATIONS AND THREATS TO VALIDITY

Our model is theoretical and has some limitations. Although the refresh based on workload conservation shows interesting results, it may not be possible to achieve in practice. If the fixed workload is too small, there might be under utilization of resources after refreshing. Also, we cover only the CPU refresh, but one has to keep in mind that all components need to be refreshed eventually. Similar studies have to be performed on the various components if one wants to perform a partitioned refresh. The CPU-refresh we suggest involves a full non-dependency between the components which is far from being the case in real components. Although some work in the industry, such as Compute Express Link (CXL) [45], lean towards more modular approaches, one has to keep in mind that a CPU-refresh involves dependency problems with the rest of the components. Finally, we consider the CPU to work at 100%

Table IV: Global PE [MWh], global GWP [tCO_2e] and global spendings [K€] of the thirteen replacement cases for the three scenarios for a length of time of ten years (from 2017 to 2026).

	Global primary energy [MWh]			Global carbon dioxide eq. [tCO_2e]			Global expenses [K€]		
	Workload	Core	CPU	Workload	Core	CPU	Workload	Core	CPU
Case 1		753.08			45.87			218.33	
Case 2	680.13	700.34	897.12	42.27	43.49	55.44	247.22	255.67	333.62
Case 3	607.37	647.79	1 041.35	38.75	41.20	65.10	278.15	294.24	450.96
Case 4	607.13	655.57	1 389.96	38.45	41.38	85.86	332.72	365.17	908.22
Case 5	625.97	661.39	1 006.18	39.80	41.95	63.05	279.95	294.96	441.06
Case 6	663.24	696.66	1 284.74	41.73	43.79	79.83	341.19	370.37	881.58
Case 7	590.25	651.90	1 777.58	37.91	41.68	110.25	426.69	480.27	1 456.18
Case 8	590.48	644.11	792.41	38.21	41.50	50.41	372.12	409.34	548.63
Case 9	573.60	640.43	1 816.60	37.66	41.80	113.88	466.08	524.44	1 546.87
Case 10	698.83	714.03	862.04	43.36	44.30	53.44	250.04	257.00	324.74
Case 11	644.77	675.18	971.20	40.93	42.81	61.08	283.79	297.71	433.19
Case 12	679.99	708.22	1 245.83	42.01	43.72	76.24	302.82	327.22	791.92
Case 13	607.09	663.56	1 738.77	38.23	41.65	106.70	389.34	438.15	1 367.53

24 hours a day. Note that this is a fictive usage of the CPU and we overestimate its energy consumption.

Another limitation of the model presented in this work is related to the data presented in table I. No information have been given by the supplier nor by waste management companies regarding the carbon footprint and energy cost of the end-of-life of CPUs. Finally, due to lack of information regarding the end-of-life and resource depletion, it was not possible to add them into our model. However, assessing factors a posteriori, after a calculation which would reflect the efficiency and the costs incurred from them, would be the ideal way to consider an effective end-of-life and resource depletion description.

VII. CONCLUSION AND FUTURE WORK

In this work, we investigate new strategies to renew components in a Data Center and evaluate them on a multi-criterion basis. Usually, servers are entirely replaced when time comes. Here, we promote partial refreshment and we propose a model of hardware renewal based on two CPUs. The model assesses the economical, environmental, and energetic impacts of the usage and manufacturing of the CPUs. We also consider an annual 2.90% failure rate.

Studying a real-world example, our CPU-renewal model shows that it is more suitable (from the environmental and energetic point of views) to replace all the old Silver CPUs as soon as the new Gold are released. However, a few years (at least 3) are necessary to soften the environmental costs of the replacement. We also study a three-CPU model in the last part of this article. Our study shows that the fixed-CPU scenario remains the worst when one wants to replace CPUs. Moreover, it appears that replacing all of the Silver by the Gold in 2019, and ignoring the 2021 Platinum is the most interesting choice for a fixed-core scenario. Our results depend on the specifications of the CPUs we have considered in this study. Thus, others CPUs might have led to different conclusions regarding how and when they should be replaced. However, our approach provides the possibility to study various replacement schemes in order to reduce carbon emissions, thus helping with the implementation of carbon-neutral resource-management strategies.

In the scope of this paper, we focus on microprocessors, but the model applies to the replacement of any hardware. We suggest the possibility to replace server hardware in Data Centers separately and not only the replacement of entire servers. Doing so, we would allow to be more selective on the hardware to replace, the amount, the replacement frequency, and the hardware specifications. Being more precise on the replacement scheme would allow Data Centers to tend to carbon neutrality, by replacing only the necessary devices. Following this idea, we are planing to apply our model to more complex infrastructures implying several hardware devices with different optimal replacement periods. The ultimate goal would be to reach the optimal point for replacement for all devices in a Data Center or any complex infrastructure involving several kinds of hardware. At first sight, this approach seems difficult to apply in practice since components are hardwired in most modern servers. Also, connection protocols tend to be made for a given generation of hardware (CPU and Memory for instance), thus forcing the replacement of memories when changing a CPU version. However, recent trends in the industry, such as Compute Express Link (CXL) [45], move in the direction of a standard to allow heterogeneous versions of CPUs and memories to connect in a very efficient way. Such a standard opens the possibility of modular replacements to optimize better the hardware replacement, as our model suggests.

Finally, another application of our model would be not only to determine the timing for replacing hardware, but also, the timing for manufacturing and distributing new hardware. For example, given a set of data centers hosting various models of servers and CPUs, would the mass manufacturing of a new generation of hardware have a global positive or negative impact? Scaling our model to allow a quantifiable answer to this question is part of the perspectives that could be studied from this work as a starting point.

VIII. ACKNOWLEDGMENTS

We acknowledge Constellation which funded and supported this research project.

REFERENCES

[1] greenIT. http://www.greenit.fr/. Accessed: 2022-04-11.

[2] Nicola Jones. The Information Factories. *Nature*, 561:163–166, 2019.

[3] Qixiang Cheng, Meisam Bahadori, Madeleine Glick, Sébastien Rumley, and Keren Bergman. Recent advances in optical technologies for data centers: a review. *Optica*, 5(11):1354–1370, 2018.

[4] Kim Hazelwood, Sarah Bird, David Brooks, Soumith Chintala, Utku Diril, Dmytro Dzhulgakov, Mohamed Fawzy, Bill Jia, Yangqing Jia, Aditya Kalro, James Law, Kevin Lee, Jason Lu, Pieter Noordhuis, Misha Smelyanskiy, Liang Xiong, and Xiaodong Wang. Applied machine learning at facebook: A datacenter infrastructure perspective. In *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*, pages 620–629, 2018.

[5] Weng Chon Ao, Po-Han Huang, and Konstantinos Psounis. Joint workload distribution and capacity augmentation in hybrid datacenter networks. *IEEE/ACM Transactions on Networking*, 29(1):120–133, 2021.

[6] Ayaz Ali Khan and Muhammad Zakarya. Energy, performance and cost efficient cloud datacentres: A survey. *Computer Science Review*, 40:100390, 2021.

[7] Stephen Lindsay, Kate Cavanagh, and Terijo Lovasz. Data centres and the australian energy sector. *CSIRO National Energy Analytics Research (NEAR) Program*, 2021.

[8] Nuoa Lei and Eric R Masanet. Global data center energy demand and strategies to conserve energy. *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*, pages 15–26, 2021.

[9] Vehbi Comert, Mustafa Altun, Mustafa Nadar, and Ertunc Erturk. Warranty forecasting of electronic boards using short-term field data. In *2015 Annual Reliability and Maintainability Symposium (RAMS)*, pages 1–6, 2015.

[10] Anisur Rahman and GN Chattopadhyay. Lifetime warranty policies: complexities in modelling and potential for industry application. In *Proceedings of the Fifth APIEMS Conference 2004*, pages 28.1.1–28.1.8, 2004.

[11] Xuejiao Han, Maoliang Hu, Zhanqing Yu, Kangsheng Cui, Lu Qu, and Xuejian Zhao. Data center new energy system under the "peak carbon dioxide emissions and carbon neutrality" background and comprehensive evaluation analysis thereof. In *2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2)*, pages 1649–1656, 2021.

[12] Zhiwei Cao, Xin Zhou, Han Hu, Zhi Wang, and Yonggang Wen. Toward a systematic survey for carbon neutral data centers. *IEEE Communications Surveys & Tutorials*, 24(2):895–936, 2022.

[13] Marcos De Melo da Silva, Abdoulaye Gamatié, Gilles Sassatelli, Michael Poss, and Michel Robert. Optimization of data and energy migrations in mini data centers for carbon-neutral computing. *IEEE Transactions on Sustainable Computing*, pages 1–15, 2022.

[14] John Judge, Jack Pouchet, Anand Ekbote, and Sachin Dixit. Reducing data center energy consumption. *Ashrae Journal*, 50(11):14, 2008.

[15] Emerson Network Power. Energy logic: Reducing data center energy consumption by creating savings that cascade across systems. *White paper, Emerson Electric Co*, 9, 2009.

[16] Arman Shehabi, Sarah J Smith, Eric Masanet, and Jonathan Koomey. Data center growth in the united states: decoupling the demand for services from electricity use. *Environmental Research Letters*, 13(12):124030, 2018.

[17] Eric Masanet, Arman Shehabi, Nuoa Lei, Sarah Smith, and Jonathan Koomey. Recalibrating global data center energy-use estimates. *Science*, 367(6481):984–986, 2020.

[18] Maria Avgerinou, Paolo Bertoldi, and Luca Castellazzi. Trends in data centre energy consumption under the european code of conduct for data centre energy efficiency. *Energies*, 10(10):1470, 2017.

[19] Scott Huck. Measuring processor power. *Intel Corporation*, 2011.

[20] Alon Naveh, Efraim Rotem, Avi Mendelson, Simcha Gochman, Rajshree Chabukswar, Karthik Krishnan, and Arun Kumar. Power and thermal management in the intel core duo processor. *Intel Technology Journal*, 10(2), 2006.

[21] Brett Battles, Cathy Belleville, Susan Grabau, and Judith Maurier. Reducing data center power consumption through efficient storage. *Network Appliance, Inc*, 2007.

[22] Carolina Koronen, Max Åhman, and Lars J Nilsson. Data centres in future european energy systems—energy efficiency, integration and policy. *Energy Efficiency*, 13(1):129–144, 2020.

[23] Jiacheng Ni and Xuelian Bai. A review of air conditioning energy performance in data centers. *Renewable and sustainable energy reviews*, 67:625–640, 2017.

[24] Alex K Jones, Yiran Chen, William O Collinge, Haifeng Xu, Laura A Schaefer, Amy E Landis, and Melissa M Bilec. Considering fabrication in sustainable computing. In *2013 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 206–210, 2013.

[25] Erik Brunvand, Donald Kline, and Alex K Jones. Dark silicon considered harmful: A case for truly green computing. In *2018 Ninth International Green and Sustainable Computing Conference (IGSC)*, pages 1–8, 2018.

[26] Alex K Jones, Liang Liao, William O Collinge, Haifeng Xu, Laura A Schaefer, Amy E Landis, and Melissa M Bilec. Green computing: A life cycle perspective. In *2013 International Green Computing Conference Proceedings*, pages 1–6, 2013.

[27] Rabih Bashroush. A comprehensive reasoning framework for hardware refresh in data centers. *IEEE Transactions on Sustainable Computing*, 3(4):209–220, 2018.

[28] Rabih Bashroush and Andy Lawrence. Beyond pue: tackling it's wasted terawatts. *Uptime Institute*, 2020.

[29] Rabih Bashroush, Nour Rteil, Rich Kenny, and Astrid Wynne. Optimizing server refresh cycles: The case for circular economy with an aging moore's law. *IEEE Transactions on Sustainable Computing*, 2020.

[30] Joseph Doyle and Rabih Bashroush. Case studies for achieving a return on investment with a hardware refresh in organizations with small data centers. *IEEE Transactions on Sustainable Computing*, 6(4):599–611, 2020.

[31] Fraunhofer IZM Bio by Deloitte. Tpreparatory study for implementing measures of the eco-design directive 2009/125/ec dg entr lot 9 - enterprise servers and data equipment task 5: Environment & economics. *European Commission, DG Internal Market, Industry, Entrepreneurship and SMEs*, 2015.

[32] Chaoqiang Jin, Xuelian Bai, Chao Yang, Wangxin Mao, and Xin Xu. A review of power consumption models of servers in data centers. *applied energy*, 265:114806, 2020.

[33] Aurélien Bourdon, Adel Noureddine, Romain Rouvoy, and Lionel Seinturier. Powerapi: A software library to monitor the energy consumed at the process-level. *ERCIM News*, 2013(92), 2013.

[34] Adel Noureddine. Powerjoular and joularjx: Multi-platform software power monitoring tools. In *18th International Conference on Intelligent Environments*, 2022.

[35] Constanța Zoie Rădulescu and Delia Mihaela Rădulescu. A performance and power consumption analysis based on processor power models. In *2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, pages 1–4, 2020.

[36] Maxime Colmant, Romain Rouvoy, Mascha Kurpicz, Anita Sobe, Pascal Felber, and Lionel Seinturier. The next 700 cpu power models. *Journal of Systems and Software*, 144:382–396, 2018.

[37] Cpu specifications die size. https://github.com/Boavizta/boaviztapi/blob/main/boaviztapi/data/components/cpu_manufacture.csv. Accessed: 2022-07-08.

[38] Carbon intensity of electricity. https://ourworldindata.org/grapher/carbon-intensity-electricity?tab=table. Accessed: 2022-07-13.

[39] Price of 1kwh in france. https://www.totalenergies.fr/particuliers/parlons-energie/dossiers-energie/facture-d-energie/les-elements-a-prendre-en-compte-pour-calculer-les-kwh-en-euros#:~:text=Pour%20bien%20calculer%20les%20kWh,Tarif%20Bleu%20pratiqu%C3%A9%20par%20EDF). Accessed: 2022-07-13.

[40] Location of intel factories. https://www.intel.fr/content/www/fr/fr/architecture-and-technology/global-manufacturing.html. Accessed: 2022-07-13.

[41] Carbon intensity of electricity in france. https://bilans-ges.ademe.fr/fr/basecarbone/donnees-consulter/liste-element/categorie/64. Accessed: 2022-07-13.

[42] Bianca Schroeder and Garth A. Gibson. A large-scale study of failures in high-performance computing systems. *IEEE Transactions on Dependable and Secure Computing*, 7(4), 2010.

[43] Sanghoan Chang and Gwan Choi. Timing failure analysis of commercial cpus under operating stress. In *2006 21st IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, pages 245–253, 2006.

[44] Dong Seong Kim, Fumio Machida, and Kishor S. Trivedi. Availability modeling and analysis of a virtualized system. In *2009 15th IEEE Pacific Rim International Symposium on Dependable Computing*, pages 365–371, 2009.

[45] Debendra Das Sharma. Compute express link®: An open industry-standard interconnect enabling heterogeneous data-centric computing. In *2022 IEEE Symposium on High-Performance Interconnects (HOTI)*, pages 5–12. IEEE, 2022.