Myths and Misconceptions Around Reducing Carbon Embedded in Cloud Platforms

Jialun Lyu
Microsoft Azure
University of Toronto

Jaylen Wang
Microsoft Azure
Carnegie Mellon University

Kali Frost
Microsoft Research

Chaojie Zhang
Celine Irvene
Esha Choukse
Microsoft Azure

Rodrigo Fonseca
Ricardo Bianchini
Fiodar Kazhamiaka
Microsoft Azure

Daniel S. Berger
Microsoft Azure
University of Washington

ABSTRACT

Major cloud providers have stated public plans to lower their carbon emissions. Historically, this has meant focusing on emissions from producing the electricity consumed by datacenters. While work and challenges remain on this avenue, research and industry are actively working on the next step of reducing carbon embedded in servers and racks. At a high level, a promising direction to reduce embodied carbon is to avoid emissions from new manufacturing, which often requires using existing components, devices, and buildings for longer. However, much of the data around carbon breakdowns and reduction opportunities remains silo-ed, leading to speculations and assumptions – both internally and externally – around the opportunities to reduce datacenter carbon intensity. We aim to clarify some of the misconceptions we have encountered.

CCS CONCEPTS

• **Hardware → Enterprise level and data centers power issues**; Aging of circuits and systems; • **Information systems →** *Computing platforms.*

KEYWORDS

Datacenter, cloud computing, carbon, sustainability, embedded carbon

ACM Reference Format:
Jialun Lyu, Jaylen Wang, Kali Frost, Chaojie Zhang, Celine Irvene, Esha Choukse, Rodrigo Fonseca, Ricardo Bianchini, Fiodar Kazhamiaka, and Daniel S. Berger. 2023. Myths and Misconceptions Around Reducing Carbon Embedded in Cloud Platforms. In *2nd Workshop on Sustainable Computer Systems (HotCarbon '23), July 9, 2023, Boston, MA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3604930.3605717

1 INTRODUCTION

The carbon impact of cloud providers spans on-site generators and staff emissions (Scope 1), purchased energy (Scope 2), and carbon embedded in chips, PCBs, racks/server enclosures, and buildings (Scope 3)[1]. As we investigate the opportunities for reducing cloud datacenter carbon intensity, we have come across some misleading statistics, myths, and assumptions – both internally and in academic publications – that can make it challenging to reason about impactful research directions. Our investigations are ongoing and have consisted of analyzing internal data, engaging with product, engineering, supply chain, and sustainability experts, engaging with the research community, and reconciling our experience with internal carbon life cycle assessment (LCA) data. The LCA spans the carbon footprint across all operations at Azure and was performed by an independent engineering consultant. The resulting carbon impact estimates are subject to confidentiality agreements, and their exact numbers may change as life cycle assessments become increasingly detailed. Consequently, we focus on a qualitative description of how such estimates impact our understanding of the challenges and opportunities in reducing cloud datacenter carbon intensity.

Our investigations have focused primarily on Scope 3 (embodied) carbon. We thus first discuss the main high-level approaches to reduce embodied carbon and then delve into myths and experiences. Our subsequent discussion summarizes known barriers and opportunities.

Limitations. The following assumptions underpin our work.

- We rely on multiple LCAs, internal per-component carbon analysis, as well as public numbers. These sources do not always agree, and individual component's carbon emissions estimations can diverge by as much as 5×. However, these differences typically do not affect the overall trends, unless otherwise noted.
- Our experiences come from the perspective of a large public cloud platform. They may but need not apply to other hyperscalers and datacenters in general.

2 APPROACHES TO REDUCE EMBODIED CARBON

At a high level, the biggest opportunity to reduce embodied carbon is to buy/manufacture fewer carbon-intensive objects. According to our LCA, compute servers, storage servers, and network servers/equipment are the main contributors to emissions, while

HotCarbon '23, July 9, 2023, Boston, MA, USA

ACM ISBN 979-8-4007-0242-6/23/07…$15.00
https://doi.org/10.1145/3604930.3605717

---

[1] We assume the scope definitions from the Greenhouse Gas Protocol [22, 36] assuming the cloud provider owns the datacenter and purchases the energy from the utility. When leasing datacenter space, energy might fall under scope 3.

buildings, power, and cooling infrastructure are smaller contributors due to their longer lifetimes.

In order to buy and manufacture fewer servers and network equipment, we see the following four broad-stroke approaches.

Resource efficiency Do more with generally fewer resources. Historic examples are reductions in overheads (e.g., in operating systems and hypervisors [14, 52], or network function virtualization) or scheduling improvements (e.g., scheduling more diverse workloads in larger server pools [42, 68, 72] or using machine-learning to help in scheduling decisions [19, 23]). This approach uniformly reduces all emissions.

Server, rack, and network lifetime extensions Keep existing deployed systems in the datacenter for longer time periods. Specifically, we refer to keeping servers in their original rack in their original datacenter location as originally deployed. This amortizes carbon released during manufacturing over a longer time span.

Component recycle Break down components from servers, racks, and networks into basic building blocks or minerals, e.g., using electrochemical recovery [44]. The recovered materials are then used during the manufacturing of new components [44, 53].

Component reuse When decommissioning a server, separate out components like CPUs, DRAMs, SSDs, NICs etc. and redeploy these components. Typically, this is not applicable to all components, e.g., the motherboard and other printed circuit boards are often not reused. Examples include reusing components in the repair of still-commissioned servers of a similar type [33, 59] or as parts in new servers [21, 33].

3 MYTHS AND EXPERIENCES

We summarize and offer our perspective on 8 myths around embodied carbon for cloud platforms that we have encountered.

Myth 1. ***Scope 3 (embodied) carbon does not matter or only matters in the distant future.*** *For example, a recent comprehensive survey [24] is based on the assumption that Scope 3 constitutes less than 3% of datacenter carbon and surveyed approaches focused exclusively on Scopes 1 and 2.*

Our LCAs indicate that Scope 1 is negligible, which also matches public statements by major cloud providers [59]. When counting only hourly-matched renewable energy, Scope 2 still accounts for the majority of carbon but not overwhelmingly so. For example, 66% of the electricity used at Google datacenters was matched with renewable energy on an hourly basis in 2021 [34]. With historic growth rates, this is likely closer to 70% today. Our LCAs indicate that with 70-75% renewable energy, Scope 3 accounts for close to half of datacenter carbon emissions. Therefore, Scope 3 emissions and embodied carbon are important factors both currently and in the near future.

Myth 2. ***Research should focus only on embodied carbon, as renewable energy will eliminate operational carbon.*** *One might get this idea from publicly released sustainability reports [17, 35, 62], and optimistic interpretations of carbon credits and growth of deployed renewable generation.*

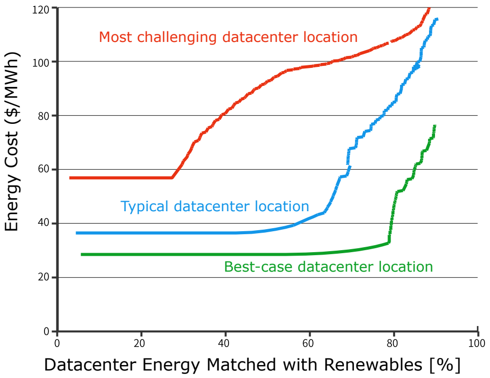

Figure 1: Energy costs grow asymptotically as the share of hourly-matched renewable energy approaches 100%. Data from RMI [29].

For example, Google has been building new renewable energy capacity to offset 100% of consumption since 2017 [35] with AWS [17] and Microsoft [62] on track to reach 100% by 2025. Since this renewable energy replaced carbon emissions and would not have been deployed otherwise, cloud providers have used these deployments to offset cloud datacenter energy consumption. However, these deployments are not necessarily in the same power grids as where datacenters are located (spatial mismatch) and do not necessarily produce power during the same hours of the day when datacenters consume power (temporal mismatch).

Building renewable generation capacity has a real impact. However, relying on others to consume the energy is a practice that can not be sustained indefinitely; it is effective in electricity grids with a low (e.g. < 40%) renewable energy mix where it can be integrated cheaply (without energy storage) and displace fossil fuel-based generators. Wind and solar are already among the cheapest forms of energy generation, and account for the vast majority of new generation capacity globally [15], so this opportunity is diminishing. The more ambitious and impactful goal is to match the datacenter's power consumption both spatially and temporally, and industry is already working towards this [34, 45]. However, reaching 100% of spatially and temporally-matched demand with renewable energy is very costly and thus challenging [28, 29, 57]. In particular, energy costs grow asymptotically as the share of hourly-matched renewable energy approaches even 90% and energy storage costs dominate (Figure 1). Thus, dealing with Scope 2 emissions remains an important research problem.

Myth 3. ***Instead of reusing components, we should prioritize recycling.*** *For example, we have faced this perspective with various internal engineering teams at Azure .*

Component reuse is significantly more effective than recycling, and server lifetime extensions are more effective than component reuse. Leaning on the idea of recycling is tempting because it reduces the complexity on the product side and can excuse a lack of upstream investment in managing embodied carbon. However, in terms of carbon, the returns from recycling are relatively slim, whereas the extra complexity to support reuse can result in significant reductions. This does not, however, mean that recycling is pointless; even slim improvements at the scale of cloud computing providers are worthwhile, and recycling provides other benefits such as the recovery of critical materials [31, 32, 44].

Among different server components, hard drives have the highest potential for recycling. However, recent life cycle assessments of hard drive recycling [44] show that reuse leads to as much as 275× higher carbon reduction than recycling. Finally, server lifetime extensions are more effective than reuse, as not all server components can be effectively reused (e.g., printed circuit boards, management controllers, and often even CPUs).

Myth 4. ***There are few opportunities to further improve resource scheduling in cloud platforms.*** *For example, the Azure scheduler Protean reached 70% packing density [42] in 2020. Recent results [18] showed improvements to 85% while also pointing out that further improvements are challenging.*

Scheduling improvements in cloud platforms typically focus on the packing density: the ratio of the number of cores scheduled for VMs to the number of cores on non-empty machines in a datacenter [42]. Improvements to packing density are increasingly limited by other scheduling objectives such as reducing noisy-neighbor interference and maintaining scheduling throughput [18].

However, scheduling a VM on a core does not mean the core is actually used. Consequently, a high packing density does not necessarily translate to high resource utilization. In fact, approximately $\frac{3}{4}$ of VMs exhibit less than 25% CPU utilization [25]. Thus, we can potentially deploy 56% fewer CPU cores if we oversubscribe scheduled-but-unused cores; other carbon-intensive resources such as DIMMs and SSDs have similar opportunities. However, achieving effective oversubscription presents a challenge for cloud platforms, as customers often overprovision VMs, and convincing them to rightsize VMs is not always successful. In addition, addressing the issue of VM opaqueness poses a significant research challenge, as it requires fusing cloud-scale and node-level telemetry [74].

Myth 5. ***Embodied carbon is high due to 3-year hardware refresh cycles at cloud providers.*** *Short refresh cycles indeed used to be a case years ago, and those numbers have somehow persisted.*

Generally, the cloud industry has moved towards longer lifetimes with the depreciable lifetime of servers and network equipment being between five and six years [43]. Table 1 summarizes some published lifetime values. Depreciable lifetime is a conservative lower bound used in financial accounting. In practice, servers are deployed for longer than their depreciable lifetime. Thus, research should focus on extending lifetime well beyond a dozen years to push the envelope compared to existing industry behavior. There are many factors that go into a decision to decommission existing servers; among them, those that can be influenced by systems research include:

1. Demand for workloads that run effectively on older hardware
2. Maintenance costs
3. Performance/Watt

These factors affect the total cost of ownership (TCO), which drives hardware deployment decisions. If a cluster consisting of older servers is underutilized due to low demand, replacing it with new hardware would re-allocate valuable datacenter power and space to in-demand hardware.

Cloud Provider	Minimum Lifetime	Date Last Updated
Microsoft	6 years	2022
AWS	5 years	2022
Google	6 years	2023
Meta	5 years	2023

Table 1: Minimum lifetimes at hyperscalers are typically 5–6 years. In practice, servers are often used for longer.

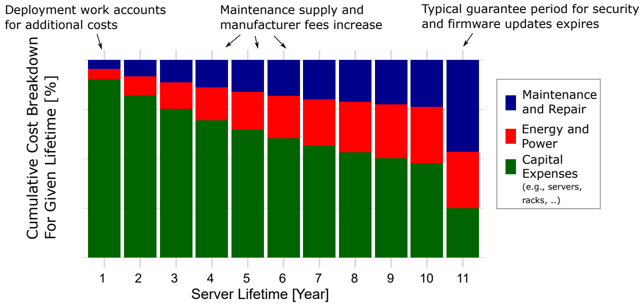

Figure 2: Cloud platform costs (TCO) are increasingly dominated by energy and maintenance costs as server lifetimes are extended. This graph is based on public TCO models where maintenance costs account for 5% of Capex per year [20, 73] and energy accounts for 6% of Capex per year [37, 38, 71]. Divergence from these numbers is indicated on top and largely affects maintenance and repair costs.

Maintenance gets more expensive over time as replacement parts become harder to source, increasing TCO over time (see Figure 2). It is also important to note that system and SoC firmware are supported (e.g., security patches) for typically only 10 years, after which maintenance costs increase significantly.

A more efficient server generation can make older hardware cost-prohibitive to continue running [69]. Research on energy and carbon breakpoints for server upgrades, for example as in Reference [48], is needed.

Myth 6. ***Manufacturers are decarbonizing quickly, and this will shrink cloud providers' downstream embodied carbon costs.*** *While some assume that hardware manufacturers and cloud providers will decarbonize at the same rate, this doesn't quite follow from public reports.*

Cloud platforms won't meet their decarbonization targets if they rely on upstream decarbonization from hardware suppliers to manage embodied carbon. While hardware providers are committing and making progress on decarbonization efforts, the rates at which this is occurring are not uniform or aligned with cloud goals. This is evident in Table 2 which shows the stated net-zero emissions target dates, or lack thereof, for seven of the largest hardware manufacturers, i.e., "fab-full" companies, (on the left) and cloud providers (on the right). When viewing decarbonization rates through such commitments, hardware manufacturers lag significantly behind cloud providers. All cloud providers except Amazon and Oracle have target dates of 2030; however, every hardware company's target date

Hardware		**Cloud**	
Company	*Target*	*Company*	*Target*
GlobalFoundries [3]	None	Alibaba [1]	2030
Intel [5]	2040	Amazon [7]	2040
Micron [10]	2050	Google [2]	2030
Samsung [8]	2050	IBM [4]	2030
SK hynix [70]	2050	Microsoft [67]	2030
Texas Instruments [6]	None	Oracle [9]	2050
TSMC [12]	2050	Tencent [11]	2030
Average	>2048	*Average*	2034

Table 2: Companies and their publicly committed target dates to reach net-zero carbon emissions by hardware (left) and cloud (right) providers. Average target dates for each group is shown, where only existing target dates are taken into account, thus why the average for hardware, where two companies have no targets, is listed as at least being 2048. Hardware manufacturers, on average, lag cloud providers' targets by over 14 years. Data obtained in May/June 2023.

is either not set, for GlobalFoundries and Texas Instruments, or set at 2050 with the exception of Intel (2040).

This divergence can even be seen in fab-less semiconductor companies, which must work closely with manufacturers. For example, Nvidia and AMD, two of the top hardware vendors, both have not set net-zero target dates [6]. If this divergence of decarbonization rates continues and cloud providers hit their renewable energy goals, embodied emissions will dominate, motivating the need for more creative and efficient downstream solutions to reduce embodied emissions.

Myth 7. ***Embodied carbon is a problem for the hardware community and less relevant for system software research.*** *This idea seems to manifest in academic publishing, where sessions and papers on sustainability appear at mainline architecture conferences like HPCA [41], ISCA [40], MICRO [64], and ASPLOS [13, 63, 65] with no sessions and few corresponding publications at OSDI, SOSP, and NSDI.*

System software plays a key role in reducing emissions overall. A breadth of work has shown the role of system software in energy efficiency [16, 26, 39, 47, 49, 56, 58, 60]. More recent work has also shown the critical role of system software in reducing embodied carbon. For example, the high cost of maintenance as well as supply-chain challenges practically limit server lifetime extensions (Myth 5). The recent Hyrax system [54] introduces graceful degradation to cloud compute servers, which enables these servers to continue running workloads even after components fail. Figure 3 shows that this approach can extend the useful life after the supply of replacement parts and maintenance become more expensive. Specifically, cloud providers often apply a threshold on how many cluster resources need to be available for scheduling workloads. The cluster is deemed not useful below this threshold and decommissioned. For a threshold of 75% and a period with good supplies of six years, Hyrax can extened useful life by 28%.

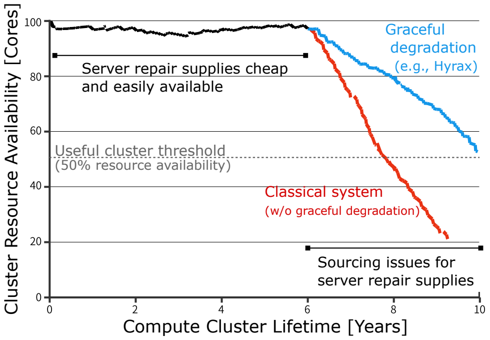

Figure 3: Resource availability over a cluster's deployed lifetime is affected by the common scenario of replacement supplies (parts and tools) becoming less available at some point, e.g., in year six. Graceful degradation approaches (like Hyrax [54]) can extend the useful lifetime of a compute cluster, e.g., from below eight to above ten years.

Another line of system work addresses the significant carbon cost of DRAM. In compute servers, DRAM dominates embodied carbon costs. Replacing DRAM with denser or lower-carbon-cost types of memory is thus a promising research direction. For example, reusing DDR4 from decommissioned servers [21] in new DDR5 servers can make a significant impact. However, this approach typically induces a different performance profile, e.g., because older types of memory need to be attached via an external memory controller, which adds access latency. Recent work has introduced multiple techniques to hide this latency [51, 55, 61], which enables DDR-reuse in practice.

Myth 8. ***Cloud servers look like __, and a uniform methodology can be applied to reduce carbon embedded in all types of servers in the cloud.*** *For example, a few authors of this paper have fallen in this trap as they tried to work across different engineering organizations, and academic papers [40, 66] might cite a specific server type (eg. [27]) to represent cloud server embodied carbon breakdowns.*

The server fleet at Azure and other cloud providers comprises multiple server types. At a high level, there are compute, storage, and networking servers, each of which has a non-negligible contribution to the embodied carbon of the fleet. Within each category, servers differ in the magnitude and relative ratios of CPU cores, memory, storage, and other resources. Some papers offer carbon breakdown examples of cloud servers where SSDs are the dominant component [40, 66], whereas such a high SSD:CPU core ratio would be atypical at Azure . Table 3 gives a breakdown for a more common class of compute servers at Azure , using publicly available server configurations [50] and carbon data [40]. We only list CPU, DRAM, and SSDs as there is no public data on NICs, common offload engines [14, 30, 52], and management controllers.

Server heterogeneity has several implications for embodied carbon research, one of which has to do with lifetime extension. Extending the lifetime of compute servers is feasible due to their failure rate not increasing over time (Figure 4). In fact, compute server annual failure rates (AFRs) tend to slightly decrease over time,

Component	Size	Quantity	Total kg CO^2e
CPU	≈600mm^2	2	14.7
DRAM	32 GB	24	49.9
SSD	960 GB	7	42.3

Table 3: Olympus class server [50] with a 600 mm^2 CPU, 768 GB of memory, and 7 TB of storage; using public carbon data [40].

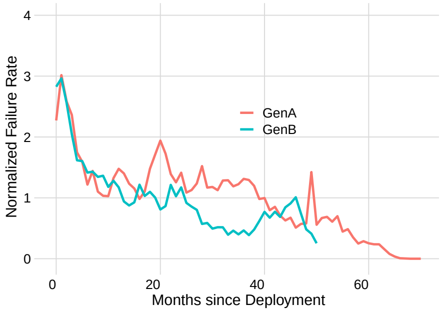

Figure 4: The AFR (Annualized Failure Rate) due to any component in a compute server at Azure does not show signs of an upward trend with age (data from [54]).

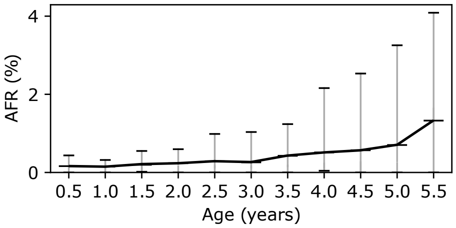

Figure 5: The AFR for hard disks shows significant upward trends after four years. Data from Pacemaker [46].

which means that server lifetime extensions have no or even positive impacts on customer experience. In contrast, storage servers see significant increases in failure rate over time. Specifically, hard drives are the dominant storage component and their failure rate increases ten to twenty-fold over just a six-year deployment time (Figure 5). Thus, extending the lifetime of storage servers is significantly more complex and requires proactively mitigating potential customer impacts. A similar implication applies when considering which components to target for reuse and how they can be reused. For example, DRAM does not typically retain customer data, whereas hard drives require deep and extensive erasures. In general, we need to develop a breadth of techniques and remain receptive to adopting concepts from various systems domains and being mindful of the variations among applications.

4 DISCUSSION

The common theme among our myths and experiences is that they span a wide range of topics. Reasoning about the subjects underlying these myths requires knowledge in carbon accounting (Myth 1), financial accounting (Myth 5), energy systems (Myth 2), recycling (Myth 3), scheduling and telemetry (Myth 4), and insights into cloud server configurations (Myth 8). Accordingly, research needs to happen within an interdisciplinary collaboration spanning multiple university departments as well as industrial partners.

Over time, systems research may arrive at a set of verified facts around carbon. We hope that our myths and experiences get discussed, cross-checked, and eventually help build part of such a community-wide fact base.

In discussing these myths, we qualified two cost barriers that will make it difficult for cloud platforms to meet decarbonization goals: asymptotic renewable costs for Scope 2 emissions, and security/support/expense-driven lifetime limits for Scope 3 emissions. For Scope 2, one way to bypass this barrier is to build systems that allow datacenter power consumption to be flexible and mold to better match supply, and some research in the systems community has already begun in this direction. For Scope 3, research directions are less clear. Component re-use offers diminishing returns compared to server lifetime extension, and recycling has even lower returns (Myth 3). Designing servers and software systems to support longer lifetimes and better re-use opportunities could be the key, and we believe this is an open problem.

ACKNOWLEDGMENTS

We thank our anonymous reviewers for their great comments and our many partner teams within Microsoft for their collaboration. We also thank our many academic collaborators including Bianca Schroeder and Akshitha Sriraman.

REFERENCES

[1] Alibaba group carbon neutrality action report. https://sustainability.alibabagroup.com/en.
[2] Climate change is humanity's next big moonshot. https://blog.google/outreach-initiatives/sustainability/dear-earth/.
[3] Environmental sustainability is imperative. https://gf.com/about-us/corporate-responsibility/sustainability/.
[4] IBM commits to net zero greenhouse gas emissions by 2030. https://newsroom.ibm.com/2021-02-16-IBM-Commits-To-Net-Zero-Greenhouse-Gas-Emissions-By-2030.
[5] Intel and the environment. https://www.intel.com/content/www/us/en/environment/intel-and-the-environment.html.
[6] Net 0 tracker. https://net0tracker.com/corporates.html/.
[7] Our carbon footprint. https://sustainability.aboutamazon.com/environment/carbon-footprint.
[8] Samsung electronics announces new environmental strategy. https://news.samsung.com/global/samsung-electronics-announces-new-environmental-strategy.
[9] Sustainability at oracle. https://www.oracle.com/sustainability/.
[10] Sustainability report | micron technology. https://www.micron.com/about/our-commitment/operating-thoughtfully/sustainability.
[11] Tencent announces plan to become carbon neutral by 2030. https://www.tencent.com/en-us/articles/2201287.html.
[12] TSMC commits to reach net zero emissions by 2050, acting on responsibility to environmental sustainability. http://pr.tsmc.com/english/news/2865.
[13] Bilge Acun, Benjamin Lee, Fiodar Kazhamiaka, Kiwan Maeng, Udit Gupta, Manoj Chakkaravarthy, David Brooks, and Carole-Jean Wu. Carbon explorer: A holistic framework for designing carbon aware datacenters. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 118–132, 2023.
[14] Alexandru Agache, Marc Brooker, Alexandra Iordache, Anthony Liguori, Rolf Neugebauer, Phil Piwonka, and Diana-Maria Popa. Firecracker: Lightweight virtualization for serverless applications. In *USENIX NSDI*, pages 419–434, 2020.
[15] International Energy Agency. Renewables 2021. 2021.
[16] Thomas Anderson, Adam Belay, Mosharaf Chowdhury, Asaf Cidon, and Irene Zhang. Treehouse: A case for carbon-aware datacenter software. *arXiv preprint*

arXiv:2201.02120, 2022.
[17] AWS. Energy transition. https://aws.amazon.com/energy/sustainability/ accessed 1/15/2023, 2021.
[18] Hugo Barbalho, Patricia Kovaleski, Beibin Li, Luke Marshall, Marco Molinaro, Abhisek Pan, Eli Cortez, Matheus Leao, Harsh Patwari, Zuzu Tang, et al. Virtual machine allocation with lifetime predictions. *Proceedings of Machine Learning and Systems*, 5, 2023.
[19] Hugo Barbalho, Patricia Kovaleski, Beibin Li, Luke Marshall, Marco Molinaro, Abhisek Pan, Eli Cortez, Matheus Leao, Harsh Patwari, Zuzu Tang, Tamires Santos, Larissa Rozales Gonçalves, David Dion, Thomas Moscibroda, and Ishai Menache. Virtual machine allocation with lifetime predictions. In *MlSys*, March 2023.
[20] Luiz André Barroso, Jimmy Clidaras, and Urs Hölzle. The datacenter as a computer: An introduction to the design of warehouse-scale machines. *Synthesis lectures on computer architecture*, 8(3):1–154, 2013.
[21] Daniel S Berger, Daniel Ernst, Huaicheng Li, Pantea Zardoshti, Monish Shah, Samir Rajadnya, Scott Lee, Lisa Hsu, Ishwar Agarwal, Mark D Hill, et al. Design tradeoffs in cxl-based memory pools for public cloud platforms. *IEEE Micro*, 43(2):30–38, 2023.
[22] Pankaj Bhatia, Cynthia Cummis, Andrea Brown, David Rich, Laura Draucker, and Holly Lahd. Corporate value chain (scope 3) accounting and reporting standard, 2011.
[23] Niv Buchbinder, Yaron Fairstein, Konstantina Mellou, Ishai Menache, and Joseph Naor. Online virtual machine allocation with lifetime and load predictions. *ACM SIGMETRICS Performance Evaluation Review*, 49(1):9–10, 2021.
[24] Zhiwei Cao, Xin Zhou, Han Hu, Zhi Wang, and Yonggang Wen. Towards a systematic survey for carbon neutral data centers. *IEEE Communications Surveys & Tutorials*, 2022.
[25] Eli Cortez, Anand Bonde, Alexandre Muzio, Mark Russinovich, Marcus Fontoura, and Ricardo Bianchini. Resource central: Understanding and predicting workloads for improved resource management in large cloud platforms. In *ACM SOSP*, pages 153–167, 2017.
[26] Miyuru Dayarathna, Yonggang Wen, and Rui Fan. Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):732–794, 2015.
[27] Dell. Dell lca, poweredge r740. https://corporate.delltechnologies.com/content/dam/digitalassets/active/en/unauth/data-sheets/products/servers/lca_poweredge_r740.pdf, 2019.
[28] Mark Dyson, Sakhi Shah, and Charles Teplin. The challenge of the last few percent: Quantifying the costs and emissions benefits of a 100% renewable u.s. electricity system. https://www.nrel.gov/news/program/2021/the-challenge-of-the-last-few-percent-quantifying-the-costs-and-emissions-benefits-of-100-renewables.html, June 2021.
[29] Mark Dyson, Sakhi Shah, and Charles Teplin. Clean power by the hour: Assessing the costs and emissions impacts of hourly carbon-free energy procurement strategies. https://rmi.org/insight/clean-power-by-the-hour, 2021.
[30] Daniel Firestone, Andrew Putnam, Sambhrama Mundkur, Derek Chiou, Alireza Dabagh, Mike Andrewartha, Hari Angepat, Vivek Bhanu, Adrian Caulfield, Eric Chung, et al. Azure accelerated networking: Smartnics in the public cloud. In *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, pages 51–66, 2018.
[31] Kali Frost, Hongyue Jin, William Olson, Mark Schaffer, Gary Spencer, and Carol Handwerker. The use of decision support tools to accelerate the development of circular economic business models for hard disk drives and rare-earth magnets. *MRS Energy & Sustainability*, 7, 2020.
[32] Gabrielle Gaustad, Mark Krystofik, Michele Bustamante, and Kedar Badami. Circular economy strategies for mitigating critical material supply issues. *Resources, Conservation and Recycling*, 135:24–33, 2018.
[33] Google. Once is never enough. https://sustainability.google/progress/projects/circular-economy/, 2018.
[34] Google. 24/7 by 2030: Realizing a carbon-free future, 2020.
[35] Google. 24/7 clean energy – data centers. https://www.google.com/about/datacenters/cleanenergy/ accessed 1/15/2023, 2021.
[36] Jessica F Green. Private standards in the climate regime: the greenhouse gas protocol. *Business and Politics*, 12(3):1–37, 2010.
[37] Albert Greenberg, James Hamilton, David A Maltz, and Parveen Patel. The cost of a cloud: research problems in data center networks, 2008.
[38] Albert Greenberg and Dave Maltz. What goes into a data center–sigmetrics 2009 tutorial. 2009.
[39] Binglei Guo, Jiong Yu, Dexian Yang, Hongyong Leng, and Bin Liao. Energy-efficient database systems: A systematic survey. *ACM Computing Surveys*, 55(6):1–53, 2022.
[40] Udit Gupta, Mariam Elgamal, Gage Hills, Gu-Yeon Wei, Hsien-Hsin S Lee, David Brooks, and Carole-Jean Wu. Act: Designing sustainable computer systems with an architectural carbon modeling tool. In *Annual International Symposium on Computer Architecture (ISCA)*, pages 784–799, 2022.
[41] Udit Gupta, Young Geun Kim, Sylvia Lee, Jordan Tse, Hsien-Hsin S Lee, Gu-Yeon Wei, David Brooks, and Carole-Jean Wu. Chasing carbon: The elusive environmental footprint of computing. In *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 854–867, 2021.
[42] Ori Hadary, Luke Marshall, Ishai Menache, Abhisek Pan, Esaias E Greeff, David Dion, et al. Protean:vm allocation service at scale. In *USENIX OSDI*, pages 845–861, 2020.
[43] Amy Hood. Microsoft earnings release fy22 q4. https://www.microsoft.com/en-us/Investor/earnings/FY-2022-Q4/press-release-webcast accessed 11/26/2022, 2022.
[44] Hongyue Jin, Kali Frost, Ines Sousa, Hamid Ghaderi, Alex Bevan, Miha Zakotnik, and Carol Handwerker. Life cycle assessment of emerging technologies on value recovery from hard disk drives. *Resources, Conservation and Recycling*, 157:104781, 2020.
[45] Lucas Joppa. Made to measure: Sustainability commitment progress and updates. https://blogs.microsoft.com/blog/2021/07/14/made-to-measure-sustainability-commitment-progress-and-updates/ accessed 1/15/2023, 2021.
[46] Saurabh Kadekodi, Francisco Maturana, Suhas Jayaram Subramanya, Juncheng Yang, KV Rashmi, and Gregory Ganger. Pacemaker: Avoiding heart attacks in storage clusters with disk-adaptive redundancy. In *Proceedings of the 14th USENIX Symposium on Operating Systems Design and Implementation*, 2020.
[47] Tarandeep Kaur and Inderveer Chana. Energy efficiency techniques in cloud computing: A survey and taxonomy. *ACM computing surveys (CSUR)*, 48(2):1–46, 2015.
[48] Donald Kline, Nikolas Parshook, Xiaoyu Ge, Erik Brunvand, Rami Melhem, Panos K. Chrysanthis, and Alex K. Jones. Greenchip: A tool for evaluating holistic sustainability of modern computing systems. *Sustainable Computing: Informatics and Systems*, 22:322–332, 2019.
[49] Fanxin Kong and Xue Liu. A survey on green-energy-aware power management for datacenters. *ACM Computing Surveys (CSUR)*, 47(2):1–38, 2014.
[50] Anand Kulkarni. Microsoft project olympus server specification. 2017.
[51] Huaicheng Li, Daniel S Berger, Lisa Hsu, Daniel Ernst, Pantea Zardoshti, et al. Pond: Cxl-based memory pooling systems for cloud platforms. In *ACM ASPLOS*, 2023.
[52] Huaicheng Li, Mingzhe Hao, Stanko Novakovic, Vaibhav Gogte, Sriram Govindan, Dan RK Ports, Irene Zhang, Ricardo Bianchini, Haryadi S Gunawi, and Anirudh Badam. Leapio: Efficient and portable virtual nvme storage on arm socs. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 591–605, 2020.
[53] Zhen Li, Luis A Diaz, Zhiyao Yang, Hongyue Jin, Tedd E Lister, Ehsan Vahidi, and Fu Zhao. Comparative life cycle analysis for value recovery of precious metals and rare earth elements from electronic waste. *Resources, conservation and recycling*, 149:20–30, 2019.
[54] Jialun Lyu, Marisa You, Celine Irvene, Mark Jung, Tyler Narmore, Jacob Shapiro, Luke Marshall, Savyasachi Samal, Ioannis Manousakis, Lisa Hsu, Preetha Subbarayalu, Ashish Raniwala, Brijesh Warrier, Ricardo Bianchini, Bianca Schroeder, and Daniel S. Berger. Hyrax: Fail-in-place server operation in cloud platforms. In *17th USENIX Symposium on Operating Systems Design and Implementation (OSDI 23)*, Boston, MA, July 2023. USENIX Association.
[55] Hasan Al Maruf, Hao Wang, Abhishek Dhanotia, Johannes Weiner, Niket Agarwal, Pallab Bhattacharya, Chris Petersen, Mosharaf Chowdhury, Shobhit Kanaujia, and Prakash Chauhan. Tpp: Transparent page placement for cxl-enabled tiered-memory. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, pages 742–755, 2023.
[56] Toni Mastelic, Ariel Oleksiak, Holger Claussen, Ivona Brandic, Jean-Marc Pierson, and Athnasios V Vasilakos. Cloud computing: survey on energy efficiency. *ACM Computing Surveys*, 47(2):pp–1, 2015.
[57] Devonie McCamey. What we know—and do not know—about achieving a national-scale 100% renewable electric grid. https://www.nrel.gov/news/features/2021/what-we-know-and-dont-know-about-achieving-a-national-scale-100-renewable-electric-grid.html, May 2021.
[58] Sparsh Mittal and Jeffrey S Vetter. A survey of methods for analyzing and improving gpu energy efficiency. *ACM Computing Surveys (CSUR)*, 47(2):1–23, 2014.
[59] Melanie Nakagawa and Brad Smith. 2022 environmental sustainability report. https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RW14sJN, 2022.
[60] Anne-Cecile Orgerie, Marcos Dias de Assuncao, and Laurent Lefevre. A survey on techniques for improving the energy efficiency of large-scale distributed systems. *ACM Computing Surveys (CSUR)*, 46(4):1–31, 2014.
[61] Amanda Raybuck, Tim Stamler, Wei Zhang, Mattan Erez, and Simon Peter. Hemem: Scalable tiered memory management for big data applications and real nvm. In *Proceedings of the ACM SIGOPS 28th Symposium on Operating Systems Principles*, pages 392–407, 2021.
[62] Brad Smith. Microsoft will be carbon negative by 2030. https://blogs.microsoft.com/blog/2020/01/16/microsoft-will-be-carbon-negative-by-2030/ accessed 1/15/2023, 2020.
[63] Abel Souza, Noman Bashir, Jorge Murillo, Walid Hanafy, Qianlin Liang, David Irwin, and Prashant Shenoy. Ecovisor: A virtual energy system for carbon-efficient applications. In *Proceedings of the 28th ACM International Conference on*

Architectural Support for Programming Languages and Operating Systems, Volume 2, pages 252–265, 2023.

[64] Soumya Sudhakar, Vivienne Sze, and Sertac Karaman. Data centers on wheels: Emissions from computing onboard autonomous vehicles. *IEEE Micro*, 43(1):29–39, 2023.

[65] Jennifer Switzer, Gabriel Marcano, Ryan Kastner, and Pat Pannuto. Junkyard computing: Repurposing discarded smartphones to minimize carbon. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 400–412, 2023.

[66] Swamit Tannu and Prashant J Nair. The dirty secret of ssds: Embodied carbon. In *HotCarbon*, July 2022.

[67] Jamie Thomson. On the road to 2030: Our 2022 environmental sustainability report. https://blogs.microsoft.com/on-the-issues/2023/05/10/2022-environmental-sustainability-report/.

[68] Muhammad Tirmazi, Adam Barker, Nan Deng, Md E Haque, Zhijing Gene Qin, Steven Hand, Mor Harchol-Balter, and John Wilkes. Borg: the next generation. In *Proceedings of the fifteenth European conference on computer systems*, pages 1–14, 2020.

[69] Amanda Tomlinson and George Porter. Something old, something new: Extending the life of cpus in datacenters. In *HotCarbon*, July 2022.

[70] user. Caring for the earth: SK hynix's green 2030. https://news.skhynix.com/caring-for-the-earth-sk-hynixs-green-2030/.

[71] Kushagra Vaid. Datacenter power efficiency: Separating fact from fiction. In *Invited talk at the 2010 Workshop on Power Aware Computing and Systems*, volume 1, 2010.

[72] Abhishek Verma, Luis Pedrosa, Madhukar Korupolu, David Oppenheimer, Eric Tune, and John Wilkes. Large-scale cluster management at google with borg. In *ACM EuroSys*, pages 1–17, 2015.

[73] Kashi Venkatesh Vishwanath and Nachiappan Nagappan. Characterizing cloud computing hardware reliability. In *Proceedings of the 1st ACM symposium on Cloud computing*, pages 193–204, 2010.

[74] Yawen Wang, Daniel Crankshaw, Neeraja J Yadwadkar, Daniel Berger, Christos Kozyrakis, and Ricardo Bianchini. Sol: safe on-node learning in cloud platforms. In *ACM ASPLOS*, pages 622–634, 2022.