The Sky is not the limit: untapped opportunities for Green Computing

COLLEEN JOSEPHSON, UC Santa Cruz, USA
NICOLA PEILL-MOELTER, VMware, USA
ZHELONG PAN, VMware, USA
BEN PFAFF, Feldera, USA
VICTOR FIROIU, VMware, USA

The Information and Communications Technology (ICT) industry emits as much carbon as the aviation industry [11], and if we continue business as usual our share of emissions will grow manyfold in the coming decade [30]. At the same time, more and more businesses are making commitments to be zero carbon or carbon neutral by 2050 or sooner. The path to zero carbon ICT has four key pillars: prioritizing renewable energy, using resources like power and water more efficiently, addressing embodied carbon, and removing institutional barriers. We discuss from a broad industry perspective the challenges and opportunities within each pillar, as well as the role ICT can play in helping other industries achieve zero carbon goals.

CCS Concepts: • **Hardware → Impact on the environment**; **Renewable energy**; • **Social and professional topics** → *Computing / technology policy*.

Additional Key Words and Phrases: carbon, sustainability, energy, green software

1 INTRODUCTION

The 2021 Intergovernmental Panel on Climate Change (IPCC) report [40] elevated sustainability to the top of public discourse. The report, based on more than 14,000 scientific research papers, spells out exactly what will happen if humanity does not rapidly get greenhouse gas emissions under control—if we don't make rapid progress towards carbon neutrality, extreme weather, destruction of animal habitats, and glacial melting resulting in the rise of sea levels will all dramatically accelerate.

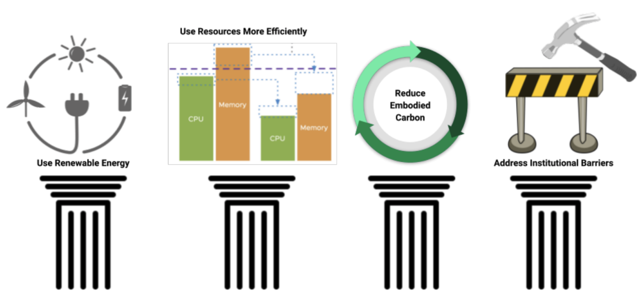

Fig. 1. The four key pillars of zero carbon ICT are prioritizing renewable energy, using resources more efficiently, addressing embodied carbon, and removing institutional barriers.

The Information and Communications Technology (ICT) industry has changed the way we work and live. Our modern economy relies on connected computation, and the growth of datacenters and communications networks shows no sign of stopping. By some models, the industry currently accounts for 4-10% of global electricity use [46], with that share projected to grow to up to 20% by 2030 [30]. Currently, electricity generation is the leading cause of greenhouse gas emissions [49]. Though the adoption of renewables is growing, 67% of the world's electricity is generated from carbon-intensive sources like coal and natural gas [38]. Due to the high proportion of non-renewable electricity generation, the ICT industry currently accounts for at least 2% of global greenhouse gas emissions, on par with the aviation industry [11].

The urgency of the climate situation, paired with the strong expected growth of the ICT industry, means that practitioners must put serious effort into keeping emissions under control. Fortunately, growing numbers of businesses are making commitments to be zero carbon or carbon neutral by 2050 or sooner. As of 2021, 38% of Global Fortune 500 Companies have set a significant 2030 emissions target [50]. But how do we get there? Focusing only on energy consumption in the sustainable ICT misses other important parts of the picture. The path to zero carbon ICT has four pillars: prioritizing renewable energy, using energy and other resources such as water more efficiently, transitioning to a circular economy to reduce embodied carbon, and removing institutional barriers.

Historically, energy consumption has been an outsized focus because reducing consumption also reduced operational costs. Now, however, arguably we must shift some effort to making sure the energy we do consume is from renewable and sustainable sources. Due to the inherently fickle nature of renewables (the sun doesn't always shine, and the wind doesn't always blow), transitioning to 100% renewable energy will take considerable changes to business-as-usual. We discuss these challenges in Section 2.

The second pillar, using resources more efficiently, requires broadening our scope. Energy efficiency is top of mind and is certainly important, but sustainability also encompasses efficient use of other resources, like water and land. Data center cooling consumes considerable water [48], and the physical infrastructure can also have non-negligible emissions when constructed with carbon-intensive materials like concrete [5]. We discuss these challenges further in Section 3.

The third pillar is addressing embodied carbon. Our current economic model is linear and based on a take-make-waste model, which creates considerable waste and pollution in favor of speed and affordability. Sustainability requires us to design to reduce resources utilized and waste and emissions during production and distribution, which comprise a major portion of the carbon footprint in our industry. To address it, the ICT industry must transition to a circular economy based on three principles: (1) designing low- or no-waste products, (2) using products and components for as long as possible, and (3) preserving or enhancing renewable resources by recovering

Authors' addresses: Colleen Josephson, cjosephson@ucsc.edu, UC Santa Cruz, USA; Nicola Peill-Moelter, VMware, USA; Zhelong Pan, VMware, USA; Ben Pfaff, Feldera, USA; Victor Firoiu, VMware, USA.

and reusing or recycling end-of-life equipment. Significant innovation in re-use and recycling will be necessary for this transition. We discuss this further in Section 4.

Finally, the fourth pillar is removing institutional barriers. The first three pillars have several technical challenges, but institutional barriers like culture, policy, and infrastructure, can often be more difficult to address. These issues cut across the other three pillars, as well as posing their own issues that we cover in Section 5.

2 RENEWABLE ENERGY

Most emissions in ICT are indirect, caused by the power used to run the business or the power used to source materials and build the hardware the business uses. Emissions can be split into three categories: Scope 1 emissions which are direct emissions from burning of fossil fuels (e.g. fuel for company vehicles), Scope 2 emissions which are indirect emissions from purchased electricity and gas, and Scope 3 which are all other indirect emissions in a company's value chain; including upstream emissions in the supply chain (e.g. emissions from the manufacture and transport of purchased goods and services, business travel and employee commuting).

As of 2018, 67% of the world's electricity is generated from carbon-intensive sources like coal and natural gas [38]. If we can transition ICT companies to operate entirely on renewable energy, this will significantly reduce overall emissions. The transition to 100% renewable power is simple in concept but challenging to execute in practice—it will require a lot of innovation and policy outside the direct control of ICT companies. As we wait for this transition to take place, however, we can take substantial and impactful steps to minimize the amount of Scope 2 and 3 emissions. The remainder of Sections 2 and 3 address how to minimize Scope 2 emissions, while Scope 3 emissions are discussed further in Section 4.

2.1 Carbon Awareness

"You cannot manage what you cannot measure." Carbon footprint reporting tools [2–4, 8] have been picking up speed in the ICT industry. Such tools provide carbon emission by time, regions, products/services, and projects. Recently, in March 2022, the SEC passed a proposal for public companies to disclose greenhouse gas emission metrics [19]. Such reports and tools can bring carbon awareness to executives, infrastructure teams, and perhaps application owners. They can drive positive change to address climate impacts of cloud workloads, thus marching the industry towards Net Zero.

For calculating a software system carbon footprint, the Software Carbon Intensity (SCI) specification [20] is a good starting point. It considers carbon emissions due to energy consumption and the embodied carbon during the creation and disposal of a hardware device:

$$SCI = ((EI) + M) \text{ per } R$$

Where E is the consumed by a software system, I is location-based marginal carbon emissions, M is embodied emissions of a software system and R is the functional unit (e.g. carbon per additional user, API-call, ML job, etc.).

Though the above SCI formula is simple and straightforward, an accurate measurement or a good estimate of each component is needed to compute the number. Consider an example VM to exemplify the work needed and challenges.

Firstly, power consumption of the VM's host is measured and the VM's energy consumption is modeled. Host hardware power sensors can be read to measure the host's power consumption. Collaboration is needed with hardware vendors to further allocate host power consumption to individual components, like CPU, memory, fan, devices, etc. Hypervisors can then attribute host power consumption to individual VMs, based on VM configuration and resource utilization.

Secondly, third-party services, such as WattTime [24] and ElectricityMap [9], provide regional and temporal marginal carbon emissions with 5-minute to hourly granularity that can be used to convert the energy to carbon for carbon-aware workload management based on time and location. Good outcomes, however, will depend on the accuracy, timeliness, and completeness of the emission factor data sets, which don't exist today.

Lastly, to incorporate embodied emissions into carbon management, a database of embodied emissions for different server models is needed. Today, only a few vendors provide this information, for example, using Dell [17] and HPE [15] product carbon footprint data. Like power consumption, embodied emission can be attributed to VMs based on their configuration and resource utilization.

Bringing carbon awareness to a multi-cloud environment and to modern applications is a new concept. Several questions are yet to be answered. What can be done beyond hosts and VMs? How can carbon emissions due to public cloud consumption best be report and analyzed? Can the accuracy of embodied carbon and scope 3 emissions in general be improved? How can the carbon emission data be improved? Can developers and operators be engaged to leverage carbon emissions to measure and improve their applications?

Carbon awareness is an initial important step to measure or estimate the carbon impact, during the era of digital transformation. This work will help the operators to monitor, operate and improve the sustainability of the IT infrastructure and modern applications. It will enable innovations in carbon avoidance or carbon reduction techniques, which will be discussed in the following sections. It will also enable matching of the measured demand in renewable energy for computation with the renewable energy generation. Such demand-generation matching presents significant challenges and opportunities, which is discussed next.

2.2 Distributed, decentralized workload shifting

Globally, renewable energy (solar, wind) generation has grown rapidly over the last decade [43]. Renewable energy has very low carbon intensity and a zero marginal operating cost, and so is a critical component to achieving zero carbon computing and should not be wasted. However, pervasive imbalances in renewable generation and the transmission capability limitations that prevent the grid from transporting energy to where it is needed causes generation to be temporarily curtailed and wasted. (Non-renewable generation is limited less by transmission limits because such plants can be built in more geographies.) A. Chien and team have documented the extent of renewable energy curtailment and that it is increasing rapidly in the U.S. and globally [25, 27, 28, 45]. Potential solutions include co-located battery storage and construction of more transmission capacity. But these approaches do not yet scale due to cost, and traditional loads cannot be physically moved.

Opportunity: On the other hand, computation workloads can be flexible in time (scheduled for delayed execution) and space (transferred across any geographical distance with limited cost), although some require real-time execution or must execute in a particular place e.g. due to network latency or data sovereignty issues. This opens the possibility of shifting workloads in time and space to take advantage in real time of any amount of excess renewable energy. Initial results show that a single datacenter that time shifts load can reduce its emissions by 19% annually [33]. Such workload shifting presents significant challenges, which we examine next.

Emerging technology: Time and space shifting of computation workloads to match renewable energy production has high potential for reducing, even eliminating, fossil fuel use in both computation and traditional uses, the latter by enabling more development of renewable energy generation without the downside of wasted energy. The first application of workload shifting has been experimentally deployed in some public cloud providers, taking advantage of their diversity of data center locations and the high capacity of the provider-owned, inter-data center connectivity, with promising results [29, 41].

Opportunity: Even after intra-provider workload shifting becomes widespread, there is high potential beyond public cloud providers, given that a large proportion of computation workloads are performed outside of public clouds, such as in private or on-prem data centers [21].

Proposed Approaches: One novel approach would be for an organization to shift its workloads across cloud providers. The emerging field of multi-cloud, such as the Sky Computing initiative [39], which is set to define standardized/ unique interfaces to all cloud providers, can become a practical framework for customer-driven workload shifting.

A non-technical challenge is the current practice of cloud providers to charge high rates for data exported from their cloud. Some providers have started to band together in waiving egress fees for their bandwidth among mutual customers [1]. This can further be addressed by market and social movements that favor the expanded use of renewable energy.

Another approach, which is the most speculative and requiring further research, is based on *cooperative computing*. Specifically, organizations in diverse time zones or geographic or meteorological conditions to pool some of their on-prem or private compute infrastructures and then migrate their workloads across these data centers to match the availability of renewable energy.

Challenges and proposed approaches: A first challenge is the security implications of sharing computation infrastructure between unrelated companies or organizations. The emerging techniques of *confidential computing* [47, 55] and *data sovereignty* [52] can help mitigate these risks.

Confidential computing may raise questions about the added computational overhead, which would result in counterproductive increase in energy usage. Nevertheless, recent advances in this area, including hardware assist in Trusted Execution Environments (TEE) have resulted in the minimization of such overhead. For example [53] shows a joint development and integration of AMD Secure Encrypted Virtualization (SEV) with VMware vSphere and vSAN, which exhibited "a low performance overhead of 1.4% on OLTP workload and 6.2% on DSS workload with SQL Server 2019." A more detailed performance analysis of memory protection (which is at the core of confidential computing) is in [36], where both AMD SEV and Intel Software Guard Extensions (SGX) show energy cost per pub/sub operation very close to the baseline without memory protection.

A second challenge is the added cost (monetary, energy, and carbon) of moving workloads due to the nonlinear power proportionality of servers (even an idle server still uses about half its maximum power, although this is improving) and the energy required to move and run workloads/data to and in a different location.

A third challenge is data gravity for stateful workloads. When working with large datasets, moving the bulk of data around is cumbersome and expensive. In such cases, we may leverage existing data replication mechanisms, in data protection solutions or distributed databases.

A fourth challenge comes from the logistics of sharing computation capacity, including distributed resource accounting and scheduling, multi-tenant priority resolution, service agreements and cross-charging, all of which raise questions of institutional barriers. Practical approaches can include existing work on cloud brokers [7], organized either by consortia of interested companies or by independent third parties.

A fifth challenge is that shifting large amounts of workloads (both in time and space) can create imbalances and instability in the power grid (generation and transmission). Large organizations participating in workload shifting could help by joining the local ISO power market to provide their own demand forecasts and to obtain real-time information and forecasts for renewable energy supply, but this approach is not available to small to medium companies.

A sixth challenge is the availability of quality, high-resolution carbon intensity data. ISOs are reluctant to provide data for business reasons, and the format and type of information varies when it is available. Recent work [18] has made inroads toward this goal, but much more work is needed to achieve uniform availability across the US. We speculate that pressure from major power consumers may help in incentivizing higher data quality and creating standardized APIs.

3 PRODUCTIVE & EFFICIENT RESOURCE UTILIZATION

Key to minimizing the energy consumption and carbon emissions for computing is maximizing productive use of IT infrastructure. Consider aviation: the more densely occupied planes are, the fewer flights are needed to meet a given level of travel demand. Before server virtualization, the average server utilization was only about 5-15% because each application ran on its own server and those applications were not busy much of the time. With server consolidation due to virtualization, server utilization has risen, but still remains below 30% [44], for several reasons: (1) oversized virtual machines (VMs) and "zombie" VMs that unproductively hoard and strand compute resources; (2) physically siloed infrastructure, time-of-day use of enterprise applications, and failover and peaking capacity all contribute to significant, unused, spare capacity; and (3) inefficient applications that use more compute, storage, and networking resources than necessary. On a related note, nonlinear power proportionality of servers also contributes to unproductive energy use as servers consume 40+% of peak power at idle. Therefore, maximizing server utilization makes the most of the server energy consumption. Imagine if server utilization could be doubled, effectively halving the amount of IT and data center infrastructure required to support the world's growing appetite for computing! Imagine the energy, carbon and cost savings!

3.1 Reducing Waste

Simplification and automation are key to reducing waste. For example, despite tools that can identify oversized VMs, technical and process issues often impede right-sizing at scale–Application and resource and business requirements must be checked before VMs can be right-sized to minimize risks.

Zombies are VMs or servers that are no longer doing productive work but still consuming compute, storage and data center resources. Zombies arise when people leave organizations and projects are ended and infrastructure is left behind and forgotten. One study found that almost 60% of physical servers and 80% of VMs were zombie across 16,000 virtualized and non-virtualized physical servers in 10 data centers [16]. Zombies that have little to no activity are easy to identify, but many zombies have residual activity such as backups, virus scanning, patch updates, that obscure whether they are in productive use. The answer may lie in life cycle management of VM and server assets. Assessing the typical activity of an asset over its lifetime can provide clues about whether there has been a dramatic change in usage and flag that asset for further scrutiny. Tracking accountability for an asset can also simplify management of that asset, answering questions about why an asset was deployed and who is responsible for it.

3.2 Shared and On-Demand Capacity

With the emergence of hybrid cloud, spare capacity for peak demand and failover can be provided by on-demand capacity in the public cloud. Solutions are already available to deploy backup capacity in the public cloud that significantly reduces failover capacity requirements. The next step is enabling the public cloud to dynamically and seamlessly absorb load once the on-premises capacity is reached while meeting performance, security, and geographic requirements. Furthermore, public cloud operations can have significantly lower energy and carbon footprints due to more energy efficient hardware, operations and adoption of renewable energy compared with on-premises operations. Where this is true, acceleration of workload migration to public cloud can be encouraged.

Historically, IT organizations supported business unit activity, applications, and projects with physically siloed IT infrastructure. Technically, this is no longer necessary with the evolution of virtual networking that can create virtual separations that can meet business requirements (resource, security, performance, legal). This can enable IT resources to be more readily shared among organizations, reducing unused capacity and minimizing IT infrastructure required to support the business. Yet siloed infrastructure is still common in on-premises environments, contributing to underutilization. Often this is due to cultural and institutional barriers that resist new ways of doing things. Section 5 adds details.

3.3 Resource-Efficient Applications

Applications that are architected, designed, and coded to minimize compute, storage, and networking resources can enormously reduce the amount of IT and data center infrastructure required. Specialization is a promising approach to making applications more resource efficient. Specialized hardware using domain-specific architectures [37] for graphics, machine learning, networking, and other major applications uses less energy than equivalent computations using general-purpose CPUs. Specialized software, such as unikernels designed to run a single application [31], tend to use fewer resources than using a general-purpose operating system [32].

3.4 Equipment Power Management

Even if average utilization is improved significantly, the non-linearity of server power (power proportionality) results in inefficient energy use during periods of low utilization and idleness. Server components that contribute most to power consumption include CPU (GPU), fans, memory and hard drives if they are not solid state. Power supplies are also a source of inefficiency. The conversion efficiency of a server power supply can decrease significantly when the server power draw drops below 40% relative to power supply's rated power, for example dropping from greater than 95% to below 80% efficiency.

Power management of servers, that is, putting servers into a lower power state, can improve energy efficiency during periods of idleness and low utilization, saving up to 50% of maximum power draw. CPUs can be put into lower P and C states when not busy. Care should be taken as there can be performance implications due to latency in returning to higher states. Fans can be programmed to spin at a rate that is proportional to the CPU temperature. Spinning hard drives can be replaced by solid state drives. Power supplies should be right-sized to be closer to server rated max power. To minimize risks, servers must be able to reliably come back online within seconds, and essential functions such as patch updates and battery charging must not be interrupted.

Unfortunately, technical innovations alone are not enough to achieve higher infrastructure utilization. There is also a need to overcome institutional barriers (Section 5) that often impede the implementation of technical solutions.

3.5 A Note About Jevon's Paradox

Jevons Paradox postulates that an increase in resource efficiency will, in the long term, result in an increase, rather than a decrease, in consumption of that resource. The International Energy Agency published a report in 2021 that shows a significant long-term reduction in energy intensity and stabilization of energy consumption with exponential increases in Internet traffic (16.9x increase) and data center workloads (9.4x increase) over the past decade [35].

Furthermore, a 2020 IDC report [54] estimates that since the introduction of virtualization technologies, the world has avoided the deployment of 130 million servers saving 2.1 GWh of energy and associated 1.2 billion metric tons of carbon dioxide emissions.

4 EMBODIED CARBON

Embodied carbon is carbon emissions due to the manufacturing of an item. Until recently, the ICT industry had not widely considered the embodied carbon of ICT infrastructure or the carbon emissions associated with its end of life, due to a lack of data, accepted methodology to accurately estimate embodied carbon, and a focus on use-phase emissions (typically scope 2 emissions) rather than supply chain-related emissions (scope 3). More recently, major IT hardware manufacturers have begun to calculate the embodied carbon of their core products. Dell provides embodied carbon for most of its product lines [17]. Dell estimates that, across its server product line, use-phase emissions are typically >80% of total carbon emissions, with embodied carbon <20%. This is consistent with HPE's estimate for its Proliant DL360 Gen10 server with use-phase emissions at 88% of total [15]. On the other hand, end user devices tend to be the opposite, with most of the carbon emissions as embodied carbon [34], because end user devices in general (1) require much less power than data center IT equipment; (2) often rely on battery power and so must be designed to be energy efficient; and (3) the embodied carbon of the components, materials and manufacturing processes have grown over time while the carbon-intensity of electricity has shrunk. Apple has estimated the life cycle emissions of its products for many years and has shown that with intentionality, carbon impact can be significantly reduced [10].

It is currently difficult to estimate embodied carbon. Broad awareness of the concept is still developing, so data is sparse. As such, many companies don't account for it because it's hard to do. As mentioned before measuring the problem is a prerequisite to any solution, so efforts like [20] are important. Once we measure the problem, however, how do we solve it?

Two main strategies are: (1) minimizing the amount of physical IT and data center infrastructure required to support workloads, as discussed in Section 3, and (2) adopting circular economy practices that extend the useful life of electronic equipment and components. Success for the latter depends on re-thinking the design and end-of-life phases to:

(1) Maximize the useful life of equipment through components that are hot-swappable, interchangeable and have cross-compatibility (for example, power supplies/cords)
(2) Minimize the consumption of virgin materials using recycled materials and facilitating the recovery and recyclability of components and materials; and
(3) Build on existing secondary marketplaces to facilitate and accelerate the repurposing of equipment, like the Junkyard Datacenter work proposes [42].

5 INSTITUTIONAL BARRIERS

Technology, tools, and information are necessary, but not sufficient. Real business requirements such as performance, security, time-to-market, risk mitigation, and legal requirements can sometimes oppose and take precedence over processes and techniques to improve energy and carbon efficiency. However, legacy institutional and cultural barriers that no longer serve a business purpose also impede progress to achieving zero carbon ICT. These barriers can include existing policies, processes, incentives/disincentives, outdated thinking, and ingrained habits. For example, consumer e-waste recycling rates are only 17% in the US [12], how can we encourage this to change? The goal of sustainable ICT requires organizational change that, like any other organizational change, starts with leadership. This includes aligning the organization by creating a vision, setting goals, aligning with business goals, developing performance metrics to track progress, implementing new policies and processes, providing training and incentivizing desired behaviors.

Internally, we have strong messaging from leadership that makes clear our commitment to sustainable innovation in engineering. However, one challenge we have encountered is that though we see general enthusiasm for sustainable practices from engineers, there is a lack of knowledge on how to be more sustainable. Issues like this are starting to be addressed through organizations like the Green Software Foundation [22], the Sustainable Digital Infrastructure Alliance [23], and other cross-industry consortiums developing best practices, tools, metrics, and standards to integrate sustainability into the software development process.

Progress in this area could be accelerated by introducing sustainability concepts in core electrical engineering and computer science curriculums. Relatively few colleges and universities offer sustainable computing courses, and none, to our knowledge, have these concepts as a required part of the curriculum.

Another barrier is resistance to change. Major infrastructure changes, even those known to be beneficial, can face resistance and have trouble reaching activation energy. Experience shows that overcoming this resistance requires that it solves a problem that is causing pain within the business along with a product or service that can be adopted to solve it. Server virtualization is one example: the trouble and expense of managing many physical machines was solved by adopting commercial virtualization software. Networking also followed this pattern, in which network virtualization software offering eased increasing administrative expense and delays of managing VLANs within an organization. We expect green computing to follow this pattern as well.

Another example is end-of-life device security policies that hinder reuse or recycling of machines, since it is perceived as more secure to destroy a device than to allow secondhand use. Barriers can also be external to an organization. Power companies, for example, resist releasing accurate and high-resolution emissions data that could enable carbon optimization because they consider it proprietary. And engineers want to make greener software, but the current tools that they use are designed primarily to optimize other aspects.

Finally, cost is another barrier that often is discussed in the context of carbon reduction and sustainability. There is a perception that dollar cost is the ultimate decision point in businesses. This is beginning to change, however. Corporations are operating with more complex cost functions that consider sustainability as an input. Due to growing social and political pressure, 38% of Fortune 500 companies have made a significant sustainability target as of 2021 [26]. This is up from 25% just 2019. We anticipate that the number of companies with sustainability targets will only increase. This, in turn, will put pressure on their partners and suppliers [51]. We hope that this will create a positive feedback loop.

6 CONCLUSION

Our modern economy depends on the ICT industry. ICT customers rely on the industry to help them meet their carbon targets. We laid out an industry perspective on the four pillars of sustainable ICT. To conclude, we issue a call to action: it is up to each organization to progress in each pillar through some appropriate combination of direct innovation, funding academic research, investing in partnerships, and advocacy. Sustainable ICT is challenging because the problems span the entire stack of hardware and software, and complex non-technical barriers can be difficult to surmount.

We see many opportunities for further research into sustainable ICT. Work is needed in load balancing of work across time and space while obeying constraints on carbon intensity along with cost and latency metrics, and honoring data sovereignty restrictions. For this purpose, identification of good candidate workloads for migration is an important sub-problem (as well as reformulating unsuitable workloads as suitable ones), which also requires modeling the carbon cost of migrating different types of workloads. Significant research is also needed to make cooperative computing across independent organizations practical: confidentiality of data and code at rest (storage) in transit (network) and in process (protected memory). Tools for creating green software, e.g. measuring, reporting, monitoring tool and tools for finding and fixing bottlenecks, are another important focus. In addition, this area is cross-cutting, so that many areas of computing will need to intersectionally consider sustainability.

We encourage each organization to face the key challenges within the scope of their core business. For challenges outside a company's core business area, several options are still available: we recommend leveraging industry groups like the Next G Alliance [13], the Information Technology Industry Council [14], the Clean Energy Buyers Association [6], the Sustainable Digital Infrastructure Alliance [23] and the Green Software Foundation [22] to dialogue with companies across areas and to jointly pursue innovation, standards and policy changes that further sustainability. Companies should also consider requiring partners and suppliers to have a zero-carbon plan or science-based target and hold them accountable. These actions will help everyone row together and in the same direction and accelerate towards a zero-carbon future!

REFERENCES

[1] Bandwidth Alliance. https://www.cloudflare.com/bandwidth-alliance/.

[2] Calculating My Carbon Footprint - Microsoft Sustainability. https://microsoft.com/sustainability/emissions-impact-dashboard. Retrieved May 18, 2022.

[3] Carbon Footprint - Google Cloud. https://cloud.google.com/carbon-footprint. Retrieved May 18, 2022.

[4] Carbon Footprint Reporting – Customer Carbon Footprint Tool – Amazon Web Services. https://aws.amazon.com/aws-cost-management/aws-customer-carbon-footprint-tool/. Retrieved May 18, 2022.

[5] Cement and Concrete: The Environmental Impact. https://psci.princeton.edu/tips/2020/11/3/cement-and-concrete-the-environmental-impact. Retrieved May 18, 2022.

[6] Clean Energy Buyers Association (CEBA). https://cebuyers.org/. Retrieved May 18, 2022.

[7] Cloud Broker: A Systematic Mapping Study. https://arxiv.org/abs/2102.12717. Retrieved May 18, 2022.

[8] Cloud Carbon Footprint - An open-source tool to measure and analyze cloud carbon emissions. https://www.cloudcarbonfootprint.org/. Retrieved May 18, 2022.

[9] electricityMap – The leading resource for 24/7 CO2 grid data worldwide. https://electricitymap.org/. Retrieved May 18, 2022.

[10] Environment. https://www.apple.com/environment/. Retrieved May 18, 2022.

[11] Facts & figures. https://www.atag.org/facts-figures.html. Retrieved May 18, 2022.

[12] Global E-Waste Surging: Up 21% in 5 Years - United Nations University. https://unu.edu/media-relations/releases/global-e-waste-surging-up-21-in-5-years.html. Retrieved May 18, 2022.

[13] Green G – Next G Alliance. https://nextgalliance.org/working_group/green-g/. Retrieved May 18, 2022.

[14] Home - Information Technology Industry Council. https://www.itic.org/. Retrieved May 18, 2022.

[15] HPE product carbon footprint – HPE ProLiant DL360 Gen10 server data sheet. https://www.hpe.com/psnow/doc/a50002430enw. Retrieved May 18, 2022.

[16] Jon Taylor and Jonathan Koomey Publish Second Comatose Server Report. https://www.anthesisgroup.com/wp-content/uploads/2019/11/Comatose-Servers-Redux-2017.pdf. Retrieved May 18, 2022.

[17] Product Carbon Footprints | Dell Technologies US. https://www.dell.com/en-us/dt/corporate/social-impact/advancing-sustainability/sustainable-products-and-services/product-carbon-footprints.htm. Retrieved May 18, 2022.

[18] Right Place, Right Time (RiPiT) Carbon Emissions Service. http://ripit.uchicago.edu/Part1/ripit-part1.html. Retrieved May 18, 2022.

[19] SEC Proposes Rules to Enhance and Standardize Climate-Related Disclosures for Investors. https://www.sec.gov/news/press-release/2022-46. March 21, 2022.

[20] Software Carbon Intensity (SCI) Specification. https://github.com/Green-Software-Foundation/software_carbon_intensity/blob/main/Software_Carbon_Intensity/Software_Carbon_Intensity_Specification.md.

[21] State of Data Center Networking 2021 Annual Report. https://go4.pluribusnetworks.com/WF-2021-StateofDataCenterNetworking-Report.html. Retrieved May 18, 2022.

[22] The Green Software Foundation. https://greensoftware.foundation/. Retrieved May 18, 2022.

[23] The Roadmap to Sustainable Digital Infrastructure by 2030. https://sdialliance.org//roadmap/. Retrieved May 18, 2022.

[24] Watttime – The Power to Choose Clean Energy. https://www.watttime.org/. Retrieved May 18, 2022.

[25] Global wind report: Annual market update. Tech. rep., Global Wind Energy Council, 2016. Documents curtailment around the world.

[26] REALITY CHECK: CLIMATE ACTION AND COMMITMENTS OF THE FORTUNE GLOBAL 500. https://www.climateimpact.com/news-insights/insights/reality-check-climate-action-and-commitments-fortune-global-500/, September 2021.

[27] A. A. Chien. Characterizing opportunity power in the California Independent System Operator (CAISO) in years 2015-2017. *Energy and Earth Science 3*, 2 (December 2020). Also available as University of Chicago, Computer Science TR-2018-07, https://newtraell.cs.uchicago.edu/research/publications/techreports.

[28] A. A. Chien and F. Yang and C. Zhang. Characterizing curtailed and uneconomic renewable power in the mid-continent independent system operator. *AIMS Energy 6*, 2 (December 2018), 376–401.

[29] A. Radovanovic et al. Carbon-Aware Computing for Datacenters. https://arxiv.org/abs/2106.11750, 2021.

[30] Andrae, A.S.G. and Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges 6* (2015), 117–157.

[31] Anil Madhavapeddy and David J. Scott. Unikernels: Rise of the Virtual Library Operating System: What if all the software layers in a virtual appliance were compiled within the same safe, high-level language framework? *Queue 11*, 11 (November 2013), 30–44.

[32] B. Duncan and A. Happe and A. Bratterud. Enterprise IoT Security and Scalability: How Unikernels can Improve the Status Quo. In *2016 IEEE/ACM 9th International Conference on Utility and Cloud Computing (UCC)* (2016), pp. 292–297.

[33] C. Zhang and A. A. Chien. Using carbon-awareness to reduce the cloud's environmental impact. Tech. rep., University of Chicago, November 2020. University of Chicago Technical Report, available from https://newtraell.cs.uchicago.edu/research/publications/techreports.

[34] Charlotte Freitag and Mike Berners-Lee and Kelly Widdicks and Bran Knowles and Gordon Blair and Adrian Friday. The climate impact of ICT: A review of estimates, trends and regulations. Retrieved May 18, 2022.

[35] George Kamiya. Data Center and Data Transmission Network. https://www.iea.org/reports/data-centres-and-data-transmission-networks, November 2021. Retrieved May 18, 2022.

[36] Gottel et al. Security, Performance and Energy Trade-offs of Hardware-assisted Memory Protection Mechanisms. https://arxiv.org/pdf/1903.04203.pdf, June 2019.

[37] Hennessy, J., and Patterson, D. A new golden age for computer architecture:

domain-specific hardware/software co-design, enhanced. In *ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)* (2018).
[38] IEA. Global Energy & CO2 Status Report 2019 – Analysis. https://www.iea.org/reports/global-energy-co2-status-report-2019, 2019. Retrieved May 18, 2022.
[39] Ion Stoica and Scott Shenker. From cloud computing to sky computing. In *Proceedings of the Workshop on Hot Topics in Operating Systems* (Ann Arbor, Michigan, 2021), ACM, pp. 26–32.
[40] IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, vol. In Press. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2021.
[41] J. Lindberg et al. A Guide to Reducing Carbon Emissions through Data Center Geographical Load Shifting. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems* (2021).
[42] Jennifer Switzer and Ryan Kastner and Pat Pannuto. Architecture of a Junkyard Datacenter. Retrieved May 18, 2022.
[43] Joel Jaeger. Explaining the Exponential Growth of Renewable Energy. https://www.wri.org/insights/growth-renewable-energy-sector-explained, September 2021. Retrieved May 18, 2022.
[44] Kelly Morgan. Saving Energy in Europe by Using Amazon Web Services. https://d39w7f4ix9f5s9.cloudfront.net/d1/80/283b833847df8ee4fe9661e0dd8f/11061-aws-451research-advisory-bw-cloudefficiency-eu-2021-r2-final-2.pdf, October 2021. (15% utilization in APAC and EU, and 18% in U.S. enterprises, page 5 [these percentages include both virtualized and non-virtualized servers]); Internal study (2019) conducted by VMware looking at maximum and average hourly CPU utilization across thousands of customers over a one-year time period.
[45] L. P. Lin and A. A. Chien. Characterizing stranded power in the ercot in years 2012-2019: A preliminary report. Tech. rep., University of Chicago, August 2020.
[46] liikenne- ja viestintäministeriö. The two faces of the ICT sector: consuming energy and materials while also paving the way to a carbon-neutral society. https://www.lvm.fi/en/-/the-two-faces-of-the-ict-sector-consuming-energy-and-materials-while-also-paving-the-way-to-a-carbon-neutral-society-1210283. Retrieved May 18, 2022.
[47] Mark Russinovich Bhatia Manuel Costa, Cedric Fournet, David Chisnall, Antoine Delignat-Lavaud, Sylvan Clebsch, Kapil Vaswani, Vikas. Toward Confidential Cloud Computing. https://cacm.acm.org/magazines/2021/6/252824-toward-confidential-cloud-computing/fulltext. Retrieved May 18, 2022.
[48] Md Abu Bakar Siddik and Arman Shehabi and Landon Marston. The environmental footprint of data centers in the United States. *Environ. Res. Lett. 16*, 6 (June 2021), 064017.
[49] Mengpin Ge and Johannes Friedrich and Leandro Vigna. 4 Charts Explain Greenhouse Gas Emissions by Countries and Sectors. https://www.wri.org/insights/4-charts-explain-greenhouse-gas-emissions-countries-and-sectors, February 2020. Retrieved May 18, 2022.
[50] Natural Capital Partners. Reality Check: Climate action and commitments of the Fortune Global.... https://www.naturalcapitalpartners.com/insights/article/reality-check. Retrieved May 18, 2022.
[51] Nicole Labutong and Vincent Hoen. How can companies address their scope 3 greenhouse gas emissions? https://sciencebasedtargets.org/blog/how-can-companies-address-their-scope-3-greenhouse-gas-emissions, May 2018.
[52] Patrik Hummel and Matthias Braun and Max Tretter and Peter Dabrock. Data sovereignty: A review. *Big Data & Society 8*, 1 (January 2021), 205395172098201.
[53] R. Nambiar. Introducing Confidential Computing for Private Clouds. https://community.amd.com/t5/business/introducing-confidential-computing-for-private-clouds/ba-p/414266, Sept 2020.
[54] Sriram Subramanian and Brad Casemore. Enabling More Agile and Sustainable Business Through Carbon-Efficient Digital Transformations. https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/company/vmware-idc-whitepaper-2020.pdf, August 2020. Retrieved May 18, 2022.
[55] Stempesta et al. Azure confidential computing products. https://docs.microsoft.com/en-us/azure/confidential-computing/overview-azure-products. Retrieved May 18, 2022.