EDITOR **DEJAN MILOJICIC**
Hewlett Packard Labs;
dejan.milojicic@hpe.com

Sustainability: Fundamentals-Based Approach to Paying It Forward

Cullen Bash, Ninad Hogade, Dejan Milojicic, and Gourav Rattihalli, Hewlett Packard Labs

Chandrakant D. Patel, HP Inc.

Sustainability has become a critical problem confronting the community. Ecological issues such as climate change and CO_2 emissions, economical frictions such as energy supplies, and socio-political issues such as wars threaten growth and equity. How can technology help?

Sustainability has always been an important topic for business and, more broadly, for humanity. With growing global recognition of the need to curb carbon emissions, it has become a strategic priority for many CEOs and boardrooms. Companies are making pledges to become carbon neutral by 2050, 2040, and even as early as 2030. Many corporations and governments around the world are trying to understand and quantify their share of carbon emissions, how to reduce their share to net-zero, and perhaps even reverse emissions to achieve a net positive impact by removing more carbon than they are emitting. Information technology (IT) can play a crucial role in all of these aspects while also improving its own carbon footprint.

In this article, we address both topics using examples from enterprise sustainability management. We dissect carbon footprint by describing what is reported and break that down into contributions from upstream and downstream operations. We present how systems practice sustainability through an example of the application of energy-aware principles to data centers.

This is the second of the four articles in the "Predictions" column of *Computer* which is in its second year. All four articles are dedicated to megatrends. The first

---

Digital Object Identifier 10.1109/MC.2022.3219173
Date of current version: 9 January 2023

one was on Metaverse, the third one will be on digital transformation, and the fourth one will summarize megatrends. Megatrends have a substantial impact on humanity, they are an evolution of multiple technological trends. Megatrends are both the sum of current trends and a guiding force of new ones.

This is also the theme of the IEEE Future Directions Committee (FDC), where megatrends are connected to FDC technology initiatives, to roadmaps, and to standards. Megatrends grew out of technology trends that the IEEE Computer Society has published for the last ten years, including in the four past and forthcoming special issues of *Computer*.

PRINCIPLES

Sustainability was defined by the UN Brundtland Commission in 1987 as the intersection of social equity, economic growth, environmental problems, and the wellbeing of current and future generations.[1] Simply stated we must preserve the quality of life for current and future generations by assuring appropriate and adequate supply-side resources to meet the fundamental demand-side needs. In this context, we introduce a fundamentals-based supply–demand framework to define and practice sustainability. We begin by representing the supply-side pool of resources as available energy.

Available energy, also called *exergy*, refers to energy that is available for performing work and is typically expressed in Joules. While the quantity of energy is conserved (first law of thermodynamics), the quality of energy (available energy) is being continually destroyed as it transitions from one form to another. Indeed, the destruction of the reservoir of available energy is akin to an hourglass with the grains of sand flowing from one half to the other. The one-way movement of grains cannot be stopped, but we have the potential—with sustainable development guided by a holistic approach—to slow the process by getting the most that we can out of every single grain of sand, or unit of available energy, in our case. We contend that all supply-side resources—materials to water—can be represented as available energy. The Joules of available energy needed to treat and distribute water, if nature desalinates the water with regularity, becomes the measure for representing water.

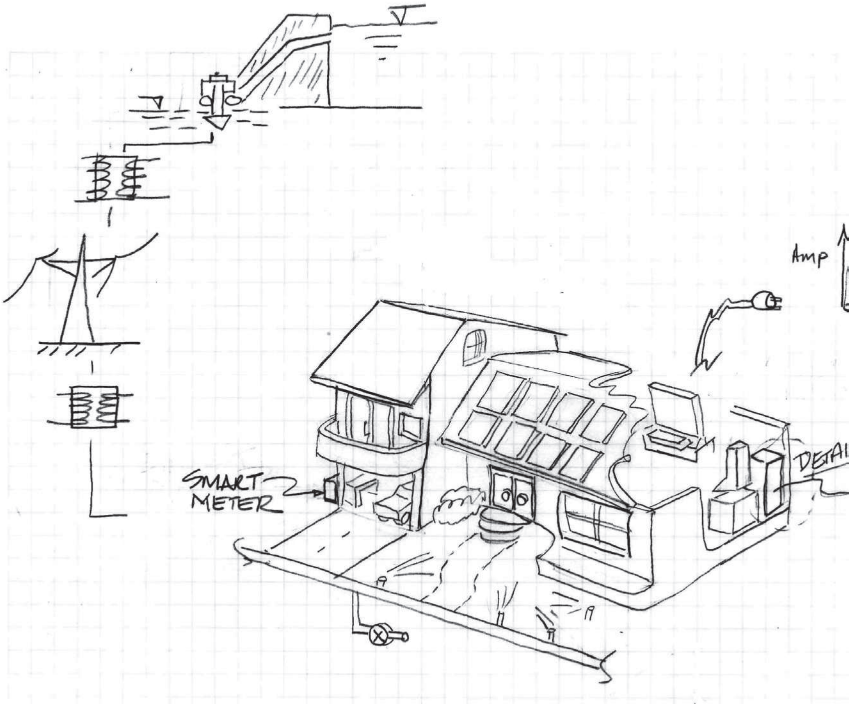

FIGURE 1. Destruction of available energy.

In applying available energy as a measure, consider the hydroelectric power station shown in Figure 1. The available energy in water—the potential energy at a given height in a dam—is converted to mechanical energy and then to electrical energy. Not all of the potential energy is converted to electrical energy. Indeed, it is instructive to examine the variety of turbines used for conversion.[2] As electrical energy is transmitted and distributed from the source (the power station) to the point of use, losses along the way in transmission and distribution accumulate, leading to significant destruction of availability prior to reaching a charging port on an electric vehicle, or a chip in a laptop. To minimize the destruction of available energy in transmission and distribution, one can consider local sources of available energy generation at or near the site of use like solar photovoltaic cells on the roof of a house. Therefore, with available energy as the measure, we introduce the following supply–demand framework.

The principles of supply side are as follows:

- Cradle-to-cradle analysis and design in minimizing the available energy in Joules required to extract, manufacture, mitigate waste, transport, operate, and reclaim components.
- The use of local resources—such as local power generation or a water microgrid—to minimize the destruction of available energy as opposed to traditional systems in which available energy is lost in the transmission and distribution over long distances from a centralized plant.
- Seek available energy in waste streams, e.g., available energy

in exhaust heat from turbines or available energy in waste from a farm or municipality.

The principles of demand side are as follows:

- Minimize the consumption of available energy by optimal provisioning of resources based on the needs of the user.
- On-demand provisioning using flexible hardware building blocks, pervasive sensing, communications, knowledge discovery, and policy-based control.
- Hardware-software codesign and management.

In the rest of this article, we discuss how these principles are applied, and how IT can help in terms of strategy and operations.

APPLICATIONS

The following examples of products, datacenters, and 3D printing illustrate the application of the supply–demand framework.

Products: For all products, from laptops to household goods, the supply-side principle of cradle-to-cradle accounts for the available energy required to extract, manufacture, clean up waste, transport, use, repurpose, and upcycle back to cradle. This type of lifecycle analysis ought to occur during design to minimize the lifetime destruction, or cost, of available energy for a product. The cradle-to-cradle cost of an example laptop is estimated to be 9 GJ or 200 L of diesel for a three-year lifetime, and six hours of use per day, in the essay titled “Joules: The Currency of Sustainability.”[4] Figure 2 divides cradle-to-cradle into three segments. Cradle to gate refers to the customer receiving the product, while gate to grave is the operational phase, and grave to cradle the reclamation phase including upcycling of a component or a product to another use. The figure shows strategies at an overview level in each phase to minimize the lifetime energy used by the item.

Data Centers and 3D Digital Manufacturing: Other larger industrial-scale applications of the supply–demand framework are data centers and digital manufacturing. Data center design and operation ought to follow all the supply–demand side principles.[5]

On the supply side

- Cradle-to-cradle lifetime analysis of all the system components should be conducted—IT, power, cooling, building complex—to minimize the lifetime available energy.
- Utilize waste streams and local sources of available energy (e.g., a microgrid with solar power, and power from waste streams such as manure from dairy cows).[5]

On the demand side

- Provision IT, power, and cooling resources based on the user’s service level agreement (SLA). This can be achieved, as described in,[15] by software-driven resource allocation and usage, and demand shaping to operate under the power production curve.

For an example of 3D digital manufacturing, consider a 3D digital print factory with one hundred 3D printers. Such a factory would need approximately 1 MW of power to run all the machines simultaneously. In addition to the 3D machinery, the factory would also contain a small data center and additional processing systems to run the operations in real time making it higher than 1 MW. Akin to the data centers, the equipment in the factory ought to be designed with an eye on minimizing the use of cradle-to-cradle lifetime available energy.

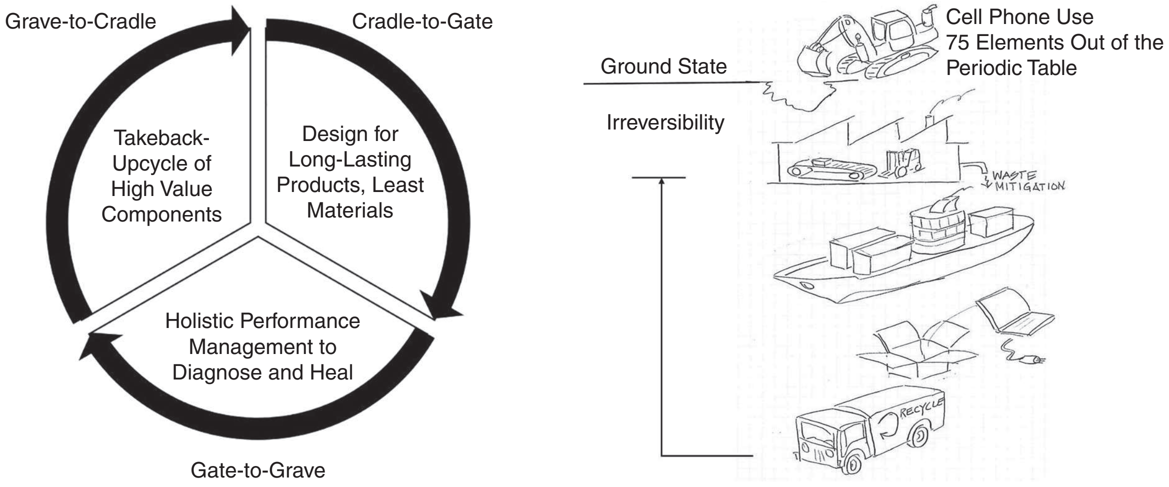

FIGURE 2. Sustainability lifecycle.

Next, on the supply-side principle of local sources and waste streams, the factory can utilize power availed locally using energy conversion means from sun and waste streams such as municipal and farm waste (biogas). The diurnal power production from the sun, supplemented with base load power from biogas, could comprise the sole source of power. On the demand side, an integrated supply–demand management software *shapes the demand in the digital factory* given the customer SLAs to stay within the supply-side power production for net zero operations.

TECHNOLOGY LIFECYCLE MODEL

Given the sustainability lifecycle, determining when to replace aging technology becomes a key consideration. With the increasingly rapid pace of technological advancement, it's tempting to want to refresh older technology with new sooner rather than later, particularly in light of large potential reductions in operational carbon footprint. Considering only the carbon emitted during the operating phase of a system (gate-to-cradle in Figure 2) ignores the upstream carbon embedded in the manufacturing and delivery process (cradle-to-gate), also known as *embedded carbon*. Indeed, knowing when to replace aging technology, or return it to use in a different capacity as part of the sustainability lifecycle, is often as important as the choice of the replacement.

The metric net positive impact (NPI) assists in the technology refresh decision by comparing the cost, in energy or carbon, of the continued operation of the older technology with that of the replacement.[3] Because the older technology has already been deployed, the upstream costs are considered sunk costs while the complete lifecycle of the replacement technology, from the embedded to the use phase, is considered. Taking the ratio of costs of continued operation of the older technology with that of the replacement gives NPI as follows:

$$\text{Net Positive Impact (NPI)} = \frac{\text{Value delivered in available energy saved}}{\text{Available energy consumed over lifetime}}$$

where the cost is represented by available energy consumed.

Similarly, the cost parameter can be replaced by carbon to provide the Net Positive Carbon Impact (NPIc) as follows:

$$\text{Net Positive Carbon Impact } (\mathrm{NPI}_c)\, T_L = \frac{\sum^{T_L}[C_{\text{old}}]}{\sum^{T_L}[TCC_{\text{new}}]}$$

where T_L indicates the time duration over which NPI_c is evaluated and is generally equivalent to the expected lifetime of the replacement system. The carbon emitted during the use phase of a system is denoted as C, and TCC is the total cost of carbon and includes both use and embedded phases:

$$\text{Total Cost of Carbon (TCC)} = [\Sigma\,(C_{\text{embedded}}) + \Sigma\,(C_{\text{use}})]_{\text{lifetime}}.$$

If the rate of carbon emissions from both systems is constant, we can write:

$$\text{Net Positive Carbon Impact}\left(\mathrm{NPI}_c\right)_{T_L} = \frac{\dot{C}_{\text{old}} \cdot t}{\dot{C}_{\text{new}} \cdot t + C_{\text{embedded}}}$$

where $\dot{c}$ is the carbon emission rate and t is time.

To evaluate the utility of NPI_c, consider the replacement of a carbon-based power generation system with that of a cleaner source of energy, like solar photovoltaic. In this example, consider a solar photovoltaic installation that either completely or partially replaces a carbon-based generation source. In the complete replacement scenario, the PV installation is sized to fully replace the carbon-based source such that $\dot{C}_{\text{new}}$ is zero. In a partial replacement scenario, the PV installation is undersized such that ongoing carbon emissions are nonzero, representing a scenario in which carbon-based sources are used when photovoltaic supplies are limited, for example, by weather or nightfall.

Monocrystalline PV cells with an embodied carbon of 2,560 kg CO_{2e}/kWp are used, where Wp is the peak power output of the cell.[6] Since actual power generation is a function of location and installation a yield of 920 kWh/kWp per year is used in this example.[6] Assuming the PV installation is replacing a power source that emits 0.316 kg CO_2/kWh, *NPIC* is calculated over the lifetime of the PV installation as shown in Figure 3. Figure 3 charts *NPIC* as a function of time for several ongoing carbon emission rates in the new system. For each case, embodied carbon of the PV installation was fixed.

Note that an $\mathrm{NPI}_c = 1$ denotes the carbon payback period (CPP) where the corresponding time T_{CPP} is the time it takes to achieve carbon payback after a system is replaced. In this example, T_{CPP} is achieved more quickly as ongoing carbon emissions are reduced, and ranges from 8.8 years at zero emissions and 14 years at the maximum emissions rate evaluated. Absent other factors, systems with low CPPs might be refreshed more quickly than those with higher CPPs. In practice, however, CPP should not be used in isolation. Rather, it is one of several deterministic factors that include economics and overall environmental benefit (i.e., net reduction in cost).

ENVIRONMENTAL STRATEGY

The definition of sustainability laid out by the Brundtland Commission is broad, but the major concern facing the globe at the present time is more narrowly focused on greenhouse gas (GHG) emissions and the resulting impact on climate change. With rising emissions, Earth's temperature has already increased by 1.1 °C since the late 1800s. In 2015, world leaders at the UN Climate Change conference in Paris met to discuss the consequences of climate change and how to mitigate the resulting impacts. The result was the Paris Agreement that set 1.5 °C in temperature rise above preindustrial levels as the limit to prevent the worst impacts of climate change.[8]

Shortly thereafter, the Intergovernmental Panel on Climate Change (IPCC) was formed and charged with setting the rate of reduction in global carbon equivalent (CO_2eq) emissions necessary to stay below the 1.5 °C maximum. In 2018, they set the goal of reaching net-zero CO_2eq emissions by 2050 and halving them from 2018 levels by 2030.[9] Industry and governments followed suit, and by 2021 net-zero carbon pledges were made by entities representing two-thirds of the global economy by gross domestic product, 61% of CO_2eq emissions, and 56% of the global population including the United States, European Union, and China.[10]

Setting goals is a good first step, but concrete action is needed to achieve results. Because environmental problems and social equity are inextricably linked with economic growth, industry must be heavily involved in developing solutions or they risk declining profits. Indeed, IDC predicts that 90% of Forbes Global 2000 companies will mandate some form of action on climate change by 2025 as a prerequisite for doing business.[11]

The ability to measure, cut, and report GHG emissions becomes increasingly important as more companies strive to attain net zero and become carbon neutral. Planning and implementing successful climate change and GHG reduction strategies requires a thorough understanding of a company's GHG emissions as a starting point. The GHG Protocol shown in Figure 4[14] is complimentary to the sustainability lifecycle model presented previously and is the most widely used set of guidelines for carbon accounting. It divides GHG emissions into direct and indirect emissions. Direct GHG emissions come from sources that are under the ownership and control of the entity or company reporting the emissions. Indirect GHG emissions are produced by reporting company but originate from sources that are owned or under the control of another entity. These two types of emissions are further divided into three scopes as follows.

Scope 1: Direct emissions originate from sources that the reporting company owns or controls.

Scope 2: The production of purchased power, steam, heating, and cooling used by the reporting company results in indirect emissions.

Scope 3: Indirect emissions are those produced by operations on resources that the reporting company does not own or control, but which it indirectly influences through its supply chain.

Industry leadership will be critical in achieving net-zero carbon by 2050 by developing their own capacity for adaptation to climate-related risks like compromised infrastructure and supply chains and by speeding up the implementation of solutions that support the shift to low-carbon across all scopes. In addition to enhancing the efficiency and sustainability of their own solutions, industries can assist their clients in using these technologies to decarbonize carbon-intensive sectors and speed up climate research. To embark on a net-zero journey companies should:

- examine both the existing upstream and downstream activities
- prioritize critical emissions from the three scope areas
- perform a materiality analysis, determine the baseline, and prioritize key performance indicators such as NPI_c and CPP (as discussed in the earlier section)
- create a roadmap for enhancing the company's ability to fulfill its emission objectives in terms of resources, best practices, and financial requirements
- develop a business intelligence reporting infrastructure/dashboard to monitor progress
- create mechanisms for governance and competency to promote buy-in across organizations
- communicate with stakeholders to create general awareness of sustainability initiatives.

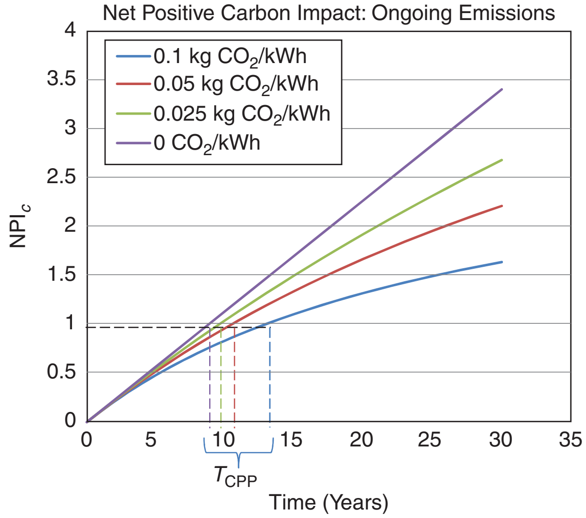

FIGURE 3. Net positive carbon impact.

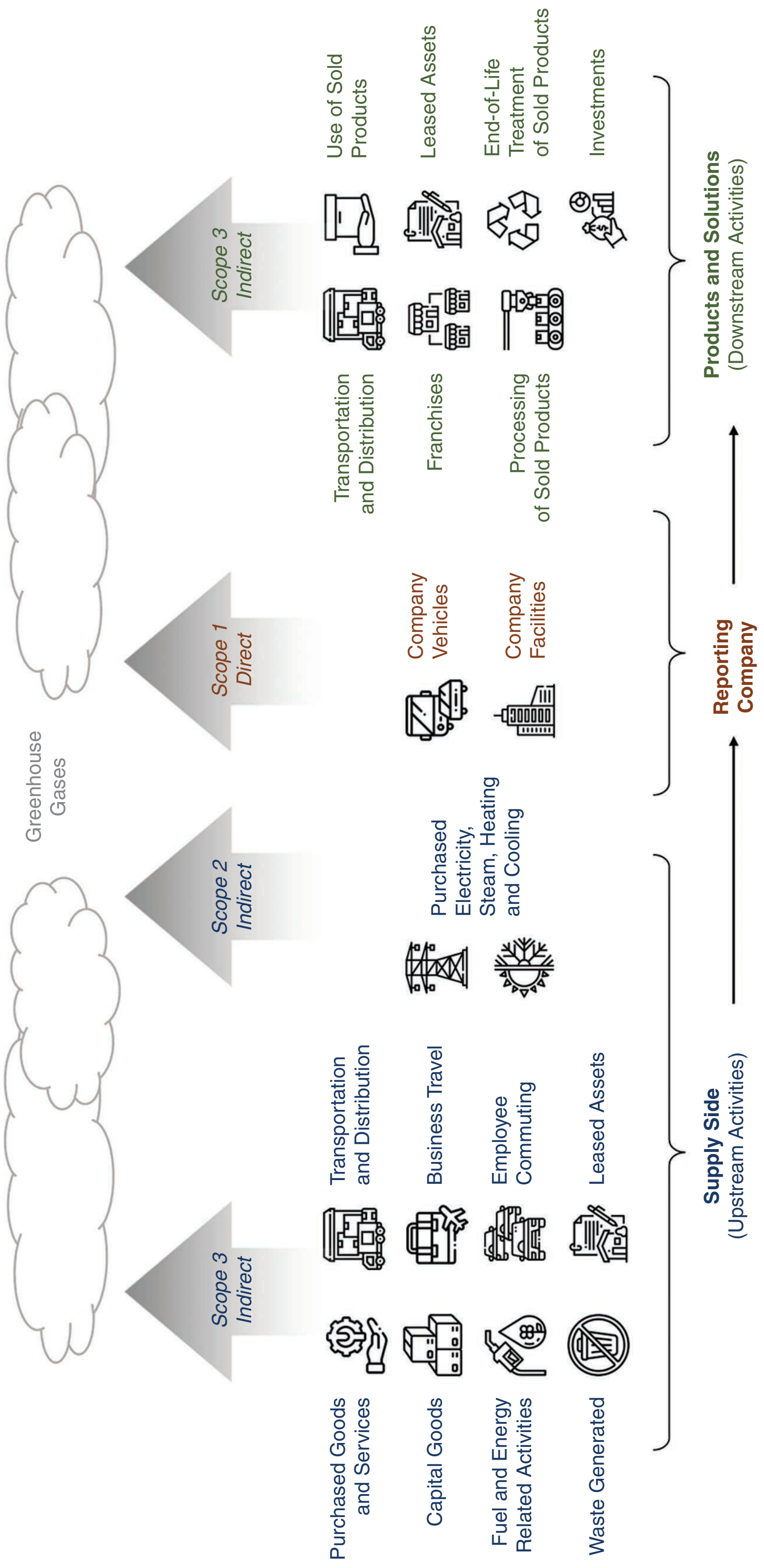

FIGURE 4. Company activities and greenhouse gas emissions.

ENVIRONMENTAL OPERATIONS: ENERGY-AWARE IT

As IT plays a key role in the management of energy, and emissions, it is important to focus on making data centers energy aware and efficient. This section will explore through the example of data centers how IT can be made energy-aware and how it can be used to improve the sustainability of other ecosystems.

The energy and materials that IT uses represent a sizable source of greenhouse gas (GHG) emissions. Data centers are estimated to contribute roughly 2% of all GHG emissions globally and use about 3% of the world's electricity supply.[13] That is roughly equivalent to the entire airline industry. This number will rise to 5%–7% of world emissions, which is alarming given what was observed during the pandemic.[13] To achieve ambitious sustainability and climate targets, businesses are challenged to improve the efficiency of their operations and lessen their environmental footprints in today's competitive economy. These initiatives frequently concentrate on an organization's IT infrastructure, which can be a major contributor to energy consumption and greenhouse gas (GHG) emissions. Additionally, the costs connected with these resources are passed on to the company in the form of energy consumption or increased demand for cooling and related equipment.

The IT industry has made it possible for today's infrastructures to be much more efficient than those of the past, but there are still plenty of opportunities to promote more sustainable and energy-efficient operations. This not only makes it possible for the IT organization to promote innovation and accomplish business goals, but it also helps it to make progress toward sustainability targets and goals.

Four aspects of effective energy-aware IT, discussed next, can help industries cut their carbon footprint. They focus on maximizing operational capacity, lowering energy consumption, and requiring the least amount of support. High performance and utilization levels enable more cost-effective IT infrastructures to employ less hardware, resulting in major cost savings including: OPEX, CAPEX, and space savings, as well as lower software and maintenance agreement license costs.

1. *Workload-Hardware Management*: The objective here is to perform the most work with the least amount of equipment by maximizing IT processing power and storage capacities while minimizing expense and resource demand. The rise in workloads has resulted in rising energy use by 10%–30% per year and will keep growing over the following years.[7] To design energy-efficient workload schedulers, it is increasingly important to understand the workloads and their energy consumption patterns. Such energy-efficient schedulers need to be aware of a hardware's performance per unit of energy and use appropriate constraints to schedule workload on the heterogeneous hardware. Data center operators can also employ sensors and firmware on their servers to provide real-time metrics to their data center information management software and enable real-time monitoring of capacity and usage. Further, such metrics should make their way into production-level technologies like Slurm, used widely for scheduling HPC applications, and Kubernetes used industry-wide for services.
2. *Energy source*: With the least amount of energy input possible, the aim is to give the highest possible degree of power, storage, and communication. IT businesses frequently struggle with power issues. Real estate or facilities executives can save money on utilities and/or use more renewable energy within their organizations. An example of IT for sustainability, smart grids can be employed to assist with demand shaping and can schedule energy usage to align with the availability of renewable energy resources. Energy efficiency can be taken into account by procurement teams in their procurement policies, where it is frequently utilized as a differentiator when awarding contracts.
3. *Resources*: The objective is to match type and quantity of material to space, power, and cooling requirements to develop systems that operate effectively inside the data center. Industries can reduce the demand for employees and support equipment. Depending on factors like age and location, data centers may consume up to half of their total electricity for cooling systems,[12] necessitating specialized techniques to reduce cooling requirements including the use of naturally occurring cooling sources that can eliminate or reduce the need for energy-intense mechanical refrigeration.
4. *Software-driven*: Using software, AI, or ML can be aimed to raise efficiency, decrease downtime, and enhance data center management procedures. This ties everything together by demonstrating how clever software drives hardware efficiency. Also, the choice of programming language and how software is written can have a big impact. An inefficient language or code can have a significant impact on performance efficiency. Software can be used to enhance efficiency and make each system more intelligent and self-sufficient. With the spread of AI and ML, software efficiency can detect the most effective performance condition for IT equipment in real-time and can reduce the energy consumption associated with hardware resources.

SUMMARY AND OUTLOOK

We have presented fundamental principles of sustainability and discussed examples of their application, technology lifecycle model, environmental strategies, and operations. They represent examples of how the world can address sustainability from the abstract to very concrete. In the rest of this section, we synthesize the key points with respect to the sustainability megatrend:

1. We need to look at the problem holistically from a fundamentals perspective (science, engineering, ecology, socio-political, economic), driven by technology solutions that can enable us to most effectively understand problems, test possible solutions and then apply them, oversee them, and evolve them as the world is changing.
2. It is never too late to start and never too little to contribute; every piece here and there accumulates, either positively or negatively.
3. Mother Earth is one, we all need to protect it. Much like an hourglass, we are continually depleting our supply-side resources which we described as available energy. A portfolio of solutions can throttle the destruction of available energy

and pay it forward. Every country, every industry, needs to be mindful of every sustainability angle.
4. Managing sustainability requires a synergistic interaction of operation technology (OT) and IT. Neither IT or OT by itself is sufficient to manage sustainability, only together can they gather insights and manage cyberphysical systems.
5. Integrating the sustainability lifecycle into the design of systems at their inception, and devising methods to extend it, is key to reducing available energy use.
6. Given the enormity of the challenge, we must create a large cadre of cyber-physical personnel, with the fundamentals of the machine age and the breadth of the cyber age.

Climate change is a fundamental threat to business growth and social equity, and a fundamentals-based approach is needed to address this challenge. The current generation can't rely on future generations to address this issue as time is running out. Solutions must come soon, and digital technologies can help accelerate them. The time is now to pay it forward to those who will come next.

The world cannot wait. We need to act! **C**

REFERENCES

1. "Our common future," The World Commission on Environment and Development (WCED) Report, also known as Brundtland Report, United Nations, New York, NY, USA, 1987. Accessed: Oct. 24, 2022. [Online]. Available: https://sustainabledevelopment.un.org/content/documents/5987our-common-future.pdf
2. S. Rao and B. B. Parulekar, *Energy Technology*. New Delhi, India: Khanna Publishers, 2005.
3. C. D. Patel, "Net positive impact: A metric for online transactions," LinkedIn, Sunnyvale, CA, USA, Jan. 22, 2020. [Online]. Available: https://www.linkedin.com/pulse/net-positive-impact-metric-online-transactions-patel-pe/
4. C. Patel, "Joules: The currency of sustainability," in *Practicing Sustainability*, G. Madhavan, B. Oakley, D. Green, D. Koon, and P. Low, Eds. New York, NY, USA: Springer Science & Business Media, 2013, pp. 117–121.
5. C. D. Patel, "Sustainable ecosystems: Enabled by supply and demand management," in *Proc. Int. Conf. Distrib. Comput. Netw.*, 2011, pp. 12–28.
6. "Embodied carbon of solar PV: Here's why it must be included in net zero carbon buildings." Circular Ecology. Accessed: Sep. 2, 2022. [Online]. Available: https://circularecology.com/solar-pv-embodied-carbon.html
7. "Data centres and data transmission networks," IEA, Paris, France, 2022. [Online]. Available: https://www.iea.org/reports/data-centres-and-data-transmission-networks
8. United Nations Framework Convention on Climate Change, "Adoption of the Paris agreement," in *Proc. 21st Conf. Parties*, United Nations, Paris, France, 2015. [Online]. Available: https://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf
9. "Summary for policymakers. In: Global warming of 1.5°C," IPCC, Geneva, Switzerland, 2018. [Online]. Available: https://www.ipcc.ch/sr15/chapter/spm/
10. "Taking stock: A global assessment of net zero targets," Energy & Climate Intelligence Unit, London, U.K., 2021. [Online]. Available: https://eciu.net/analysis/reports/2021/taking-stock-assessment-net-zero-targets
11. "IDC FutureScape: Worldwide future of digital infrastructure 2021 predictions," IDC, Needham, MA, USA, Oct. 2020. [Online]. Available: https://www.idc.com/getdoc.jsp?containerId=US46470820&pageType=PRINTFRIENDLY
12. "2020 Data Center Industry Survey Results." Uptime Institute. Accessed: Oct. 27, 2022. [Online]. Available: https://uptimeinstitute.com/2020-data-center-industry-survey-results
13. "Carbon emissions of data usage increasing, but what is yours?" Climate Neutral Group. Accessed: Oct. 27, 2022. [Online]. Available: https://www.climateneutralgroup.com/en/news/carbon-emissions-of-data-centers/
14. "FAQ." Greenhouse Gas Protocol. Accessed: Oct. 27, 2022. [Online]. Available: https://ghgprotocol.org/sites/default/files/standards_supporting/FAQ.pdf
15. Z. Liu et al., "Renewable and cooling aware workload management for sustainable data centers," in *Proc. 12th ACM SIGMETRICS/PERFORMANCE Joint Int. Conf. Meas. Model. Comput. Syst.*, London, U.K., Jun. 11–15, 2012, pp. 175–186, doi: 10.1145/2254756.2254779.

CULLEN BASH is a vice president and lab director at Hewlett Packard Labs, Milpitas, CA 95035 USA. Contact him at cullen.bash@hpe.com.

NINAD HOGADE is an engineer at Hewlett Packard Labs, Fort Collins, CO 80528 USA. Contact him at ninad.hogade@hpe.com.

DEJAN MILOJICIC is a distinguished technologist at Hewlett Packard Labs, Milpitas, CA 95035 USA. Contact him at dejan.milojicic@hpe.com.

GOURAV RATTIHALLI is an engineer at Hewlett Packard Labs, Milpitas, CA 95035 USA. Contact him at gourav.rattihalli@hpe.com.

CHANDRAKANT D. PATEL is a chief engineer at HP Inc. in Palo Alto, CA 94034 USA. Contact him at chandrakant.patel@hp.com.