clean-IT

Towards Sustainable Digital Technologies

Summary

Abstract

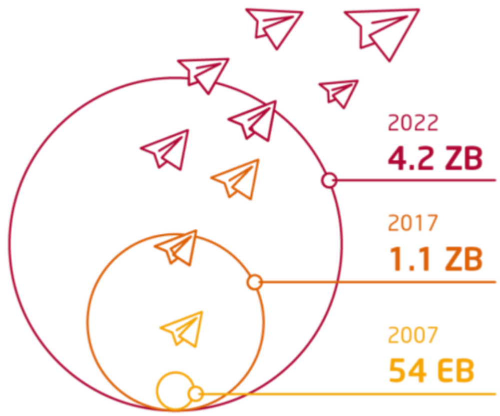
Digital technologies are indispensable to achieving sustainable development goals and reducing carbon emissions in many sectors. Yet, IT itself has an immense energy requirement for its countless devices, data centers, applications and global networks. With Covid-19 the shift to digital in living, learning and earning has brought us closer to the time when digitization will become the climate hazard No. 1. Since digitization has many important benefits, policy must strategically prioritize research to: deliver digitization with less energy consumption; increase algorithmic efficiency and effectiveness so that people and the planet thrive; and make “Sustainability by Design” the new paradigm in digitization worldwide.

Challenge: Digital Technologies on Their Way to Becoming Climate Hazard No. 1

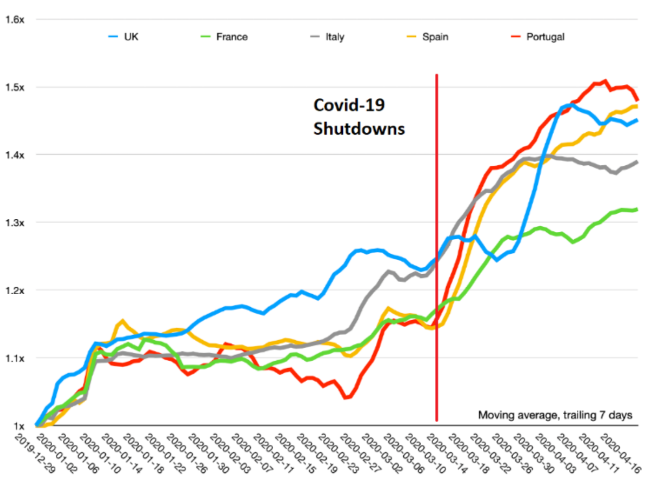
During the last decades, digital technologies were celebrated as the clean counterpart to old-fashioned "dirty" manufacturing, agriculture and the energy production industries. It was believed that digital devices, products and services, due to their immaterial nature, do not contribute (or contribute very little) to global pollution by wasteful consumption of material resources and the emission of greenhouse gases. This is, however, an erroneous view. All digital devices and applications contribute significantly to the global carbon footprint. Even though neither computers, tablets and smartphones – nor even data centers – have chimneys, the amount of carbon emissions caused by digital technologies has become a threatening climate issue.

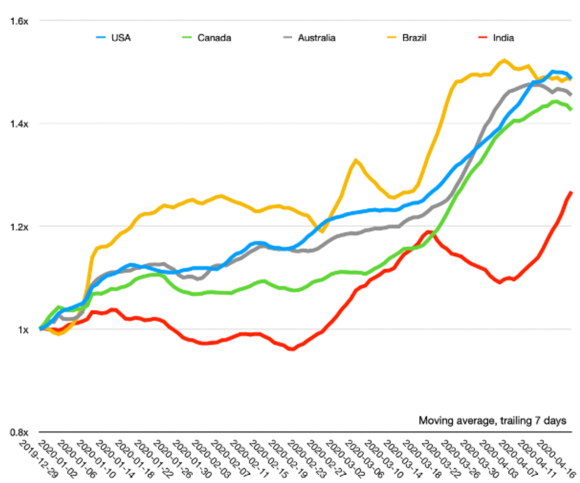
Global Internet Traffic on the Rise

All data traffic requires energy. The total amount of annual internet traffic has risen exponentially during the last few years and continues to rise steeply. The International Energy Agency calculates that while in 2007 only 54 exabytes (54 billion gigabytes) of data were transferred over the internet, this amount increased by a factor of 20 in 2017 to 1.1 zettabytes (1.100 billion gigabytes). This organization estimates that the annual data traffic will quadruple by 2022, reaching 4.2 zettabytes.[1]



Considering the current Covid-19 pandemic, the estimation of the IEA will be a massive underestimation of the escalation of internet traffic in the next years. Just in the last few weeks since the shutdown of major global economies, internet traffic has increased by 50 % due to social distancing and the rapid expansion of internet use in the workplace and for casual digital social interactions.[2] We cannot expect internet use to decrease after the pandemic. The crisis has essentially accelerated the already existing trend in "New Work" and digital interaction, which would have occurred in any case.

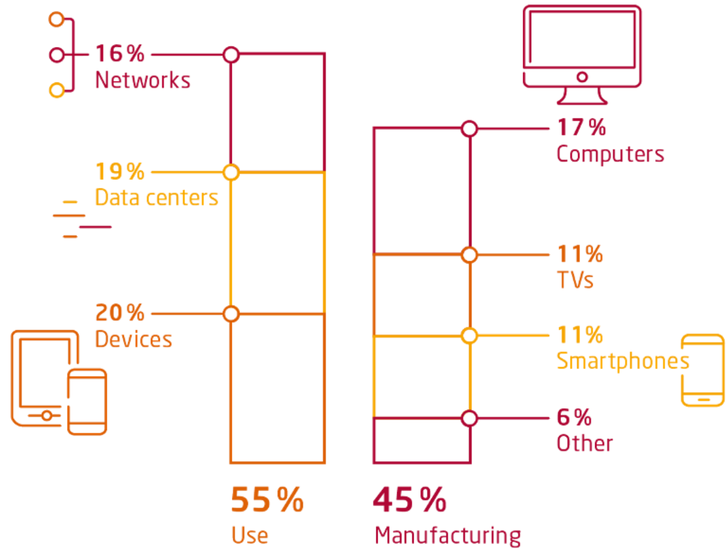


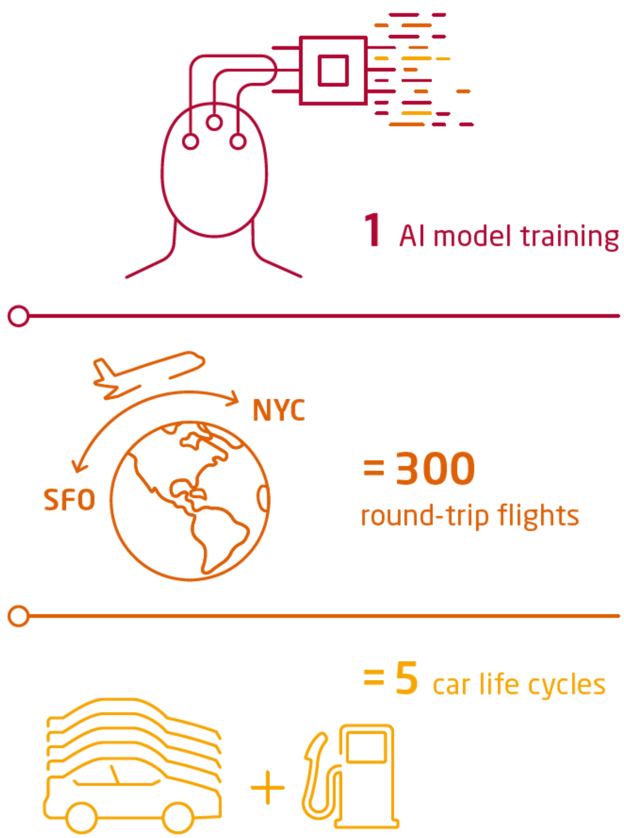


[1] IEA 2017, Global Internet Traffic.

[2] Cloudflare 2020, Internet Traffic in Corona Times.

Rapidly Increasing Software Use
A study of "The Shift Project" shows another interesting trend. Already in 2017,[3] energy consumption by use of digital technologies surpassed that of the production of digital devices by more than 5 %, with a continuously rising share. This suggests that measures to decrease the carbon footprint of IT need to focus increasingly on making the use of digital technologies more energy-efficient rather than on the resource-efficient production of digital devices. Of course, the former must not exclude the latter.

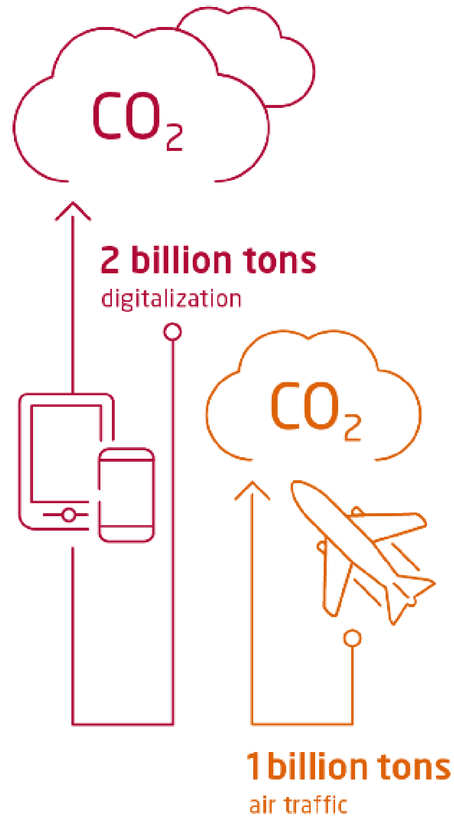




Energy Consumption of the Latest Digital Innovations
Researchers at the University of Massachusetts Amherst have studied the life cycles of training several common AI models and found that the energy consumption, and therefore the carbon emissions, of developing advanced neural networks are significant. The training of one specific AI model required about 300 tons of carbon dioxide equivalents, which equals the carbon emissions of the life cycle of five cars including fuel or 300 round-trip flights from New York City to San Francisco.[4]

[3] The Shift Project 2019, Energy Consumption in the IT Sector.
[4] Stubell et al. 2019, Carbon Footprint of Digitization.

Inefficient Software on the Way to the No. 1 Climate Hazard
Already today, the total carbon emissions of digital technologies surpass those of global air traffic by a factor of two. In 2019 all air traffic combined accounted for about one billion tons of carbon emissions and 2 % of overall emissions.[5] In the same year, digital technologies emitted about 2 billion tons, or about 4 %, of all human-induced carbon dioxide.[6] The relatively small share of air travel and digitization in global CO2-emission at the moment might be misleading. While air traffic has reached its peak, awareness of its negative impact on climate is widely recognized and we are already on track to reducing its carbon footprint; but the same is not true for digital technologies. Humanity is just at the beginning of a massive acceleration of digitization on a global scale, which does not yet consider the energy-efficiency of software. If this trend is ignored, digital technologies will become the most significant contributor of greenhouse gases and therefore the No. 1 climate hazard in the near future.



Green IT Is Not Enough
The issue of the increasing carbon footprint of IT has been recognized for some time now and has led to various initiatives that can be summed up under the label "Green IT. The primary focus of this movement is the reduction of the waste of natural resources during production and the use of digital devices[7] as well a call for "digital sobriety."[8] Already in 1992 the US Environmental Protection Agency (EPA) and the EU Commission (in 2003) introduced the "Energy Star" label for energy-efficient ICT devices. Despite this, the energy consumption of digital technologies continues to rise steeply. The reason for this increase is partly because the testing schemes of the Energy Star label have not been rigorous enough. Another major problem is that the increasing share of carbon emissions caused by the use of software and digital applications is not accounted for by efficient hardware. Algorithmic efficiency remains a blind spot in most of the Green IT initiatives.

Digital sobriety is also not likely to support the quest for a carbon-free planet. Digital technologies and their various innovative applications, such as big data, AI, blockchain, etc. are essential to decreasing carbon emissions in other sectors like energy production, manufacturing and agriculture. Reducing the use of digital technologies in those sectors would increase total carbon emissions across all sectors. In addition, other benefits of digital technologies would not come into effect to achieve other sustainability goals, such as fighting poverty, pandemics such as Covid-19 and promoting literacy on a global scale. Digitization is the answer to many sustainability challenges. Therefore, it is necessary to use more IT with less energy. This can only be done if software development follows a novel paradigm, which we call "Sustainability by Design". It focuses first on an awareness of the digital carbon footprint and especially on the impact of wasteful algorithms. Since the use of digital technologies already represents the biggest

[5] ATAG 2019, Carbon Footprint of Aviaton.
[6] The Shift Project 2019, Carbon Footprint of the IT Sector.
[7] OECD 2009, Green-IT.
[8] The Shift Project 2019, Digital Sobriety.

share of the digital carbon footprint--and will continue to rise steeply--it is necessary to focus on making algorithms more efficient. The trade-off between precision/speed/data throughput and energy consumption must be brought into balance by becoming a core principle of software design.

Why clean-IT and "Sustainability by Design"?

Humankind is still in the nascent phase of applying an innovation that will change the lives of all people on the planet. With the introduction of digital applications to almost every imaginable aspect of human life, we will see an increase in the use of digital technologies that is not yet comprehensible. At the edge of the massive roll-out of such innovations, the repercussions of their use are barely tangible. However, examples from the past can guide us to not repeat our mistakes in applying new technologies.

When plastics were invented it was a groundbreaking innovation in the field of new materials. A world without plastics would be less prosperous and innovations in many other areas would not have been possible. However, the increased use of plastics poses a threat to life on earth today because of the high level of difficulty in recycling this material. We see the oceans and land we inhabit choking on plastic waste. Since the inception of an awareness campaign, less plastic is used and more importantly, more environmentally friendly and recyclable plastics have been invented. However, had the scientists and designers behind plastic development been mindful of its potential environmental impact from the onset, we would not be facing the massive (and very expensive) crisis of today.

Another example outlines the case for water waste in fountain systems. Early fountains used water without recycling, requiring a large reservoir at a high place to allow water to flow with sufficient pressure. The result was the gorgeous spectacles of beautiful fountains such as those in Versailles or the Summer Palace of Peter the Great near St. Petersburg. Today the Internet has only reached half the world's population, with the rest left behind as a result of the high cost involved. Yet the spread of the Internet could follow its own course of evolution just as fountains did. When electricity became available, with pumps and filters, the same water could be re-used, and with water waste brought under control, water fountains were erected for the enjoyment of all.

These two case studies reference the importance of following a design-led approach that promotes spending more time exploring the problem space before rushing into the solution space. A practice that enables us to really understand the system in which the challenge (and the pending solution) lies, includes the knowledge of its complexities and relationships. The journey can assist us in understanding the unintended consequences of our actions. With digital systems and software specifically, it is very much the same: both are currently being designed in a manner that does not take their environmental impact into account. Today, especially during the Covid-19 crisis, we are designing digital technology with a similar mindset as was prevalent in the designers of the first fountains - in this case, not water but energy is used wastefully or the object is wasteful in itself in the same way as plastics are. Ignorance of better methods unnecessarily increases our carbon footprint.

But this is avoidable. Software, which is based on algorithms, can take many forms to produce the same outcome. Often unnecessarily complicated programming or software design causes higher energy consumption when compared to algorithms that are more efficient. Innovative software architectures can achieve the same or slightly lower precision or data throughput, while saving enormous amounts of energy. It matters how many lines of code are used to run a software program, because even if algorithmic redundancy or bugs have no impact on the effectiveness of a software application, every algorithmic operation consumes energy. Therefore, the less code is necessary to execute a digital program, the better the performance and energy consumption.

To solve the paradox of *more from less*, new algorithmic paradigms must be put into practice. The principle of "**Sustainability by Design**" needs to become the very foundation of software development.

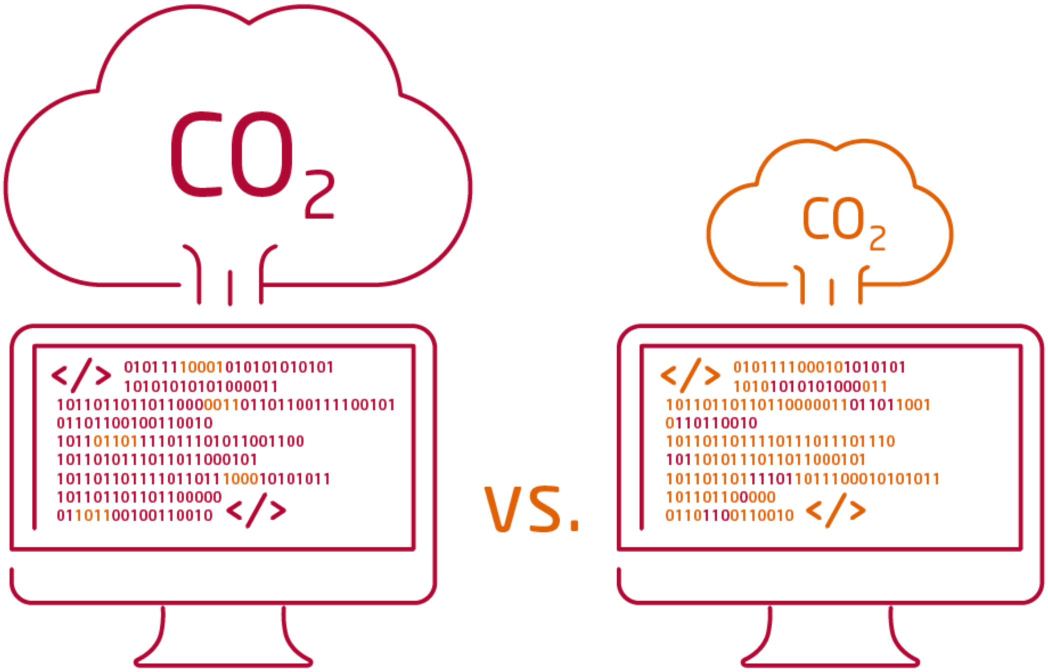
The following examples of the latest research in computer science on AI algorithms and data center operations illustrate that clean-IT and Sustainability by Design can already be achieved.

Examples of clean-IT

Heuristic Algorithms
Computer scientists and algorithm engineers develop software that aims at solving different classes of problems. In other words, any software problem or problem solved by software has 1) to be effective and 2) to be absolutely correct according to the internal logic of the topic. To achieve 100 % solutions, a certain amount of computational effort is necessary. In the real world that means that algorithms for "best solutions" are vast and complicated most of the time or that some sorts of problems cannot be solved at all, because especially complex problems require computation time outside of any reasonable scale, e.g. in the case of complex climate models or traffic prediction. The energy consumption of any algorithm is determined mostly by the runtime of the algorithm to find its solution. Therefore, complex computational algorithms with long runtimes consume a lot of energy in performing their task.

An important field of research in algorithm engineering focusses on tackling the issue by taking into account the trade-off between accuracy and runtime. Such algorithms work with heuristics, short-cuts or randomized code for achieving a solution close to the optimal solution, but requiring much less runtime. Such "second-best" algorithms, which most of the time come very close to optimal solutions, reduce the runtime of algorithms by orders of magnitude between 100-10,000, depending on the class of the problem.

Experiments at HPI have shown that the application of heuristic algorithms for the optimization of submodular functions, which can be implemented to optimize traffic, use raw materials in production or allocate goods in markets, reduce the runtime in comparison to traditional algorithms by a factor of 288. While traditional software takes two days to compute the solution for this problem, the heuristic algorithm takes only ten minutes and therefore reduced the energy consumption of the computing device to a mere 0.35 % of the original setup.



Clean Data Profiling

One of the major tasks of data engineers involves the problem of automatically organizing data in a meaningful way so that datasets become useful for a variety of applications in the "smart world"of artificial intelligence and analytics of all kinds. "Pure" data is without value. To have an impact it is necessary that data is structured into reasonable categories (exploration of the kind of data), normalized (structured homogenously), is cleaned from redundancies etc. Meta data plays a major role in this kind of task, which helps to organize data for value creation. In this sense, the organization of datasets is foundational to all data driven digital products and services.

One key aspect to data profiling is Unique Column Combinations (UCC). They identify entities in datasets and support different data organizing activities. Until recently, it was only possible to discover UCCs for moderately sized datasets with considerable run time effort. For large datasets UCCs were not discoverable due to runtime and memory limitations. Researchers at HPI have developed a novel UCC discovery algorithm ("HPIValid"), which drastically reduces the runtime of UCC discovery by orders of magnitude and at the same time reduce the memory footprint in comparision to state of the art algorithms like "HyUCC".

Across different moderately sized datasets HPIValid performed 5-100 times faster by using up only 5-20 % memory on average with decreasing efficiency on larger datasets. However, while the HyUCC algorithm was not able to discover UCCs on extremely large data sets, HPIValid was able to perform the calculation within a reasonable runtime and memory usage, therefore enabling value creation on massive datasets.[9]

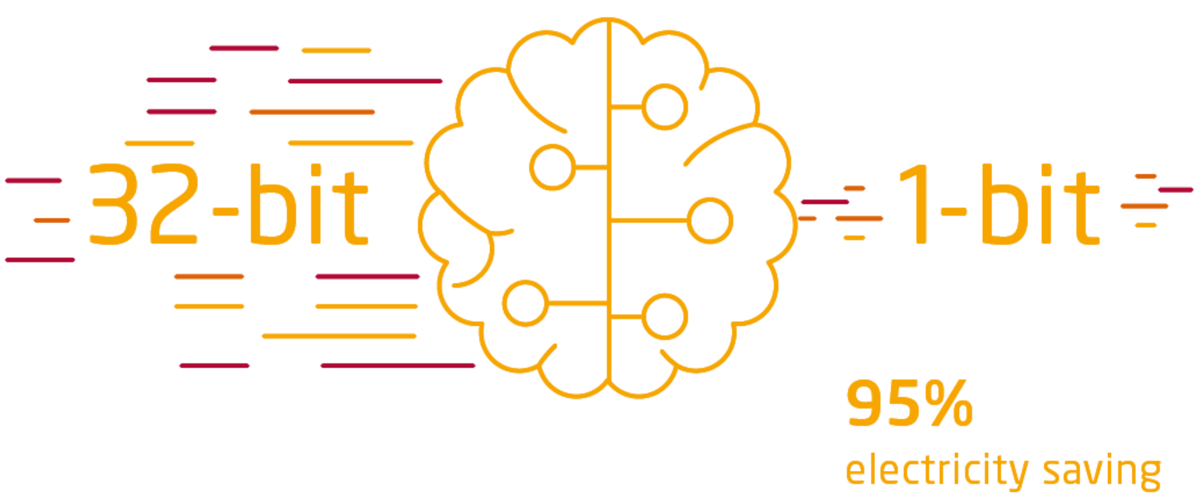
[9] Birnick et al., 2020, HPIValid.

Clean-AI

While the best AI systems train neural networks based on 32-bit algorithms (e. g. ResNet), the procedure can also be carried out with "binary neural networks" (1-bit algorithm). This drastically reduces the effort in the individual calculation steps and immediately leads to energy savings by a factor of 20. Although binary neural networks are currently about 5 % less accurate than those of AI systems of global players, this reduction decreases energy usage by 95 %. With AI applications used millions of times a day, this value adds up to significant levels.

Model Name	Accuracy	Model Size	Operations
BinaryDenseNet28	60.7 %	4 MB	$2{,}58*10^8$
BinaryDenseNet37	62.5 %	5.1 MB	$2{,}70*10^8$
BinaryDenseNet45	63.7 %	7.35 MB	$3{,}43*10^8$
ResNet (32-bit)	69.3 %	46,8 MB	$1{,}61*10^9$

The table shows the significant decrease of model size and operations of three variations of binary neural networks compared to full-precision 32-bit ResNet neural networks and a moderate loss of accuracy.[10]



[10] Bethge et al 2019, Clean-AI.

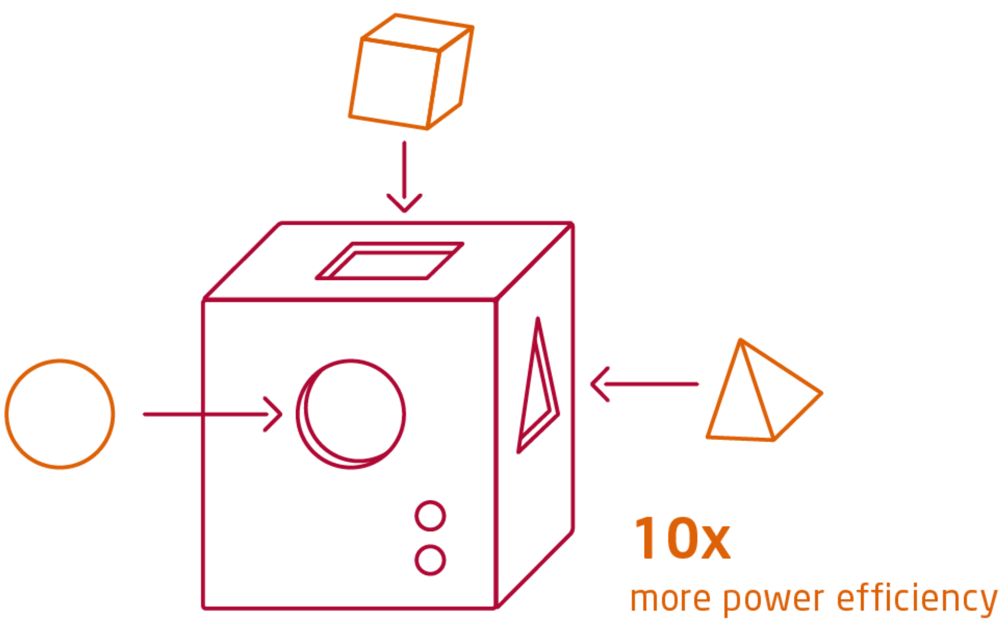
Energy-Aware Heterogeneous Computing

Next-generation data centers are embracing an increasingly diverse landscape of accelerators and hardware architectures, each offering advantages for certain algorithm classes or application domains. Unfortunately, today's software widely ignores this degree of heterogeneity. By executing workloads on the best-suited hardware, power efficiency can be improved significantly. In a preliminary evaluation conducted in a research seminar on energy efficiency, using FPGA accelerators instead of general-purpose processors boosted power efficiency by more than a factor of 10 for weather simulation models (see Table [n]). Similarly, Qasaimeh et al. have demonstrated a 20 fold improvement in energy efficiency when using FPGAs for certain Computer Vision tasks.[11] In contrast to these examples, tasks heavily relying on floating point arithmetic often show better performance and power efficiency characteristics on GPUs.[12]

Device	Problem Size (MCells)	Throughput (MCells/s)	Power Consumption (W)	Power Efficiency (MCells/Ws)
NVidia Tesla K20Xm	256	1127.88	60	18.80
Intel E5-2630 v4	256	1435.48	85	16.89
Xilinx XCKU060	1024	2209.35	9.5	232.56

In this special case a weather simulation was run on different processors with different results. For this specific task the Xilinx processor performed with higher energy efficiency than other processors. However, other tasks require other specialized hardware. How to route the right tasks to the right processors is the question that needs to be further investigated.

With its diverse hardware landscape, the HPI Future SOC Lab makes core aspects of next-generation data centers already available today, thus providing the perfect platform for investigating energy-aware computing strategies based on heterogeneous computing resources.



[11] Qasaimeh et al., 2019.

[12] Cong et al., 2018.

References

- ATAG 2019: Carbon Footprint of Aviaton, https://www.atag.org/facts-figures.html.
- Bethge et al 2019, Clean-AI, http://openaccess.thecvf.com/content_ICCVW_2019/papers/NeurArch/Bethge_BinaryDenseNet_Developing_an_Architecture_for_Binary_Neural_Networks_ICCVW_2019_paper.pdf
- Birnick et al., 2020, HPIValid, https://hpi.de/news/2020/vldb.html?tx_extbibsonomycsl_publicationlist%5Bcontroller%5D=Document&tx_extbibsonomycsl_publicationlist%5BfileName%5D=VLDB_Enumeration_with_partial_information.pdf&tx_extbibsonomycsl_publicationlist%5BintraHash%5D=bc9f648c76182d69b229f2a93ae3d45c&tx_extbibsonomycsl_publicationlist%5BuserName%5D=puma-friedrich&cHash=6ba0903332d88f3a8b81ab51b2421fba
- Cloudflare, Internet Traffic in Corona Times, https://blog.cloudflare.com/recent-trends-in-internet-traffic/
- Cong et al. 2018, Energy-Aware Computing: https://doi.org/10.1109/FCCM.2018.00023
- IEA 2017, https://www.iea.org/reports/digitalisation-and-energy
- OECD 2009, Green-IT, http://www.oecd.org/internet/ieconomy/42825130.pdf
- Qasaimeh et al. 2019, Energy-Aware Computing: https://ieeexplore.ieee.org/document/8782524
- Stubell et al. 2019, Carbon Footprint of Digitization, https://arxiv.org/pdf/1906.02243.pdf
- The Shift Project 2019, https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report_The-Shift-Project_2019.pdf
    - Energy Consumption in the IT Sector
    - Carbon Footprint of the IT Sector
    - Digital Sobriety