Green computing, or how to slow down the imperceptible contamination of ICTs

Alberto Prieto[1], Beatriz Prieto[1] and Juan José Escobar[2]

[1]Department of Computer Engineering, Automatics and Robotics. University of Granada (Spain)

[2]Department of Software Engineering. University of Granada (Spain)

Basque Conference on Cyber Physical Systems and Artificial Intelligence (BCCPS&AI)
European project CybSPEED (MSCA-RISE-2017 #777720).

U.P.V. E.H.U.
Grupo de Inteligencia Computacional

Green Computing or Sustainable ICT

- It is the study and practice of designing, manufacturing, using, and disposing of computers, servers, and its hardware like monitors, printers, storage devices, and networking and communications systems in order to efficiently and effectively consume energy with minimal or no impact to the environment
  - We are not going to consider the environmental aspects related to the manufacture, recycling and disposing

- Dhaini, M., Jaber, M., Fakhereldine, A., Hamdan, S., & Haraty, R. A. (2021). Green computing approaches- A survey. Informatica, 45(1)
- IBM (2022) Learn how green computing reduces energy consumption. https://www.ibm.com/cloud/blog/green-computing#:~:text=Green%20computing%20(also%20known%20as,consumed%20by%20manufacturers%2C%20data%20ce

Context

- Green Computing is part of one of the greatest challenges of today's society, consisting of reducing energy consumption
- In general, society is unaware that ICTs constitute a relevant field of electric energy consumption, having a great impact on greenhouse gas emissions
- Scientists, engineers and professionals must actively participate in the challenge of reducing it
- In addition to environmental reasons, reducing power consumption:
  - has strong economic implications, and
  - improves the autonomy of the many devices that use batteries, such as smartphones, mobile devices and elements of the Internet of Things

Content

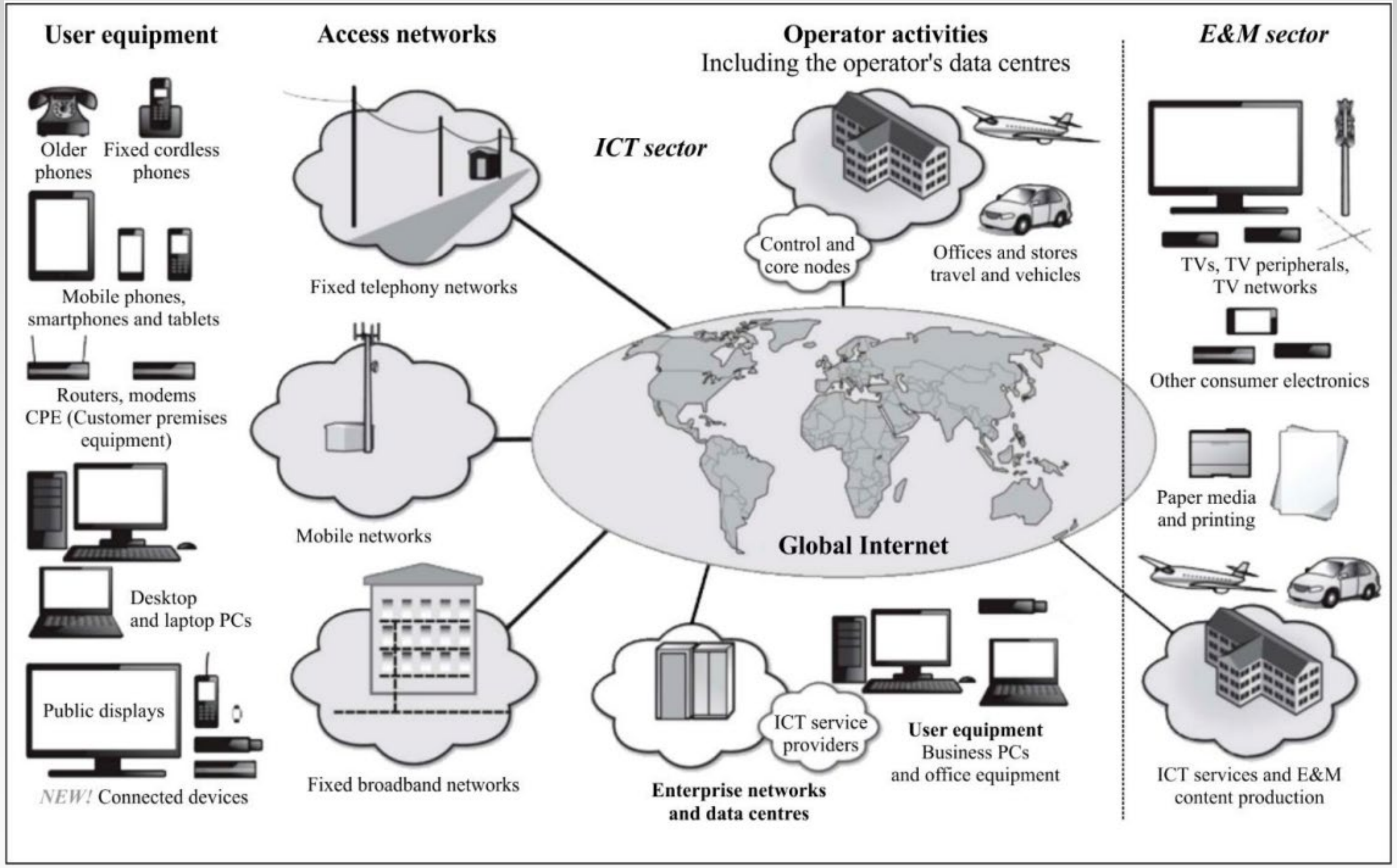
- Analysis of the contribution of ICT to energy consumption
- How ICTs affect the environment
- Average consumption of electric power per bit
- Some data and practical situations
- Procedures for reducing energy consumption in ICT
- Conclusions

Analysis of the contribution of ICT to energy consumption

- The US Semiconductor Industry Association states:
  - Although world power production grows linearly, the demand for electricity from computers does so exponentially
- In the worst case, ICTs could contribute up to **23% of global greenhouse gas emissions** by **2030**
- If the trend continues, the electric power consumption of the large amount of technological equipment **will exceed the world electric power production by 2040**, which would not be enough to power all the computers in the world

- V. Zhirnov ,Semiconductor Industry Association and the Semiconductor Research Corporation, Rebooting the IT Revolution. (2015)
- A.S. Andrae y T. Edler. (2015). On global electricity usage of communication technology: trends to 2030. Challenges, 6(1), 117-157
- A. Burgess, T. Brown. By 2040 there may not be enough power for all our computers, Manufacturer, 17 Aug 2016
- C. Freitag, M. Berners-Lee, K. Widdicks, B. Knowles, G. Blair, A. Friday. The climate impact of ICT: A review of estimates, trends and regulations. arXiv preprint arXiv:2102.02622. (2022)
- J. Malmodin y D. Lundén. (2018). The energy and carbon footprint of the global ICT and E&M sectors 2010–2015. Sustainability, 10(9), 3027

IoT sectors involved in energy consumption



Greenhouse gas emissions trajectories for the information and communication technology sector compatible with the UNFCCC Paris Agreement. International Telecommunication Union (ITU). Recommendation ITU-T L.1470

How ICTs affect the environment

- Direct effect
- Indirect effect
- Tertiary effect

J. Desjardins. (2018). What happens in an internet minute in 2018. Visual Capitalist.
https://www. visualcapitalist. com/internetminute-2018

The direct effect, in first instance, is due to:

- The great proliferation and global **increase in the number** of electronic **devices**, transmission networks and data centers connected to the Internet.

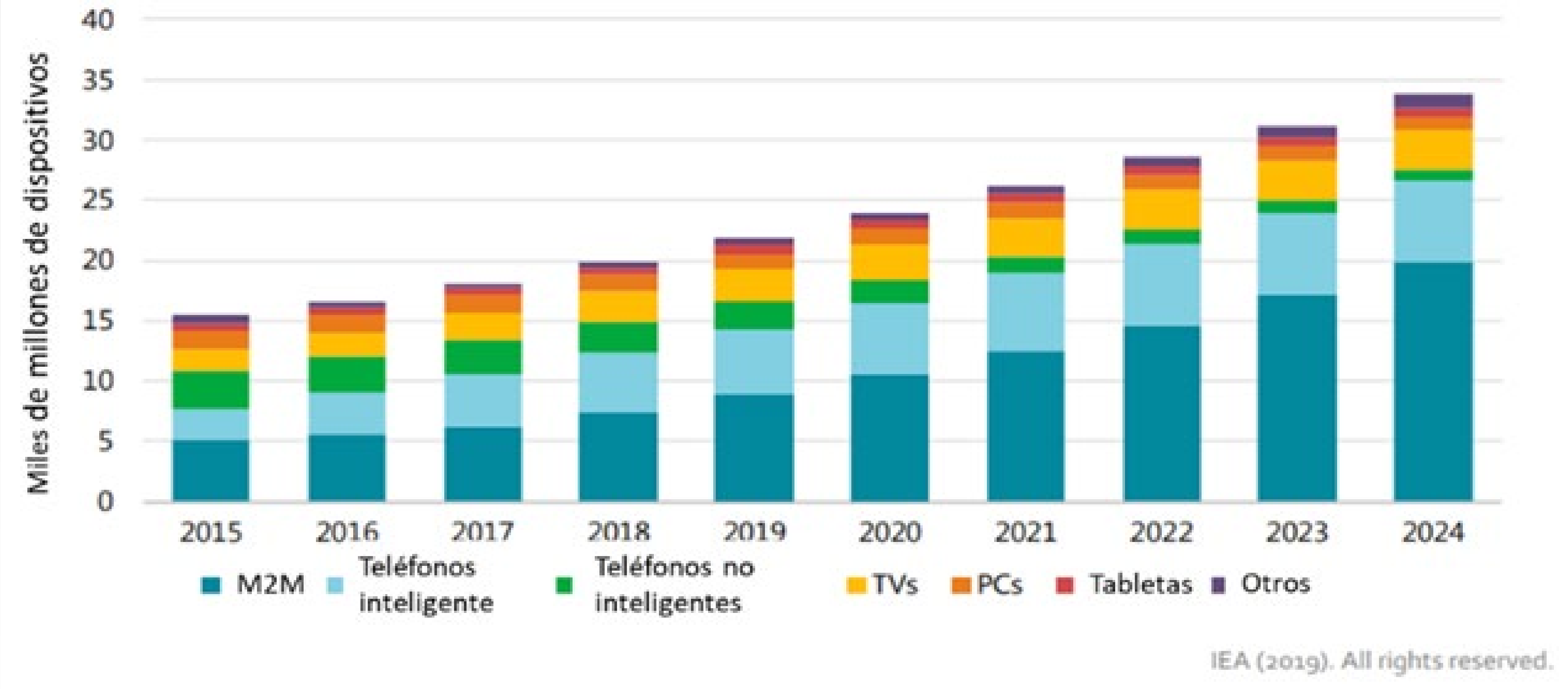


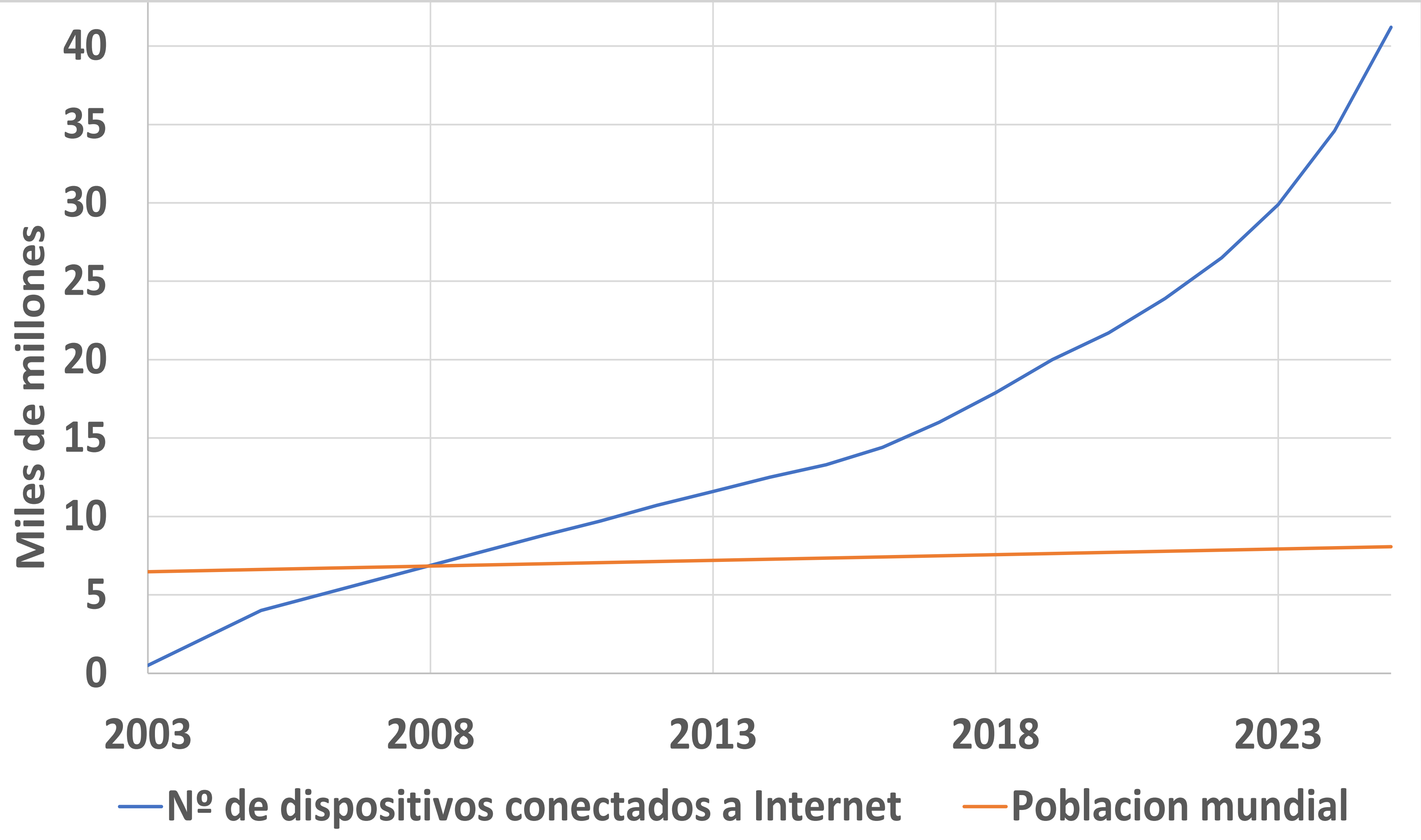
Figure done by *International Energy Agency* based on the work of T. Barnett y colaboradores (2019) and Cisco (2016)

The direct effect is also due to:

- The **increase of applications** that we constantly use both in routine tasks (smartphones, emails, social networks, ...), as well as in traditional computing systems (PCs → HPC)
- New applications that require new devices that, although individually have very low consumption, given their enormous quantity, their global contribution to consumption is very significant

Example of fields of application: IoT

- It can be considered that IoT was born when the number of devices connected to the Internet exceeded the total number of inhabitants of the Earth (end of 2008)

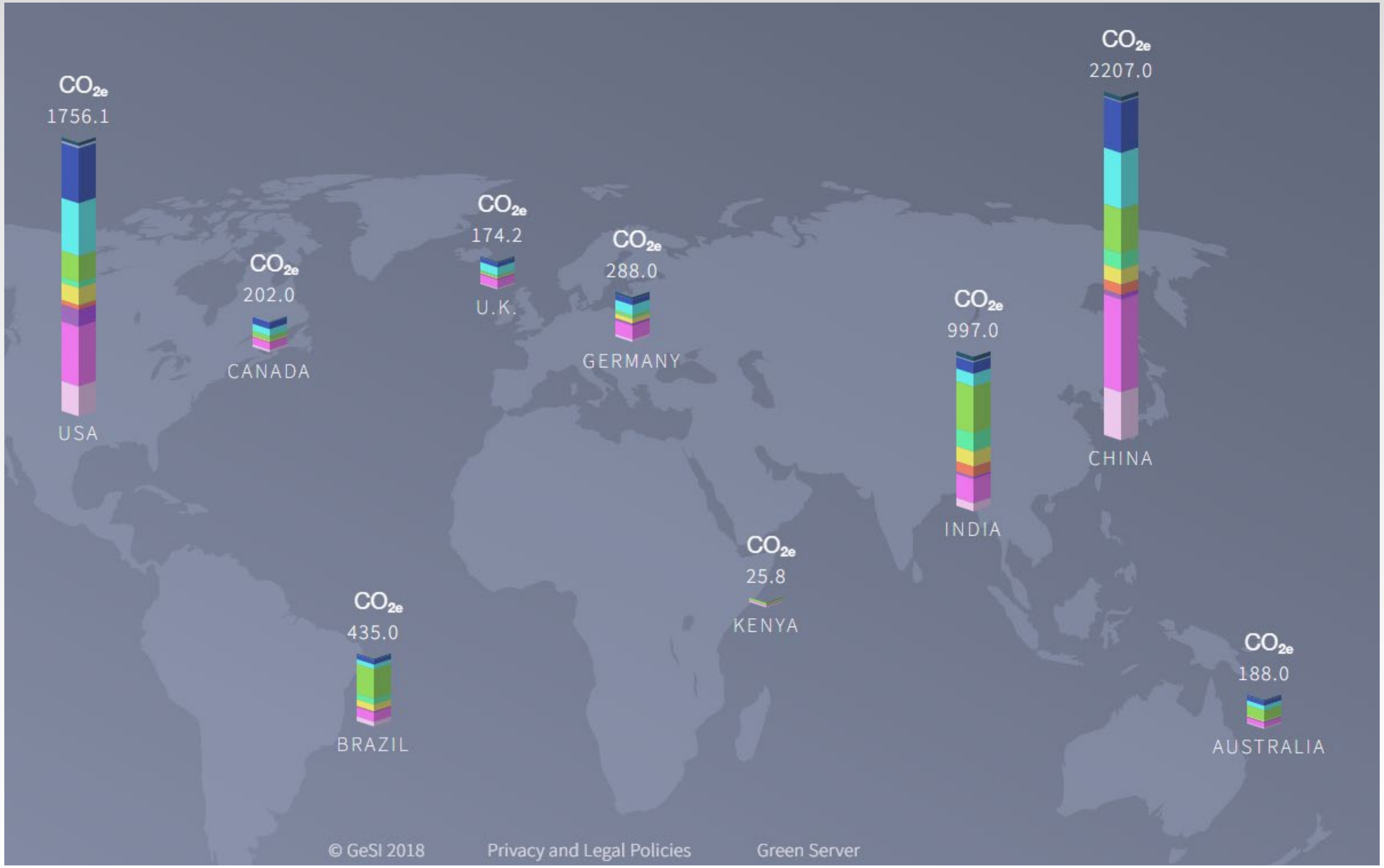


Indirect effect

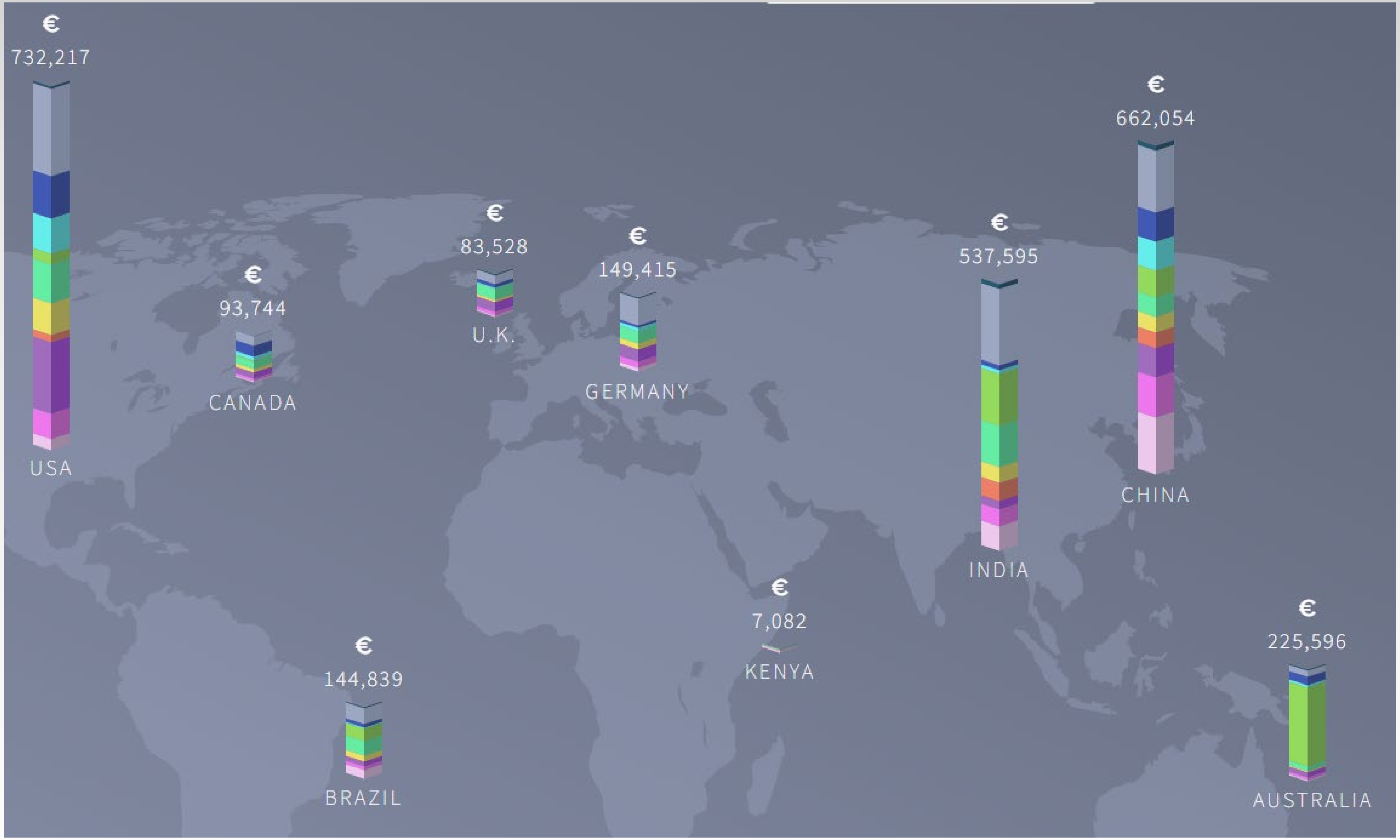
- It is caused by ICT applications that ease the improvement of efficiency and the **reduction of primary energy consumption** in very diverse sectors such as: construction, industry, transport and commerce, by providing intelligent solutions
- **It's good for the environment**
- In other words, the increase in ICT consumption comes to a large extent from its reduction in other sectors, moderating, as a total balance, global consumption
- **Objective:** identify the different ICT applications in buildings, transport and industry that result in a reduction of energy consumption

Indirect effects: reduction of CO_2 emissions in millions of tons, thanks to ICT

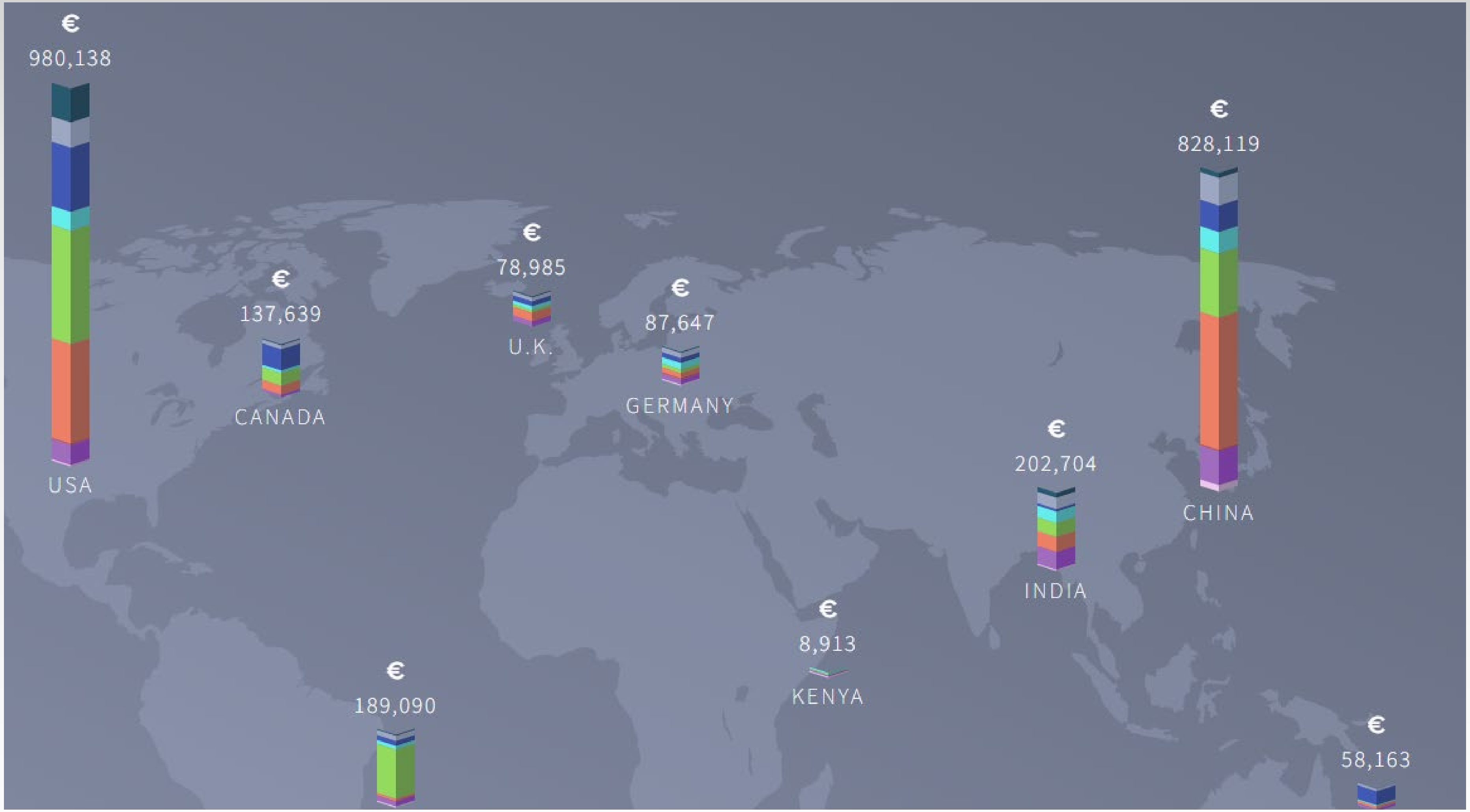
- https://smarter2030.gesi.org/explore-the-data/



ICT- enabled cost-saving (in million €) (not calculated for each use case)



ICT- enabled revenues (in million €)

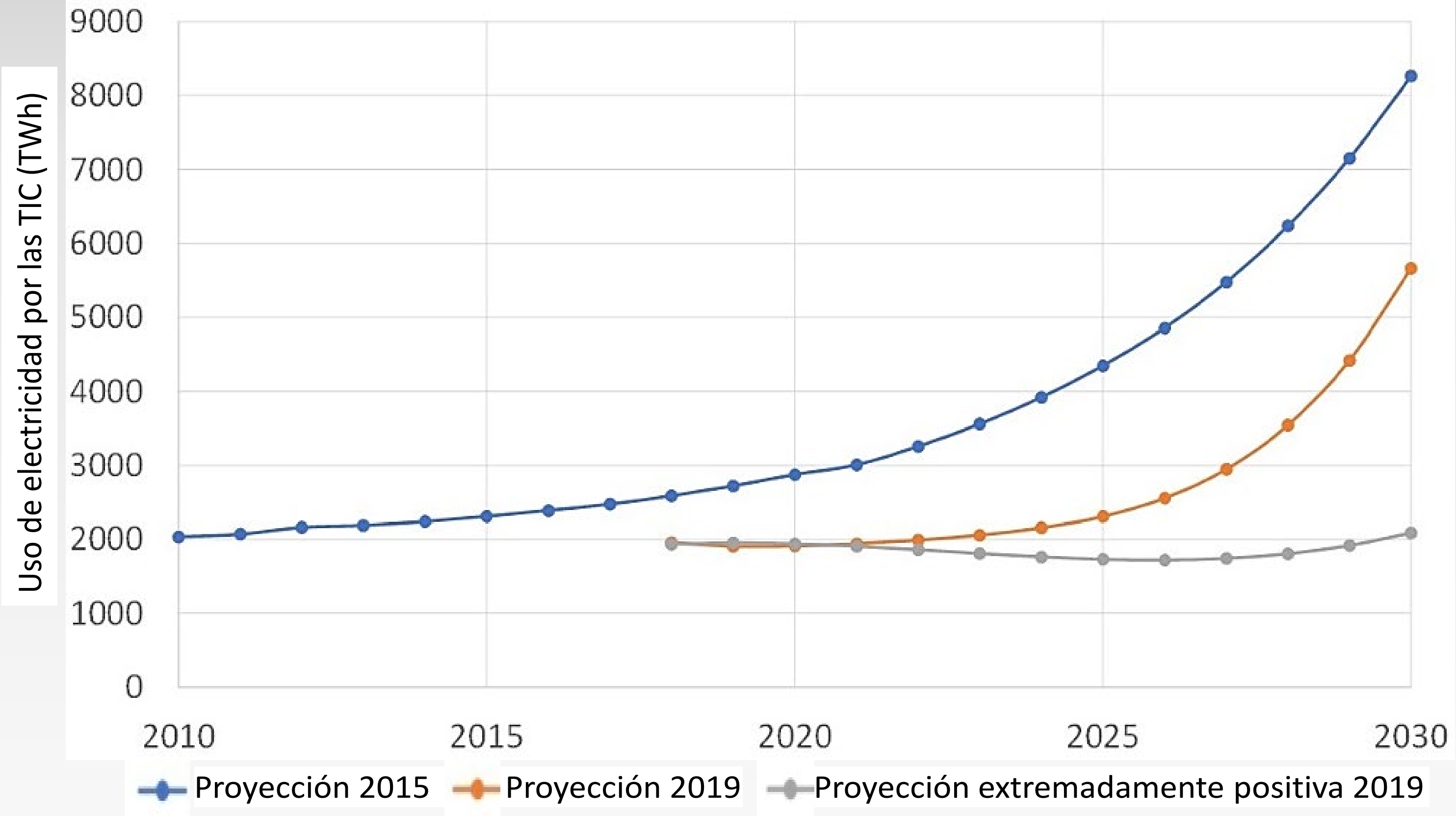


Tertiary effect (rebound effect)

- It is a phenomenon that occurs as ICT services are more useful, cheaper and energy efficient, as this increases our digital lifestyle, which produces a rebound effect: ICT equipment consumes less, but is used much more
- Globally it has a **negative consequence**
- Estimates show that the possible rebound effects due to digitization vary from **10% to 30% of higher power consumption**, data that varies according to the sector, technology and end use

GeSI. Global e-Sustainability Initiative. Accenture strategy SMARTer2030-ICT solutions. (2015) https://smarter2030.gesi.org/downloads/Full_report.pdf

Andrae and Edler projections on the use of electrical energy by ICT in TWh per year (2015 and 2019)



$$P = \frac{E}{t}$$

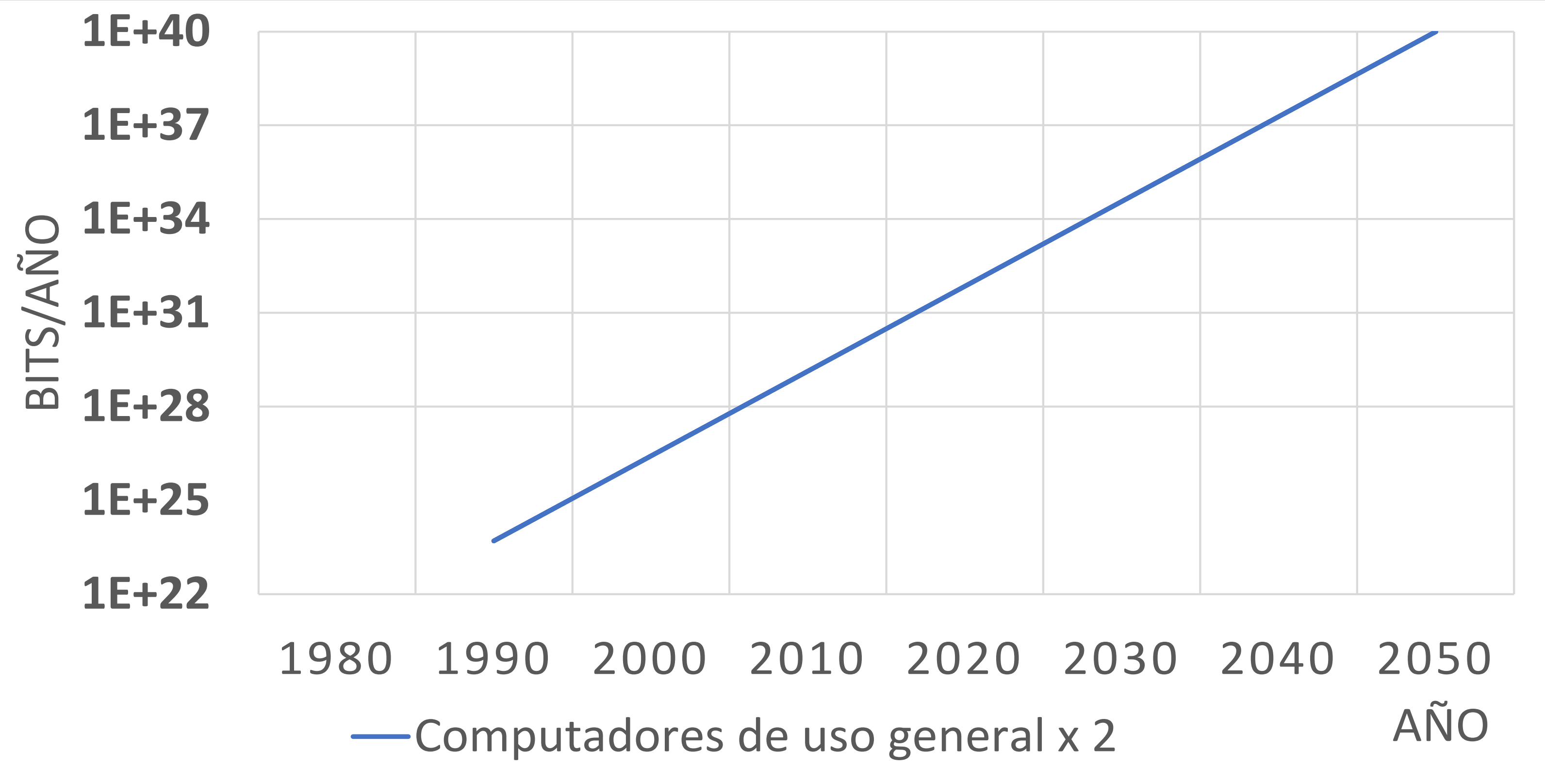
1 W·h = 3,600 Joules

- A.S. Andrae y T. Edler. (2015). On global electricity usage of communication technology: trends to 2030. Challenges, 6(1), 117-157. https://doi.org/10.3390/challe6010117
- A.S. Andrae (2019). Comparison of several simplistic high-level approaches for estimating the global energy and electricity use of ICT networks and data centers. International Journal, 5, 51. DOI: 10.30634/2414-2077.2019.05.06

Measure energy consumption. It is generally estimated:

- In computing during a time *t* (day, year, etc.):
  - Volume of computations (bits, instructions, etc.) x average consumption for each computation
- In information traffic
  - Volume of information transmitted (bits) x Average consumption per bit

Estimates by various authors of the number of bits computed



- F. Roccaforte, F. Giannazzo y G. Greco. (2022, Enero). Ion Implantation Doping in Silicon Carbide and Gallium Nitride Electronic Devices. Micro 2(1) pp. 23-53)
- Semiconductor Industry Association and the Semiconductor Research Corporation, Rebooting the IT Revolution: A Call 547 to Action. (2015)

Theoretical minimum energy consumed per bit. Landauer's principle

- At constant temperature in a physical system it is verified:

$$S_2 - S_1 = \frac{E_{1\to 2}}{T}$$

  - S: entropy, E: Energy, T: absolute temperature
  - Entropy measures disorder or uncertainty
  - The entropy change is equal to the energy exchanged divided by T
  - One-bit processing → entropy decrease → heat dissipation

- Boltzmann entropy (S) formula:

$$S = k_B \cdot ln\, W$$

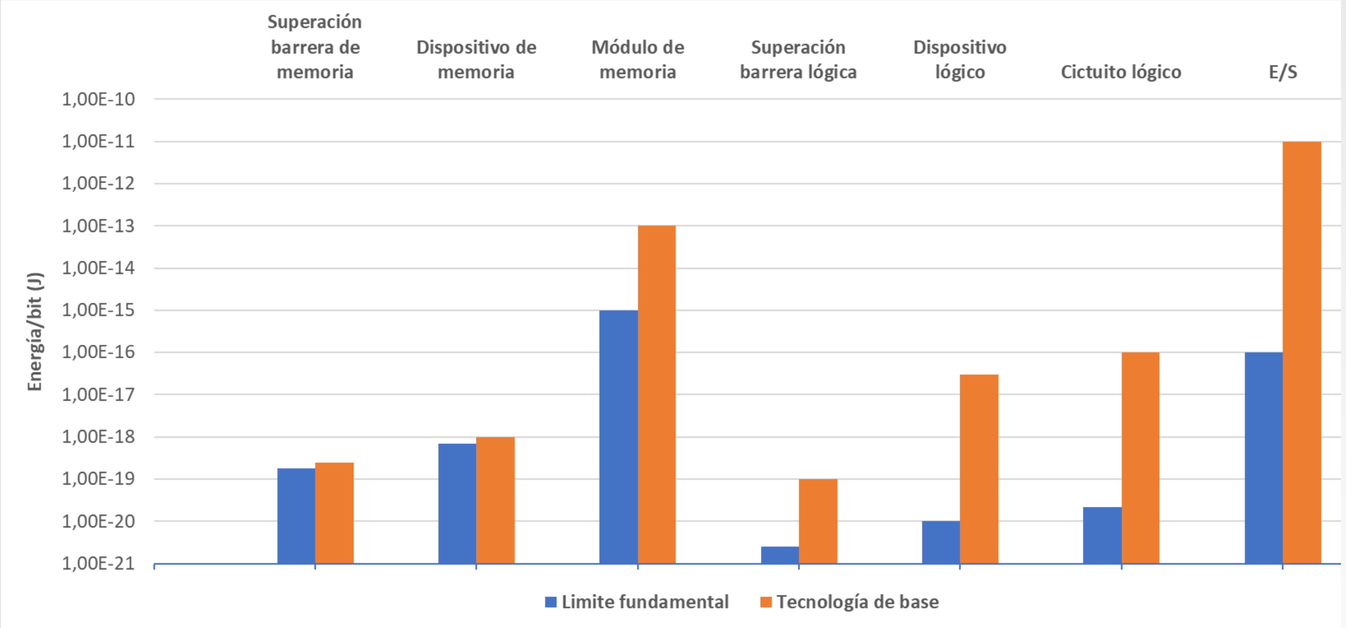
  - $k_B \approx 1.38\times10^{-23}$ J/K; W is the number of states. S= E/T

$$E = k_B \cdot T \cdot ln\, 2$$

- At a temperature = 20º; **energy per bit processed ≈ *$3\cdot10^{-21}$ J/bit***

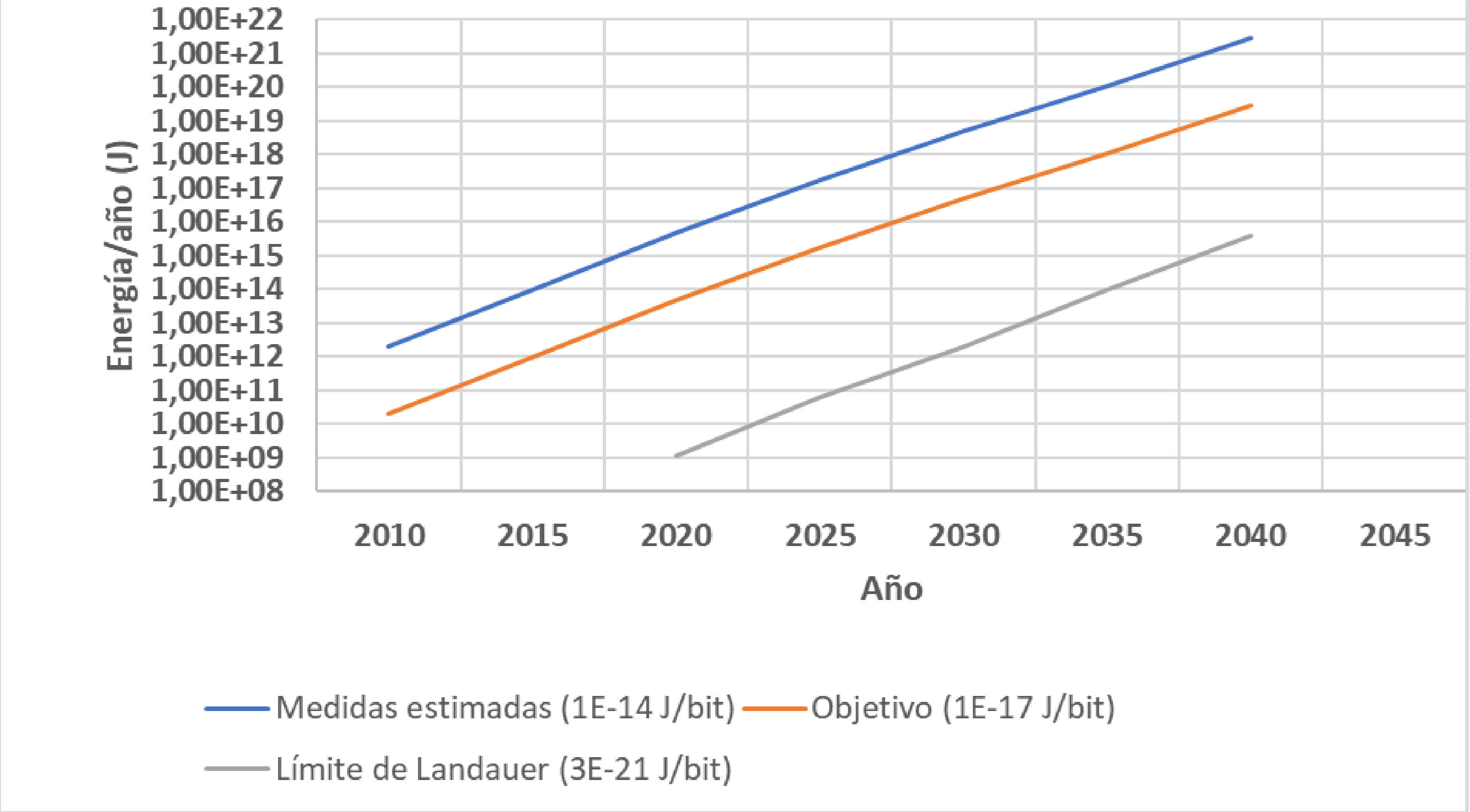
V. Zhirnov et al. published (2014) an article where:

- With real data, they obtain the evolution of the consumption per bit of different binary elements (logic devices and memory elements) considering the consumption of individual transistors and microprocessors and the dynamics of the physical processes in the different components (capacitive, resistive, etc.)
- They conclude:
  - The minimum power required per bit switching is around ≈ **10^{-14} J/bit**
  - Estimated value as an achievable target ≈ **10^{-17} J/bit**



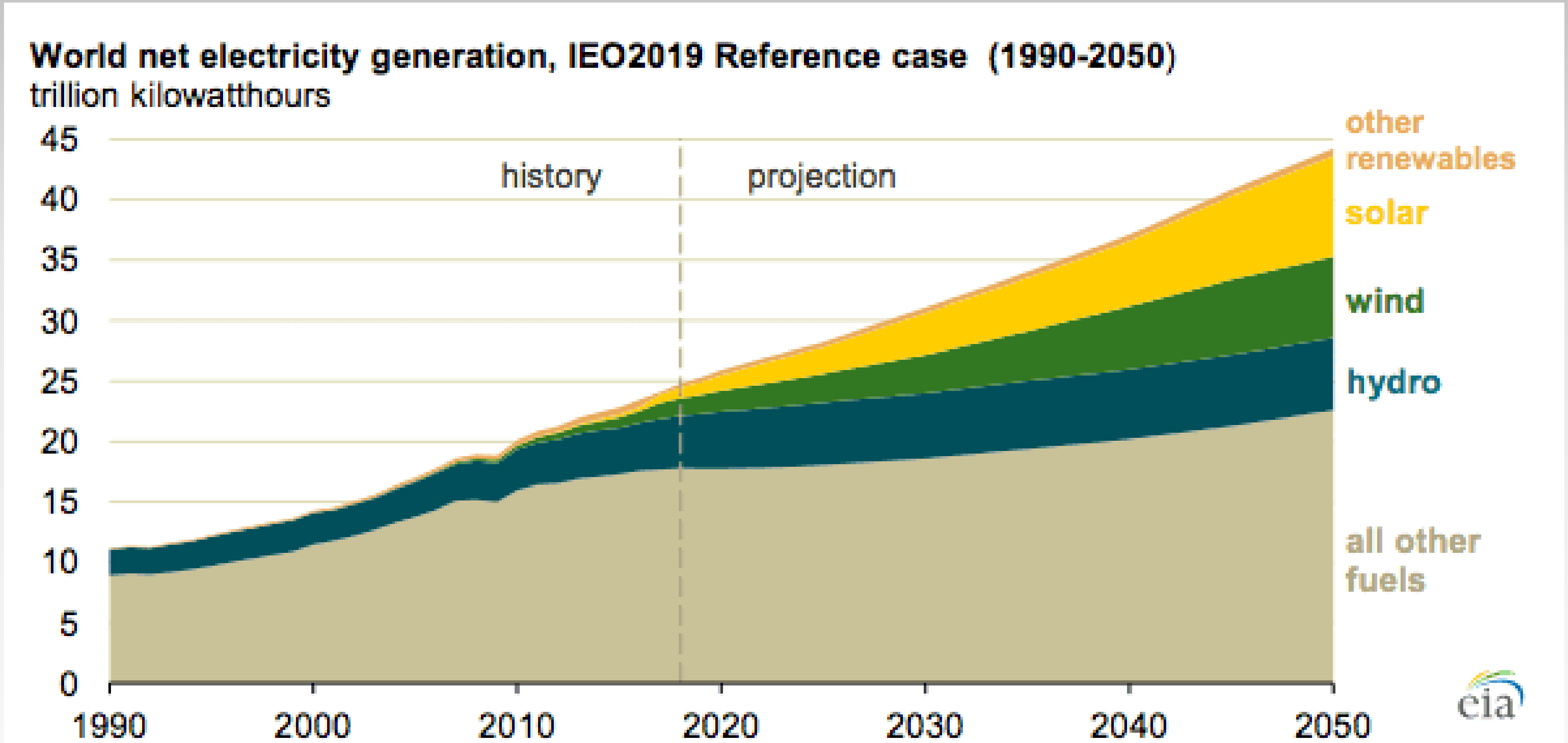
- V. Zhirnov, R. Cavin y L. Gammaitoni. (2014). Minimum energy of computing, fundamental considerations. In ICT-Energy-Concepts Towards Zero-Power Information and Communication Technology. IntechOpen

Global electrical energy consumed by computing in one year



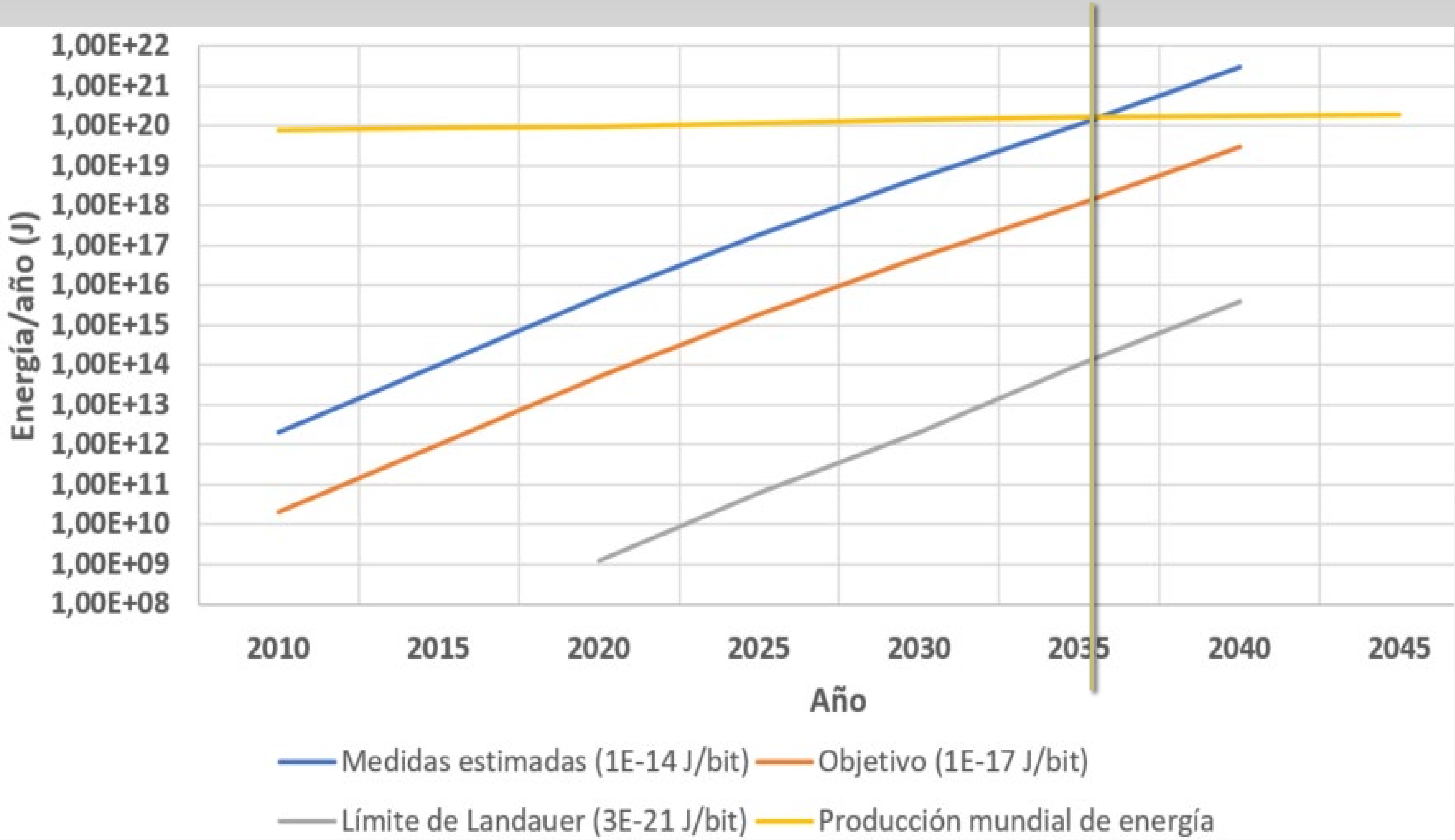
- Value calculated from real data (2014) ≈ **10^{-14} J/bit**
- Zhirnov estimate as a goal to achieve ≈ **10^{-17}J/bit**
- Landauer limit ≈ **$3{\cdot}10^{-21}$ J/bit**

- Semiconductor Industry Association and the Semiconductor Research Corporation, Rebooting the IT Revolution: A Call 547 to Action. (2015)



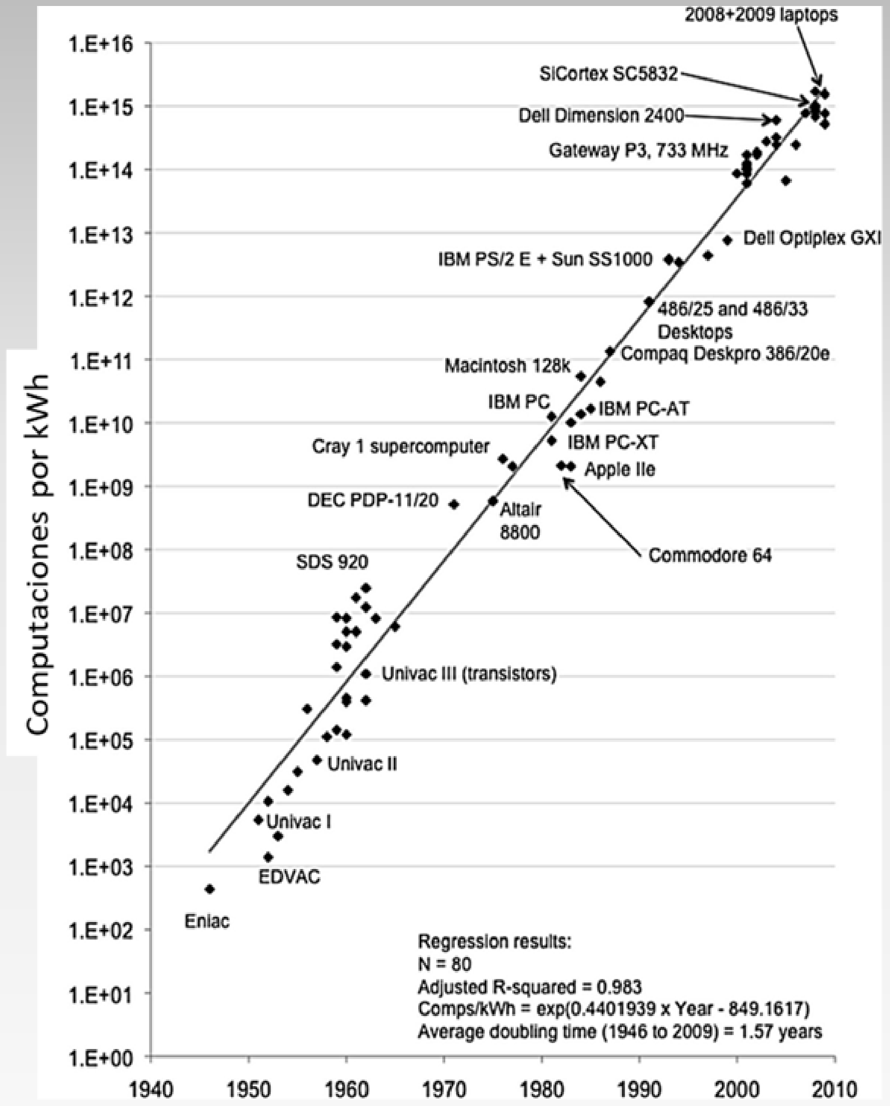
Source: U.S. Energy Information Administration, International Energy Outlook 2019

In the worst case, by 2035 there will not be enough energy…



Koomey's Law (2010)

- The number of computations per joule of energy dissipated doubles approximately every **1.57** years. (2011 → 2.6)
  - This Law was fulfilled with great precision (R^2=98%) with data taken between years 1946 and 2009



Fuente: Jonathan Koomey

- J. Koomey, S. Naffziger. (2015). Moore's Law might be slowing down, but not energy efficiency. IEEE Spectrum, 52(4), 35
- J. Koomey, S. Berard, M. Sanchezn y H. Wong. (2010). Implications of historical trends in the electrical efficiency of computing. IEEE Annals of the History of Computing, 33(3), 46-54

Consequences of Koomey's Law

- If energy efficiency continues to double every 2.6 years, and considering in 2014, a consumption of 10^{-14} J/bit, the Landauer limit will be reached in **2036**
- In reversible computations, the lower limit of power consumption is estimated by the Margolus–Levitin Theorem: $6 \cdot 10^{33}$ operations per second per joule of energy. This limit, applying Koomey's Law, would be reached approximately within 125 years (**2145**)

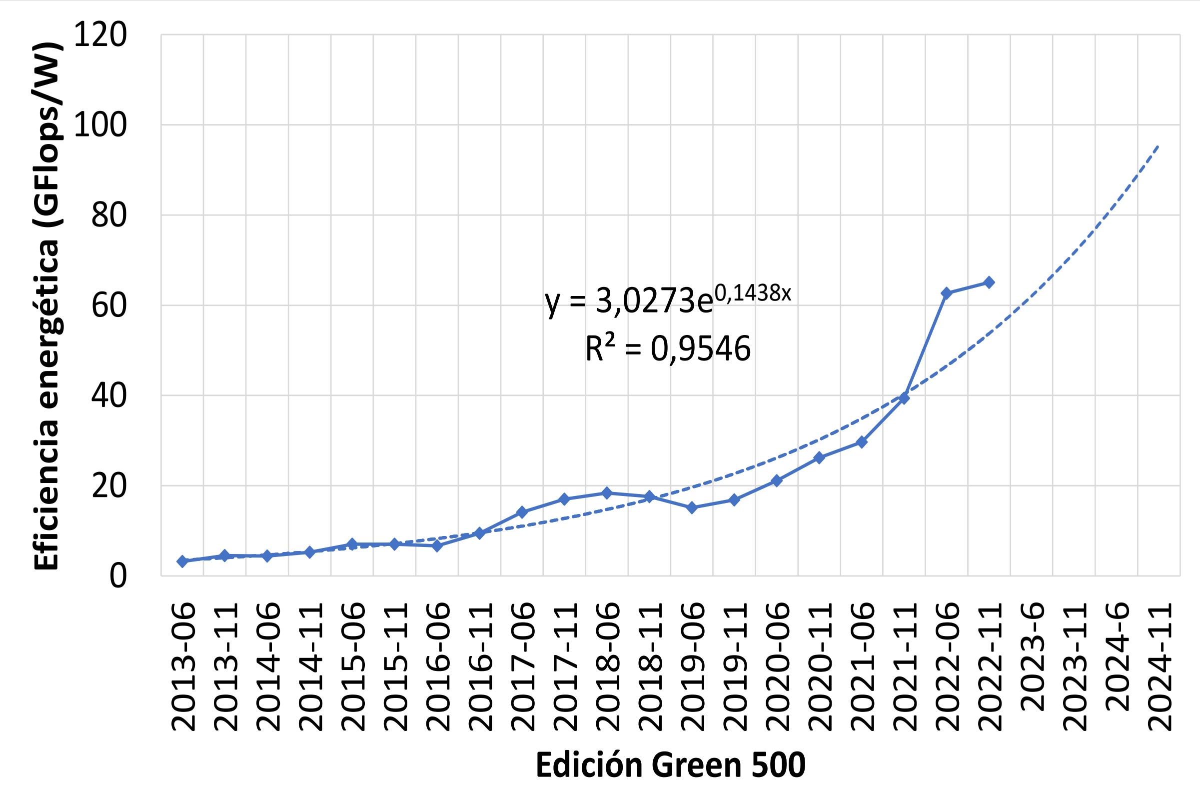
Energy efficiency and Green500 data

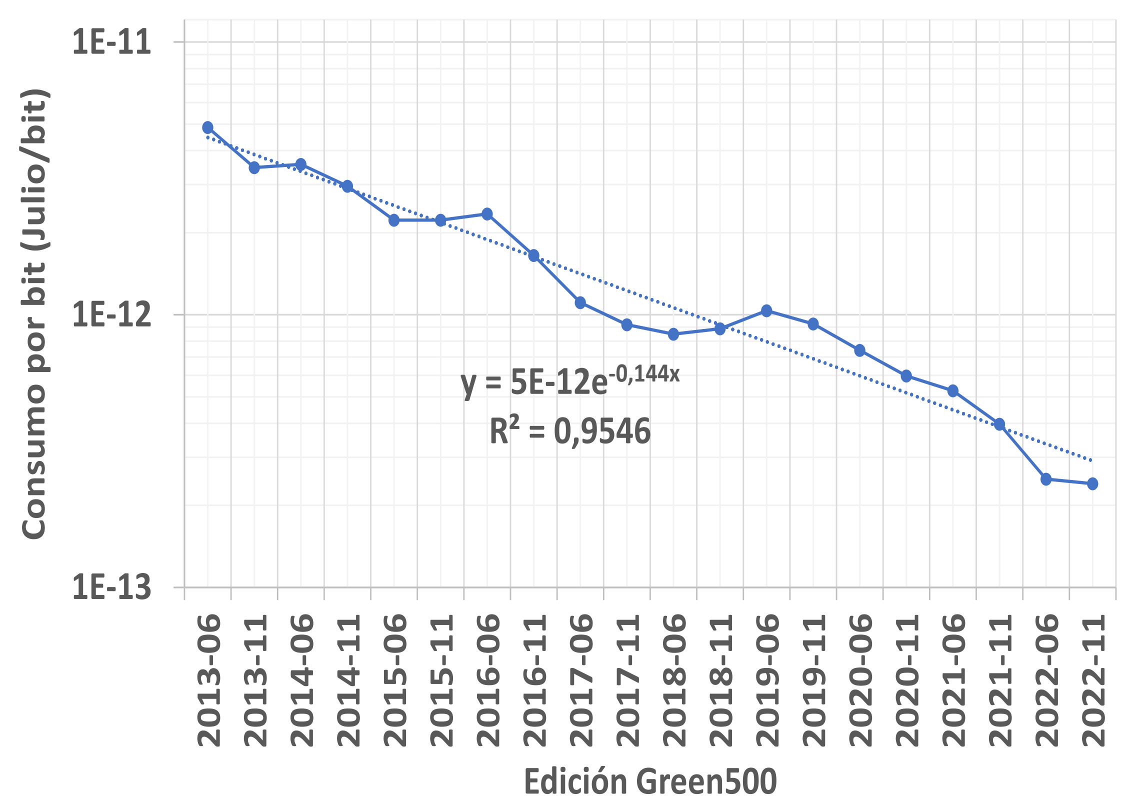
$$EE = \frac{Computer\ performance}{Electrical\ power} = \frac{Instructions/s}{Energy/s} = \frac{R_{max}}{P} = \frac{FLOPS}{Watts}$$

Green500 Data

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	255	**Henri** - ThinkSystem SR670 V2, Intel Xeon Platinum 8362 32C 2.8GHz, NVIDIA H100 80GB PCIe, Infiniband HDR, Lenovo Flatiron Institute United States	8,288	2.88	44	65.396
2	34	**Frontier TDS** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	120,832	19.20	309	62.684

Evolution over the years of energy efficiency (Flops/W) and consumption per bit of the first supercomputer on the Green500 lists





- Own elaboration based on data obtained from: https://www.top500.org/lists/green500/

Comparison

	2014 (J/bit)	2023 (J/bit) Landauer
Our calculations:	$3.5 \cdot 10^{-12}$	$2.9 \cdot 10^{-13}$
Zhirnov estimate:	10^{-14}	$8.3 \cdot 10^{-16}$
Zhirnov objetive:	10^{-17}	$8.3 \cdot 10^{-19}$
Landauer limit:	$3 \cdot 10^{-21}$	$2 \cdot 10^{-21}$

Analysis of energy efficiency in PCs and processors

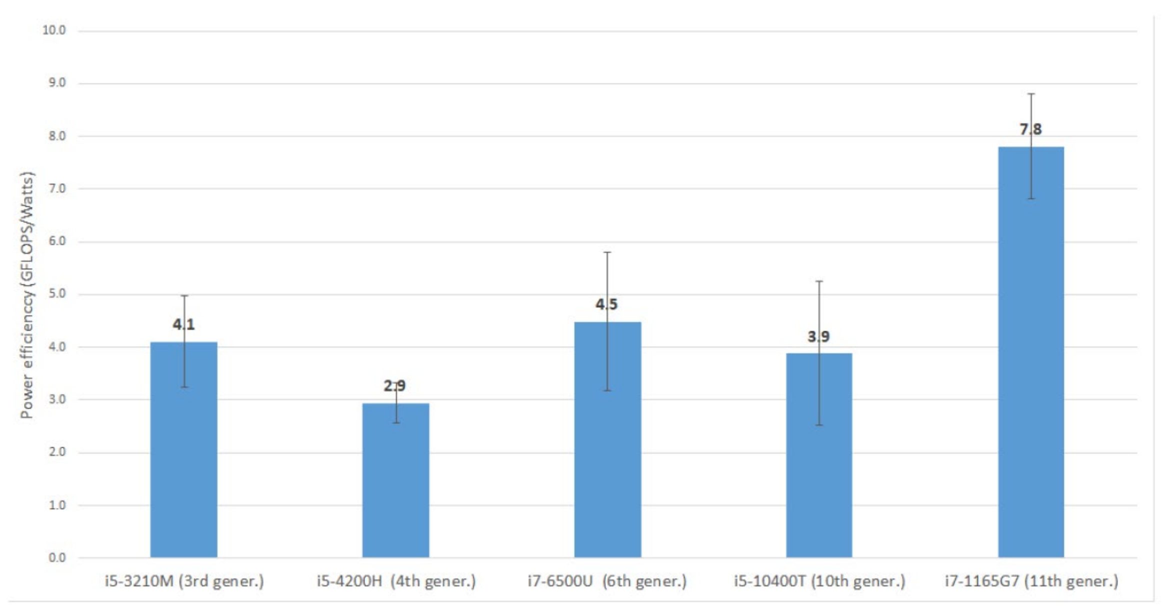


Figure 15. Power efficiency of the microprocessors measured during the Linpack run.

Table. 12. Position, in terms of power efficiency, of the under test systems within the Green500 list [34].

Platform	Position in Green500	Pos. in TOP500	Rmax TFlops	Power (KW)	Power Efficiency (GFlops/W)
MN-3 - MN-Core Server, Xeon	1	301	2,181.2	55	39.379
SSC-21 Scalable Module - Apollo 6500 Gen10 plus, AMD.	2	291	2,274.1	103	33.983
Etc.					
***SUT5. ASUS 2**			**0.148**	**0.033**	**4.434**
JOLIOT-CURIE SKL - Bull Sequana.	79	113	4,065.6	917	4.434
Etc.					
***SUT2. Toshiba 1**			**0.064**	**0.036**	**1.780**
Nuri - Cray XC40, Xeon	155	251	2,395.7	1,359	1.762
Etc.					
***SUT4. HP**			**0.112**	**0.069**	**1.633**
HKVDPSystem – Sugon TC6000, Xeon	162	355	1,979.0	1,216	1.627
***SUT3. ASUS**			**0.075**	**0.093**	**0.799**
***SUT1.SONY**			**0.036**	**0.046**	**0.792**
Thunder – SGI ICE X, Xeon	173	156	3,126.2	4,820	0.649
Etc.					

* System Under Test

- Prieto, B., Escobar, J. J., Gómez-López, J. C., Díaz, A. F., & Lampert, T. (2022). Energy Efficiency of Personal Computers: A Comparative Analysis. Sustainability, 14(19), 12829

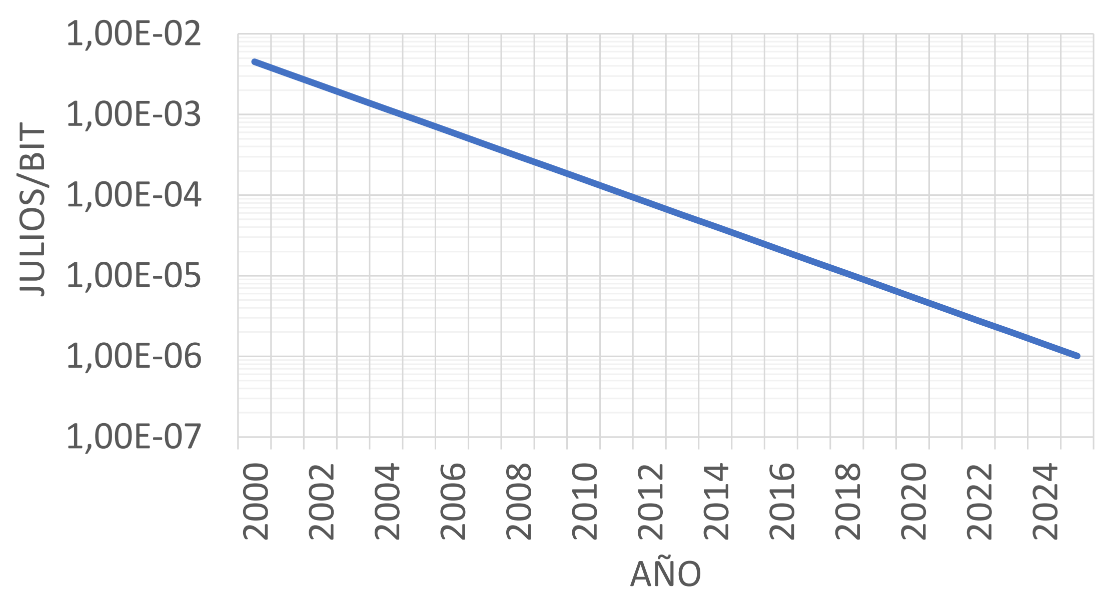
Estimation of energy consumption in Internet traffic

- It is very difficult since it depends on several different factors:
  - The transmission channel (atmosphere, cable, optical fiber, etc.)
  - The distance between emitter and receiver
  - The data flow (transmission speed)
  - In the case of a transmitted message, it also depends on the coding used, the type of modulation, etc.

- J. Aslan, K. Mayers, J. G. Koomey y C. France. (2018). Electricity intensity of internet data transmission: Untangling the estimates. Journal of industrial ecology, 22(4), 785-798

Estimates from Aslan et al.

- **Consumption per bit is halved approximately every 2 years** (in processing every 2.6 years)
- The average transmission energy of 1 bit in 2020 through the Internet was of the order of **$2.77 \cdot 10^{-6}$ J/bit**, which means:
  - ≈ $2.8\ 10^{11}$ times greater than that of one-bit processing (10^{-17} J/b, Zhirnov)
  - ≈ 10^{7} times higher if we take into account the energy efficiency of supercomputers obtained by us ($2 \cdot 10^{-13}$ J/bit)



• J. Aslan, K. Mayers, J. G. Koomey y C. France. (2018). Electricity intensity of internet data transmission: Untangling the estimates. Journal of industrial ecology, 22(4), 785-798

Some data and practical situations

Queues about confidentiality in emails from the University of Granada

- **Official text recommended by the General Secretariat of the UGR**:
  - *Este mensaje ha sido generado desde una cuenta de la Universidad de Granada para los fines propios de la institución. Su contenido se considera información confidencial, por lo que queda informado de que su utilización, divulgación o copia sin autorización no está permitida. Si usted ha recibido indebidamente el correo le rogamos que advierta de ello por esta misma vía al remitente y proceda a su eliminación. Cualquier incidencia relacionada con la recepción de nuestros correos electrónicos y en particular las relativas a la seguridad y confidencialidad pueden ser comunicadas a protecciondedatos@ugr.es. Para más información al respecto, puede consultar nuestra política de privacidad*
  - *This message has been generated from an e-mail address of the University of Granada for the institution's own purposes. Its content is considered confidential information, so it is informed that its unauthorized use, disclosure or copying is not permitted. If you have improperly received the email please warn the sender of this, same way and proceed to its removal. Any incident related to the receipt of our emails and in particular those related to security and confidentiality may be communicated to protecciondedatos@ugr.es. For more information, please refer to our privacy policy.*

1,285 characters (not counting links)

UGR emails (2021-2022):

- 6,819,020 messages sent to the Internet from central or departmental mail servers
- 34,671,665 messages coming from the Internet and delivered in the UGR PAS/PDI inboxes
- 22,278,138 messages coming from the Internet and delivered in the inboxes of UGR students

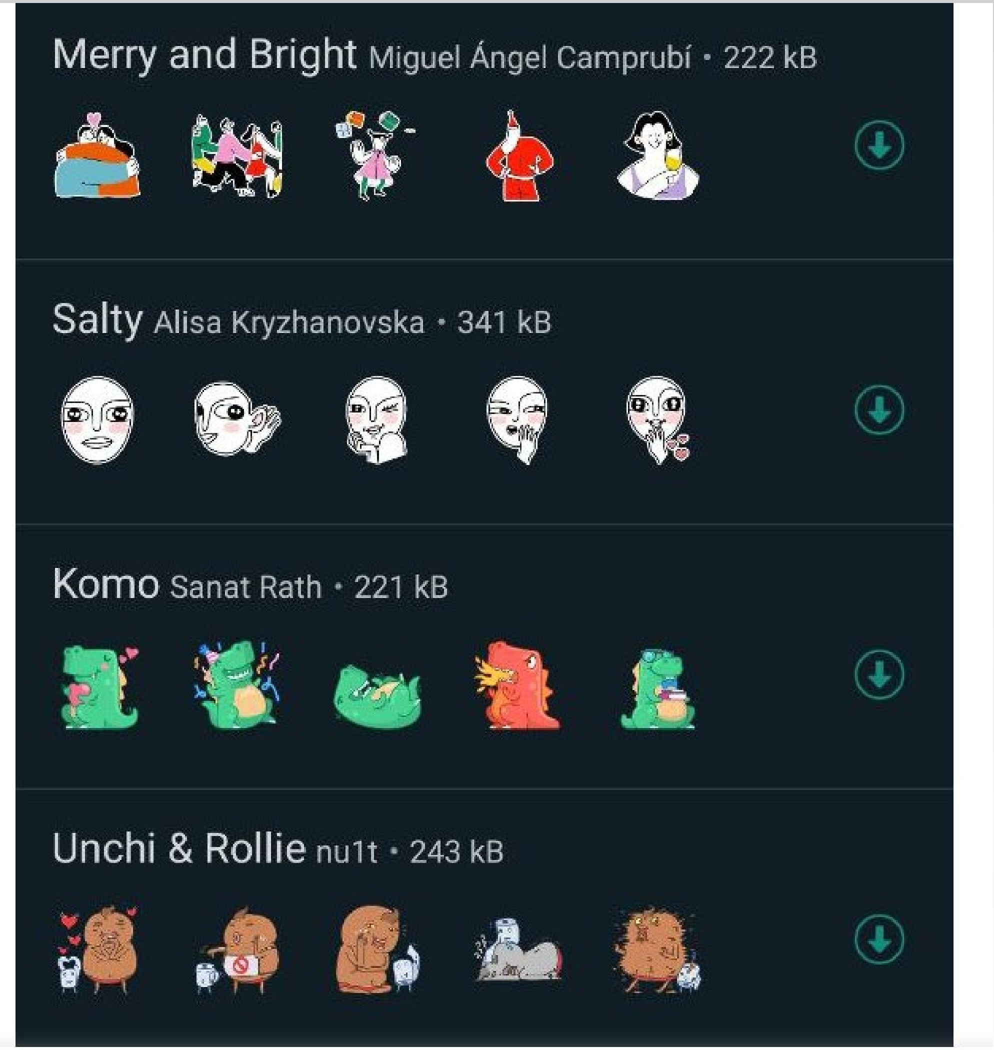
		Messages sent	Messages received
Nº characters/email		1,285.00	1,285.00
Nº bits (UTF8)/character		16	16
Annual messages		6,819,020	56,949,803
Nº of annual bits		1.40199E+11	1.17089E+12
Energy per transmitted bit (2022)	Jules/bit	3.00E-06	3.00E-06
Total energy	Jules	4.21E+05	3.51E+06
Total energy	KWh	1.17E+02	9.76E+02
Electric MIX	Kg CO2/KWh	0.351	0.351
Carbon footprint for emails queue	KgCO2	4.10E+01	3.42E+02

Half a ton of CO2 per year!

Emotics, emojis and stickers (WhatsApp)

- Emoji: 4 bytes
- Stickers ≈ 50 KB, <100 KB → 0.120 g of CO_2
- 100,000 million WhatsApp daily
- If 1 in 40 is a sticker of 50KB → **109,500 T of CO_2 per year**

Para que os hagáis una idea, **en el caso del *pack* "Chummy Chum Chums", oficial de WhatsApp, sus 16 *stickers* solo pesan 3,7 megas**, por lo que están muy lejos de ese máximo que tenemos disponible (1MB/pegatina).



La contaminación silenciosa

ALBERTO PRIETO ESPINOSA
Academia de Ciencias de Granada

Todo proceso de transferencia o movimiento de datos, entre móviles, computadores, etc, consume energía, y muchos dispositivos, entre los que se encuentran los supercomputadores, están funcionando las 24 horas del día, siendo los consumos muy altos

Uno de los retos más importantes de la sociedad actual es reducir el consumo de energía con el objeto de mantener o hacer posible la sostenibilidad de nuestro planeta. Por hacer referencia a nuestro contexto, la Unión Europea tiene como una

y de los programas que es mucho mayor. Así, cuando enviamos un correo electrónico, además del consumo inherente a la transmisión de los bits hay que añadir el del programa que me permite editar, enviar, recibir y visualizar los emails.

Hasta ahora los parámetros que se utilizaban para medir las prestaciones de un

A. Prieto. La contaminación silenciosa. Ideal, 17/12/2020, p. 24

ChatGPT Training

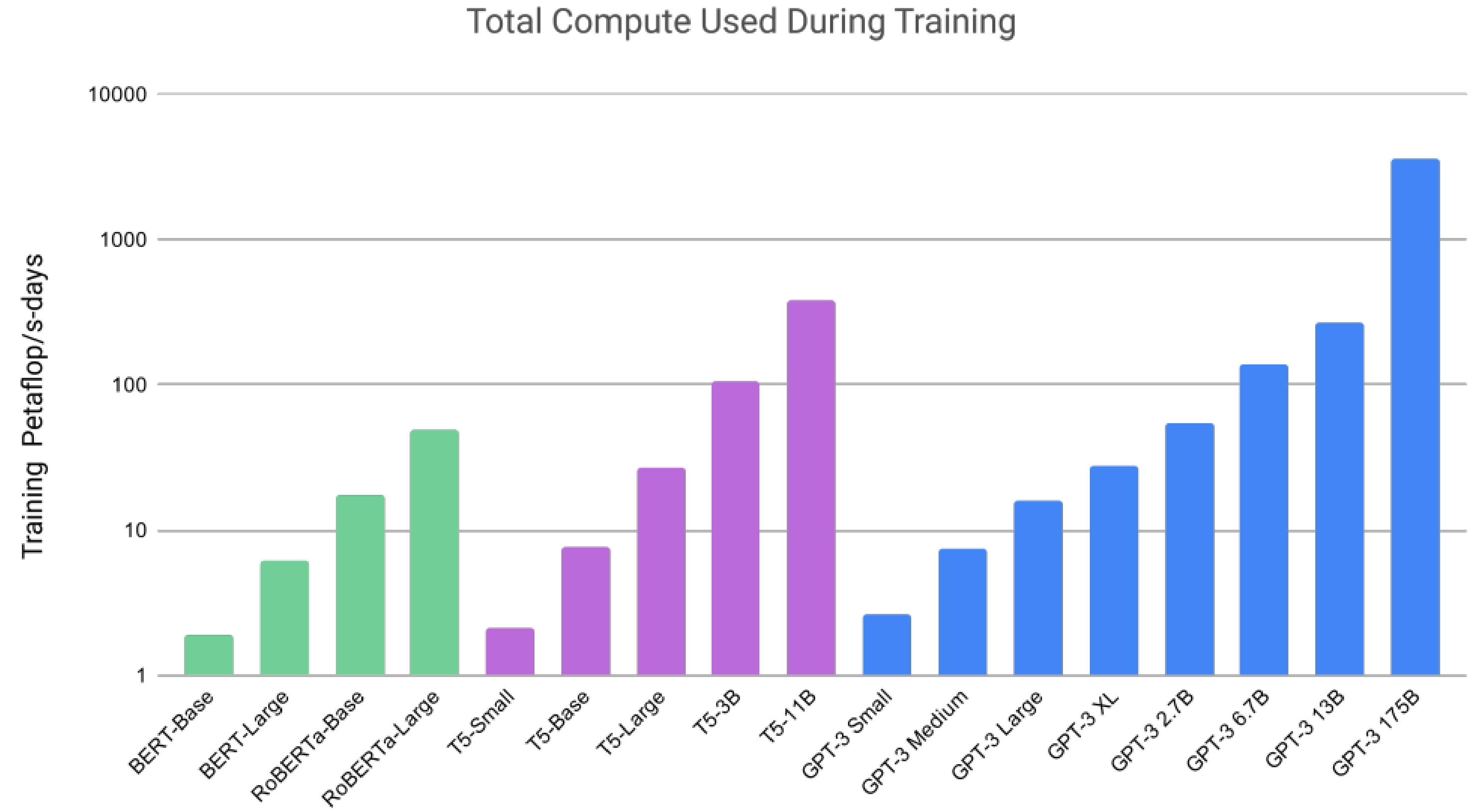


Figure 2.2: Total compute used during training. Based on the analysis in Scaling Laws For Neural Language Models [KMH+20] we train much larger models on many fewer tokens than is typical. As a consequence, although GPT-3 3B is almost 10x larger than RoBERTa-Large (355M params), both models took roughly 50 petaflop/s-days of compute during pre-training. Methodology for these calculations can be found in Appendix D.

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in neural information processing systems, 33, 1877-1901. Cited by 12,240

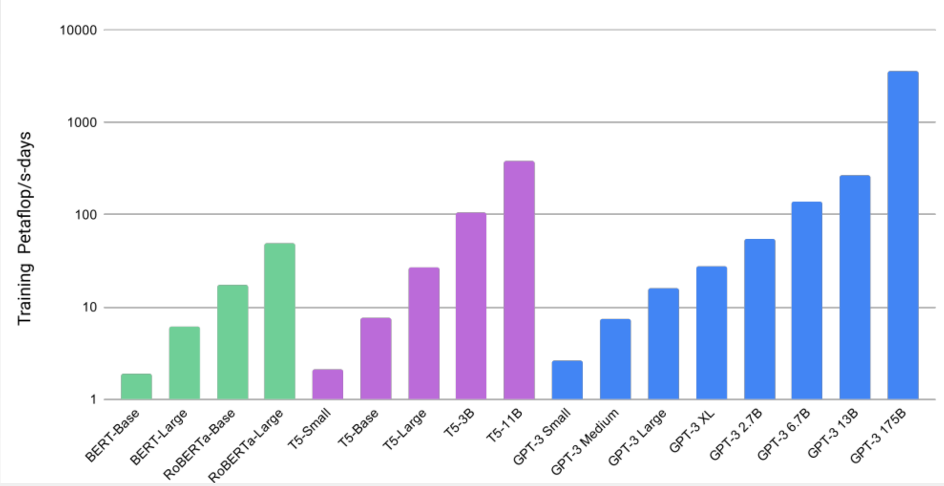
If training were done on the first supercomputer of the TOP 500 and Green500 (June 2023) ...

Green500 Data

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	255	**Henri** - ThinkSystem SR670 V2, Intel Xeon Platinum 8362 32C 2.8GHz, NVIDIA H100 80GB PCIe, Infiniband HDR, Lenovo Flatiron Institute United States	8,288	2.88	44	65.396
6	1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, HPE DOE/SC/Oak Ridge National Laboratory United States	8,699,904	1,194.00	22,703	52.592

ChatGPT training

	1st TOP500 (June 2023)	1st Green500 (June 2023)
	Frontier	Henri
PetaFlops-day	3,000.00	3,000.00
PetaFlops-s	2.59E+08	2.59E+08
Gflops-s	2.59E+14	2.59E+14
Energy Efficiency (GFlops/watts)	52.592	65.396
Energy (w·s = J)	4.93E+12	3.96E+12
Energy (KW·h)	1.37E+06	1.10E+06
MIX (Kg CO2/KWh)	0.351	0.351
Carbon Footprint (Kg CO2)	480,529.36	386,445.65
Rmax (PFlops)	1,194.00	2.88
Exection time (dd:hh:mm:ss)	02:12:18:05	06:16:00:00



480.5 ton of CO2 (2 dd 12 hh) vs 386.4 ton (6 dd 16 hh)

ChatGPT access once trained

6.3 Energy Usage

Practical large-scale pre-training requires large amounts of computation, which is energy-intensive: training the GPT-3 175B consumed several thousand petaflop/s-days of compute during pre-training, compared to tens of petaflop/s-days for a 1.5B parameter GPT-2 model (Figure 2.2). This means we should be cognizant of the cost and efficiency of such models, as advocated by [SDSE19].

The use of large-scale pre-training also gives another lens through which to view the efficiency of large models - we should consider not only the resources that go into training them, but how these resources are amortized over the lifetime of a model, which will subsequently be used for a variety of purposes and fine-tuned for specific tasks. Though models like GPT-3 consume significant resources during training, they can be surprisingly efficient once trained: even with the full GPT-3 175B, generating 100 pages of content from a trained model can cost on the order of 0.4 kW-hr, or only a few cents in energy costs. Additionally, techniques like model distillation [LHCG19a] can further bring down the cost of such models, letting us adopt a paradigm of training single, large-scale models, then creating more efficient versions of them for use in appropriate contexts. Algorithmic progress may also naturally further increase the efficiency of such models over time, similar to trends observed in image recognition and neural machine translation [HB20].

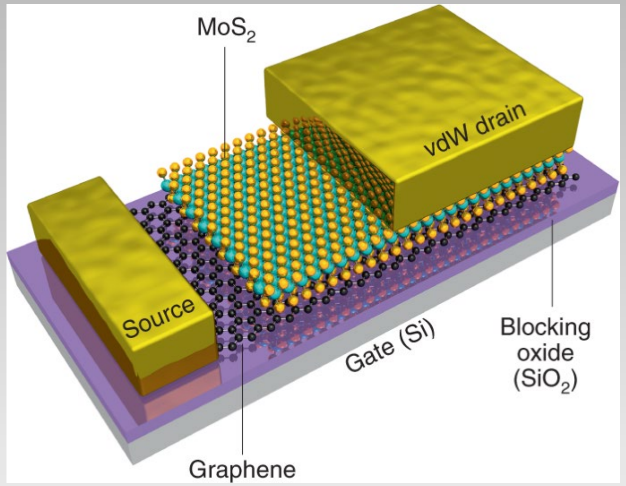
0.4 KW·h x 0.351 = 0.14 Kg CO_2 vs 450 T CO_2

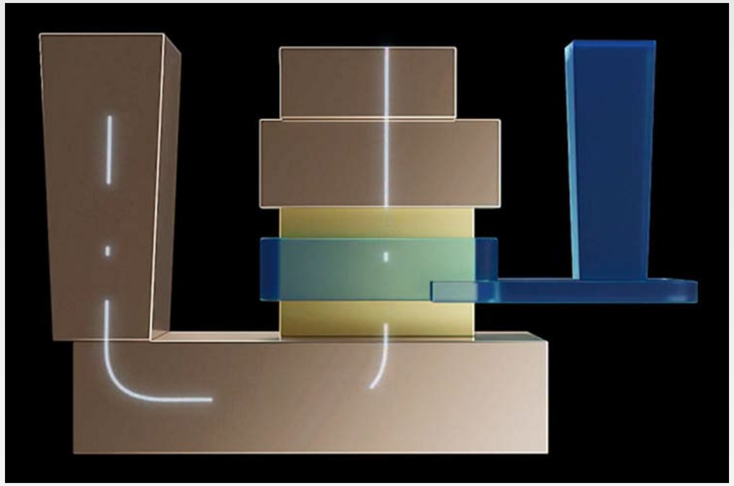
ICT energy consumption reduction procedures

- Fortunately, the most pessimistic forecasts regarding power consumption associated with ICTs are not being fulfilled, being lower than expected
- The ICT industry is aware of the problem and financial resources and active policies are being invested to reduce the increase in consumption, both in the manufacture of new products and in the consumption inherent to their use
- Actions to reduce consumption:
  - **A. Technological improvements in electronic components and devices**
  - **B. Management and planning of the use of resources**
  - **C. Scale changes**

A. Technological improvements in electronic components and devices, …

- Changes in the devices and in the internal architecture of the microchips
  - IBM prototype of IC that makes it possible to vertically stack the transistors
  - Increased integration density and reduced energy consumption (estimated up to **85% less**)





IBM's VTFET with a vertical channel (yellow) and gate-all-around (blue). Contacts are brown and the white line shows current flow.

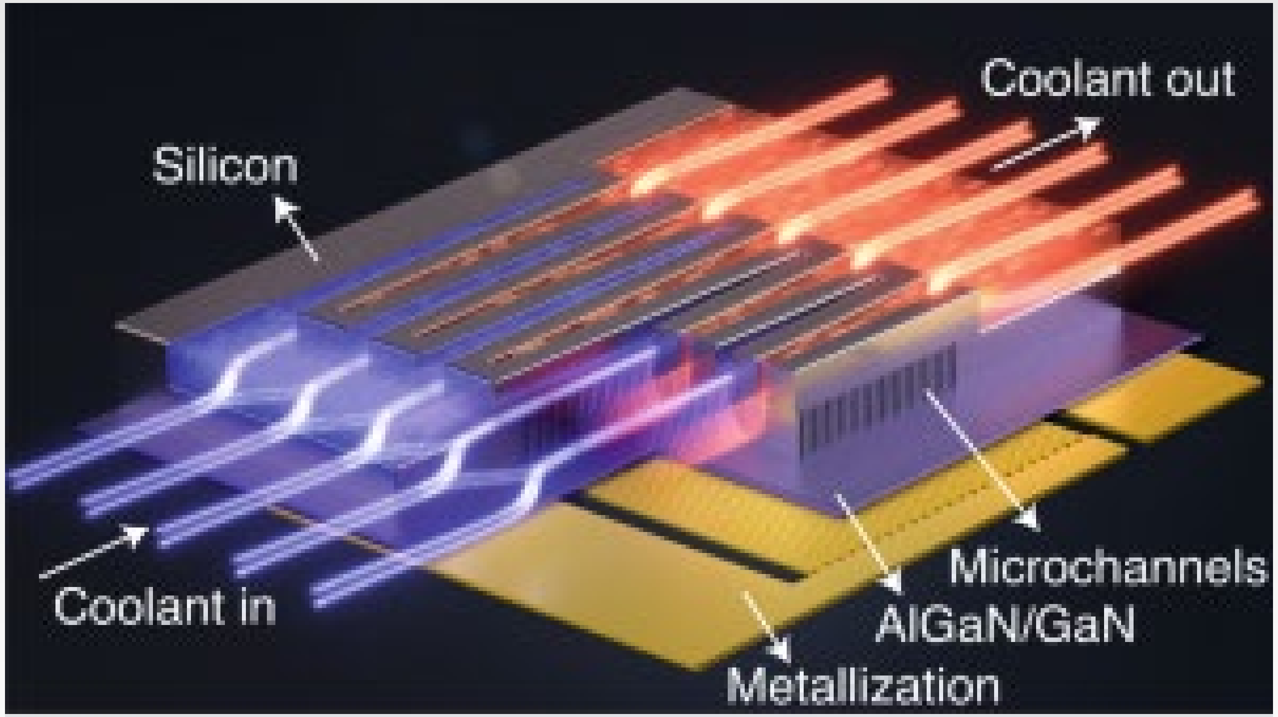
- Zhang, J., Gao, F., & Hu, P. (2021). A vertical transistor with a sub-1-nm channel. Nature Electronics, 4(5), 325-325
- Steve Bush, (14 diciembre 2021) IBM beats finFETs with vertical CMOS at IEDM. Electronics Weekly.com

… technological improvements in electronic components and devices …

- Inclusion of power management functions within the CPUs, which depending on the workload, dynamically switch between different power states (standby mode, for example)
- Development of processors for specific use or functions, such as GPUs and TPUs
- AC/DC switching power supplies. Introduction of new materials, such as gallium nitride and silicon carbide, which allow designs at higher frequencies

… technological improvements in electronic components and devices …

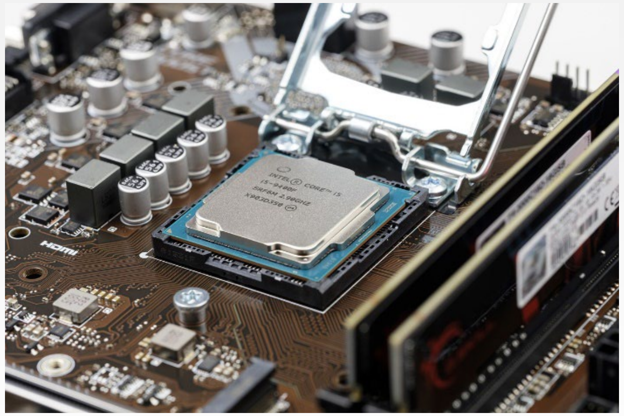
- Direct integration into the chips of microfluidic cooling systems that replace external fans



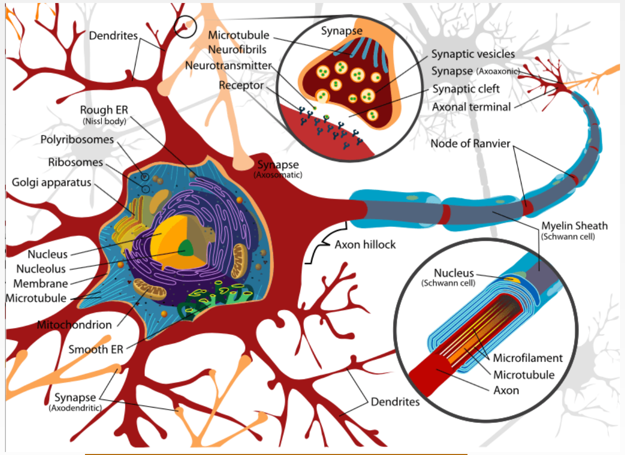
- Varnava, C. Chips cool off with integrated microfluidics. Nat Electron 3, 583 (2020). https://doi.org/10.1038/s41928-020-00494-5

... technological improvements in electronic components and devices

- Changes in the technology of other devices. HDD → SSD, power consumption reduction more than **50%**
- Neuromorphic computing
  - The human brain is one of the most energy efficient systems since it consumes a power of about 25 watts (less than half of a laptop PC) having 86,000 million computing elements (neurons)
    - I consider that the very low energy consumption is due more than the architecture to the underlying technology (material) → instead of silicon, gallium arsenide, etc. that of the brain is of a biological nature (biochemistry, cells, tissues,...)



pixabay



Wikimedia Commons

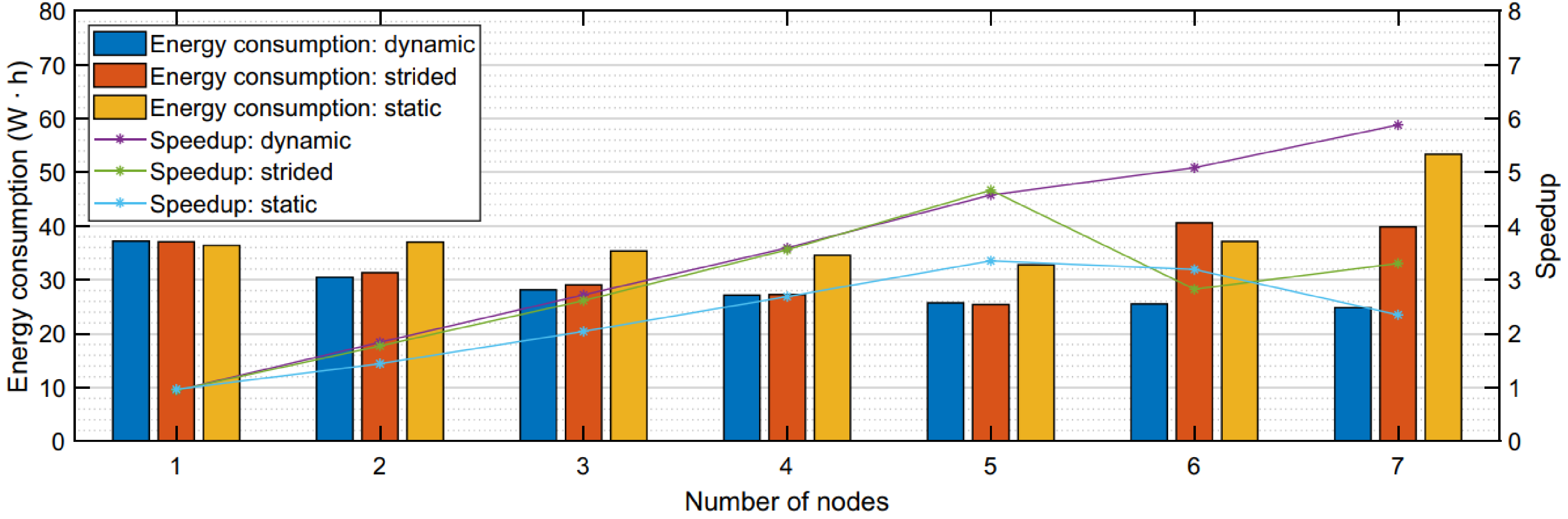
B. Management and planning of the use of resources:

- Use different systems trying to reduce global energy consumption, such as:
  - Making the resources (servers, memory systems, etc.) that are not necessary at a given moment enter standby or suspend or hibernate modes
  - **Dynamic Voltage and Frequency Scaling** (DVFS)
    - Run slowly programs that do not require a very short response time
    - If the clock frequency is roughly halved, **the execution time is doubled, but the power consumption is reduced by a quarter**
  - Executing, as far as possible, the applications within **"off-peak hours"** where the production of electric energy from clean sources is greater because wind production is higher due to the wind or corresponds to hours where solar radiation is greater. Double benefit:
    - Reduction of the economic cost of the energy necessary for the execution of the programs
    - The use of alternative energies is encouraged

C. Scale changes

- The proliferation of smartphones and small mobile devices leads to a reduction in power consumption as each of them offers a multitude of functions and services that were previously performed by independent consumer devices
- Planning and assignment of tasks to the available hardware resources **taking into account their power efficiency**. In particular, the parallelism of applications and programs should be exploited seeking the best possible efficiency
  - In many cases, efficient resource allocation requires redesign of applications and algorithms
- Computational offloading: Computationally intensive processes are offloaded (endorsed) to an external platform, which can be anything from a hardware accelerator to a cluster system, or cloud resources
    - It is only beneficial when a large amount of computing is required with relatively little amount of communication
- Merger or transformation of midsize data centers to **hyperscale data centers** (Google Cloud, Amazon Web Services, Microsoft Azure, OVHCloud, or Rackspace Open Cloud), where power consumption is much better managed

Example of considering energy consumption in an application: EEG classification with 3,600 features, using mRMR+KNN : the results depend on the distribution of work among the cluster nodes



(a) Speedup and energy consumption when increasing the number of computing nodes

The parallel implementation with 7 nodes reduces the consumption of the sequential to 13.38% !

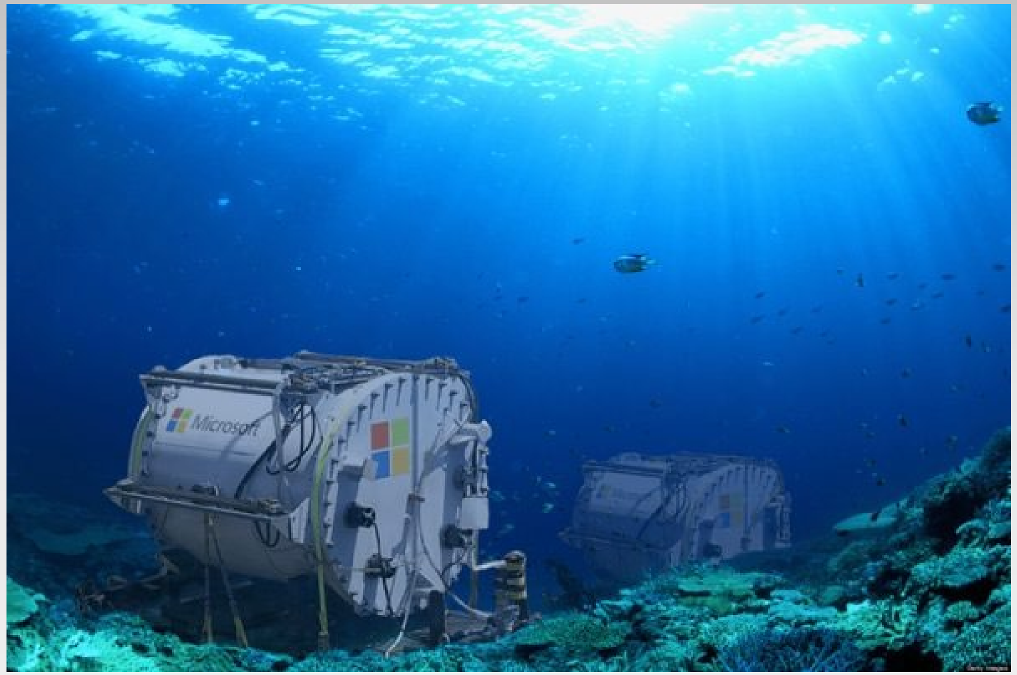
JJ. Escobar, F. Rodríguez, B. Prieto, D. Kimovski, A. Ortiz, M. Damas (2023). A distributed and energy efficient KNN for EEG classification with dynamic money saving policy in heterogeneous clusters. Computing.

Google cooling plant in Hamina (Finland)

- Data centers account for up to 5% of greenhouse gas emissions, and around 40% is from their cooling systems
- The Hamina center uses the water from the icy seas of the Gulf of Finland to cool all its facilities

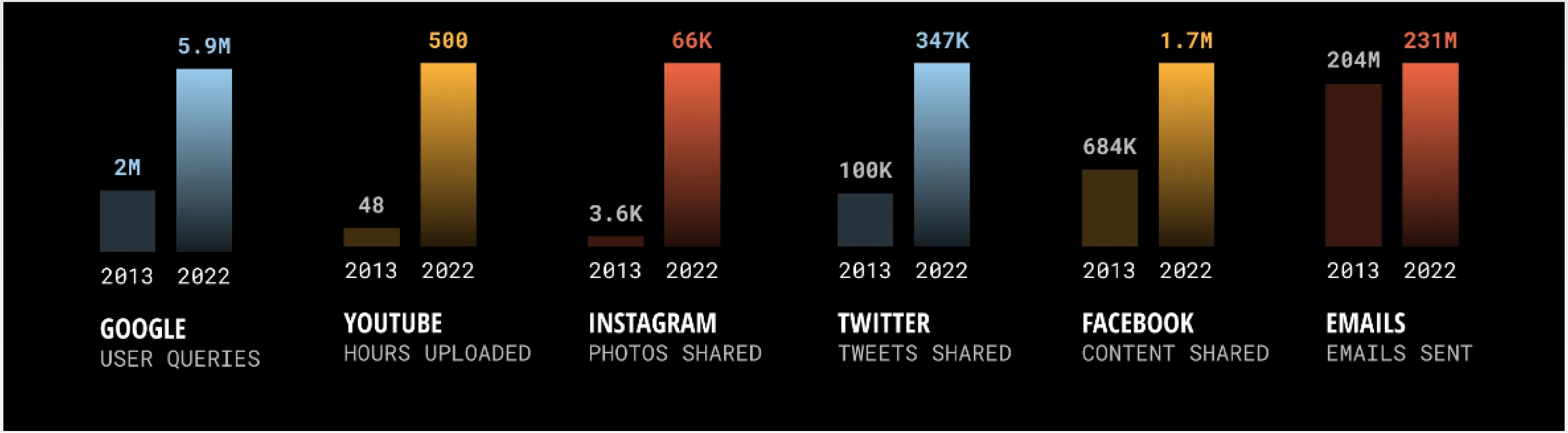
Microsoft Project Natick

- Study of the feasibility of placing data centers underwater, close to coastal communities
  - It involves submerging 864 servers in a submarine-like container that will be powered by a mix of nearby land-based renewable energy, including solar and wind power



Conclusions …

- Data never sleeps, so that **every minute**:



Informe Data Never Sleeps 10.0, realizado por Domo. https://www.domo.com/data-never-sleeps

… conclusions …

- It is estimated that by year 2030 information technologies will consume approximately 13% of the world's electricity, and by 2050 that of data centers will be about three times the total amount of power generated in Japan
- Data storage, on the other hand, would reach 163 Zettabytes by 2025
- The reduction of power consumption in the field of ICT is a transcendental question, and must be faced from very different fields (computer engineering, software engineering, dissemination, teaching, etc.)

… conclusions.

- **Energy consumption** should be considered as a **performance measure** as important as the computational performance of circuit designs, architectures, programs, and applications
- Information on the energy efficiency of the systems should be included in scientific and technical publications, which is not usually done
- On the one hand, there are environmental and economic reasons, but also the need to improve the autonomy of devices that use batteries
- **We must all contribute, from our respective fields, to the challenge of achieving the sustainability of our planet**

Acknowledgements

- We would like to thank our colleagues for their valuable collaboration on these issues:
    - Miguel Damas
    - Antonio Díaz
    - Christian Morillas
    - Jesús González Peñalver
    - Andrés Ortiz (UMA)
    - Francisco Gil (UAL)
    - Francisco Illeras
- Work partially financed by the Ministry of Science, Innovation and Universities (Spain), through excellence projects
    - PGC2018-098813-B-C31, PID2022-137461NB-C31

Thank you so much for your attention!