SYNTHESIS REPORT

Energy and digital technology: mutual challenges

ASSOCIATION NATIONALE RECHERCHE TECHNOLOGIE

FUTURIS

THE POWER OF COLLECTIVE INTELLIGENCE

MAY / 2022

LES CAHIERS FUTURIS

Chaired by Olivier Appert, member of the Académie des technologies
and Gérard Roucairol, honorary chairman of the Académie des technologies
Richard Lavergne, co-chairman
Denis Randet, co-chairman
Mihaja Razafimbelo, rapporteur
Clarisse Angelier, ANRT, publication director

This work received financial support from FutuRIS subscribers:

AI CARNOT, AIR LIQUIDE, AMPIRIC - AIX-MARSEILLE UNIVERSITE, ANR, BERGER-LEVRAULT, BOUYGUES, BRGM, CEA, CNRS, CPU, EDF, ENGIE, FACEBOOK, GE HAELTHCARE, INSERM, INSTITUT MINES TELECOM, INRIA, INSTITUT PASTEUR, IRIS SERVIER, MINISTERE DE L'EDUCATION NATIONALE ET DE LA JEUNESSE, MINISTERE DE L'ENSEIGNEMENT SUPERIEUR, DE LA RECHERCHE ET DE L'INNOVATION, NOKIA, REGION PAYS DE LA LOIRE, SHNEIDER ELECTRIC, SNCF, TOTAL ENERGIES

ACTIVE MEMBERS OF THE WORKING GROUP

ALAIS	**Jean-Christophe**	Air Liquide
ALVAREZ-HERAULT	**Marie-Cecile**	Ancre/INP Grenoble
ASTIC	**Jean-Yves**	RTE
AUVERLOT	**Dominique**	France Stratégie
AZAIS	**Philippe**	CEA
AZARMAHD	**Telman**	EdF
BACHA	**Seddik**	INP Grenoble
BAREUX	**Gabriel**	RTE
BERTHOMIEU	**Nadine**	Ademe
BERTIERE	**François**	Académie des Technologies
BERTRAND	**Arnaud**	ATOS
BONIFACE	**François**	GRDF
BOUYER	**Etienne**	CEA
BULLIER	**Guillaume**	CRE
BURTIN	**Alain**	EDF
BUSATO	**Guillaume**	RTE
CHAILLOT	**Christophe**	Orange
CLAUSSE	**Marc**	Ancre-Allistene/INSA Lyon
DAUVE	**Sebastien**	LETI-CEA
de WARREN	**Nicolas**	Uniden
DUPONT DE DINECHIN	**Benoit**	Kalray
FAUCHEUX	**Ivan**	CRE
FERRASSE	**Jean-Henry**	Aix-Marseille Université
FERREBOEUF	**Hugues**	The Shift Project
FOURNEL	**Guillaume**	CRE
GAME	**David**	RTE
GELENBE	**Erol**	Académie des Technologies
GEOFRON	**Patrice**	Université Paris Dauphine
GILLET	**Guillaume**	Engie
GIRAUD	**Guillaume**	RTE
GRELLIER	**Christian**	Bouygues
GUILLOT	**Françoise**	Safran
HARTMANN	**Joël**	STMicroelectronics
HAUET	**Jean-Pierre**	Comité scientifique d'Equilibre des Energies
KAMIYA	**George**	IEA/EXD/SIO
KHERROUF	**Samira**	Ademe
LABRY	**Delphine**	Ministère des Armées
LATROCHE	**Michel**	CNRS
LEBLANC	**Elvire**	CEA
LECAILLE	**Aurélien**	GRTgaz
LEMOINE	**Fabrice**	Ancre/Université de Lorraine
L'HELGUEN	**Eric**	EMBIX
LUCCHESE	**Paul**	CEA
MASSINES	**Françoise**	CNRS
MAZAURIC	**Vincent**	Shneider Electric
MAZINGAND	**Hervé**	Technip Energies
MENTING	**Jos**	Engie
MONCOMBLE	**Jean Eudes**	Conseil Français de l'énergie/WEC
MONTAGNE	**Xavier**	Ministère de l'Enseignement supérieur et de la recherche
MORTELETTE	**Tiphaine**	Leyton
ORGERIE	**Anne-Cecile**	Ancre-Allistene/CNRS
POSTEL-VINAY	**Grégoire**	Ministère de l'Economie et des Finances
PRIEM	**Thierry**	CEA
PROULT	**David**	CEA
QUEHEN	**Audrey**	Engie
ROCHE	**Nicolas**	Enedis
SCHMITT	**Michel**	CGE/Ministère de l'Economie et des Finances
SLAOUI	**Abdelilah**	CNRS
THIBAUT	**Muriel**	IFPEN
TISSERAND	**Isabelle**	Laposte
VANDELAER	**Olivier**	Engie

CONTENTS

INTRODUCTION

This note provides a summary of the work undertaken in 2021-2022 by a group gathering companies, research organizations and public authorities. The group has been meeting since 2017 at ANRT, chaired by Olivier Appert, to support the National Energy Research Strategy (SNRE). Work is devoted to analysing the technical challenges of the energy transition. The analyses made by the group are pragmatic, concrete, and correspond to the experience of its members.

The group's work has highlighted the importance of digital technologies, in particular to organize and operate the electricity network after the introduction of intermittent renewable energy. Overall, these technologies constitute one of the ways of tackling climate change. However, since the Paris Agreement on climate requires reducing CO_2 emissions by -7.6% by 2050, it is imperative to analyse the environmental footprint of all sectors of activity, throughout the lifecycle. And digital technologies are no exception: the upstream extraction and processing of minerals, the electricity consumption and emissions related to manufacturing and operating equipment and to treating information, the efficiency of the comprehensive recycling of materials downstream in the cycle, all call for vigilance. It is essential that the benefits of digital technology should not be overshadowed by its environmental impacts. Digital technology is like cholesterol: it can be good or bad. This year, the group therefore looked at the energy consumption of digital processes, which is clearly a problem since the growth of the capacity requirements for computing and exchanges appears limitless.

Some studies have announced catastrophes, but the prophets of doom exaggerate: they take no account of the progress in digital technology, including how it is used. The problem remains, but as our work shows, less so in sectors that apply digital than in the digital sector itself. We have seen that the sectors of electricity production, automobile, aviation and building are highly dependent on energy. Everywhere, digital technology can be used to improve the organization of work, reduce energy consumption, and improve the way products operate. And it is particularly necessary in the key sector of electricity.

Without doubt, although electricity consumption still enters into the cost price in both manufacturing and operations, it is interesting to reduce it, but it is not a stumbling block.

On the other hand, it is a stumbling block for the digital industry: heat dissipation is now limiting performances, from integrated circuits to supercomputers. And since we can no longer count on Moore's Law and continued miniaturisation, we need to find new solutions.

But these solutions involve a real revolution. It will involve questioning the traditional separation between the manufacture of universal processors, which can be bought anywhere, and that of systems. For example, the cars of the future will be designed by carmakers that master the design of embedded processors, and that have access – at the same time as controlling their confidential information – to supercomputers capable of developing the artificial intelligence models involved.

What we can expect is a total transformation of value chains. Our work gives an idea of what will occur in some industrial sectors, but all sectors will be affected.
This revolution can create opportunities but, if we do not react now, risks giving the control of our industry to countries that will command the new interaction between electronics and products.

France and Europe face a double problem:

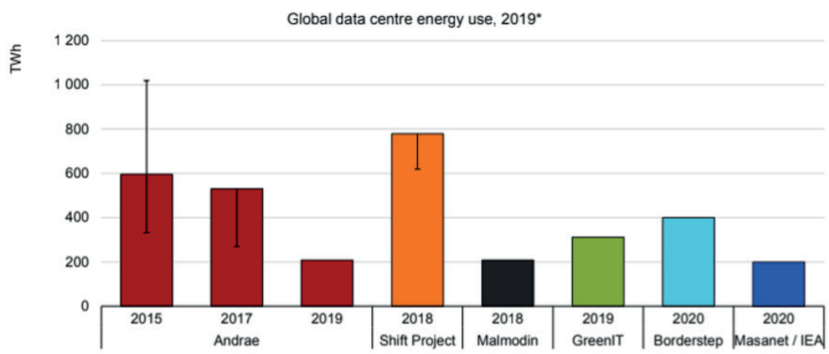
- Controlling the entire digital chain.
- Ensuring that the sectors that apply it possess skilled staff to conceive – in liaison with European digital industry – the digital technology used in the design, operation and maintenance of the materials that they produce.

01

ESTIMATED EVOLUTION OF THE ENERGY CONSUMPTION OF DIGITAL

The world's internet traffic multiplied by over 1,000 from 2000 to 2019. However, a recent IEA analysis shows that the energy consumption and footprint of emissions from information technologies and communication have remained stable, because their energy efficiency has improved enormously. This has disproved catastrophic forecasts like the one published in Forbes in 1999, according to which by 2010 half of US electricity production would be absorbed by the digital economy and the internet.
In reality, forecasts are difficult due to the rapid evolution of demand, the efficiency of ICT, and the way we use them. Methodologies and hypotheses differ depending on the organization. Extreme results, which receive wide coverage, result from simplistic models. For example, for the consumption of data centres, these models give results over twice as high as the probable range (200-350 TWh).

Figure 1: Estimations of data centre energy consumption



The game will be changed by the continued growth of needs and the arrival of emerging technologies (5G, the internet of things, artificial intelligence and machine learning, blockchain, etc.). To contain consumption and emissions, we will need to make considerable investments in research and development. And while their direct footprint will need to be reduced as much as possible, the biggest impacts of digital technologies on emissions will come from their applications in other sectors (electricity, transport, building, etc.), even though their usage could ultimately reduce the total consumption and emissions of these sectors.

In France, according to a report by the Senate, in 2019 digital technology had a carbon footprint of 15Mt, which is 2% of the domestic carbon footprint, and devices alone represented 81% of this footprint.

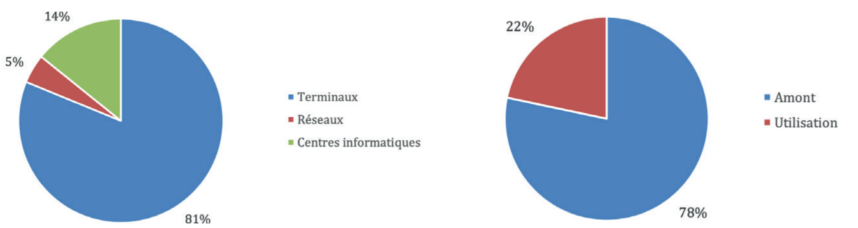


Figure 2: Carbon footprint of digital in France in 2019. Source French Senate - Citizen 2020

Proactive scenarios, for example those proposed by ADEME or the Ministry for the Ecological Transition, forecast a strong reduction in electricity and digital consumption by 2030. This will in fact depend on several factors, such as the choices made on 5G coverage, and the development of the internet of things and edge computing.

The Conseil général de l'économie has made a different forecast for 2030, by extrapolating data from ARCEP (Autorité de régulation des communications électroniques et des Postes).

	Consommation 2018 (GWh)	Variation du parc	Variation du temps d'utilisation	Efficacité énergétique	**Consommation 2030 (GWh)**
Téléviseurs	10 000	-0,6%	-3,2%	-5%	4 996
Smartphones	240	+1,9%	+4,5%	-5%	255
Ordinateurs-tablettes	7 510	0%	+4,5%	-7%	4 924
Box	5 030	0%	0%	0%	5 030
Data centers	7 870	+10,1% (pour moitié)		-4%	10 096
Réseaux	5 100	+10,1% (pour moitié)		-4%	6 543
Total	**35 750**				**31 843**

Tableau 1 : 2030 forecast – hypotheses and results. Source CGE 2020

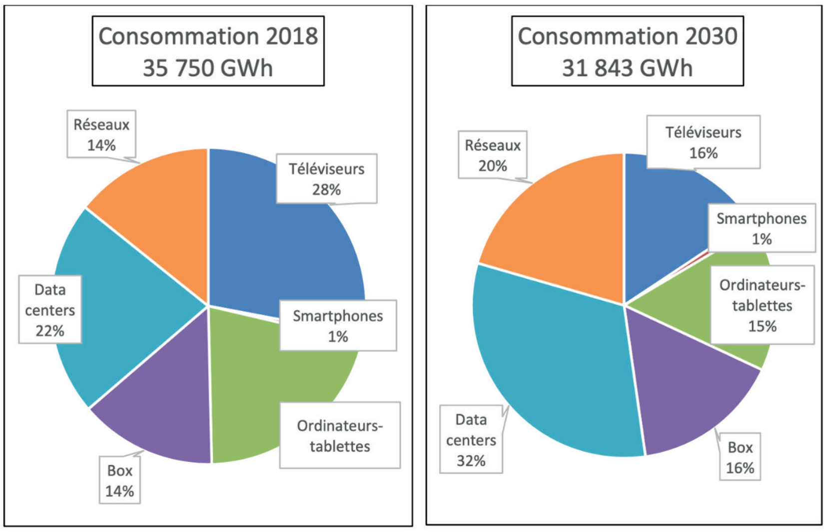


Figure 3 : Breakdown of electricity consumption
Source : « Réduire la consommation énergétique du numérique », Ministère de l'économie et des finances – Décembre 2019

A general drop is expected, but the consumption of data centres and networks is increasing significantly.
Video is particularly concerned, since the move to 4K resolution will increase consumption by 10%.
However, energy-saving measures are indirect because computer technologies have not been equipped to be measured. This will need to change.

Another point is the manufacturing of equipment, which currently represents 40% of the global digital impact. Unfortunately, few detailed studies exist on this subject, except for a 2018 study by Jens Malmodin (Ericsson) and Dag Lundén (Telia Company) entitled "The Energy and Carbon Footprint of the Global ICT and E&M Sectors 2010–2015", which features the graph below giving estimations of the carbon footprint of the sectors of ICT (Information and Communication Technology) and E&M (Entertainment & Media) in 2015 :

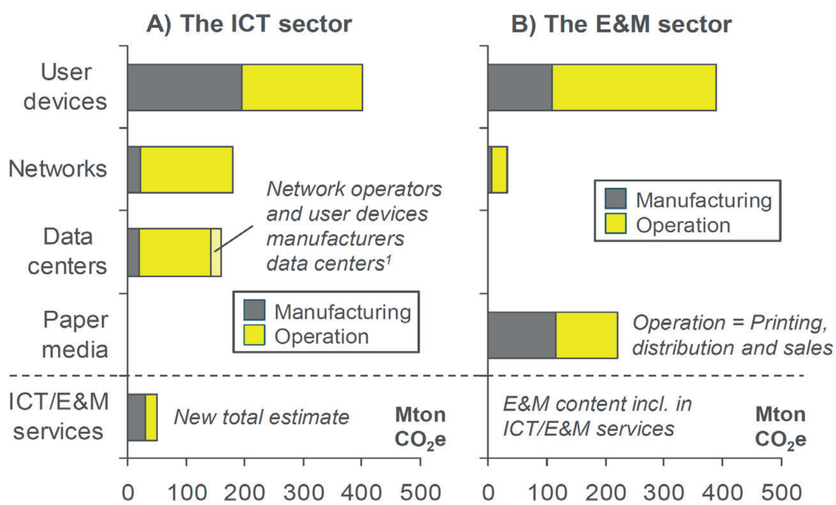


Figure 4 : Estimated carbon footprints of ICT (Information and Communication Technology) and E&M (Entertainment & Media) in 2015

The number of devices (telephones, tablets, laptops, televisions) and more generally the internet of things (IoT) amounts to billions and is rapidly increasing. Because most of them are manufactured outside of Europe, the means of direct action are limited to labels indicating carbon footprints.
Progress in the sustainability of equipment reduces the impact of their manufacturing. Similarly, repair and recycling need to be promoted more strongly. Regulations and standards exist, such as on programmed obsolescence[1]. Another option concerns the limitation of resources consumed thanks to a law on the circular economy.
The interesting point here is that precise data concerning devices and their recycling seem to be lacking in the digital field. The Conseil Général de l'Économie has in fact recommended that organizations involved in the ecological transition (Ademe) work with the electronic communications regulator (Arcep) to define and publish a corpus of common minimal data (equipment stock and digital services) by sector and get it to be included in companies' CSR assessments.

At this stage, we can make the following points:

- Traffic will continue to grow considerably.
- The measurement of digital energy consumption is unsatisfactory.
- The manufacture of equipment currently represents 40% of the overall impact of digital technology.
- Forecasts vary greatly due to different estimations of productivity gains.

1- Article L. 213-4-1.-I. of the French consumer code: The regulation applies via a refusal of non-removable batteries, which was the object of a European directive transposed into French law by the decree of 10 July 2015, and the obligation for information on the availability of spare parts (article L. 111-4 of the French consumer code).

02 PHYSICAL AND TECHNOLOGICAL LIMITS

THERMODYNAMIC POINT OF VIEW

The connection between energy and information was initially identified in the paradox of Maxwell's Demon (1867). The solution to this paradox led to the clarification of the thermodynamic role of information technologies and to establishing the equivalence between missing information and entropy. In fact, the role of information technologies is to increase the knowledge that we possess on a given system. From a thermodynamic point of view, this ambition goes against the natural evolution of an isolated system, to the point that, to respect the second principle of thermodynamics, the acquisition of information achieved by this process is bound to be less than the degradation of high-quality energy into heat. The thermodynamic machine that achieves this acquisition is the processor: to increase the knowledge of a given system or reduce its entropy in an equivalent manner, it consumes high-quality energy (of electric origin) and dissipates it in a thermostat, generally conserving the energy transfers, according to the first principle of thermodynamics. Since the dual machine here is the Szillard engine (1929), the thermodynamic classification suggests viewing the processor as a cooling machine whose efficiency is defined by a performance coefficient. The latter is useful to compute the thermodynamic service rendered by digitalization at the end of a basic operating cycle. In order to dispose of the processor for the next basic calculation, its operating cycle "closes" on an erasure mechanism (Landauer's principle[2]) that can be used to relate the value of the information acquired to the energy allocated to it by the polarization of electronic logic circuits:

- Currently, the performance coefficient achieved by CMOS technology to assign a memory is around 10^{-5}.
- With spin electronics or spintronics[3], this performance coefficient is multiplied by 2 or 3, although this technology is not yet capable of proposing a complete computer.

2- Formulated for the first time in 1961 by R. Landauer of IBM, the energy required to erase a single bit of information is at least kB TLn2, where kB is the Boltzmann constant and T is the temperature of the physical system considered. This empirically postulated value was confirmed by experiment in 2012 by a team at the Ecole Normale Supérieur in Lyon.

3- Spintronics or spin electronics is the use of a fundamental property of particles known as spin for information processing (by application to a magnetic field). The information is read through the electron's electrical charge.

MOORE'S LAW AND KOOMEY'S LAW

Since the arrival of CMOS technology and the microprocessor (1960), the progress made in the capacity to integrate transistors and increase the processing power have improved the performance of processors to such an extent that:

- The number of microprocessor transistors on a microchip doubles every two years (Moore's Law, 1971).
- The number of calculations per Joule (energy unit) dissipated doubles about every 18 months (Koomey's Law, 1965).

Although these are not physical laws, but only empirical extrapolations, these predictions have turned out to be valid:

- From 1971 to 2001, the density of transistors doubled every 1.96 years.
- From 1946 to 2009, the number of calculations per Joule dissipated approximately doubled every 1.57 years.

Consequently, computers have become smaller and less expensive while increasing in rapidity and power.
Nevertheless, Landauer's Paradigm, which establishes the minimum amount of energy required to erase one bit of information (2.80 ZJ at room temperature) constitutes a physical limit of Koomey's Law, whose extrapolation should be reached at the end of the 2020 decade with the current computer architecture. The limits of Moore's Law are described below. In fact, the two laws started to show signs of slowing down around the middle of the 2000 – 2010 decade.
To push back these limits, other calculation paradigms have been proposed, such as by C. Bennett of IBM in 1973, who drew from the distinction between physical entropy and computational entropy:

- Physical entropy is associated with the fluctuations of a macroscopic system to achieve the equiprobability of its micro-states; the heat released is related to variations in these fluctuations when the system evolves and the irreversibility is due to an imposed evolution in a finite time.
- Computational entropy is associated with the loss of information; half of the losses are due to irreversibility (erasure of input memory), and the other half are due to the dissipation of transistor commutation circuits polarized by voltage steps.

In this situation, the following may be expected:
Processors are slowed down by polarization using sufficiently slow voltage ramps – provided that memory is sufficient and that the intermediate computation steps are maintained to be able to reversibly "rewind" the execution of the full computation and recuperate the energy associated with the erasure mechanisms. To obtain the result in a comparable time would require applying parallelization. While the possibility of performing reversible computing generally involves making a compromise between the material footprint (memory capacity and processors) and the energy footprint, it constitutes an intrinsic quality of quantum computing (see 2.2).

1/ THE SLOWING DOWN OF MOORE'S LAW

For over forty years, Moore's Law has characterized the exceptional growth of microprocessing power. This empirical law results from the capacity of semi-conductor manufacturers to double the number of transistors on silicon microchips every eighteen months. However, for the last few years, maintaining this pace has become considerably more difficult. The limits of dissipated power and the density of integration make it increasingly complex to reduce the size of transistors.

In terms of integration, new problems are emerging. The extreme density of transistors (tens of millions per mm2 in 5-nanometre technology) requires piling up more and more metallization layers to interconnect them, which tends to reduce some of the benefits of increasing performance by extending the time it takes to propagate signals in the circuits. The impossibility of continuing to regularly reduce the critical lithography dimensions in 2D requires moving on to innovative 3D integration solutions for transistors. In addition, the semi-conductor industry is currently coming up against another physical limitation: nanoscopic transistors are getting close to the size of an atom (0.1 nm). At this scale, the behaviour of particles is described by quantum physics, whose non-deterministic laws invalidate the expected operation of the transistors. Although for decades miniaturisation alone has been sufficient to stimulate innovation leading to the invention of new uses and thus ensuring their growth, the future of the semi-conductor industry now needs to look beyond Moore's Law.

2/ QUANTUM COMPUTING

One of the motivations behind quantum computing is the risk that digital technology is coming up against the energy wall.

A quantum system is characterized by its "state". This is represented by a unit vector /E> in a Hilbert vector space (on the complex number). This vector is initially immobile, but under the influence of external action, it evolves over a unit sphere in a determinist manner. This is the solution to Schrödinger's equation, which is valid as long as nothing is measured.

A measurable quantity is represented by an operator (called an "observable"), which on this space has eigenvectors and real eigenvectors.

We can anticipate the potential results of measuring it, i.e. precisely these eigenvectors, with the probability of square modules of the components of the vector on the basis of these eigenvectors. For this measurement to make sense, we need to obtain the same result if we do it again; this is guaranteed by the fact that when we do the measurement, the E vector disappears to become the eigenvector corresponding to the eigenvalue that we measured.

A quantum computer is made from qbits. A qbit is a 2-state quantum system, represented by a vector in a two-dimensional space. Its state can be written as a/+> + b/->, and is called the "coherent" state. Two qbits are represented by a 4-dimensional vector space: a/++> + b/+-> + c/-+> + d/-->, and this state is called "entangled". For example, if we go from 2 qbits to 10 qbits, we move in a 1,000 dimensional space.

Algorithms change this state vector. They are made up of a series of "quantum gates" that modify it step by step. A quantum gate is a radiofrequency pulse similar to those involved in nuclear magnetic resonance, on one or more qbits.

At the end of the process, we obtain the result by measuring something.

The advantage of quantum computing is that each operation is done on the whole state vector, and therefore impacts all base states at once. It is the equivalent of a massively parallel computation. Take the example of looking for a subscriber's name starting with his or her telephone number: after first creating a state that combines the names of all of the subscribers, an operator called "Oracle" can change the sign of the coefficient of the subscriber who has the initial telephone number. The next step involves applying a series of quantum gates to concentrate the state of the system as much as possible on the one whose coefficient has just been inverted (continuation of Grover's algorithm). The state of the system is then "almost certainly" the state sought. To be certain, the same operation needs to done again or done in parallel on two identical quantum computers.

In practice, a qbit can be an isolated electron with a spin or ½ spin kernel, a Josephson junction, ions or ½ spin atoms trapped at very low temperature, Bose-Einstein condensates, quantum dots, or atom nuclei spins within a molecule, provided that there are two base states.
Quantum computers actually exist in up to 5 or 10 qbits. Beyond that level, the slightest interaction between the quantum computer and the outside can be considered as a measurement and creates a decoherence that leads to errors. In reality, for most specialists, quantum computers are only of interest when they have at least a hundred qbits.

Solutions to reduce errors could involve working at very low temperatures and, if possible, with the minimum number of connections between the inside and the outside of the cryostat; or calculating faster than the thermal noise; or lastly, correcting errors as for standard computers.

Concerning the energy consumption of quantum computing, two main reasons confirm that it is less energy-consuming than standard computing. The first reason is that the computing requires much fewer steps. The second is that the "gates" are reversible and only transfer recuperable energy. These advantages remain theoretical.

Quantum computing could theoretically

- Factor very high numbers (cryptography: Shor's algorithm).
- Find the configuration of the minimum energy of a complex molecule in chemistry or biology.
- Optimize highly complex systems: logistics, finance, all kinds of networks, particle physics, etc.
- Accelerate the operation of artificial intelligence.

Due to the way it operates, a quantum computer is often optimized for a single type of usage (e.g. simulated annealing). At some point, quantum computers will probably be used for their performance in some very specialized computations, but not in the near future.

03 DIGITAL INFRASTUCTURES. STATE OF THE ART AND EVOLUTION

1/ COMPUTER SYSTEM PARAMETERS

THE MAIN USAGE PARAMETERS OF COMPUTER SYSTEMS ARE THE FOLLOWING:

1. Processing power, expressed as the number of operations per second. This depends on:

- The frequency

Historically, up to now this has been constantly increased by making ever-smaller integrated circuits (Moore's Law), but this has now reached its limit at 4-5 GHz.

- The number of integrated circuits per processor, and the number of processors.
- The calculation method, with the recent introduction of specialized processors that boost the processing power without increasing the frequency (including the implementation of parallel processing).

Nevertheless, specialized hardware also involves designing circuits and software, while universal processors could be mass manufactured separately. In addition, as a result of specialization, and more particularly parallelization, up to 90% of existing software could be outdated.

2. The intensity of traffic, or speed, expressed in the number of simultaneous messages.
This is related to the computing method and the organization of the system. Operations that require high power can only be processed on supercomputers, or at least grouped computers.

- Part of the traffic is the communication between these computers with terminals, telephones, smart devices, etc.
- Another part is the consequence of using a distributed algorithm like in a blockchain, which involves replacing central validations with validations by the users themselves
- Yet another part is video streaming. This is growing with the definition of images. Should 4K be the limit?

3. Latency, or the time it takes to transmit information from one point to another.
Latency directly depends on the distance travelled by the information.

4. Energy consumption
There are two sorts:

- Manufacture and recycling of materials.
- Europe has established standards that manufacturers draw inspiration from, particularly for their 10-year targets.
- Operations.

Both are difficult to measure:

- The first one because complete cycles are difficult to track, in particular given that most manufacturing takes place in Asia and generates high levels of CO_2.
- Consumption from operations depends on the computing configurations, which constantly vary in a non-detectable way. This area really merits research.

5. Algorithmic complexity
Depending on the quantity of data to process, this complexity is expressed in terms of memory size or the number of operations used by a computer program that implements an algorithm. When the algorithm is distributed, in other words, when it results from the execution by independent, asynchronous actors communicating by messages, then the number of messages exchanged adds to the complexity.

6. Platformization
The hosting of data in relatively powerful computers (data centres) allows users to access processing power that they could not normally afford. And the combination with a large amount of data opens the way to new services. However, platformization has a flipside: the ownership of data, or at least access rights. To escape the stranglehold of the GAFA on their data, Europeans got together to create the labelling system GAIA-X. The GAFA are part of it; fortunately, because it would be costly without using their base. However, it is major European industrialists that will act to guarantee our independence. The Germans have clearly understood this.

7. Architectures, centralization-decentralization compromises, systems optimization
Computing power and platformization tend to lead to centralization, while lower latency and exchanges lead to decentralization.
The compromise takes the form of architectural choices, with tiering and networking of centres and terminals:

- FOG computing, which is the architecture of 5G, with staged clouds
- Edge computing, which involves doing as much local processing as possible (in association with sensors-actuators, e.g. in autonomous cars, in which edge computing is in fact indispensable to react fast enough, but artificial intelligence models have to be charged from supercomputers).

2/ NEW ARCHITECTURES

2-1 / FOG COMPUTING: THE NEW TELECOMMUNICATIONS ARCHITECTURE OF 5G

The current architecture is in transition due to the insertion of an intermediate layer, FOG computing (millions of processors with local networks, connected to data centres, but that also manage end-of-network connections with mobiles, connected objects, surveillance sensors, smart grid sensors, industrial sensors/actuators, and soon connected vehicles, etc.). FOG computing is done using compact, reliable computers, like Intel's NUCs that can also act as "SDN routers", interconnected by local networks with Ethernet or optic cables that are part of the hardware structure, and small servers to store data and offer services. Its role is to feed software that supports applications (mobiles, home automation, security and surveillance, road traffic management, games, etc.), which at the end of the network are connected to "terminals" (sensors, actuators or mobiles), while ensuring fast access to the system.
FOG services are highly profitable for operators. Currently, FOG is different for each type of service, for example depending on whether the client wants security or speed.

Each specific FOG is a set of machines: on the one side are the customers (which represent ends of chains, like applications) and on the other side are the services. This structure also has an internal network (outside the public internet) and including an SDN controller and several NUCs. The SDN controller creates connections dynamically to process the requests sent to the FOG.
Energy consumption is measured, for example in an NUC, by measuring the current.

The SDN controller plays a role in stabilizing the system. When overloads occur, the response time increases and the SDN controller modifies the connections to stabilize the response time.
A compromise is always made between the service quality and the energy consumption. The optimum solution is to choose the best path between the different servers and networks.

In conclusion, FOG computing brings considerable progress (IoT, connected cars) in enabling specialized supports. In addition, the computer network has become more complex. The challenge now is to set up measuring tools to understand the energy consumption of the entire system. This will involve experimenting and pushing the connection of measuring tools. Lastly, control algorithms in SDN controllers lead to judicious, shared choices in selecting the required performance (either energy consumption or response time).
Moreover, it is fundamental to note that this telecommunications architecture is above all that of a computer system distributed between several interconnected clouds. Whether traditional telecoms actors (operators and equipment manufacturers) are capable of mastering this paradigm shift is questionable, as is their future competitiveness vis-à-vis the GAFA.

2-2 / EDGE COMPUTING AND IOTS, CONTINUUM COMPUTING

Computing is no longer carried out at a central point (cloud or a central data centre), but also in a fragmented manner, taking into account the location of data production. The processing – or pre-processing – can be done in edge computers, as close as possible to the place where the data are created. When it involves using data that are different, due to their location or production, the processing has to be shifted towards a more central cloud. For artificial intelligence and machine learning, the elaboration (training) of the model of the phenomenon to study takes place in a centralized manner in the cloud (often via a supercomputer installed for the purpose), in order to benefit both from considerable computing power and to be able to cross data from very different origins. Once the model is created, it can be loaded in devices whose behaviour will be subject to the model. This is in particular the case for the implementation of autonomous cars.

2-3 / THE END OF THE UNIVERSAL PROCESSOR: THE MOVE TO SPECIALIZATON AND PARALLELISM

Until around 2005, thanks to the miniaturization of transistors according to Moore's Law, it was possible to increase the performance of microprocessors while reducing their energy consumption. After that point, the frequency could no longer be increased because, beyond 4 to 5 GHz, too much heat had to be dissipated.
In order to increase performance to match requirements, there are however two possible ways: the specialization of processors, and parallel computing, which can also be combined.

The problem is programmability. Since part of the application code is done in the specialized component, and because using massive parallelism requires reworking existing algorithms, the application portfolio will need to be redeveloped to benefit from all of the new capacities of this kind of processing. The more specialized a component, the more rapid and energy-efficient it is, but it becomes harder to program; for example, the Fortran codes written 30 years ago are no longer useable.

Parallelism can be implemented at several levels:

- processors;
- integrated circuits, which can now gather up to several hundred processors (in the US these circuits are called sockets);
- servers, where several sockets can coexist;
- large systems, like supercomputers or the cloud, which host thousands of servers in a local network.

Most of the application portfolio is made up of sequential or weakly parallel programs. If applications are to benefit from parallelism, programs will need to be rewritten, and the problem-solving methods (algorithms) that they implement will need to be redesigned. At least 90% of the existing software will need to be updated.

3/ SYSTEMS OPTIMIZATION

Sizing energy systems involves obtaining temporal data (energy demand/purchase price/energy sales/climate profile of the site) and technological data (catalogue of energy converters, catalogue of short- and long-term storage means). These data are then employed to determine the optimum installation, such as production units, storage means, size of units and usage planning. The aim is to avoid oversizing and conflicts between systems, and at the same time meet demand. This is a major issue for energy operators.

To set a target, the standard criterion is a compromise between the cost of the system and its performance. Today, environmental criteria are also considered, including the carbon footprint. The idea is to bring together energy sciences and environmental sciences to work out how to organize all of these environmental criteria. Environmental and social constraints involve a difficult equation. And the high number of criteria requires greater computing power. The solution is to look for a reasonable optimum (close to the real optimum) in a reasonable time (which depends on the actors and industrial sector). Hazards also need to be taken into account in the decision (relating to resources, demand, economic conditions, etc.). This integration can require prohibitive computing times. The aim is to find alternative methods. Modelling at each time step (e.g. 10 min. for heat or 10 s for electricity) results in an excessively large problem. The volume of data need to be reduced by creating typical data (typical clusters).
To model a physical system, the choice of the physical precision can be adjusted by taking into account the uses of all components. The greater it is, the better account the model will take of the operation (e.g. degradation of the system, etc.), but the resolution time will be long.
It is thus crucial to communicate with digital experts to establish either data models through learning, or so-called hybrid models, which are simpler physical models coupled with data models.

A first avenue of progress involves succeeding in measuring energy consumption.

The current figures on general digital consumption are on the rise. Most studies indicate that this has been the case since the early days of digital. The increased traffic and number of devices override the significant improvements made in energy efficiency, in particular due to the internet of things (IoT). However, it is complicated to obtain a clear idea of digital consumption. Several misconceptions need to be challenged, such as the notion that computer equipment that is not being used consumes little energy. An unused server continues consumes up to 50% of its full energy consumption. And the same is true for most internet infrastructures, which are rarely used to saturation point.
Identical servers do not consume the same energy. They have the same performance but do not consume in the same way, with a difference ranging from 10% to 15%. Producing the energy profile of a server based on its CPU performances is also incorrect. The CPU usage rate is not a good indicator.
It is complicated to obtain the energy consumption of an application or particular service because several sub-services operate at once, communicate between each other and use different servers, which means they have different consumption levels. In addition, some servers are not really available for measurements.
To optimize this consumption, it is therefore necessary to have models deployed on both public and private clouds to attempt to estimate the consumption of services. This is done on increasingly distributed architectures. Theoretical calculations are incorrect due to stacked hardware and software layers. This energy consumption needs to be understood, along with its distribution, to be able to design information systems that correspond to user needs.

The second avenue involves designing “just enough” operating systems. For example, concerning the digitalization of energy, which objectives need to be considered: performance, data security, real-time efficiency, fault tolerance (redundancy, overprovisioning, how to optimize a new or already built infrastructure). To achieve this, it is important to develop analysis and decision-making tools that take all constraints into account. This requires a high level of interdisciplinarity and exchanges. Information systems will be constrained from an energy point of view, in terms of power peaks, or intermittent renewable energy supply.

The third avenue concerns improving the energy efficiency of infrastructures. As new generations of mobiles pile up, their consumption will need to be compensated by making gains on uses in other sectors.

The final avenue concerns sobriety to reduce energy consumption. This more sober approach could involve making digital more “green”, analysing the life cycle of materials and improving recycling, looking more closely at the internet by prioritizing the most useful sensors and digital intelligence, and also looking at the questions of the societal acceptability of this sobriety.

The overall framework could be borrowed from the European Commission, which in July 2020 published its “energy system integration strategy”, defined as the coordinated planning and operation of the energy system considered “as a whole”, combining all energy vectors, infrastructures and consumption sectors. The aim of the strategy is to result in effective, affordable, deep decarbonization of the European economy.
The three main elements of energy system integration are:

- A “circular” energy system, centred on energy efficiency (e.g. production installations combining heat and electricity; the reuse of some types of waste and residue and waste heat generated from industrial processes and data centres; energy produced from biowaste).
- Greater electrification of suitable sectors (e.g. use of heat pumps for heating or low-temperature industrial processes, electromobility, electric furnaces in some industries, etc.).
- Use of renewable and low-carbon fuels, including hydrogen, for sectors in which direct heating or electrification are not the best solution or impossible, inefficient, or too expensive. For example, natural gas in a transition phase between coal and low-carbon solutions; renewable and low-carbon hydrogen in industrial processes, or even for heavy-duty road and rail transport; synthetic fuels for air and maritime transport (biomass, biogas, etc.).

Concretely, energy system integration (ESI) comprises:

- A multi-vector energy approach to guarantee decarbonization, supply security, resilience, and cost control.
- A focus on the principle of energy efficiency.
- Decarbonization of energy demand, mostly via electrification (nuclear for those that choose it, renewable electricity and gas).
- ESI is also key to facilitate the penetration of renewable electricity in the energy system (such as via P2G and G2P). In addition, it includes optimization of existing infrastructures (acceptability, costs, etc.) to ensure flexibility and back-up.
- Use of gas and renewable gas, in particular when electrification is too difficult or costly.
- A changing, then decentralized, energy model, emergence of local energy systems and “energy communities” alongside the traditional centralized model.
- Increased interfacing / call for digitalization, which also brings risks (cybersecurity).

4/ SECURITY AND RESILIENCE

The first requirement, in particular for the electricity network, is to get broken-down systems working again as quickly as possible.
Another requirement is to resist cyberattacks. Cybersecurity is a major issue for digital technology. Several international discussion groups exist. At European level, the question of cybersecurity aims to share best practices and standards. Today, the energy system is no longer linear, and data flows passing through different actors are complex. Another issue on the table is coupling between different types of energy, like gas/electricity, etc.
This interdependence spreads risks, from user to system. The EU project provides a consistent overall approach to this management of risk. Cybersecurity undoubtedly brings additional costs, but it is essential to ensure that the entire system continues to operate.

With different systems in coexistence, intersectoral communication is indispensable to control and optimize interdependencies. From data exchange to materials, intense, real-time management of digital communication is required to balance systems. A decentralized control system (e.g. edge-to-cloud) will bring advantages but also new potential cyber-weaknesses. Currently, given the highly diverse transpositions of regulations by Member States, the EU is attempting to impose regulatory constraints on industrialists. The EU directive NIS 1.0 on the resilience of critical entities reveals the wide differences in implementation between the Member States. In preparation for the NIS 2.0 version, the Commission has launched a series of consultations due to terminate in early 2022.

In 2019, the SolarWinds attack revealed that the vulnerability of the supply chain is now a major challenge for industrial cybersecurity. A system assumed to be trusted can propagate vulnerabilities to other systems. The SolarWinds supply chain attack also highlighted the existence of major weaknesses in the energy sector, i.e. the lack of preparation and low systemic resilience.
Although different recommendations have been aimed at “clients” and “suppliers”, they are often not implemented (regulations) or applied (lack of awareness, costs and efforts, etc.).

The main observations and recommendations related to cybersecurity are the following:

- The rapid pace of change brings new vulnerabilities that must be prepared for fast. Security measures are still lagging behind the threats.
- Surveillance systems equipped with total visibility are lacking.
- Cybersecurity measures come at a cost, and it is absolutely crucial that all system actors/operators make an adequate response. Companies should devote 10% to 15% of their IT budget to protection against data violations and attacks. Additional expenditure of a similar amount could be devoted to industrial cybersecurity.
- Intersectoral partnerships are indispensable to manage the cyber challenge facing the entire energy system.
- It is indispensable to have a standardized regulatory approach in the EU; the EU NIS 2.0 is a step in the right direction.
- The cybernetic risk to the supply chain needs to be dealt with.
- A cyber-certification system for critical technologies is vital.
- Our labour market lacks qualified people. Industrial cybersecurity is still all too often brushed over in academic studies. Significant investments should be put into theoretical and practical training.

04 SECTORAL ANALYSIS

1/ ELECTRONICS, IT, TELECOMMUNICATIONS

An understanding of digitalization begins with the material components of a digital system.

France has the potential to fulfil the value chains of the digital transformation:

- Manufacture of components by CEA-LETI and ST Microelectronics.
- Kalray, specialist in parallel processors.
- OVH data centres.
- Supercomputers and quantum learning machine by Atos.
- The telecommunications developed by Orange.
- Supply chain control, e.g. by Atos and Kalray.

1-1 / COMPONENTS; CEA-LETI & ST MICROELECTRONICS

The manufacture of electronic components is based on two main elements:

- The life cycle, with all that it involves in terms of materials, supply, and the capacity to produce while consuming as little as possible.
- Data and the contribution of artificial intelligence to increase processing based on their analysis.

Industrialists are mobilized on electronic device consumption (life cycle assessments). The CEA-LETI is extremely motivated to work on these subjects. The aim is to reduce this type of consumption by a factor of 1,000. To achieve this, the company proposes a 9-point action plan. Each area of action also concerns a technology: the first concerns semi-conductors, the second focuses on architecture, and the last level concerns embedded software closely related to the targeted application.

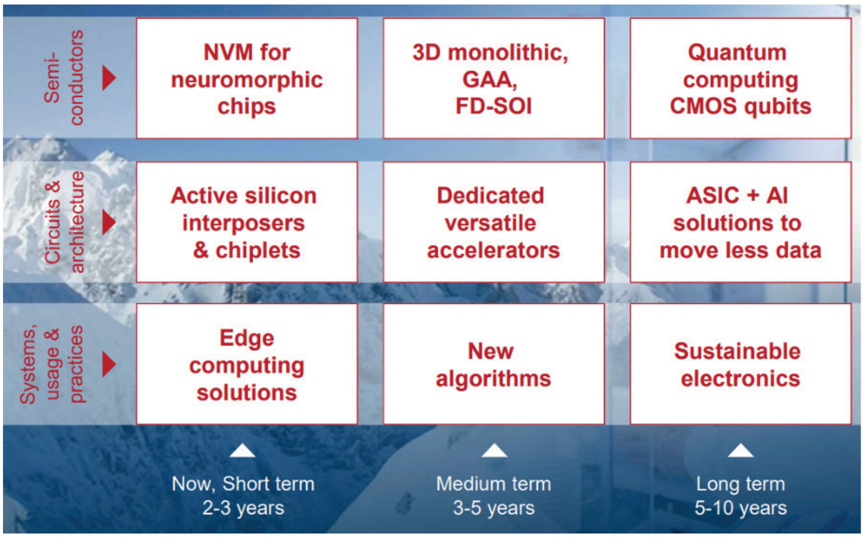


Figure 2 : The 9 focus areas of the proposed action plan (source CEA-LETI)

Today, to tackle the problems defined requires taking a systems approach on these three levels.
The first level, which is the short term, involves neuromorphic computing. Artificial intelligence is treated on a very large scale on the cloud. The ambition is to be capable of setting up embedded intelligence circuits at the level of an object linked to the internet of things. In terms of computing performance and energy performance, we are still a long way from the capacities of a bee's brain, and a lot of work remains to be done. We can attempt to modify the architectures and bring the computing unit and the memory unit as close as possible, because from an energy point of view, the movement of data represents 90% of processor consumption.
At the moment, we can use non-volatile memory, which works with CMOS technology. Several comparison criteria are defined, such as programming power, endurance, writing speed, etc. Resistive memory and 0-electric memory are being explored since they possess largely relevant criteria compared to the frequently used flash memory. This already results in much lower levels of consumption.
Still in the short term, other means of reducing consumption are possible. For example, developing a silicon on insulator (SOI)-type substrate and using a technology known as "fully depleted SOI (FD SOI)" to make microcontrollers; this intrinsically offers a very interesting alternative in terms of consumption, in particular for embedded applications.

In five years, the savings in the relationship between computing power and energy consumption could be 100-fold.

Today, STM electronics is working at 28 nanometres; the CEA-LETI is preparing 10 nanometres, which will be on the market in 2025-2026. Below 10 nanometres, the interest of SOI is difficult to justify. TSMC has already commercialized 7 nanometres. The CEA-LETI is exploring the concept of stacked nanosheets that could make it possible to go to 5 nanometres and further still.

STMicroelectronics has a strong presence on the internet of things (IoT) market, where energy consumption is already a real problem. Internet traffic will increase dramatically with the arrival of 5G, which will allow IoTs to multiply. Exchanges of information could amount to 25 to 50 ZB by 2025 (1 ZB=1021 octets), and 500 ZB by 2030. The artificial intelligence employed by IoTs takes place in the cloud or data centres, via software that runs on large computers. STMicroelectronics estimates that data centres could consume 25 to 50 times more energy in 2030, which represents 100 to 300 Twh. Some architectures are attempting to rein in this increase, but the most efficient method is to maximize local processing on data produced locally, which is known as edge computing.

A neural network model is trained by central learning to benefit from all of the circulation data accumulated by the biggest number of vehicles possible and for as long as possible (which requires very high archiving capacity and processing power). This model is then loaded into each vehicle to be able to react in real time thanks to specific processors embedded in the circulation events. This so-called neural inference phase takes place locally and is therefore a type of edge computing.
STMicroelectronics is aiming at AI processing in edge computing, with a target of 80% edge by 2030. In a car, there are in fact a great deal of inference tasks involved: vocal recognition, object detection, anomaly detection, etc. Currently, sensors do not carry out any processing. The aim is therefore to act at the level of each component to:

- Associate sensors with local processing of information.
- Equip microcontrollers with integrated communication systems and embedded security; set up local neural inference (neural networks, information learning and processing).

To optimize consumption, STMicroelectronics employs new technologies, including FDSOI integrated circuits and phase-change memory (pcm), which can be associated with neural networks to carry out partially analogue rather than simply binary processing.

Another challenge is to increase the memory capacity to carry out operations. Today, energy consumption is mainly due to communication between processors and memory. Yet graphic processors appear to have the capacity to process information faster than standard CPUs while consuming less energy by processing it directly in the memory.
Software needs to be integrated into microprocessors to carry out local processing, because it is associated with neural networks, and because memory is added at the level of the microcontrollers. The ultimate aim is a self-powered system.

1-2 / KALRAY PARALLEL PROCESSORS

Kalray produces multicore processors. Using AI, computing is carried out close to data using accelerators, i.e. edge computing. There are two main players in this area: Intel and Nvidia (focused on accelerating AI)). There is a continuum that goes from the IoT to data centres and that is increasingly sensitive due to the issues raised by 5G.
The general motivation behind edge computing is energy consumption, because computing takes place locally. The fact of working close to data sources also significantly reduces latency.
Kalray is well versed in edge computing, in particular for autonomous driving: data management using sensors (camera, movement sensors, satellite position) (a stage called perception during which computing needs to be fast), localization, prediction, decision-making (with a standard control-command technique). A great deal of computation is done locally and therefore requires considerable communication (with ultra-low frequency).
At the level of computing machines, Kalray works with homogeneous multicores (data centres), which are processors fitted with multiple cores. They are a lot easier to programme but changing scale is difficult because of sharing outside the core. Duplications are made with groups of cores called 'manycores'. The difference between a multicore and a manycore processor is the quality. The former are not necessarily specialized, and the latter are more effective.
Kalray builds manycore processors based on programmable, non-specialized cores like GPGPUs. Kalray processors are mostly used in edge computing, many of them in neural inference, with class programming (like java, c++), but also in machine learning models (high-level language).

This type of development is included in the European project EPI (European Processor Initiative). The second stage will involve aiming at more complex computation through GPGPU and then at modern applications. These second-stage applications will then need to be subject to a feasible application. Lastly, Kalray wants to create an alternative to GPGPU and move beyond the limits of today's technology.

1-3 / DATA CENTRES, OVH DEVELOPMENTS

Data centres are particularly electricity-intensive. In late 2020, the EU published a report on the energy consumption of the cloud. It is interesting to note that among the scenarios presented, the worst-case scenario featured the strongest growth of electricity consumption... independently of the activities deployed (whether positive or negative for climate action).

Yet, still according to the European Commission (in another report), the data market in Europe could amount to up to 6% of GDP during one decade (with, mechanically, a high need for data centres). In other words, the debate is complex, and the roll-out of this type of equipment (and the growth of the resulting consumption) is seen either as a threat, or as a factor to transform the EU economy.

We observe that the problem lies at local level and global level, with a problem relative to the conditions for inserting data centres in local systems (electric systems, and district heat networks).

A good illustration is to compare electronics with lighting: we moved from oil lamps to candles, with very mediocre light yields, and then incandescent lamps with 1% to 2% yields, and finally LEDs, with light yields of up to 25%. The latest shift took 30 years. In comparison, data centres are at an even lower level than oil lamps and we can only hope that in the decades to come, the same progress will take place as for lighting.

The basic observation is not necessarily alarming. For the 2010 decade, we note a stabilization in energy consumption by data centres. This observation can be explained by the development of "hyperscale" data centres, which are much more efficient than previous generations.

The challenge is to determine whether, in the 2020 decade, it will be possible to pursue this stabilization. Without doubt, making the equipment comprised in data centres more efficient is not enough. Instead, it will require inserting data centres, based on a "sector coupling" vision, into the electricity network (supplying different services to balance the electricity system). It is also important to consider the conditions of electricity supply, in particular through PPAs (Power Purchase Agreements), in order to guarantee a renewable energy supply (and in return, a contribution to financing renewables).

Despite these different levers, the concentration of data centres puts some electric systems under pressure and requires additional infrastructure investments (both in the energy and telecommunication domains, in particular for optic fibre networks). Different forecasts illustrate this phenomenon in Europe: by 2030, data centres could represent around 20% to 30% of the electricity consumed in Denmark and Ireland (depending on the national electricity operators), countries that have developed highly aggressive strategies to attract them. In 2019, the Netherlands established a moratorium on new data centres.

It is essential for the debate to move on, given the European ambition for data centre climate neutrality by 2030. Interestingly, a broad coalition of actors is currently mobilized around the "Climate Neutral Data Centre Pact", which aims at this target.

OVHCloud, which supplies data centres, has defined its objectives based on five points:

1. Power usage effectiveness (PUE). The aim is to reach a PUE of 1.3 for all new data centres by 2025, and 1.3 by 2030 for all signatories (including old data centres, with a PUE of 1.4 for hot zones).
2. Carbon-free energy, so that by 2030, 100% of energy used is decarbonized, with freedom in the choice of energy mix. And to reach 75% by 2025 (for operators with power greater than 50 kW); one of the avenues is to relate consumption to green means of production and create a relationship between energy consumers and producers. This is what OVH does for the servers it installs. The problem is not specifically French, and OVH is also working with Italy, India and North America.
3. Water usage. The debate is inevitable: while in France accessing water is still easy, it is subject to constraints elsewhere and it is important to act on this now.
4. Re-use (recycling) of all equipment (servers, etc.).
5. Reuse of energy. However, this commitment is limited because: using this heat depends on a data centre's environment; the local interest is variable, without forgetting lobbying for district heating; and more heat is emitted in summer than in winter.

1-4 / SUPERCOMPUTERS, MADE BY ATOS

Examples of cases where supercomputers are essential include weather forecasts, climate modelling, underground exploration, and all kinds of digital modelling (high-energy physical, crash tests, genome sequencing, health and conception of new drugs, risk analysis, defence, etc.). In addition, supercomputers are required to compute learning models based on data oceans, like in the case of autonomous cars before downloading the models into the vehicles themselves.

The challenge in the coming years is exascale computing (at least 1018 double precision operations per second, which is one hundred times more than currently).

The supercomputers developed by the ATOS group are at the cutting edge of global performance.

The 2013 iPhone 4 has the same performance of 1.6 gigaflops (floating point operation per second) as the 1985 Cray 2. A MacBook Air totals 230 gigaflops. The Atos Sequana supercomputer has a performance of 44 million gigaflops.

Supercomputers are like huge radiators. Their energy consumption can reach 20 MW, and they weigh up to 200 tonnes; each component produces extreme heat, up to 300 watts today and 800 watts in two years' time. They can have up to 10 million cores, with a maximum computational performance of up to 400 petaflops (1 petaflop = 1015 operations per second). Their energy performance diminishes with size and, beyond a certain number of nodes, no progress is possible.

Once again, supercomputer supply can be part of a PPA (Power Purchase Agreement), in order to guarantee a renewable energy supply (with in return, a contribution to financing of renewables). Similarly, Atos has entered into research partnerships to set up supply deals for green hydrogen.

1-5 / THE QUANTUM PROGRAM

A platform called the Quantum Learning Machine (QLM) is being developed at Atos. This platform uses a server with a huge shared memory (up to 48 terabytes of RAM) to carry out all computation in random access memory. It also integrates work carried out with quantum research institutions. The aim is to model the quantum noise of physical devices to successfully simulate a real quantum processor with its noise, or to remove the noise to simulate a perfect computation. In addition, work is being done on quantum algorithms to directly connect them to a supercomputer.

The Quantum Safe Program studies quantum supremacy, which can undermine current decryption methods. This program aims to resist future quantum attacks.

The key areas involve acquiring quantum knowledge, quantum program learning, and program testing. The aim is to optimize quantum algorithms on contemporary computers.

1-6 / TELECOMMUNICATIONS

The current impact of information and communication technologies in carbon emissions is 3.7% of the overall impact. The increased demand for services is exponential, but carbon emissions remain stable.

The key players in these technologies are all committed to reducing emissions, especially in the telecoms sector. About 50% of private purchases made as part of Power Purchase Agreements (PPAs) are by digital players. Analysts do not consider these investments in their carbon impact, nor the efforts made in new data centres with free cooling. The energy demand is much lower than forecast, and the entire sector is reducing its carbon footprint.

Operators have made the following commitments:

- By 2025, 30% reduction in CO2 compared to 2015, mainly thanks to renewable energy (purchased or produced). Energy consumption growth can also be combined with decarbonization. In addition come energy-saving programmes, with hardware progress and algorithmics.
- By 2040, carbon neutrality. Incompressible residual emissions will be compensated by carbon sinks.

Currently, Orange data centres represent 9% of the total consumption of the telecommunications sector.

Orange consumes almost 5.3 TWH, 2 of them in France. 85% correspond to the network and information processing. Orange data centres represent 9% of the total consumption of the telecommunications sector.

Over the years, Orange has been accumulating carbon sinks. It also plans to obtain supplies from specialized actors in one of two possible ways: as a product (including intermittence) and from the actor, which can balance it with other sources. This counters the risk of intermittence, but is more

expensive.
The aim is to make users move to fixed systems. This will require knowing how to connect usage and architecture with consumption, the key being to coordinate an entire ecosystem in a global manner.

1-7 / SUPPLY CHAIN CONTROL

Currently, 75% of components used by European manufacturers are produced outside Europe. The European Processor Initiative (EPI) EU programme, which Atos and Kalray participate in, aims to give Europe the capacity to create its own processors. The 2030 target is to be 65% to 70% autonomous. Twenty-seven partners are working on a processor that will be available in 2023. It will integrate security elements and allow effective learning for artificial intelligence; a key challenge will be containing its energy consumption.
For integrated circuits, without massive aid, ST cannot keep up with the race for ultimate miniaturization led by the Asian giants Samsung and TSMC. Intel, which seemed to have given up, has decided to get back in the game and to pursue a more open policy involving more dialogue with its clients, and probably a new factory built in Europe.

2/ ELECTRICITY PRODUCTION AND SUPPLY

In the domain of electric energy, digital technology already has an established place. It is vital to the electricity level, and plays a key role in flexibility. For EDF, the objectives of digitalization are to improve the business process, innovate in connection with company strategy (CAP 2030 for EDF SA), create new services and business lines, improve the level of employees (skills development, expertise), and construct a robust technological base.
Concerning data, the aim is to develop IoTs, in-house digitalization, smartphones for different activities, drones for surveillance, and Linky objects. In terms of sharing, AI is omnipresent, with automation, and the concept of dataspace. Lastly, digital twins and virtual reality are also being developed (as part of training).

In addition, the values and requirements central to digital are taken into account, i.e. ethics, cybersecurity, and storage issues.
In general, digital technology facilitates the operations of the extended enterprise; it helped us all get through the Covid-19 crisis by pursuing our activities remotely.

2-1 / PRODUCTION

All of the functions required for production are impacted by digital technology: the security of operators (secure 4G in power stations), deployment of IoTs (listening, visualization), aid for operators, certification of tests and productions for all control systems, etc.
Edge computing is also moving towards a more responsible approach, bringing a local production and operating capacity. It speeds up decision-making and acts on network saturation.

Concerning nuclear power, the aim is to increase the lifespan of the park in total security, build power stations for tomorrow and dismantle sites. For new nuclear pressurized water reactors (EPRs), digital twins are used in the construction phase, along with surveillance and ageing models of structures.

The ambition for hydropower is to increase economic and industrial performance, improve the

control of plants and develop new capacities.
For renewable energy (solar and wind), supervision and maintenance operations are centralized in a secure cloud, and all data are sent to a renewable data lake to enable a comparison between actual and estimated production, to have performance indicators in real time, and thus constantly optimize.
VVPs (Virtual Power Plants) virtually control a number of production sites. Blockchain is employed to trace the source of energy and guarantee traceable energy (green or otherwise) for consumers.

2-2 / UPSTREAM-DOWNSTREAM OPTIMIZATION

Increasingly, a real-time approach is taking hold, with closer European integration, and an emphasis on secure transactions. One means is to develop AI systems to reduce time steps and deal with intermittence issues.
Regarding distribution, EDF is trying to establish itself as a trusted third party by adapting networks to new uses and decentralized production. The contribution of digital is most obvious on the industrial side (supervision, maintenance, optimization of the entire network, IoTs, 5G), but also in the creation of a platform to disseminate and provide reference data on energy production and consumption. This makes it possible to become a true data operator in order to decarbonize the entire sector, including industrial, personal and public.
.

2-3 / COMMERCIALIZATION AND ENERGY SERVICES

The challenge here is to provide quality supply and commercial performance, while being reactive in terms of availability (with the hike in energy prices imposed on the market); then to supply services to support the ecological transition and deal with global warming based on consumption control and energy efficiency; and lastly to develop energy management services (based on data through AI, or by developing digital twins or connected objects), and services devoted to reducing energy consumption.

This involves strong requirements:

- A trusted service: taking the example of clouds, requirements have moved to EU level, through the labelling policy and the GAIA-X project.
- Concerning cybersecurity: the aim is to establish transparent exchanges with actors, but at the same time the principle of privacy by design is applied in the General Data Protection Regulation (GDPR).
- The issue of responsible digital technology is crucial here: whether it concerns IT for Green or Green IT, the aim is to minimize data storage, taking an “ethical” approach to using AI.

2-4 / ELECTRICITY SUPPLY

RTE, the French transmission system operator for electricity between 400 kV and 63 kV, manages the standard network featuring interconnections with manufacturers and interconnections with the major renewable energy producers.
The electricity grid operates like a three-point infrastructure based on voltage, the current in each line, and the overall frequency of the network.

The main phenomena that the system tries to avoid are the following:

- Frequency collapse, involving an imbalance between production and consumption.
- Voltage collapse, mostly resulting from a distance between production and consumption.
- Cascading overload of electricity lines.

The distribution of current flows depends on the geographic distribution of production equipment and consumers; it also depends on the impedance of elements in the network, related to the characteristics of the conductors; and lastly on the topology of the network, which represents its actual intelligence (e.g. ring network).

The network operator guarantees that the flow remains stable in a normal regime and a when degraded by a facility outage. Facility outages (at the line, station or generating unit) can be simulated to anticipate system responses. Undeniably, renewable energy significantly increases the range of power flow uncertainties.

As a result of the energy transition, the electricity system is undergoing significant developments, if not a revolution. Means of renewable energy production are growing rapidly. However, they are difficult to control due to high levels of uncertainty regarding sun and wind forecasts. RTE must therefore manage and adapt the network in real time. Digital technology means that several measures can be implemented. For example, RTE can carry out peak-shaving on renewable energy for a few hours per year. This improves the sizing of the network: shaving 0.3% of the overall renewable energy volume will save an estimated 7 billion euros on network investments by 2035.

Another solution consists in changing the control architecture of the electricity network. Until now, control has taken place in two layers: at the electricity station (local protection, time in seconds), then at central level (for a duration of 10 mins). The mass of information is getting difficult to manage in real time, leading to the suggested installation of an intermediate layer, which will be a control layer. This layer will manage sub-sections of the network, from 5 to 50 stations for closed-loop management. This will resolve problems like congestion, for example. The operator will no longer be a pilot but a navigator, with the shift to a forecasting approach.

The implementation of this control layer takes the form of new adaptive automatons on zones. They can be used to manage peak-shaving. Given the time constants of around one minute, automatic actions can be put in place. These automatons can also be used to create an interface with producers. The aim is to have about 200 automatons by 2035.

Automatons can be used for the following:

- They can alter the topology of the network, for example by opening a circuit breaker to close the network. This action is free, but increases the risk of a blackout.
- Automatons can also use renewable energy production erasure to allow for a more precise modulation of production, such as by controlling the maximum power that each station can integrate. This action is not free, because producers are compensated for the production that they would have had, but the risk of imbalance on the network is low.
- The third use of automatons is the flexible solution, by using smartwire batteries to slightly alter the impedance of a facility in the electric distance and the direction of flows. This option requires significant financial investment, but it cheaper to use.

With these tasks, automatons integrate their actions into a model of the electricity system that allows them to produce simulations. They are provided with the cost of these tasks and the zone constraints (such as the transit current limits in each facility, maximum battery capacities, load-shedding constraints). All of these elements are the parameters of the optimization made by the simulator. Model predictive control is applied, based on a closed-loop principle. The state of the network is then analysed according to the different actions.
These automatons are installed using a classic centralized architecture: with sensors and means of action at stations and at producers', the data are sent to a centralized computer in data centres. Work is being done to design a hybrid edge-to-cloud mode.

These cyber-physical systems raise the following challenges:

1. Finding suitable methods and tests for these closed-loop controls using co-simulation, such as associating an electric network simulator with a telecom network simulator.
2. Determining the scope of these different automaton action zones to establish robust, upgradable criteria, and to manage interactions between systems (overlapping or influence).
3. Highlighting the potential of cloud-to-edge auto-adaptive systems to get the most out of them: bandwidth savings, operational latency, resilience, robustness in the face of loss of stations.

In addition, research is extending to other challenges: voltage management can be used to size the network differently compared to transit management, and this can therefore be integrated into modelling.

3/ AUTOMOBILE INDUSTRY

Digitization has brought down the consumption of vehicles. From 1990 to 2000, the average consumption of a car per km decreased by 6.5%. Over a longer period (1990-2010), the corresponding figures are -17% over the average consumption.
Dieselization has improved figures by another 15%. The introduction of controlled direct injection (1991) led to a 26% improvement in the consumption of a diesel car, thanks to the very targeted control of injection using increasingly complex computers and much greater pressure in the engine. This requires controlling and optimizing a large number of variables, one of which is consumption, along with emissions of pollution (nitrogen oxide and particles).
Electronic injection began much earlier for petrol. In 1984, 2Ko of memory was already being used, leading to a gain of 15%. Then direct injection arrived, bringing additional savings of 10%. Moreover, automatic transmissions consume 20% to 25% less than manual gear boxes.
The latest step in digitalization concerns optimization of the fifteen or so possible combinations involving the combustion engine, clutch box, and electric engine, known as hybrid motorization.
Digitalization has been beneficial for the entire automobile fleet, but has not prevented an increase in the proportion of CO_2 emissions from road transport in Europe.

Stricter emissions standards for new vehicles will result in an increasing share of electric or rechargeable hybrid vehicles, to the detriment of conventional combustion vehicles. Their share of the market is likely to reach 40% by 2030. However, energy consumption levels are unlikely to change as a result.

Nevertheless, the impact of shared vehicles could be high: the total number of km travelled could drop by 30% to 40%. Extreme simulation scenarios (e.g. Lisbon) show that maintaining trains and subways and replacing buses with robot-taxis, shared or otherwise, could lead to the same degree of mobility as today with a tenth of the fleet, a third of the number of vehicles on the road, and a number of parked vehicles divided by fifteen. The time spent in transport would also be reduced.

However, the perspective of shared self-driving vehicles remains distant, probably over ten years away, as mentioned in the Pichereau Report. The problem is to be able to guarantee that accidents will not occur. Since vehicles are technically piloted using artificial intelligence, enough experience needs to have been accumulated of the situations that a vehicle is likely to encounter. Hundreds of thousands of cars equipped with recorders are on the road, with variable tracking equipment and levels of autonomous control. To build up enough learning to avoid accidents, massive simulations are carried out using data from driving scenes that are either recorded or generated from "blank papers". These simulations also test how the system reacts to rare events.

The processors that control cars are neuromorphic. They are embedded, which is the only way to take driving decisions quickly enough. However, the statistic models that they employ have to be periodically refreshed, and only very powerful computers can provide them.

US manufacturers Tesla and Waymo are the most advanced. Tesla has chosen to develop its own processors and computers. Waymo is planning on producing a self-driving car equipped with 5 lidars (remote sensing and telemetry method similar to a radar), 27 cameras, 6 radars and enormous computing power.

In conclusion, digitization, which has revolutionized engines, will have a direct secondary impact on consumption. The major transformation will involve assisted driving. An extreme level of shared autonomous vehicles could considerably reduce the number of vehicles and kilometres travelled.

This perspective is still a long way off and will require technical progress coupled with social and political change. However, the shift has begun, already obliging carmakers to establish much closer relationships with the design of electronic components and computers.

In terms of R&D, the Pichereau Report[4] made noteworthy recommendations to increase efforts, especially in the following areas:

- AI (data, knowledge, algorithms and systems) for automated and connected vehicles, and autonomous transport vehicles and systems.
- Interactions between the driver and the automated system (e.g. CNRS).
- Development of secure, trustworthy, verifiable and explainable artificial intelligence.
- Cybersecurity and secure data exchanges for automated vehicles.
- Electric and electronic architecture of vehicles.
- Object detection software.

4- Le déploiement européen du véhicule autonome : Pour un renforcement des projets européens [European roll-out of autonomous vehicles: the need to strengthen EU projects], not translated, by Damien Pichereau

4/ AVIATION INDUSTRY

The main uses of digital are simulation, flight equipment, data analysis and predictive maintenance.

Simulation goes as far as getting aeroplane designs to “fly on the ground”. The time and cost savings in the design, and the increased energy efficiency of planes are considerable. However, the computing requirements are enormous, and Airbus uses some of the most powerful computers available. It is important to remember that this has a positive impact on the environment because it leads to more frugal aircraft.
Flight equipment comprises a growing number of electric engines that act on increasingly thin control surfaces.
Thanks to data analysis, Airbus can offer predictive maintenance services that are increasingly part of its activity, in the form of the Skyways platform, used by numerous airlines. Airbus has sufficient power to have maintained control of this platform, which it developed with Amazon.

The potential contributions of digital technology to reduce the carbon footprint of air traffic are not negligible, in particular regarding optimization of flight paths: one of the levers for decarbonizing the air sector is to precisely employ meteorological modelling to get the most out of winds and make considerable savings during flights. This optimization also requires high processing power.

5/ BUILDING AND LAND PLANNING

The combination of digital technology, construction and urban planning features several aspects:

- Building information modelling (BIM) is a system gathering all parties involved in a construction, including in the design, building, and control of the works. It can be defined as a digital model used to share information in real time in order to ensure that the production conforms with the design. The system can also be used to manage the building as efficiently as possible over time thanks to a data library. Digital twins can also be employed to model the building’s behaviour and consumption, or test out different architectural configurations (with the aim of improving design to correspond to climate and solar gains, or sizing constraints like seismic requirements). BIM also takes technical, legal and training constraints into account.
- Urban service platforms manage buildings and provide an opening to new urban operators. With smart meters, load shedding can be managed and done remotely to ensure the production and supply of electricity at the scale of the building (cables, etc.).
- The last aspect involves integrating buildings at the scale of a neighbourhood.

The smart grid has changed the scale and tends toward the neighbourhood. Feedback on experience shows that analysis and optimization work is required. In 2021, a building was controlled using a tool that recuperates energy production and consumption in real time.
Concretely, three levels need to be treated: the network infrastructure, the management of all information systems, and the urban planning design.

Digital and energy actors are working together with state services to pursue three objectives:

- a technological objective, concerning the network interconnection, e.g. for load shedding;
- an economic objective, to obtain a variable economic model for smart grids;
- a regulatory objective in order to, with partners, explain to state services how to develop energy economic models. On this point, several studies kicked off in 2015, and in 2021 a decree on regulations was adopted.

With the aim of controlling in real time and visualizing the consumption of a neighbourhood, the first step was carried out on a private parcel: an urban island comprising an office building, a residential building, and a high-rise residence. This stage of the project was delivered in 2015 and showed a 60% drop in overall consumption. The second step was at neighbourhood scale; geothermal power was employed and overall energy consumption dropped further still. The third example, in a neighbourhood of the city of Lyon, led to the concrete installation of an urban service platform at the level of an island. This platform is guaranteed over time by an urban contract.

In conclusion, three key subjects sum up the challenges and difficulties facing the new neighbourhood ecosystem: *the standardization of computer protocols, digital twins, and flexibility.*
Apart from electronics-computing, these applicative sectors reap all the benefits of digital technology:

- Computer-aided design, which saves time and money (e.g. in aeronautical engineering, the replacement of wind tunnels by simulators).
- Digital twins and maintenance.
- Company organization (EDF).

From an energy point of view, the main advantages come from users:

- Reduced consumption (cars, planes, buildings).
- More efficient usage: if autonomous cars become widespread, we will save 30% to 40% more energy.

“SMART GRID AND ECO-NEIGHBOURHOODS: FEEDBACK ON EXPERIENCE”
THE EXAMPLES OF BATIGNOLLES AND ISSYGRID

The challenges overcome in the Batignolles example were the following:

- Different actors from the digital, energy and building sectors were involved.
- The field of play was extended, with a residential and tertiary part covering almost 300,000 m2 .
- Investment in the project came from equity, with a very low budget for the price of a model, showing that large-scale operations can be done at a reasonable cost.

For Issygrid, although the oldest buildings were under ten years old, obtaining general information (from the electricity network, housing, residential) was subject to difficulties. For example, no less than 14 systems needed to be interconnected, and not all of them were standardized following the same protocol. This difficulty was overcome, despite technical issues and residents’ wariness of using energy monitoring (about 50% of inhabitants). The first load-shedding tests were controlled using the standardized digital system. Concerning the energy profile, renewable heat met 85% of requirements, compared to 50% in Paris.

With the aim of improving performance, sensors were added with the help of network operators: CPCU (urban network) Eau de Paris (deep geothermal), Enedis (electric grid), and social housing landlords. Thanks to this work, information could be obtained, even at the scale of private individuals (with consent agreements, 90% agreed). Then a digital model of the neighbourhood was built to understand annual and monthly consumption, which illustrated the successful operation of geothermal power; similarly for the district heating network, although the globalization did not work as well.

Another observation was that collective self-consumption is more efficient than individual self-consumption. This however calls for discipline in using these means at specific times and requires tracking inhabitants (here thanks to blockchains already in place).

05. CONCLUSION AND RECOMMENDATIONS

Digital technology will play a key role in the energy transition towards a low-carbon economy. Although also subject to the obligation of being energy efficient, the electricity consumption it requires, which involves increasingly lower levels of carbon, should not be an obstacle However, the considerable growth of computing and equipment requirements – which seem to be endless – is on a rising trend, for both manufacture and operations. The impact of manufacturing is not restricted to the electricity consumed. The entire chain is concerned, from the extraction of materials to recycling. Unfortunately, most of this manufacturing takes place outside Europe, particularly in Asia, but we can at least prescribe recycling and prohibit programmed obsolescence.
The introduction of digital technology is associated with the design of new products, new services, more efficient business organization, and the modification of uses. All of which reduce energy impacts, even if that is only one of the targets pursued by companies. These firms respect environmental requirements, and sometimes attempt to go further, but they are mostly concerned by their cost prices (including the cost of energy) and their clientele.

ELECTRONICS, IT, TELECOMMUNICATIONS

In this area, energy consumption is becoming an obstacle for production and operating systems, while computing and transmission needs continue to increase, and the miniaturization of integrated circuits is reaching its limits (Moore's Law).
Moore's Law, which suggests that computational progress becomes increasingly efficient and cheaper, and the industrial concentration that it has led to in terms of microprocessors, has fostered the separate development of software and equipment. This separation has triggered the impressive dissemination of digital technology in most activities, and the protection of users' application portfolio. Nevertheless, it appears that employing specialized computing devices implemented by classic electronics or quantum computing currently constitutes the only potential avenue for continuing to improve performance, and therefore the quantity of possible applications, along with controlled energy expenditure (see the report by the AdT: Les technologies matérielles supports du numérique future [the hardware technologies behind the digital future], not translated, February 2022). The downside is that software becomes dependent on hardware and has to explicitly take its physical constraints into account.
This requires a veritable revolution for integrated circuit manufacturers, systems architects and programmers. The introduction of parallelism means that 90% of software needs revising. At the same time, the development of artificial intelligence, and the employment of specialized hardware accelerators instead of general usage processors, impose much closer interactions between manufacturers and their clients.
This is a radical shift, but it is probably the price to pay for digital and energy to work together and offer sustainable progress to the whole of society. One of the most important aspects is the change required from programmers, in terms of skills, location, and working methods. This set of challenges will put European resources and competences to the test in all aspects of the game, with strong R&D requirements. For Europe, which has been lagging behind, it is a new opportunity, provided we know how to use it.

For integrated circuits, STMicroelectronics, the leading European manufacturer that closely relies on the CEA/LETI for research, is well positioned on the internet of things. Its objective is to attain 80% edge by 2030, integrating software into microprocessors by associating them with neural networks and adding memory. However, without massive support, ST will not be able to keep up with the race for ultimate miniaturization, led by the Asian champions: Samsung and TSMC.Intel had seemed to have given up, but recently decided to scale up and pursue a more open policy involving more dialogue with their clients, and probably the construction of a new factory in Europe. For the rest of the digital chain (parallel processors, servers, supercomputers, quantum operators), Europe must maintain and extend a strong level of support.

The European Processor Initiative has set a target for 2030 of 65% to 70% European supplies for all of its equipment; this target also includes energy consumption. The EuroHPC programme is aiming to maintain Europe and France within the top four global leaders of computing power. In addition, France is the only country in Europe that is active on the entire chain, from ST to Atos, and including Kalray and OVH. This gives the country a responsibility of leadership in constructing a common position.

In telecommunications, the event of the moment is the move to 5G, which will open the internet of things. Although more energy efficient, the inevitable considerable increase in traffic will need to be dealt with to avoid a hike in CO2 emissions. One way is to purchase low-carbon electricity, but, like for electronics as a whole, we will need to act on the electricity consumption of systems.

ELECTRICITY

It is mainly thanks to the introduction of digital technology that electricity production and supply are facing the introduction of intermittent renewable energy and the accompanying decentralization. These profound modifications, which affect technology, the organization of major operators and their relationships with government, and market regulations, require a great deal of research by operators and public research bodies.

EDF, along with its subsidiaries, wants to be the data operator to decarbonize the entire sector.

AUTOMOBILE

Electronics were introduced into cars thirty years ago with the idea of improving engine yields and comfort.

Developments have never stopped. However, we have reached a turning point that calls for new electronics employing artificial intelligence and clouds connected from the local edge in the car to the central cloud, i.e. autonomous cars. This shift is likely to profoundly change usages and bring to an end the rational utilization of equipment that currently spends too much time idle. With shared autonomous cars, we could save 30% to 40% of manufacturing energy. The changes made to manufacturing and operations will also be considerable; this is the move into the internet of things: more interactions with manufacturers of integrated circuits and sensors, and more local processing (edge computing). All of which requires new skills.

AVIATION INDUSTRY

Simulation can go so far as to get aeroplane designs to “fly on the ground”, saving considerable time and design costs and improving the energy efficiency of aircraft. The computing required is enormous, and Airbus employs some of the most powerful computers on the market.

The other major use of electronics is data analysis for predictive maintenance. Airbus is sufficiently powerful to have retained the control of the “Skyways” service platform used by numerous airlines, despite having developed it with Amazon.

BUILDING INDUSTRY

The move to digital twins (BIM) helps to organize a dispersed profession. At the same time, it raises issues regarding the standardization and ownership of data. One challenge is to design energy ensembles at neighbourhood scale, which can lead to a 60% reduction in consumption. Self-consumption is also more efficient than at the individual level. However, the risk is that data leaders could monopolize the relationship with local authorities and impose standardization based on profits rather than urban planning.

RECOMMENDATIONS

1/ Consider all of the causes of digital system consumption, from manufacturing to operations

- The large differences between studies come from shortfalls in analysis. To ensure that these studies can be genuine decision-making tools, this needs to be remedied. The solution is to standardize the measuring points of a digital system and how they result in an overall measurement. Research is needed, for example in the largely unexplored domain of the relationship between algorithms, in particular distributed algorithms, and consumption (where it is already clear that the quantity of peer-to-peer exchanges like those involved in bitcoin and blockchain algorithms cannot be made general, since this quantity grows like the square of the number of participants).
- This overall knowledge will enable an optimization of the cost-performance-environmental quality of systems, including the rules and protocols for exchanging data, partly thanks to integrating data centres.
- "Just enough" systems should be designed to consider the set of constraints as closely as possible.

2/ Anticipate the new interaction between hardware and software

The performance and volume required to process information will continue growing, at a point when the technology of integrated circuits is reaching the limits of miniaturization. The way forward is specialized operators, in other words close interaction between the manufacturers of devices and components. France needs to take advantage of this new set-up by:

- Investing in R&D on new-generation computer technologies.
- Training programmers to interact closely with hardware designers and to be able to renovate existing software, while ensuring overall cybersecurity.
- Redeveloping the software application portfolio to benefit from the contribution of specialized components and massive parallelism.
- Supporting neural networks, neuromorphic computing, and the interpretation of artificial intelligence, targeting assisted or shared assisted vehicles (Picherau Report), without neglecting socio-economic support.

3/ Vis-à-vis Europe, build solidarity and promote our strong points

Asians and North Americans are building up their forces with the objective of domination, or at least autonomy. For energy, the question is now clear, but it will take years to capitalize on our strong points. For digital technology, with the new hardware-software interaction, the future of a large part of our industries and services is in the balance. France has many assets, in particular its presence throughout the digital chain.

- Support the European Processor Initiative, setting at target of 65-70% European supply by 2030
- Similarly for the Euro HPC initiative
- Establish common practices and standards for cybersecurity in Europe. Regulate the diversity of national transcriptions of EU rules.
- Support the Climate Neutral Data Centre Pact
- Make recyclability an obligation and prohibit programmed obsolescence

ASSOCIATION NATIONALE RECHERCHE TECHNOLOGIE

33 rue Rennequin
75017 - PARIS
Tél : +33(0)1 55 35 25 50
com@anrt.asso.fr
www.anrt.asso.fr