



HAL
open science

The HiPEAC Vision 2017

Marc Duranton, Koen de Bosschere, Christian Gamrat, Jonas Maebe, Harm Munk, Olivier Zendra

► **To cite this version:**

Marc Duranton, Koen de Bosschere, Christian Gamrat, Jonas Maebe, Harm Munk, et al. (Dir.). The HiPEAC Vision 2017. Duranton, Marc; De Bosschere, Koen; Gamrat, Christian; Maebe, Jonas; Munk, Harm; Zendra, Olivier. HiPEAC network of excellence, pp.138, 2017, 978-90-9030182-2. hal-01491758

HAL Id: hal-01491758

<https://inria.hal.science/hal-01491758v1>

Submitted on 17 Mar 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HiPEAC Vision 2017

HIGH PERFORMANCE AND EMBEDDED ARCHITECTURE AND COMPILATION

Editorial board:

Marc Duranton, Koen De Bosschere,
Christian Gamrat, Jonas Maebe,
Harm Munk, Olivier Zendra

This document was produced as a deliverable of the H2020 HiPEAC CSA under grant agreement 687698.

The editorial board is indebted to Dr Max Lemke and to Dr Sandro D'Elia of the Technology & Systems for Digitising Industry unit of the Directorate-General for Communication Networks, Content and Technology of the European Commission for their active support to this work.

Design: www.magelaan.be

Source cover picture: Willyam Bradberry/Shutterstock

© 2017 HiPEAC

January 2017

ISBN 978-90-9030182-2

FOREWORD

HiPEAC's domain (High Performance and Embedded Architecture and Compilation) – computers, programmable systems, processors, microcontrollers etc – is evolving very rapidly and has a far-reaching impact on our everyday life. The aim of this 2017 HiPEAC Vision is to highlight some of the ongoing evolutions in this domain and to outline a number of recommendations to steer it. The main findings of the Vision can be summarized as follows:

The computer is disappearing from view, yet is becoming embedded in the very fabric of everyday life. Change can be seen at every level: not only in the content of our interactions with computers, which are evolving from answers to computational problems to interactions through social networks, but also in the forms in which we interact, which are shifting from letters and digits to sounds, gestures, images and movies.

Yet we expect more change to come: programming will become learning for the machine, and interactions with computers will be augmented by virtual and augmented reality and modelled as interactions between humans. All this is made possible by the use of Artificial Intelligence-based techniques. We expect systems not only to observe, but also to interact with the physical world and to control it. The most visible developments by early 2017 are Intelligent Personal Assistants and Advanced Driver Assistance Systems; the latter is evolving into the autonomous driving car.

These developments pose significant challenges to the HiPEAC community.

Today, we need systems to be secure. Tomorrow's systems must take this a step further: we must be able to trust them, that is to say, their control of the physical world must be within safe limits, not causing harm and keeping private data private.

Questions relating to energy are more of a challenge than ever before: we mainly rely on the cloud for compute-intensive tasks, which in itself demands energy efficient servers to keep its carbon footprint within acceptable limits. However, in order to stem the flood of data from the Internet of Things, we must employ intelligent local data processing on remote devices that use minimal energy. Perhaps we do not need the bleeding edge technology node for these devices but can make do with 'older',

but more cost-efficient IC-technology. This may well require us to break away from the traditional Von Neumann architecture and to rethink device technology.

As the landscape of devices and systems changes and the application of these systems quickly broadens into cyber-physical systems, it is the right time not only to evolve but also to completely reinvent the basic concepts of computing.

In systems design, to master complexity we need methodologies which enable composability and interoperability of components, and we certainly need to add AI-based techniques and tools to help us to achieve this. We should be able to design predictable systems from unreliable components, for example, by using run-time assistance to guarantee reliability.

For all these challenges, one very important principle remains: we can only achieve efficient solutions if we adopt a holistic approach to the development of cyber-physical systems, an approach in which all disciplines come together and are regarded as first-class citizens.

Yet in Europe, we seem to be facing these challenges with an ICT workforce whose growth is not keeping pace with change, and with a level of investment that is lower than that of China, Japan and the US.

This document identifies several areas of possible improvement to help Europe face the challenges to come. One key route to improvement would be a more complete coverage of the aspects of the development of new solutions. We observe that there is insufficient financial stimulation following the initial development phase, with the scaling-up phase of a new initiative receiving less support. We also note that, although short-term and long-term developments are sufficiently supported through project funding, middle-term (five year) developments, the key grounding stage success for all innovations, are not. We also remark that the fragmentation of funding is harmful to progress.

'Grand challenge'-style competitions, well-balanced in their objectives, may be a way to attract more companies and new students to the domain, and to bring the role and importance of ICT to the attention of the general public in Europe.

CONTENTS

INTRODUCTION	4		
EXECUTIVE SUMMARY	6		
1. PART 1: RECOMMENDATIONS	9		
1.1 ON COGNITIVE SYSTEMS	9		
1.2 GENERAL CHALLENGES	9		
1.2.1 GUARANTEE TRUST	9		
1.2.2 IMPROVE PERFORMANCE AND ENERGY EFFICIENCY	9		
1.2.3 MASTER COMPLEXITY	9		
1.3 TECHNICAL CHALLENGES	10		
1.3.1 ENSURE SECURITY, SAFETY AND PRIVACY	10		
1.3.2 MASTER PARALLELISM AND HETEROGENEITY	10		
1.3.3 LEVERAGE PREDICTABILITY BY DESIGN	10		
1.4 INCREASE HOLISTIC VIEW	10		
1.5 REINVENT COMPUTING	10		
1.6 INCREASE THE EUROPEAN ICT WORKFORCE	11		
1.7 RESEARCH POLICY RECOMMENDATIONS	11		
1.7.1 ON COMMERCIALIZING THE RESULTS OF EU PROJECTS	11		
1.7.2 ON PRODUCT DEVELOPMENT LIFECYCLE SUPPORT	11		
1.7.3 ON GRAND CHALLENGES	11		
1.7.4 ON SYNERGIES WITH DATA INFRASTRUCTURES, HPC AND COMMUNICATIONS	11		
1.7.5 ON ECOSYSTEM BUILDING	11		
2. PART 2: RATIONALE	13		
2.1 INTRODUCTION: ENTERING IN THE CENTAUR ERA	13		
2.2 STRUCTURE OF THE DOCUMENT	14		
2.3 SOCIETAL TRENDS	15		
2.3.1 EVOLUTION OF SOCIETY	15		
2.3.2 SECURITY CHALLENGES	16		
2.3.3 PRIVACY EROSION	17		
2.3.4 ENVIRONMENTAL DEGRADATION AND THE ENERGY CHALLENGE	18		
2.3.5 EDUCATION	20		
2.3.6 SELF-SUFFICIENCY IN ICT	22		
2.3.7 AGEING POPULATION OF EUROPE	23		
2.3.8 THE SHARING ECONOMY	24		
2.3.9 EFFECTS OF DIGITAL TECHNOLOGY ON THE BRAIN	24		
2.3.10 JOB MARKET CHANGES	26		
2.4 MARKET TRENDS	28		
2.4.1 GENERAL TRENDS	28		
2.4.2 CANNIBALIZATION OF DISCRETE DEVICES	29		
2.4.3 SATURATION OF THE PC AND MOBILE PHONE MARKETS	29		
2.4.4 FOG AND EDGE COMPUTING	30		
2.4.5 THE NEW PROCESSOR LANDSCAPE	30		
2.4.5.1 SMARTPHONES DRIVE THE DEVELOPMENT OF PROCESSOR ARCHITECTURE	31		
2.4.5.2 VERTICALIZATION IS STILL CONTINUING	31		
2.4.5.3 SEVERAL COUNTRIES DEVELOP THEIR OWN PROCESSORS	32		
2.4.5.4 THE 'MAKER' OR DO-IT-YOURSELF MOVEMENT	35		
2.4.5.5 IMPACT AND PROPOSED COURSE OF ACTION	36		
2.4.6 OPEN SOURCE HARDWARE AND SOFTWARE	37		
2.4.6.1 OPEN SOURCE SOFTWARE	37		
2.4.6.2 OPEN SOURCE HARDWARE	37		
2.4.6.3 IMPACT AND PROPOSED COURSE OF ACTION	38		
2.4.7 CRYPTOGRAPHY AND SECURITY	38		
2.4.7.1 CLASSICAL CRYPTOGRAPHY	38		
2.4.7.2 HOMOMORPHIC ENCRYPTION	39		
2.4.7.3 POST-QUANTUM CRYPTOGRAPHY	39		
2.4.7.4 BLOCKCHAIN AND NEW DIGITAL CURRENCIES	40		
2.4.7.4.1 BEYOND BITCOIN	41		
2.4.7.4.2 THE INFLUENCE OF BITCOIN ON COMPUTING	41		
2.4.7.5 SIDE CHANNEL ATTACKS	41		
2.4.7.6 IMPACT AND PROPOSED COURSE OF ACTIONS	42		
2.4.8 COMMUNICATION/RELATION WITH THE ENVIRONMENT: VIRTUAL AND AUGMENTED REALITY	42		
2.4.8.1 AUTOMATIC TRANSLATION	42		
2.4.8.2 HEAD-UP DISPLAYS IN CARS	43		
2.4.8.3 AUGMENTED REALITY GAMES	43		
2.4.8.4 SMART GLASSES	43		
2.4.8.5 GAMES WILL BECOME INDISTINGUISHABLE FROM REALITY	44		
2.4.8.6 IMPACT AND PROPOSED COURSE OF ACTIONS	45		
2.4.9 FROM IOT TO CPS TO 'DISAPPEARING COMPUTER'	45		
2.4.9.1 IOT AND CPS	45		
2.4.9.2 HIGH PERFORMANCE COMPUTING: AN ENABLER FOR NEW CPS APPLICATIONS	48		
2.4.9.3 SELF-DRIVING VEHICLES AND DRONES	49		
2.4.9.3.1 SELF-DRIVING VEHICLES	49		
2.4.9.3.2 DRONES	51		
2.4.9.4 IMPACT AND PROPOSED COURSE OF ACTIONS	51		
2.4.10 HEALTH AND AUGMENTED HUMAN: TOWARDS 'CYBORGS'?	52		
2.4.10.1 ON-BODY CYBER-PHYSICAL SYSTEMS	52		
2.4.10.2 IN-BODY CYBER-PHYSICAL SYSTEMS	52		
2.4.10.3 IMPACT AND PROPOSED COURSE OF ACTIONS	53		
2.4.11 ARTIFICIAL INTELLIGENCE AND COGNITIVE, SMART DEVICES	53		
2.4.11.1 DEEP LEARNING	55		
2.4.11.2 DATA ANALYTICS	58		
2.4.11.3 IBM'S WATSON AND NATURAL MAN-MACHINE INTERACTION	58		
2.4.11.4 THE 5 TH RESEARCH PARADIGM?	59		
2.4.11.5 THE DECLARATIVE OR PARENTING SYSTEMS	60		
2.4.11.6 OUR NEW ASSISTANTS: VIRTUAL OR ROBOTS	60		
2.4.11.6.1 VOICE CONTROLLED PERSONAL ASSISTANTS	60		
2.4.11.6.2 COMPANION ROBOTS	61		
2.4.11.7 IMPACT AND PROPOSED COURSE OF ACTIONS	62		
2.5 TECHNOLOGICAL TRENDS	64		
2.5.1 TIME TO REVISIT THE BASICS: VON NEUMANN, NEURAL			

NETWORKS AND QUANTUM COMPUTING	64	2.6.1.3 EUROPEAN EXASCALE SOFTWARE INITIATIVE	97
2.5.2 THE SILICON ROADMAP	65	2.6.1.4 RETHINK BIG	98
2.5.2.1 CURRENT STATUS	65	2.6.1.5 ECSEL	98
2.5.2.2 WHAT IS COMING NEXT?	67	2.6.1.6 ARTEMIS SRA	99
2.5.3 THE NON-SILICON ROADMAP	73	2.6.1.7 NEXT GENERATION COMPUTING ROADMAP	99
2.5.3.1 PRINTED ELECTRONICS	73	2.6.1.8 EUROLAB-4-HPC	100
2.5.3.1.1 THE ROADMAP OF THE ORGANIC AND PRINTED ELECTRONICS ASSOCIATION	73	2.6.1.9 CPSOS RESEARCH AND INNOVATION AGENDA	101
2.5.3.1.2 MARKET OF PRINTED ELECTRONICS	73	2.6.1.10 ICT-ENERGY	101
2.5.3.1.3 PRINTED SENSORS AND LOGIC	73	2.6.1.11 ITRS	102
2.5.3.1.4 PRINTED INTEGRATED SMART SYSTEMS (ISS)	73	2.6.2 ACTIONS IN OTHER COUNTRIES	102
2.5.3.1.5 IMPACT AND PROPOSED COURSE OF ACTIONS	74	2.6.2.1 US - SEMICONDUCTOR RESEARCH CORPORATION (SRC)	102
2.5.3.2 NON-VOLATILE MEMORY TECHNOLOGIES	74	2.6.2.2 US - NATIONAL STRATEGIC COMPUTING INITIATIVE	102
2.5.3.3 CARBON-BASED TECHNOLOGIES	75	2.6.2.3 US - A NANOTECHNOLOGY-INSPIRED GRAND CHALLENGE	103
2.5.3.4 QUANTUM COMPUTING (QC)	76	2.6.2.4 US - NATIONAL SCIENCE FOUNDATION (NSF)	103
2.5.3.4.1 UNITARY QC VS. SIMULATED QC	76	2.6.2.5 JAPAN	103
2.5.3.4.2 RECENT BREAKTHROUGHS	76	2.6.3 EUROPEAN POSITION (SWOT)	104
2.5.3.4.3 POST-QUANTUM CRYPTOGRAPHY AND MACHINE LEARNING	76	2.6.3.1 STRENGTHS	104
2.5.3.4.4 PROGRAMMING THE QUANTUM	76	2.6.3.1.1 HIGH QUALITY EDUCATION	104
2.5.3.4.5 THE EUROPEAN QUANTUM COMPUTING INITIATIVE	77	2.6.3.1.2 LARGE NUMBER OF PHDS	104
2.5.3.5 IMPACT AND PROPOSED COURSE OF ACTIONS	77	2.6.3.1.3 LARGEST PUBLICATION AND CITATION COUNT OF THE WORLD	104
2.5.4 THE ARCHITECTURE ROADMAP	77	2.6.3.1.4 RESEARCH AND TECHNOLOGY ORGANIZATIONS (RTOS)	105
2.5.4.1 ARCHITECTURES FOR PREDICTABLE COMPUTING	77	2.6.3.1.5 LARGEST MARKET IN THE WORLD	106
2.5.4.2 ACCELERATORS	78	2.6.3.1.6 LARGE EMBEDDED MARKET	107
2.5.4.2.1 RECONFIGURABLE ACCELERATORS	78	2.6.3.1.7 COMMON MARKET	107
2.5.4.2.2 ACCELERATORS FOR NEURAL NETWORKS	80	2.6.3.1.8 VARIETY OF RESEARCH FUNDING INSTRUMENTS	107
2.5.4.2.3 QUANTUM ACCELERATORS	81	2.6.3.1.9 DECENT LEVEL OF PUBLIC FUNDING OF R&D	107
2.5.4.3 IN-MEMORY COMPUTING	81	2.6.3.2 WEAKNESSES	107
2.5.4.4 IMPACT AND PROPOSED COURSE OF ACTIONS	81	2.6.3.2.1 WEAK ACADEMIA-INDUSTRY LINK	107
2.5.5 EVOLUTION OF MASS STORAGE	82	2.6.3.2.2 STRONG IN RESEARCH, BUT NOT IN COMMERCIALIZATION	109
2.5.5.1 MAGNETIC DISKS	82	2.6.3.2.3 EU ICT CONTRIBUTES LESS TO THE GDP THAN IN OTHER ADVANCED COUNTRIES	110
2.5.5.2 SOLID STATE DISKS (SSD)	83	2.6.3.2.4 EUROPE LACKS ADVANCED FOUNDRIES	112
2.5.5.3 IMPACT AND PROPOSED COURSE OF ACTION	83	2.6.3.2.5 EUROPE LACKS VENTURE CAPITAL CULTURE	113
2.5.6 EVOLUTION OF COMMUNICATION	84	2.6.3.2.6 LACK OF ICT-WORKERS	113
2.5.6.1 WIRELESS COMMUNICATIONS.	84	2.6.3.2.7 FRAGMENTATION OF FUNDING	113
2.5.6.2 WIRED COMMUNICATIONS	85	2.6.3.3 OPPORTUNITIES	114
2.5.6.3 THE TRIANGLE: COMPUTATION - COMMUNICATION - STORAGE	85	2.6.3.3.1 THE END OF MOORE'S LAW	114
2.5.6.4 IMPACT AND PROPOSED COURSE OF ACTIONS	86	2.6.3.3.2 EMBEDDED SYSTEMS, IOT, CPS	114
2.5.7 A NEW SOFTWARE CRISIS AND COURSE OF ACTION	86	2.6.3.3.3 CYBERSECURITY	114
2.5.7.1 A CHANGING LANDSCAPE	86	2.6.3.3.4 SOLUTIONS FOR SOCIETAL CHALLENGES	114
2.5.7.1.1 A CHANGING DEVICES LANDSCAPE	86	2.6.3.4 THREATS	114
2.5.7.1.2 A CHANGING SYSTEMS LANDSCAPE	87	2.6.3.4.1 FINANCIAL CRISIS	114
2.5.7.1.3 A CHANGING USERS LANDSCAPE	88	2.6.3.4.2 SATURATING MARKETS	114
2.5.7.1.4 A CHANGING DEVELOPMENT TOOLS LANDSCAPE	88	2.6.3.4.3 COMPUTING INITIATIVES IN CHINA, RUSSIA, JAPAN	115
2.5.7.2 THE PRODUCTIVITY CHALLENGE	88	2.6.3.4.4 CHINA IS BUILDING A HUGE PATENT PORTFOLIO	116
2.5.7.3 THE CORRECTNESS CHALLENGE	90	2.6.3.4.5 POLITICAL INSTABILITY (BREXIT, IMMIGRATION CRISIS, ...)	116
2.5.7.4 THE PERFORMANCE CHALLENGE	92		
2.5.7.5 THE DATA CHALLENGE	93		
2.5.7.6 THE HOLISTIC CHALLENGE	95		
2.5.7.3 IMPACT AND PROPOSED COURSE OF ACTIONS	96		
2.6 THE POSITION OF EUROPE IN THE WORLD	97	3. GLOSSARY	117
2.6.1 OTHER ROADMAPS	97	4. REFERENCES	121
2.6.1.1 ETP4HPC	97	5. PROCESS	135
2.6.1.2 PRACE SCIENTIFIC CASE	97	6. ACKNOWLEDGEMENTS	136

INTRODUCTION

This document deals extensively with ‘societal trends’, ‘market trends’ and ‘the position of Europe in the world’; these are the areas in which we have found the most changes since the HiPEAC Vision 2015. The ‘technological trends’ sections only contain the new topics not detailed in the previous version of the document.

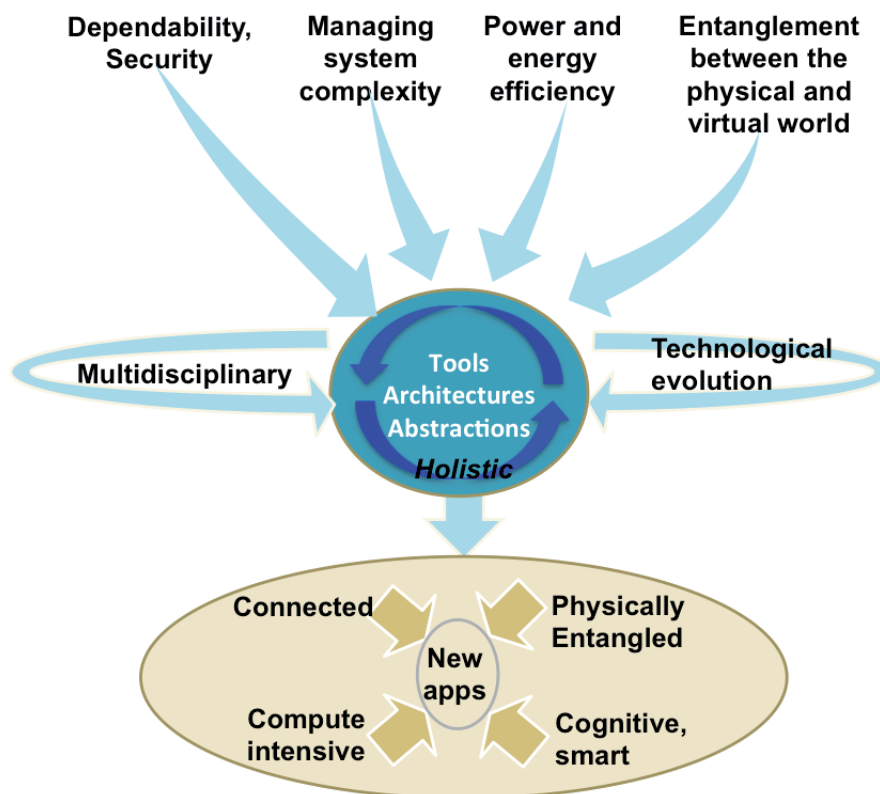
This document could be used as input for the Horizon 2020 calls in 2018/2019, leading to research activities up to 2022-2023, and to commercial exploitation in the 2025-2030 timeframe. Hence, the challenge is to try to find out today which type of computing systems, new approaches, new technologies, new systems and applications will be important ten years from now!

REMINDER: THE RECOMMENDATIONS OF THE HIPEAC VISION 2015

The ultimate goal of our community is to develop Tools, Architectures and Abstractions to build the next generation of killer applications. These applications will be characterized by four elements:

- They will be compute-intensive, i.e. they will require efficient hardware and software components, irrespective of their application domain: embedded, mobile or data centre;
- They will be connected to other systems, wired or wireless, either always or intermittently online. In many cases they will be globally interconnected via the Internet;
- They will be physically entangled, which means that they will not only be able to observe the physical environment they are operating in, but also be able to control it. They will become part of our environment;
- They will be smart, able to interpret data from the physical world even if that data is noisy, incomplete, analog, remote, etc.

All future killer applications will have these four characteristics, albeit not all to the same extent.



Our community has to provide the next generation of tools, architectures and abstractions required to build these killer applications efficiently and correctly. Building them will require taking into account several non-functional requirements such as energy, time, security and reliability. New computational models will be needed in some cases, such as neuromorphic architectures, Bayesian computing, pseudo-quantum computing, and statistical/probabilistic computing.

Potential ways to tackle these challenges are to:

- Develop approaches (= architectures, tools, abstractions) that take into account **non-functional information** (such as temperature, energy consumption and management, wear, ageing, errors) at all levels of the applications, making it possible to make decisions throughout all levels of the stack rather than only at the lower levels. This should be performed in a way that ensures a high level of interoperability (thereby developing – de-facto – standards) and security (keeping in mind potential misuse of this information);
- Develop methodologies that enable **combining multiple computation paradigms** in a single system (e.g. Von-Neumann, streaming, distributed, reactive, neuromorphic, Bayesian computing, pseudo-quantum computing, statistical/probabilistic computing). In particular, these methodologies must ensure quality, testability and reliability of the results of these systems;
- Further develop the path of **reactive systems**, in particular by applying knowledge from the cybernetics domain concerning the use of feedback loops to stabilize dynamic, complex systems;
- Develop formalisms, methodologies and tools for “adequate precision computing”, or more generally to deal with a “**desired level of Quality of Service**”: tools that take into account power, security, and time, which use the design-by-contract paradigm, which can reduce over-specification, and which can be used both with predictable and with reactive systems;
- Develop approaches that enable domain specialists to express only **what** needs to be done, while tools and computers take care of **how** to transform this knowledge into an efficient computing system representation and execution;
- Further develop design space exploration methods and tools;
- Develop approaches to validate complex systems composed of black or grey box components.

Due to the complexity and the large variety of problems, a single tool from one provider will certainly not solve the problem. Instead, we need a coherent framework of interoperable and complementary tools from multiple sources. For example, Eclipse [151] has created such an ecosystem.

EXECUTIVE SUMMARY

The Information and Communications Technology (ICT) domain is evolving rapidly and, in some areas, the lifetime of products is shorter than the time between two HiPEAC Vision documents. Each time we start a new version of the document, we expect a minor evolution from the previous document, and each time we discover that it was an incorrect assumption. This time it has been true again: while the main challenges of the HiPEAC Vision 2015 remain valid and have even increased in importance, new challenges are ahead of us.

The insights of this HiPEAC Vision 2017 are summarized in Figure 1:

Computers are disappearing from our view. They take on new forms, not only those of smartphones and tablets, but also as cars, smart meters, thermostats, and so on. They communicate

with their users not only through keyboards and alphanumeric display screens, but also using voice, sound, pictures and video, closely resembling human interaction. **We are entering the Artificial Intelligence (AI) era.** This will not only change how we interact with machines, but it will also redefine how we instruct a machine what to do: less programming and more learning.

The function of the computer is shifting, from computational tasks providing answers to numerical problems, to humans and computers working together (what we call **the beginning of the Centaur Era***), computers augmenting reality to assist humans, or even creating virtual worlds for humans to explore: the **cyber-physical entanglement** between the physical and virtual world. Computers will increasingly interact with the physical world, not only to get information from it, but also to control it. Such systems are called Cyber-Physical Systems (CPS). There is a **expan-**

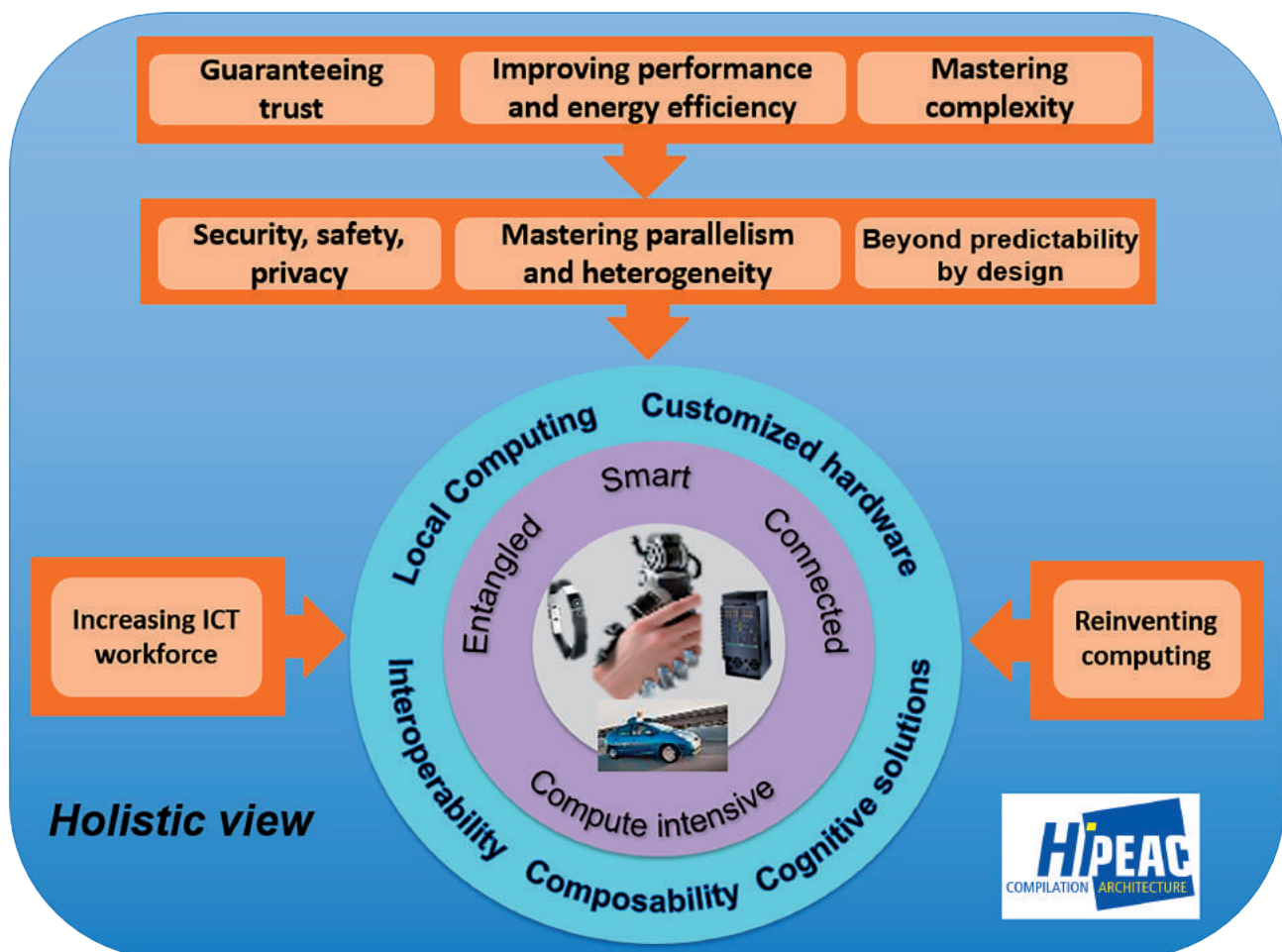


Figure 1: Main challenges and recommendations of the HiPEAC Vision 2017

* In Advanced Chess, a 'Centaur' is a man/machine team. Advanced Chess (sometimes called cyborg chess or centaur chess) was first introduced by grandmaster Garry Kasparov, with the objective of a human player and a computer chess program playing as a team against other such pairs (from Wikipedia).

sion from security into safety and trustability: because of the direct control of physical devices, a malfunction of a computer, due to a programming error, hardware failure or a hacker, could have lethal consequences. Humans need to trust the machines, not only by behaving in a correct and predictable way, but also by keeping sensitive information about the human confidential. Therefore, **enforcement of security and privacy** are of paramount importance.

For compute-intensive tasks, we will still be using the cloud, and that means that connectivity is crucial, yet **local processing** is also becoming increasingly important to stem the flood of data produced by the Internet of Things (IoT) or to cope with the constraints of safety or privacy. The increasing computational requirements are making computer system architects look for **accelerators for specialized tasks, diverting** in many cases from the traditional **Von Neumann architecture**.

Energy efficiency of computing systems remains a major challenge for the coming years, and not only to decrease their environmental footprint: without a significant improvement in energy efficiency, Exaflops computers will not be economically viable and the myriad of small (battery-powered) computing devices will not be successful due to their lack of autonomy.

As the cost per transistor is no longer decreasing, and even appears to be rising, we might see **diversified tracks for using silicon technology:** many designs will not use the latest technology node, but the more mature (and cheaper) one. Only high performance systems will require a very expensive and aggressive state of the art technology node. It is also the right time to **revisit the assumptions** that drove the semiconductor and computer industry for decades, and to challenge its explicit and implicit assump-

tions in order to open new tracks and new approaches and to **eventually reinvent computing**.

With the flood of new systems and new system architectures, increasing attention must be paid to **composability and interoperability** between systems. The complexity of the new systems will be so high that human designers will only be able to master it with the help of computers using AI-based techniques. Innovative approaches will be required to **ensure that the systems will do what they are supposed to do**, both at the functional and at the non-functional level (e.g. timing requirement and reliability). We need to develop design techniques that go **beyond predictability by design** and allow the **building of reliable systems from unreliable parts**.

Holistic approaches, implying multi-disciplinary techniques, will be needed in order to meet all the requirements in term of trustability, efficiency and cost.

Yet, at the same time, we notice in Europe a lack of new ICT workers slowing down innovative digital initiatives. The contribution of ICT to the GDP in Europe is significantly less than that in the US, Japan or China. Compared to other regions, less venture capital is made available in Europe for information technology.

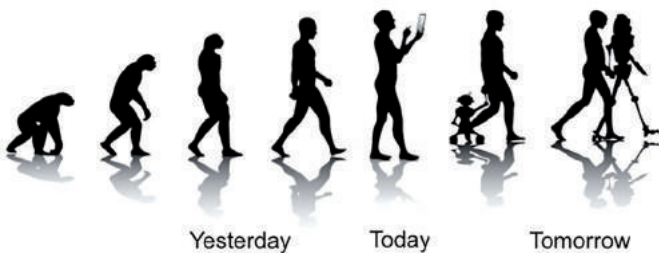
Related to these trends, we notice that European citizens have little understanding of information technology (IT), and little appreciation for its societal values.

The following sections will provide more details of the challenges, actions and recommendations resulting from the analysis of the societal trends, the market trends, the evolution of the technology and the position of Europe in the world that are detailed in Part II of this document.

PART 1: RECOMMENDATIONS

1.1. ON COGNITIVE SYSTEMS

Cognitive systems will be a cornerstone of the next generation of smart systems and at the core of many future businesses. The HiPEAC community will not directly develop them, but will use them and help end users to use them. We are concerned with providing the basic technologies (software ‘engines’, hardware accelerators and tools, e.g. machine learning for compilation). Cognitive systems will have an impact on hardware design (especially for embedded systems/edge computing) and we should develop the next ‘machine’, a computing platform for optimally supporting the future kind of AI-based smart applications. We should also apply AI techniques to help solving our problems, e.g. managing complexity, writing correct code and even generating new hardware architectures using a process similar to *generative design*.



Source: Dreamstime, Dimitri Skvorcov

1.2. GENERAL CHALLENGES

1.2.1. GUARANTEE TRUST

Challenge: Develop systems for real-time/safety-critical applications; trustable systems

We must not only be able to trust the systems we build (security and privacy), but also be able to trust that the systems we depend on reach their assigned goals within safe boundaries (safety).

Solutions: Investigate runtime mechanisms to ensure that systems will reach the assigned goal within (safe) borders. New technical solutions should be developed to ensure privacy and keep personal data confidential, while being able to use that data to enhance services and applications. In order to achieve all these challenges, new concepts are required. This challenge requires hardware/software co-design.

Remarks: This should be at the core of the HiPEAC community: to provide hardware and software that guarantee performance for real-time and safety-critical applications. A lot of progress has already been made in this domain, but perhaps more “re-thinking

the basics” will be required to define more substantial and sustainable solutions. For example, one potential approach could be to add a real-time mode, or safety-critical mode, where performance can be traded for guaranteed requirements.

1.2.2. IMPROVE PERFORMANCE AND ENERGY EFFICIENCY

Challenge: Performance is, besides speed, composed of non-functional aspects such as power and energy consumption, reliability, time requirements, etc. The performance of all these aspects is as important as that of speed. Low power and reduced energy consumption remain challenging (for HPC, but especially for IoT). If the number of internet devices is going to triple by 2020, so will the energy needed to run all these devices [421]. Energy is also a major challenge for servers, the number and size of which will increase so as to handle all the increasing computing needs. One key challenge, besides reducing energy consumption, is to ensure energy proportionality with the computing load.

Solutions: The basic concepts of mainstream computing systems need to be revisited in view of the new requirements, such as computing system power and reaction time guarantees. A holistic view of the complete compute stack is required to apply the solutions at all levels. This is particularly true for energy, which should be considered at system level (e.g. making local computations instead of transferring data; communication is the main source of waste of energy), device level (use of more energy efficient architectures, perhaps as dedicated coprocessors), language level (avoiding data communication, ensuring locality of data), and even at the algorithmic level (using *adequate computation* models, for example avoiding floating point computations where integer computations are sufficient).

1.2.3. MASTER COMPLEXITY

Challenge: Hardware and software complexity is growing, especially with heterogeneous systems. It threatens the understanding of the systems we are designing. An increasing number of systems is already considered *no longer completely understandable* [433].

Solutions: Cognitive systems are a potential solution to mitigate the complexity issue. Rethinking the basics of computing may also be a way to solve it. Interoperability (standardization) and composability are essential ingredients: systems should dynamically exchange their capabilities and should be able to adapt their interface to the other system’s needs.

Remarks: The IT community should embrace techniques of cognitive systems for solving its own problems, e.g. managing code generation, complexity of heterogeneous systems and designing

optimized architectures (with approaches similar to “*generative design*”).

1.3. TECHNICAL CHALLENGES

1.3.1. ENSURE SECURITY, SAFETY AND PRIVACY

Challenge: With an estimated cost of more than US\$2 trillion per year by 2019 [257], security is going to become more critical a challenge than ever before. As ICT will move to CPS, safety requirements will become also more and more preeminent.

Solutions: Increase holistic support for security. Fundamental hardware and software solutions for providing system-wide security should be further developed, including using the last innovations in cryptography (like *homomorphic encryption*). New approaches should be developed to protect the privacy of users, but still allowing the service providers to enhance their services (e.g. by using *differential privacy* techniques).

Remarks: If criminals hack our computers, it might be in part because we do not have the required number of people to secure them. Security should also be one of the preoccupations of the HiPEAC community, albeit at the compute node level and in terms of low-level tools.

1.3.2. MASTER PARALLELISM AND HETEROGENEITY

Challenge: This challenge has several sides:

- **The rise of accelerators:** In the context of the ever-more numerous heterogeneous systems, accelerators are now becoming a major, booming area and their use is likely to expand in the near future. Indeed, they provide an effective way of responding to the ever-increasing demand for more computing power, especially with regards to AI, as well as a way to somewhat limit the increase in complexity. We believe non-Von Neumann type accelerators, alongside more traditional processors, are going to be a key strategic asset in the future.
- **Interoperability and composability:** Systems must be able to interoperate, while guaranteeing correct operation at the system level. However, system architectures are becoming deeply heterogeneous, as they include processors, coprocessors, hardware modules that have a fixed function alongside ones that are reconfigurable, the latter either at coarse-grained level (CGRAs) or fine-grained level (FPGAs), etc. Ensuring their interoperability and composability is a significant challenge that requires new research.

Solutions: It is important to develop approaches that allow to easily combine different computational models, and a smooth transition from one model to another, both at the software level and at the hardware level (coprocessors).

¹ Generative design is a technology that starts with design goals and then explores all of the possible permutations of a solution to find the best option. The process lets designers generate brand new options, beyond what a human alone could create, to arrive at the most effective design.

In the ever-increasing quest for computing power, *in-memory computing*, through specific applications, frameworks and strategies can be an efficient and cost-effective way of improving the performance of ICT systems, especially related to Big Data.

Remarks: We recommend research supporting accelerator development, accelerator applications, and in-memory computing, with strong EU industrial involvement, including CGRAs and neural network (*Deep Learning*) accelerators.

1.3.3. LEVERAGE PREDICTABILITY BY DESIGN

Challenge: Increase transparency and predictability in ICT. Predictable computing is becoming necessary to overcome the software crisis and solve the various related challenges.

Solutions: Increase transparency across all levels of ICT systems. This is far from being the case at present. We thus consider it important to support research efforts in the direction of predictable computing, both at the conceptual and tooling levels, including architectures. Due to the complexity challenge, predictability should perhaps not be absolutely pursued at the conception time, but enforced dynamically during the lifetime of the systems (“*Beyond predictability by design*”).

1.4. INCREASE HOLISTIC VIEW

Challenge: The requirements for ICT systems increase in all domains of performance: speed, time predictability, power, energy, trustability. These conflicting requirements add to an increasingly complex landscape, especially with regards to heterogeneity in ICT systems.

Solutions: The only way to practically cope with this complexity, to fulfil these requirements and to create optimized systems is to create and develop system development methods and tools that take into account the whole range of constraints, functional as well as non-functional (time, power, energy, trustability) as first-class citizens, in a holistic yet transparent and manageable manner, across all layers, software as well as hardware. A multidisciplinary approach is important to gather all the knowledge required. Tools inspired from cognitive computing can be developed to help find optimal solutions from the numerous constraints.

1.5. REINVENT COMPUTING

As the classical Von Neumann model of computing loses its validity due to hardware developments in diverse areas such as processor architectures, memories, accelerators and communication, the accompanying software models lose their validity as well. The complexity of developing ICT systems becomes the source of another software crisis. Therefore, we think it is time to revisit the basic concepts of computing which is an update of the previous “We need new solutions”. For example, computing systems were not originally designed to dynamically interact with the world in a natural manner. Similarly, more and more “non-Von Neumann” solutions (like Deep Learning accelerators, Quantum machines) will co-exist side by side with more classical architectures.

We recommend rethinking the 'traditional' software stack, from hardware driver through operating system to application. We believe it is time to revisit and re-assess its fundamental assumptions in the light of all of its elements.

The new ICT landscape, with the end of 'Dennard scaling', the complexity challenge, the trustability challenge, and the strong rise of CPS that entangle two worlds, requires us to revisit the basic concepts both of hardware and software. This will offer a strong and unique opportunity for disruptive changes where Europe can (re)take the lead.

1.6. INCREASE THE EUROPEAN ICT WORKFORCE

It is vital to significantly increase the number of European ICT workers.

The core asset is people, and we observe a loss of interest from students in the core ICT technologies. There is already a shortage of several hundreds of thousands, and this shortage is only going to increase in the coming years. If this trend continues, Europe will progressively lose its knowledge and any remaining stronghold in this domain, and will end up as purely a customer of technologies developed on other continents, especially in the case of the high-end technologies.

- As potential future employment is one key driver for students to apply to study in a domain, there should be enough attractive European companies; they should be encouraged to develop new, innovative technologies, for example in processor design, as it has been done in China, Russia, India, etc. The drivers for financing these innovations could be pre-commercial procurement from cities or administrations, for example, and considering the strategic and sovereignty aspects.
- The Arduino/BBC:bit initiative / Raspberry Pi Zero and the right software is a good way to bring more people to ICT. Simple programming languages for kids (such as the Scratch language from MIT) also helps.

1.7. RESEARCH POLICY RECOMMENDATIONS

1.7.1. ON COMMERCIALIZING THE RESULTS OF EU PROJECTS

It is often observed that 'Europe seeds, others harvest': the groundwork for new developments is often carried out in Europe, but non-European companies capitalize on these foundations. We observe that this is often caused by the lack of entrepreneurial spirit, the lack of incentives for researchers to commercialize their results, and the lack of venture capital to quickly scale up start-up companies. We recommend continued support for companies, especially for the scale-up of ICT start-ups into larger companies. With the leaking of innovative developments to non-European companies comes the migration of knowledge out of Europe. We therefore recommend stimulating the development of high-end systems, in order to keep strategic expert know-how in Europe.

1.7.2. ON PRODUCT DEVELOPMENT LIFECYCLE SUPPORT

The Future and Emerging technologies (FET) instrument is focused on long-term research, whereas the Leadership in Enabling and Industrial Technologies (LEIT) instrument focuses more on the short term. Projects that deal with the medium term are the ones that should make a difference at this time. A transparent coordination of funding on various levels of maturity would be welcome. For example, FET, LEIT (and the Electronic Components and Systems for European Leadership initiative (ECSEL)) could have similar calls in terms of content, but with a differentiation in terms of maturity. At the same time, we recommend the limiting of the fragmentation of public funding so as to reach critical mass in order to make projects that could have real impact.

1.7.3. ON GRAND CHALLENGES

We recommend the launch of long-term, disruptive and ambitious 'European Grand Challenge' competitions in the ICT domain, similar to the current Flagships projects, yet more dynamic in term of consortium composition and open for competition between teams to get the best solution(s), and having meeting points where good solutions may be merged and next steps re-launched. The structure should be agile and the research focused on *solving* the challenge. These challenges should be sufficiently realistic to be accomplished in less than a decade, and broad enough to require active cooperation between various disciplines. It will require new research, multiple technologies, innovations and concrete engineering work to be successful. The objective needs to be clear and measurable and lead to a *clear and visible achievement*, (e.g. having a humanoid robot that can go alone from the Eiffel Tower to Gare du Nord in order to board the train to Brussels and then go to the Atomium). This will also add public visibility.

1.7.4. ON SYNERGIES WITH DATA INFRASTRUCTURES, HPC AND COMMUNICATIONS

The HiPEAC community should increase its collaboration with ETP4HPC and BDVA, mainly at the compute node level and in terms of low-level tools. The following topics are of common interest:

- Making processors more powerful, and managing complex system software;
- Scalability of systems across physical boundaries of boards, racks, data centres;
- Tooling (compilers, debuggers, virtualization and so on);
- Analyse the constraints of new application domains to define new hardware/software acceleration, processing near memory or in-memory.

The HiPEAC community must also link up with existing network technology projects, defining and developing projects to shape the future IoT.

1.7.5. ON ECOSYSTEM BUILDING

We need to enable the emergence of a strong European computing ecosystem. It is important to leverage the creativity of the

maker community, young entrepreneurs and start-ups, and of the sharing economy.

We need to enable the emergence of a strong European ecosystem of SMEs that use ICT. Europe should facilitate paths for enabling technologies (embedded computing, design implementation and test facilities, system software and tools development centres) and innovators to implement their smart digital ideas. Professional networks are key ingredients for bringing people together and for addressing citizens: it is important to maintain a platform on which all the stakeholders in computing can meet, network and discuss the future of computing in Europe. This platform should facilitate access to talent and enabling technologies from all over Europe. It should support entrepreneurs, SMEs,

start-ups and academics. Other pivotal tasks include stimulating collaboration and mobility, vision building, reaching out to the public, disseminating results in the European computing community and beyond, developing platforms and interoperability approaches allowing development costs to be shared and favouring re-use and customization for specific needs. Since ecosystem building is a slow and labour-intensive process, it cannot be done in the context of a single project, but requires continued financial support either by the community itself or by society. Therefore, we recommend that the European Commission continues to invest in ecosystem building in Europe. We believe that it is the only way to maximize the economic impact of the research efforts made in, in particular, SMEs and academic research labs.

PART 2: RATIONALE

2.1. INTRODUCTION: ENTERING IN THE CENTAUR ERA

'Competition has been shown to be useful up to a certain point and no further, but cooperation, which is the thing we must strive for today, begins where competition leaves off.'

Franklin D. Roosevelt

'Forget humans versus machines: humans plus machines is what will drive society forward.'

Dr. John Kelly, senior vice president of IBM Research



Source: razum/Shutterstock

In the 2015 HiPEAC Vision document, the main highlight was *'The end of the world as we know it'*; new directions now seem to be regarded as very important in the domain in which HiPEAC works. This time, we feel that, in order to solve the complexity challenge and the software crisis, and to open more opportunities, both in research (we propose what we call *the 5th research paradigm*) and in the market, humans need to work side by side with more intelligent systems, define the goals and having the final decision, but also collaborate closely in the process, and not impose to the machine how we think problems should be solved. This is what we call the *'Centaur era'*², fuelled by the advances in process optimization, robotics, machine learning etc.

The *Centaur era* could be characterized in the following ways:

- Cooperation between man and machine (for example, in interactive HPC simulations), humans in-the-loop (requiring adequate GUI, etc.);

² In Advanced Chess, a 'Centaur' is a man/machine team. Advanced Chess (sometimes called *cyborg chess* or *centaur chess*) was first introduced by grandmaster Garry Kasparov, with the objective of a human player and a computer chess program playing as a team against other such pairs (from Wikipedia).

- The machines should adapt to the environment, including humans (e.g. in cobotics where robots are safe to work closely with humans). There is a real entanglement with the physical world;
- The machines should communicate with humans in a human-natural way (e.g. voice controlled assistants, IBM's Watson system);
- It should be a service-oriented approach, for example without explicitly programming machines (even if the 'core software' of the machine could be designed using the classical explicit programming approach).

From the human point of view, *trust* in machines is essential for their success. Therefore, we should be able to understand what a machine is doing and why, and be sure about the safety of machines, and within which limits they can operate.

One of the drivers that could lead to the *centaur era* is the ever-increasing complexity of systems. It is difficult to make things simpler (current 'RISC' processors are as complex as the CISC processors) without removing features, so we should delegate complexity to be managed by ICT systems. A new System-on-Chip (SoC) comes with several thousand pages of datasheet. We often add more complexity to systems for greater energy efficiency, for example. In fact, the *'World Depends on Technology No One Understands'* [117, 208]. Beside the centaur approach, mitigation techniques, developing abstractions and education (because people do not grasp the complexity of technological stuff) should be further developed.

We are also at a crossroads from the technology point of view: the limit of scaling seems to be real this time, and new technologies are in the starting blocks, but not yet ready, and will not entirely replace the 'old' silicon-based systems. They will again be employed for accelerators, helping performance and power consumption. It is possible that two tracks will be followed: one using cost-optimized, lower-density technology, e.g. for IoT devices, and very high-end technology, only affordable for large markets or where performance cannot be reached with other approaches. We also think it is the *right time to question the implicit assumptions* of our current systems, which were designed with initial constraints that do not exist anymore. We need to refactor, re-engineer and re-invent computing (the US initiative along these lines is called *'rebooting computing'*, which, for us, is not the right term: if you reboot without changing the system, you will end up in the same situation). For example, here is a short list of what could be reassessed:

- Systems were designed with only performance in mind, assuming that energy and time are irrelevant. For CPS and IoT systems, these assumptions do not hold true anymore;

2.2. STRUCTURE OF THE DOCUMENT

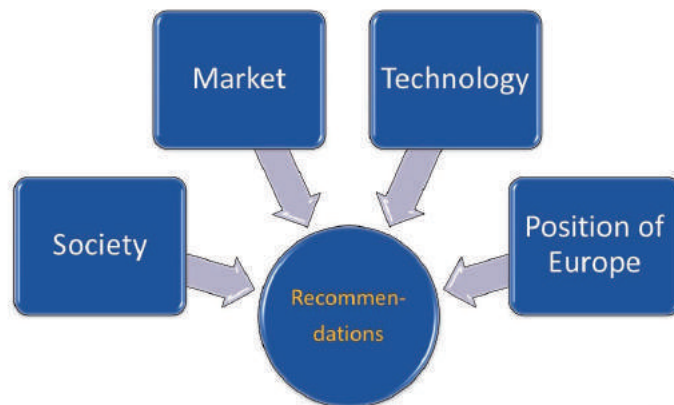


Figure 2: Structure of the HiPEAC Vision 2017

- Digital (i.e. values coded in exact binary encoding) is assumed to be superior to other data representations and processing (analogue computations, approximate computations, spike-based computations and so on);
- It was assumed that adding complexity is not a problem in order to get improved performance or energy efficiency. We can trade performance or energy efficiency for reduced complexity. Reducing complexity might be required to guarantee predictability or security.

It is also very clear that safety has become more and more pressing: everything that we do now will have safety implications. For example, any software could end up in a CPS, such as a self-driving car. As a result, the frontier between a 'safety critical' and a 'traditional' computing system is blurring.

Therefore, the challenges ahead are that computing systems will need to interact with the physical world to understand what is going on, to take decisions, and to help us in our everyday lives. To allow this to happen, we need to trust the systems, with all the issues this entails: we need systems that are reliable, that will not crash, and that are resilient to attacks by hackers. We also need systems that are highly energy-efficient: embedded computing in Internet devices or sensors will need extremely long battery life or to be able to scavenge energy from the environment; super-computers and servers will need to consume less energy and dissipate less power so they can be run affordably.

The HiPEAC Vision uses analyses of social trends, the evolution of the market and technological constraints and advances, and a SWOT analysis (Strengths, Weaknesses, Opportunities and Threats) of Europe in order to generate recommendations and directions that should be followed by the HiPEAC community. As the world is evolving very rapidly in our field, the information presented might no longer be accurate at the moment of reading the document. This rapid evolution is also the main reason why the successive HiPEAC Vision documents are different, even if some recommendations are still present from one to another. Furthermore, the process of creating the document is still done by humans, with their subjectivity. Therefore, if you have comments or remarks, please feel free to contact us at vision@hipeac.net in order to further improve the next version of the document.

2.3. SOCIETAL TRENDS

'Se vogliamo che tutto rimanga come è, bisogna che tutto cambi.'
'If we want things to stay as they are, things will have to change.'
Giuseppe Tomasi di Lampedusa, *Il Gattopardo* (1958)

2.3.1. EVOLUTION OF SOCIETY

The only constant we observe is change, and this change keeps accelerating. Although most of the observations from the 2015 Vision still hold, the concerns in 2014 about privacy after Edward Snowden's disclosure of the activities of the NSA, GCHQ and other national intelligence agencies is, in 2016, overshadowed by a concern about internal security due to a number of terrorist attacks in Europe since then. Despite statistics showing that violence is decreasing, that for every person killed in a terrorist attack, 100 people are killed violently in car accidents, and that Europe is still one of the safest places on earth, a large part of the population does not perceive it as such.

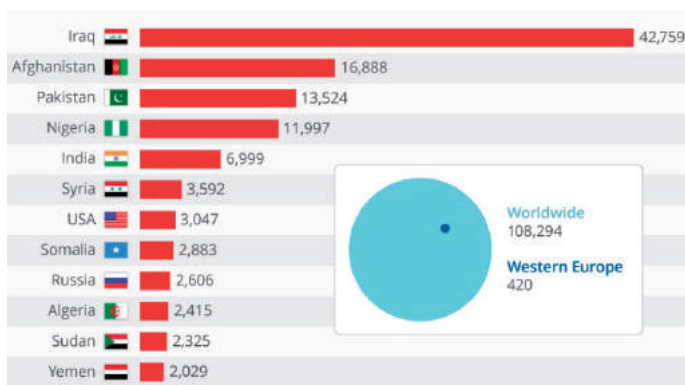


Figure 3: Victims of terror attacks outside western Europe
Source: Statista Charts based on the Global Terrorism Database

According to a McKinsey study [287], 65-70% of the households in advanced economies had a flat or falling income in the period 2005-2014. This was caused by a number of factors including: the cost of an ageing population; the destruction of jobs due to the relentless automation of industry and to offshoring manufacturing; the cost of the effects of climate change; austerity measures imposed by governments; and increased inequality. In some countries like the UK and Greece, real wages fell by 10% in the period 2007-2015 [183]. Many people feel that, for the first time in many decades, the younger generation might have a lower standard of living than the generation that is retired, or is about to retire. Many people are worried about their future retirement benefits. For most people, the single most important challenge for the European society is how to preserve the current way of life in this changing world. There seem to be two visions on how to do this: via an open or a closed society [371].

WHY THE COST OF LIVING IS POISED TO PLUMMET IN THE NEXT 20 YEARS [382]

Over the last 50 years we have witnessed a steady reduction of the cost of living, a trend that has been witnessed globally.

1. Transportation: air travel is now affordable for most people in developed countries, and almost everybody has access to a car. In the future, self-driving cars might lead to 'car as a service' at a fraction of the price we pay today for owning a car;
2. Food: the cost of food has fallen more than 50% over the last 50 years due to more efficient production techniques and cheaper transportation;
3. Healthcare: thanks to an extended health care system, basic health care is affordable for most people. Even expensive operations are covered by health care insurance for most people. People stay active long after retirement;
4. Housing: the quality of houses has improved dramatically over the last 50 years (insulation, safety, facilities, ...). The average number of people per household has decreased;
5. Energy: there is an abundance of energy at affordable rates;
6. Education: more people than ever before in human history receive a good education and an increasing share of learning is informal via the Internet. Dictionaries and encyclopaedias are now freely available on the Internet;
7. Entertainment: the advent of digital economies has resulted in new remuneration models (advertising, streaming, crowd-funding) that can result in lower prices for music, video, and games.

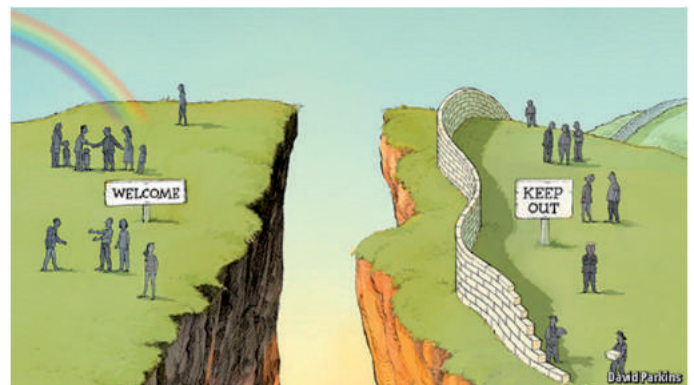


Figure 4: Open versus closed society
Source: David Parkins

The majority of the benefits of globalization has gone to a tiny fraction of the world's population [210]. The financial crisis has led to austerity measures all over Europe, while no one appears to have been thoroughly punished for its causes (except for the Greeks). People lose their income and look for explanations in the changes they see around them. Some react by wanting to turn back the clock by leaving the European Union, by closing the borders for immigrants and by focusing mainly on internal affairs. They hope that the resulting closed society will bring back 'the good old days'.

The people in favour of a closed society want to return to a society when the aforementioned challenges were not yet as problematic as they are today.

People in favour of an open society often recognize the same problems. However, they believe that in order to survive in this rapidly changing world, we need to adapt and to find new solutions, to move forward. The old approaches may have worked in the 20th century, but are not fit for the problems of the 21st. Closing borders may reduce immigration, but does not address its causes. Additionally, with many European countries currently facing an ageing population and eventually a declining population, they may actually need immigrants to sustain their society.

Furthermore, closing borders is no solution for the problems of pollution, climate change, international finance, energy and food security, and refugees; these things can only be solved at a global level. Many believe it is better to embrace an open society, and to alleviate the negative side-effects of such a society by having a solid social security system, by adapting laws to the realities of modern financial and business practices, and by making sure that immigrants get treated fairly so that they can contribute to society as soon as possible.

If several major G8 countries decided to opt for a closed society and consequently reduce their international commitment (contributions to United Nations, NATO, reduce foreign aid and investments), other countries would quickly try to fill the void to increase their international influence, leading to even more political change [87, 309].

After many years of opening up and integrating the countries of the European Union, we might enter a phase in which this integration will stall and politicians will focus more on protectionism and isolationism.

The divide between pro-open and pro-closed cuts across traditional left- and right-wing political parties, across generations, race/ethnicity, gender, income and education levels. That makes it very difficult for traditional political parties to adapt to this new sociological reality. The pro-open politicians may have an increasingly hard time swimming against the tide.

RELEVANCE FOR COMPUTING

For the computing systems industry, a closed society is, however, not an option. Computing as we know it today is a product of, as well as a creator of, the globalized society. Computing is a key enabling technology for a global economy. Protectionism and isolationism will slow down the sharing of ideas and innovations. In Europe, the digital single market initiative aims to tear down regulatory walls between the 28 national markets. The abolition of roaming costs in Europe in June 2017 is one example of this initiative.

According to Michael Curtis [499, 1002] the introduction of the IBM PC contributed to the fall of the USSR. Russian leaders understood that, in order to compete with the west, they had to introduce personal computers, but that they would no longer be able to control their citizens as soon as these citizens had access to computers, networks and printers. Mikhail Gorbachev introduced Glasnost in 1986 (only five years after the IBM PC was launched in 1981), eventually leading to the fall of the Berlin Wall in 1989.

2.3.2. SECURITY CHALLENGES

The situation in countries bordering the European Union is definitely less stable now than it was a decade ago. Whereas until a couple of years ago, the European Union acted as if it could ignore these problems outside of its borders, recent history shows that they are increasingly affecting internal European affairs: an unstoppable stream of refugees from the Middle East and Africa trying to enter the European Union, an unstable political situation in Turkey that gets exported to some European countries, and a number of terrorist attacks in European cities inspired by the so-called Islamic State. Some politicians use war rhetoric in order to mask their inability to address these issues directly, and to build support for more investments in internal security. Soldiers and heavily armed law enforcement officers are now a common sight in many European cities.

In parallel with the increase in physical attacks, there is also a surge in cyber-attacks. This is a logical consequence of the fact that a major part of modern society has a critical dependence on its cyber infrastructure (banking, communication, businesses and utilities to name but a few). Stealing information is now as lucrative as robbing a bank, only less dangerous for a robber because it can be done from a distance. Disrupting a global cyber infrastructure can have a serious impact on society and on the economy. Disclosing classified information can have serious political consequences as demonstrated by the multiple *-leaks incidents.

The number of cyber-attacks is increasing rapidly. According to Juniper Research, the estimated global annual cost of cybercrime is estimated at over US\$2 trillion dollars by 2019, four times greater than in 2015, and 16 times higher than in 2013 [257]. Many people are amazed at how apparently easy it is to hack mail servers of political parties, and to bring down government and company websites. Governments are increasingly worried about attacks by organized crime (including terrorists), and state-sponsored attacks. The latter two are difficult to fight because they have access to huge resources [116].

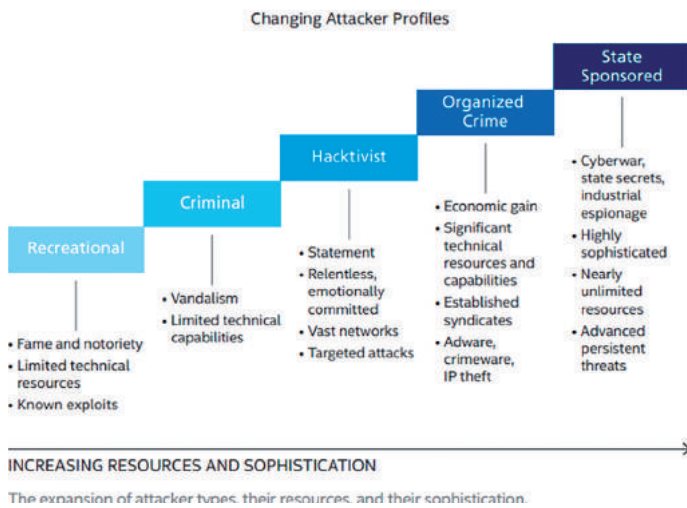


Figure 5: Changing profiles of cyber-attackers
 Source: McAfee Labs Threats Report, August 2015

RELEVANCE FOR COMPUTING

Computing systems are a key enabling technology to build sophisticated surveillance systems, to analyse the ExaByte+ of data that is created daily, and to design intrusion detection systems. The expectation is that companies and governments will invest heavily in security systems over the coming years in the hope of improving their security performance and of bringing down the cost of security. This will require a number of changes in legislation (about the use, and limits thereon, of cameras, databases, encryption, and so on by individuals, companies, governments and so on). The computing systems community will have to be ready to respond with technical solutions such as privacy by design, more advanced image processing, stronger encryption, tools to help protect anonymity, data analytics on structured and unstructured data, removal of single points of failure in infrastructure, and accurate sensors for the detection of explosives, drugs, and other dangers.

2.3.3. PRIVACY EROSION

People continue to be willing to give up privacy in order to facilitate convenience (like registering on a website via social media) and perceived immediate gains (like access to free music in return for a registration on a website; a traffic-jam-free trip by disclosing the origin and the destination of a trip; sharing location and pictures in order to be able to play Pokémon Go [138]). People also still seem to be ready to give up some of their privacy and even some of their liberties in the belief that this will help the government stop terrorists and increase their safety.

Very few people have a deep understanding of information technology and the business models of the Internet companies. This fact - in combination with the lack of clear understanding of privacy and how it is affected by modern technology - leads to situations in which they behave in ways that might compromise their privacy without them being aware of it [59].

People disclose personal information in their network through social media. Posting holiday pictures on Twitter tells the whole world (including burglars) that one is not at home. Pictures can

be downloaded, modified and used for commercial purposes without permission. Pictures from a hospital room reveal to insurance companies and future employers that somebody had a medical problem. Career information on LinkedIn tells future employers what somebody's skills are, whether that person likes to work abroad, whether somebody's family could afford a prestigious private university. The fact that Microsoft bought LinkedIn at \$65 per profile (six times the price Facebook paid per WhatsApp user) means that Microsoft believes it can create a lot of extra value for its shareholders by combining the LinkedIn information with other information it already owns about its hundreds of millions of users.

Very few users could imagine that one day a company like Microsoft might correlate their activity in Office 365 with the kind of job or the set of skills they have and base its commercial strategy on this information. E-commerce websites can already use a delivery address to check the value of the house where a customer lives on real-estate websites, and to check whether the house has a garage, a garden or a pool and therefore send customized advertisements to that customer. Many people have gone through the experience of when, hours after searching for information about a particular city, several websites start showing adverts for hotel rooms or rental cars in that city. There is a documented case where a shop started sending adverts for pregnant women to a teenage girl before her parents had found out about her pregnancy [255]. Advanced image processing techniques allow for the mining of information from social media by linking objects, animals and people. Even if two people are never seen together in the same picture, if they can be identified in two separate pictures walking the same dog, there is a strong presumption that they know each other. Some people find it convenient; other people find it upsetting that this information, which was not meant to be shared, is apparently circulated around the world and that it turns out to be very difficult, if not impossible, for an Internet user to avoid unintentionally disclosing information and to control how it is used.

There are multiple definitions of privacy. In the 19th century, privacy was defined as the 'right to be left alone'. A more modern definition is that privacy is the 'control one has over the information about himself/herself'. It is necessary that doctors maintain medical records about their patients, but nobody expects the doctor to share this information with third parties (medical privacy) unless this was required for medical treatment. We expect the same behaviour from financial institutions (financial privacy), websites (Internet privacy) and voting systems (political privacy). In 2015, the top 10 healthcare data breaches [250] in the US concerned 111 million medical records (if the records had been of unique patients, this would have affected 35% of the US population). There are several major data breaches per day worldwide. Whereas Internet companies can be obliged to implement the right to be forgotten, this cannot be enforced for stolen data. Once the data is stolen, there is no way for the owner of the data to control its use anymore. This can, in some cases, have far-reaching consequences such as identity theft.

Current privacy laws are not adapted to the latest technological evolutions. Normally, law enforcement officers need an official search warrant to enter a house, and to check the contents of a computer hard disk. In several countries, it is already common practice to confiscate the smartphone of the driver involved in a car accident in order to check whether he or she was using it at the wheel. Since the distinction between a smartphone and a computer is very small, one could consider the smartphone an extension of the home environment, meaning that nobody should be forced to hand one over without a search warrant.

One notable case is the FBI-Apple dispute on the unlocking of the iPhone 5C used by one of the shooters involved in the 2015 San Bernardino attack. The FBI could not unlock the phone, and ordered Apple to unlock the system. Apple refused to do so because it did not want to create software that could undermine the security of one of its products. The FBI went to court, but eventually withdrew its request after it received help from a third party to unlock the phone. The encryption controversy drew a lot of public attention. About 50% of the American public supported the FBI's stance while the rest supported Apple's stance, or did not have an opinion. People supporting Apple and Google in the encryption controversy state that encryption only affects less than 0.1% of criminal investigations. At the same time, half of all thefts in the US are of mobile devices (more than 3 million). If smartphone encryption became universal, that very encryption would also protect the privacy of millions of smartphone users, and make smartphones less attractive for thieves [158].

According to Apple, Google and their supporters, the benefits of encryption massively outweigh its disadvantages, in the same way that the benign uses of cars, knives and fire greatly outweigh their use to commit crimes. One could also reverse the argument: if one day an effective technology was developed to break encryption (and quantum computing might be able to do it), it would create a massive global disruption because the majority of contemporary security technologies are based on the hypothesis that encryption is practically unbreakable if implemented well.

Privacy is clearly eroding, and most people seem to accept this. Most people are already under surveillance for a large part of the day (smartphones tracking location, access control systems in companies, hundreds of cameras in public places, licence plate recognition, Google Street View filming the street, tourists taking pictures with people in the background and posting them on the Internet, and so on). Soon, devices connected to the IoT will start exchanging messages without asking us. When the navigation system in a car detects that somebody is driving home, it could tell the home automation system to prepare the house for arrival. Voicemail could suggest to a caller to call again half an hour later. If we want these technologies to become mainstream, we will have work on a better definition of privacy [176]. It is clear that users will have to become more aware of privacy, that there is an urgent need for a (global) legal privacy framework and that computing systems will have to support better privacy mechanisms. One interesting evolution is the adoption of differential privacy (that's to say, adding noise to hide information about individual

records, but not destroying the validity of statistical conclusions) by Apple [187]. Also to preserve privacy, Apple runs recognition tasks locally on the client devices, and not on its servers. Although preserving privacy will almost certainly be more expensive than giving it up, it is to be hoped that it will remain affordable for everyone, and not become a luxury good for the wealthy.

RELEVANCE FOR COMPUTING

The computing community should take privacy seriously and work on solutions to prevent breaches. Two obvious solutions are: (i) to reduce the attack surface of computing systems by making them simpler. Modern computing systems are so complex that it is almost impossible to properly secure them; and (ii) to give all ICT workers basic privacy training to make them aware of the privacy impact of their design decisions.

2.3.4. ENVIRONMENTAL DEGRADATION AND THE ENERGY CHALLENGE

The European Union (EU) has an ecological footprint that is about twice the biocapacity of its surface area [238]. This means that the EU currently uses two Europes to support its lifestyle. It also means that Europe depends on solid trade and a good relationship with a sufficient number of countries willing to share their resources with us, even if they are scarce.

Most of Europe enjoys a moderate climate. That means that the availability of food and water are generally not considered critical challenges, and that there are currently no areas with important shortages. However, this might change in the future if precipitation patterns change and excessively wet or dry weather decrease crop yields or introduce new pests, for example. Studies show that a warmer climate might increase food production in the northern regions of Europe, and decrease it in the southern regions. Europe as a whole turns out to be less vulnerable to climate change than other regions, such as desert areas.

The situation is more dramatic in the Middle East and in North Africa. For example, in Syria, there is already a severe multiyear drought that started in 2007, caused 1.5 million people from rural communities to move from their home and contributed to the destabilization of the country. Kuwait recorded in July 2016 a world record temperature of 54°C [341]. Many places in the world have recorded record temperatures in several months in 2015 and 2016. The record high temperatures in 2016 are more alarming than those in 2015 because they cannot be explained by El Niño anymore. If the multiyear drought continues in the Middle East and Africa, experts expect serious drought-fuelled conflicts in those regions. Hence, although food and water might not be a direct challenge for Europe, it might create an indirect challenge in the form of increased migration pressure.

You think migration is a challenge in Europe today because of extremism, wait until you see what happens when there's an absence of water, an absence of food or one tribe fighting against another for mere survival. [354]

John Kerry, U.S. Secretary of State



Figure 7: Vertical harvest, Wyoming, US

Source: Vertical Harvest

These events are consequences of global warming or, more accurately, of global climate change, which is principally caused by human use of fossil fuels.

The energy challenge is a considerable one for Europe. Fossil fuels are finite and harmful for the environment, and the green alternatives do not yet offer the same energy security 24/7. In order to meet climate and energy targets, we need to invest in improved energy efficiency and in increased use of renewable energy sources [34]. A solution for affordable electrical energy storage (like a battery that can store the energy needed to run a household for a couple of days) would definitely help to make distributed green energy production a valid alternative to large-scale fossil or nuclear power plants [112].

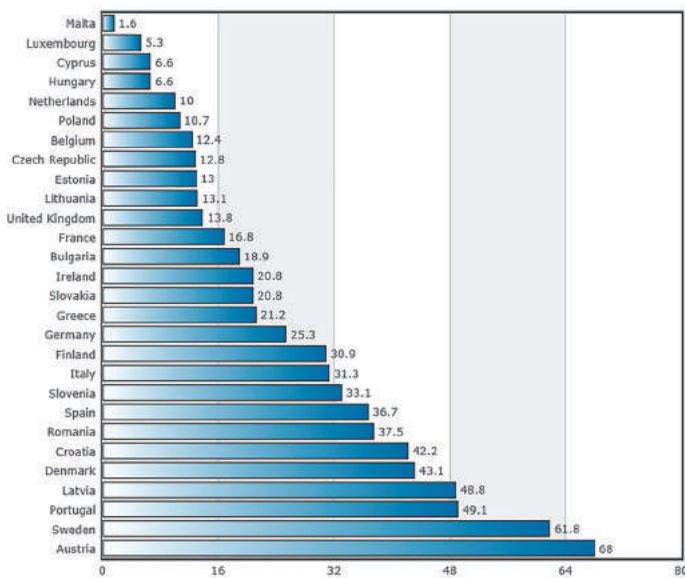


Figure 6: Percentage of electricity produced from renewable sources in European countries

Source: Eurostat

The availability of abundant cheap green energy could help to solve several other challenges. Stimulating the use of electric vehicles could reduce the pollution caused by transportation, vertical farms [235] could produce food year-round in urban areas, and drinking water could be produced from seawater or from waste water.

RELEVANCE FOR COMPUTING

For all of the reasons above, decreasing energy consumption is a major challenge for ICT, as it is for many industries.

Furthermore, since electricity is the power source for ICT devices, low electricity consumption is a crucial element for its acceptability in society, in active mode (see Figure 8), e.g. in servers, the “cloud” and HPC centres [356], but also in standby mode [421].

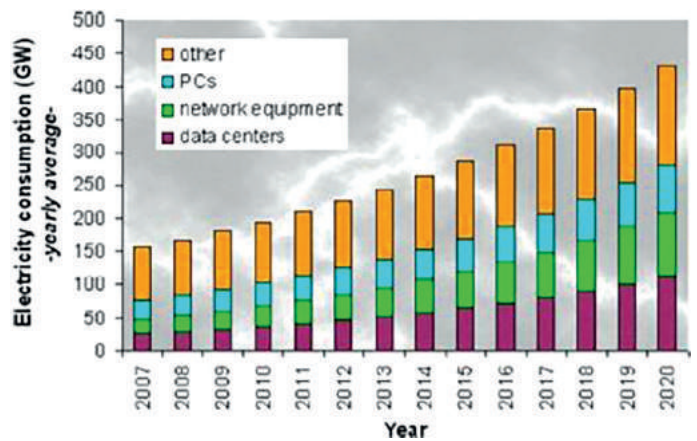


Figure 8: Growth of energy consumption of ICT devices

Source: [19]

IoT will also have a significant impact on the global energy consumption in the future. For example, replacing ‘old’ passive switches – which have absolutely no power consumption in standby mode – by intelligent switches connected by Bluetooth or Wi-Fi – which will have a very low but nonzero energy consumption in standby mode – could have a significant global impact, because the very low standby energy will be multiplied by billions of devices that are likely to be in use in the future.

Most of the energy spent in computing is transformed into heat. Cooling is a major cost for data centres and HPC centres, and adds to the electricity requirements, since most cooling is electricity based. To avoid this, some companies decide to move their data centres in cold areas, or in areas where the energy is cheap.

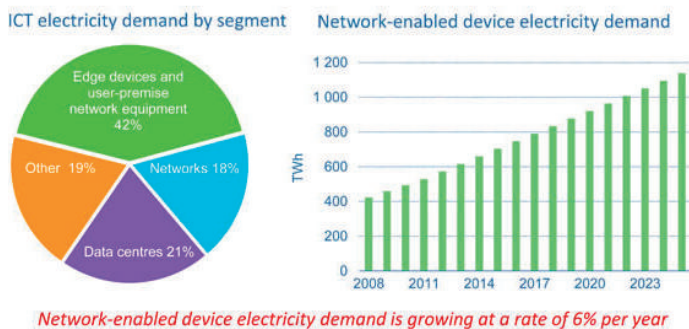


Figure 9: Growth of electricity demand from network-enabled devices

Source: [356]

In addition, loss of electrical power also leads to anxiety. According to [272], 90% of people feel anxiety or even panic when the battery of their smartphone is getting empty. Range anxiety is making some people reluctant to buy an electric car, despite of all its advantages [18]. IoT devices will not succeed if they must have their batteries changed too often: imagine having 20 smart sensors in your house, and each one needs to have its battery changed or recharged every month.

Computing is thus both a huge energy consumer and a key enabling technology for saving energy (in production, in heating and cooling, in transmission, in generation). The IoT will more than triple the number of connected internet devices by 2020. This will entail a yearly increase of 7% in the power consumed by Internet devices (considerably greater than the forecast 3% global increase in electrical power consumption per year). The Internet could consume as much as 20% of world energy use by 2030 [284]. According to SIA and SRC [323], computing might consume more energy than the world can produce by 2040.

Hence, low-power computing will remain a critical and significant challenge for the decade to come, and a radical improvement in the energy efficiency of computing is required.

2.3.5. EDUCATION

The ICT sector generates 25% of total business expenditure in Research and Development (R&D), and investments in ICT account for 50% of all European productivity growth [81]. The fact that the ICT sector simultaneously represents only 4.8% of the European economy shows that it is a very efficient sector, given that with such a small overhead it can generate large benefits for everyone. A workforce of 7.5 million ICT workers is behind these achievements. However, most countries in Europe are unfortunately currently facing a shortage of ICT workers, which is holding back innovation in computing [37].

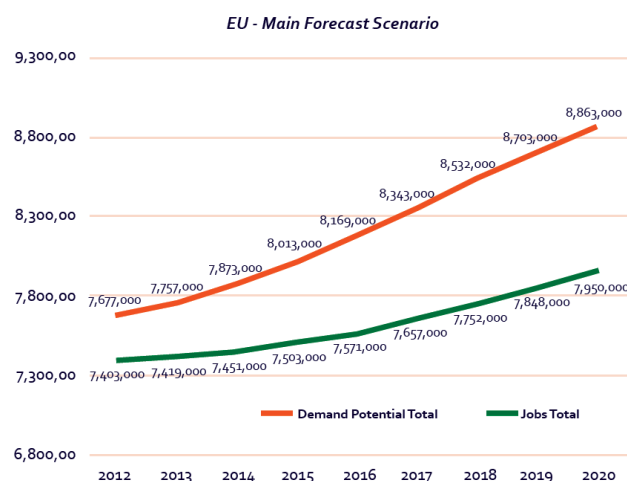


Figure 10: ICT workforce development and ICT worker demand potential in Europe 2012-2020

Source: [37]

The European Union had a structural shortage of 500,000 ICT workers in 2015, and this shortage is estimated to grow by 80,000 per year to around 900,000 by 2020. If more ICT personnel were available, industry could employ 300,000 extra workers, which could lead to more innovation. Unfortunately, enrolments for ICT degrees are more or less stable, and there is a delay of 3-5 years between the initial enrolment and graduation. Hence, there is no possibility of closing this skills gap of 1.2 million workers by 2020. This will have an impact on innovation and on the digitization of European industry (CPS, the IoT, smart factories, autonomous cars, and so on).

Focusing more on foreign students could increase ICT student enrolment numbers. Large European countries like the UK, France and Germany already attract more than one million international students in total (which is slightly more than the US), and almost half of these students study for a degree in a STEM subject (Science, Technology, Engineering, Maths). Hence, there is huge potential.

For the students studying in the US, the top three countries of origin are China, India and South Korea [276]. These countries provide a large proportion of international students in all developed countries. Over the last few years, the number of Chinese international students has increased quickly due to the growing number of middle-class families who can afford to send their only child abroad. However, more Chinese students than ever are deciding to return to China after graduation (currently about 80%; and more women than men) [214]. In the UK, 88% of all foreign students return to their home country [260]. In the future, the influx of Chinese students might stagnate, or even decrease due to the shrinking number of young adults, economic slowdown and the improved quality of Chinese universities [179].

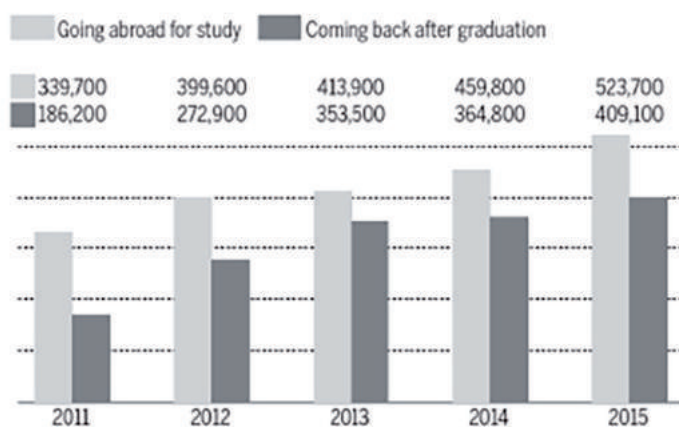


Figure 11: Proportion of Chinese students returning home after studying overseas 2011-2015

Source: Ministry of Education, PRC

There are multiple reasons for going back home [41, 265].

- In order to stay in the US or in Europe, they need to find an employer that is willing to sponsor a work visa. However, for STEM jobs, this is not a big problem (whereas for holders of a degree in Arts it is difficult to find such a sponsor);
- Students can often find a better paid job in their home country (based on their prestigious overseas degree) than in their country of study where they have to compete with all the other graduates with similar degrees (and the disadvantage of having to deal with all the administrative issues of work visas). Some companies negotiate lower salaries in return for the burden of having to do all the paperwork. The trend towards a more closed society could also encourage foreign students to return to their home countries;
- The current generation of Chinese students were born under the one child policy. This has the advantage that an increasing number of middle-class parents can afford to send their one child to a European or American university, but it also explains why they are very eager to get their child back in China after graduation.

Many countries now have programmes to try to convince their outgoing students to come back after graduation [292]. In Europe, there are schemes including Research Council Starting Grants, European Research Council Advanced Grants and Marie Skłodowska-Curie Actions. Other regions have similar programmes. This makes it more difficult for the host countries to retain the graduates that were trained there.

There is, of course, the untapped resource of female students. According to the US National Center for Education Statistics, the six bachelor's degree programmes with less than 20% women in 1971 all evolved to have a more gender-balanced situation by 2011, with two exceptions: engineering and computer science. Computer science seemed to be on the right track in the 1970s but, after 1983, there was a steep decline in the number of computer science bachelor's degrees conferred to women: the 2011 level was the same as that of 1974 [314].

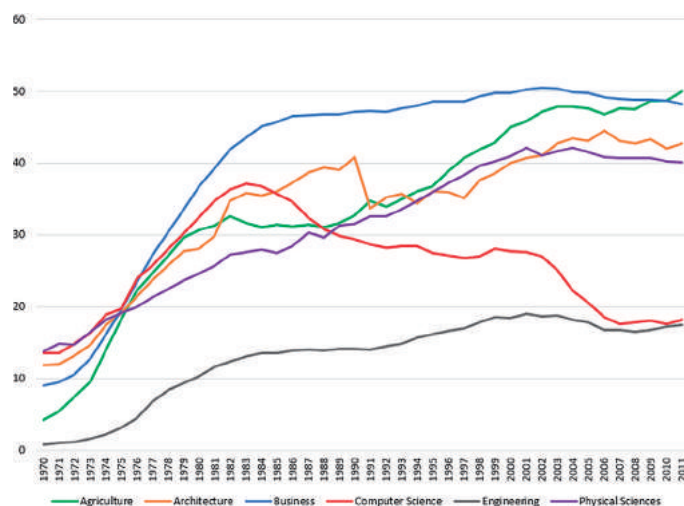


Figure 12: Percentage of bachelor degrees conferred to women in the US

Source: Randy Olson with data from NCES

Worldwide, there are a significant number of initiatives to encourage girls and young women to study engineering and computer science, but these seem to have limited effect or are small-scale. One notable case is Harvey Mudd College (800 students) where in 2016 more women than men graduated in computer science (55% in 2016, compared to less than 15% in 2008). The important change made was to make the courses more inclusive so as to make female students feel more at ease [63]. Another notable example is School 42, a private initiative by French billionaire Xavier Niel, who wants to offer free computer science education to the most talented students in Paris and in Silicon Valley. One of the distinctive aspects of School 42 is the pedagogical approach: no teachers, no classrooms, but peer reviews, code projects, internships and gamification instead [136]. All these experiments are worthwhile, but the big challenge is to scale them up to millions of students per year (not all of them being extremely talented, or from families able to afford expensive degrees).

An approach which scales and is affordable for many is Massive Open Online Courses (MOOCs). In 2011, MOOCs became instantly popular by opening up Stanford University's course on Artificial Intelligence by 'rock star' professors Sebastian Thurn and Peter Norvig to 160,000 students. Immediately, several initiatives were launched: Coursera, Udacity, edX and several smaller ones. Completion rates turn out to be generally below 10%, which is much lower than for traditional courses. It seems that MOOCs have currently passed the peak of inflated expectations, and are slowly evolving towards the plateau of productivity [100]. Many universities are currently offering a handful of MOOCs in domains in which they excel. It is often considered a marketing instrument for a traditional degree course.

RELEVANCE FOR COMPUTING

Efforts to attract and retain international students in higher education should never lead to reduced efforts to attract local students. Only a minority of foreign students will eventually decide to stay, and to contribute to the local economy. Even without international graduates, a country should be able to produce grad-

uates in numbers sufficiently high to sustain society and the economy.

In 2015, foreign students contributed US\$30.5 billion to the US economy [277], and £2.3 billion the UK economy (about 50% in tuition fees and 50% in living costs) [260]. In some other European countries, foreign students pay the same (low) tuition fees as local students, which means that the former are subsidized by the guest country. This could, in some cases, be justified for students from developing countries (it could be considered overseas aid), but it is hard to defend for students from middle-class families in other countries, especially if they return home after graduating. In that case, they take away resources that could be used to train local students. It is important that these students pay the real cost of their training.

2.3.6. SELF-SUFFICIENCY IN ICT

“People who are really serious about software should make their own hardware.”

Alan Kay

ICT is a strategic asset for a country from the viewpoints of both sovereignty and the economy. High performance computing allows the simulation of military devices and planes, cars, pharmaceutical products and many more things. It is also important to store data and to analyse what is going on in different communication networks as was highlighted by the revelations of Edward Snowden. Therefore, all major countries want to control a large part of their ICT infrastructure to avoid being blocked in their development by other countries.

The US is currently the dominant provider of computing solutions with CPUs (Intel) and GPUs (NVIDIA), which are used to build high performance computing and servers. Components can be banned from export under several US rules, such as the International Traffic in Arms Regulations (ITAR), which controls the export and import of defence-related articles and services. The US Department of Commerce prevented Intel and NVIDIA (but also AMD, IBM for their processors and HP for its optoelectronic devices) from shipping the processors required for the upgrade of the Chinese Tianhe-2 supercomputer citing worries over nuclear weapons related research [202]. As a consequence, China developed, over the span of only three years, a completely new system, including a very energy-efficient computing chip. The resulting supercomputer, the Sunway TaihuLight, reached the top of the TOP500 list of most powerful supercomputers on the LINPACK benchmark in June 2016 with 93 petaflop/s (quadrillions of calculations per second) [181]. It superseded the Tianhe-2, which was the first-placed supercomputer in the last six TOP500 lists. The operating power consumption of the Sunway TaihuLight is 15.37 MW, or 6 Gflops/W, which makes it third in the green500 list [267].



Figure 13: Sunway TaihuLight

Source: Xinhua [40]

Japan is also targeting Exaflop computing, and the “post-K” computer, designed by Fujitsu, will similarly use a homemade processor, based on the ARM architecture (the previous architecture supported by Fujitsu was based on the SPARC architecture).

We observe that in the short time since the last HiPEAC Vision, some countries have gone from having only intentions to having real plans and fully operational systems. Their architecture is either based on brand new designs (like the Chinese ShenWei SW260), on MIPS (Russian Baikal-T1) or on ARM (Japanese future Fujitsu chip for HPC or Chinese FT-2000/64). China, Russia, Japan and India are actively developing processors either for desktop computers, servers, HPC or even embedded devices (more details will be given in section 2.4.5.3). Regardless of whether this is related to the revelations of Edward Snowden or not, there is a growing movement away from well-established US computing platforms such as those of Intel, Google, Apple and Microsoft either to avoid bans on accessing core components, or because of fears that hardware and software might have spyware deeply implanted.

RELEVANCE FOR COMPUTING

We can repeat a statement of the last HiPEAC Vision that, to our knowledge, in this domain ‘No state-funded or EU-funded initiatives exist in Europe, yet. The opening up of the CPU-market is, however, an opportunity for Europe to jump in, as it clearly shows that information technology is not tightly bound to one computing platform anymore. Open architectures, where the code can be reviewed and the design audited, may play a major role in this climate. It is already the case with the ARM architecture in the embedded/mobile domain, but for example IBM with its OpenPOWER initiative [58] is also following this trend, and MiPS is now part of Imagination Technologies in UK. The RISC-V Instruction Set Architecture (ISA), from the University of California at Berkeley, is a standard open architecture that can be used by industry [64]. Finally, the Leon (SPARC instruction set) is also an example where people have access to the netlist of the implemented core [262].’

Besides the fear of ‘undocumented features’ and sovereignty drives, there are other reasons why it is important to keep control of (part) of the ICT ecosystem: innovation, and keeping local

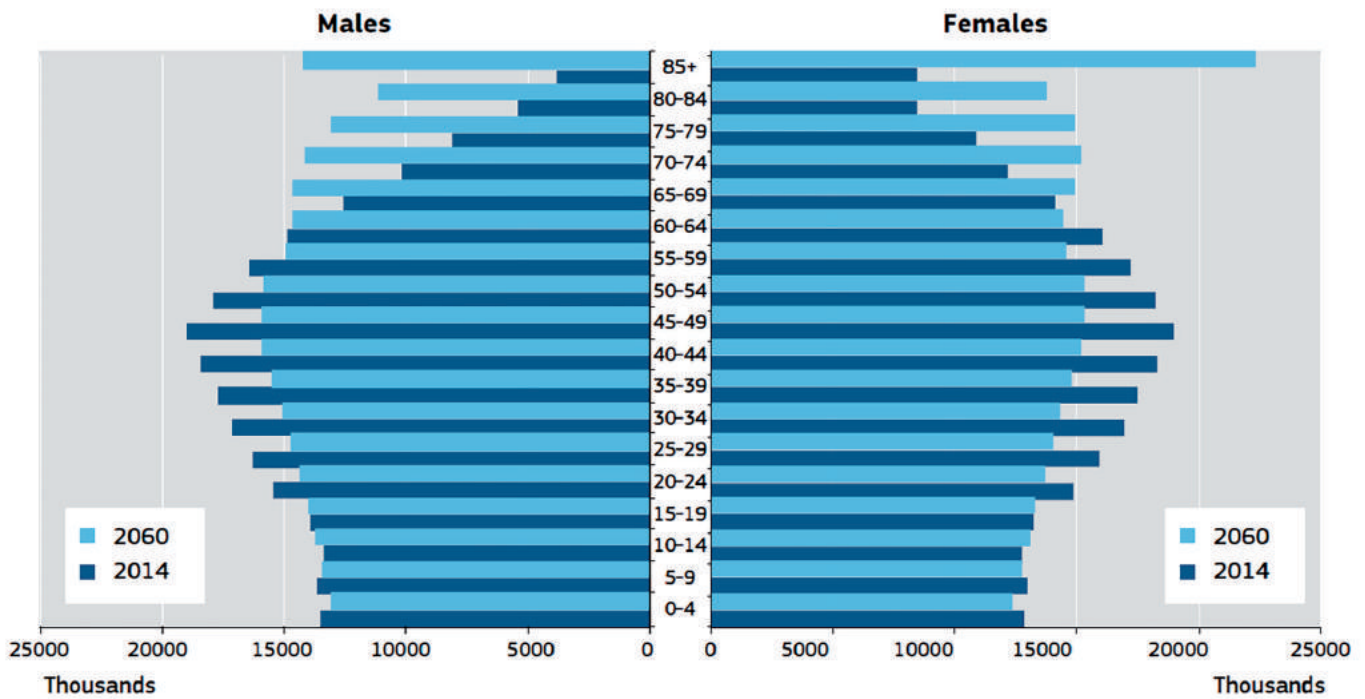


Figure 14: EU population demographic forecast to 2060
Source: European Commission

knowledge. For example, if there was no more microprocessor development in Europe, the good ideas from European research and development teams either will not turn into real products or will only be used by non-European companies. A European company that has a good differentiating IP will have a low chance of implementing it in a non-European chip (unless the company gets bought). Similarly, it is very demotivating for European researchers to not have their results used locally or even at all: they will move away, and/or the domain will dry up in Europe, leading to a loss of knowledge in this field. It is the same for engineers: if there is no local employment, graduates will move abroad and new students will no longer choose the degree, leading to an evaporation of know-how.

This is especially true for hardware-related development, but less so in the case of software (due to the open source model where the source code is accessible and where people from everywhere can contribute).

2.3.7. AGEING POPULATION OF EUROPE

The ageing population presents significant challenges to European economies and welfare systems. The demographic transition is viewed as one of the biggest challenges facing the EU [38]. The age distribution of the European population is transitioning from a traditional pyramid into a pillar. This has far-reaching consequences:

- The proportion of the population who are of working age (20-64 years) will decline from 61% in 2013 to 51% in 2060, which means that every working person will have to support more than one non-working person (child, student, retired, ill);

- The retired will, by 2060, represent almost 30% of the population, and more than a third of the people voting in elections. This will have political consequences;
- With 30% of the population over 65 years, and 12% over 80 years, the economy will have to adapt too: more leisure services for those in good health, and health and care related services for the fragile and dependent. By 2020 there will already be two million vacancies in health and social care in Europe [79].

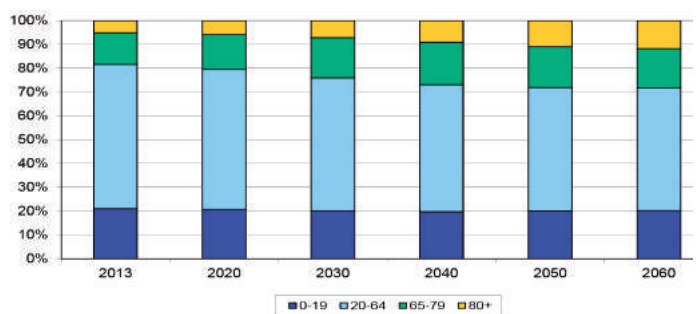


Figure 15: Evolution of the different age groups in the European population
Source: [38]

The so-called Silver Economy covers new market opportunities related to the rights, needs and demands of the growing population over fifty. The Silver Economy is one of the fastest-growing markets, and hence creates a major opportunity for new jobs and growth. Examples of sectors expected to benefit significantly from the Silver Economy are: cosmetics and fashion, tourism, smart homes supporting independent living, service robotics, health (including medical devices, pharmaceuticals and eHealth) and wellness, safety, culture, education and skills, entertainment,

personal and autonomous transport, banking and relevant financial products.

RELEVANCE FOR COMPUTING

ICT can play an important role in the Silver Economy: ICT can help older people to stay healthy, independent and active at work or in their community. The European Commission funds several research projects in the field [384].

Examples are:

- Telemedicine, telehealth, and mobile health applications, a growing multi-billion Euro market;
- Smart homes and smart cities for independent living. Smart homes integrate ICT and communication devices that anticipate and respond to the needs of residents;
- Medical innovations to efficiently treat the growing number of chronic diseases. There will be a growing need for advanced and affordable medical equipment for use at home (blood pressure devices are already available for home use; home dialysis machines will become common too). The healthcare market is the second biggest embedded systems market after automotive, with an annual growth of more than 5%;
- The market for service robots for domestic tasks will grow as the technology matures and becomes affordable. Japan is currently the forerunner in this market. Its society seems to be more eager to accept this new technology than that of Europe.

2.3.8. THE SHARING ECONOMY

The term 'sharing economy' has been in use since 2010 and has multiple definitions. The sharing economy provides access to products, services and talent without ownership being a prerequisite. Users, providers, lenders or borrowers often have peer-to-peer contacts mediated by the Internet.

Examples of sharing are:

- Agriculture: sharing a vegetable garden
- Finance: crowdfunding
- Travel: Airbnb, 'couch surfing'
- Labour: co-working spaces
- Books: libraries
- Transportation: carpooling, car sharing, bike sharing, Uber
- Computing: cloud computing, open source hardware and software
- Digital rights: copyleft, Creative Commons

The sharing economy is not without controversy. Companies engaging in a large-scale sharing business like Uber encounter fierce opposition from the established players in the real economy (in the case of Uber, taxi drivers' unions, taxi companies and the cities selling licences; in the case of Airbnb, residential districts complaining about noisy tourists). Small-scale sharing initiatives are tolerated by the traditional players as long as they do not have too much of an impact on their bottom line.

A concern about the sharing economy is that it is not well regulated. If the service is of low quality, it is difficult to complain or to get compensation; in some cases, safety regulations are applied creatively, permits are not obtained, or taxes are not paid. The

question is whether large-scale sharing companies like Uber and Airbnb are still part of the sharing economy, or are just a growing international taxi company or hotel chain with a different business model. The question is also whether they will survive when they are forced to comply with the same laws or regulations that the local players are, and have to take full liability for all the services that they offer worldwide. At the time of writing, Uber is still struggling to break through in many countries in Asia [47, 232], and is losing money at a very high rate (\$2 billion in 2015).

RELEVANCE FOR COMPUTING

Open source software and open hardware are the emanations of the sharing economy in computing. In some areas like high performance computing and mobile computing, open source software is an important business model. For laptop and desktop computers, it is not the dominant business model but Microsoft recently decided to make some of its previously proprietary software open source. It is yet unclear what the dominant model will become in emerging new markets such as the IoT, CPS, and self-driving cars.

2.3.9. EFFECTS OF DIGITAL TECHNOLOGY ON THE BRAIN

The effects of digital technology on humans and society has been studied extensively, and there are both positive and negative effects. One positive consequence is the gain in business productivity. Customers have access to online information, they can make online appointments and buy goods and services without having to queue, physical meetings can be replaced by virtual meetings, collaboration tools allow people to work together efficiently and form the basis of the paperless office. Furthermore, on a personal level, it is now easier to keep in touch with friends and family members via social media. People from poorer countries who cannot afford to travel now get access to the first world digital resources like online courses (MOOCs). Disabled and old people can participate fully in social networks because their participation is not constrained by their limited mobility; this helps them maintain or develop cognitive abilities. Children get access to a virtually unlimited source of information about all possible topics leading to a lot more informal learning, including foreign languages.



Figure 16: Evolution of society over a century
Source: Méta-Media

However, there are also some side-effects.

The smartphone has in no time become part of what people are. People feel incomplete if their smartphone is not within reach, they increasingly need it to perform important cognitive tasks. For teenagers, it is part of their personality. They would rather give up owning a car, a television, or a private swimming pool than give up a smartphone. Some people would never date a person with a crack in the screen of his or her smartphone because they believe that somebody who does not care for his or her smartphone will not care for people either [308].

Some people use cyber technology for more hours per day than Olympic athletes train in their discipline. The consequence is that their brains have adapted to this new environment and, in some cases, become addicted to it. They suffer from communication addiction, meaning that they feel anxious when they are not connected (nomophobia – no-mobile-phone-phobia), that they compulsively check incoming messages (easily more than one hundred times per day and even at night; messages raise the dopamine levels in the brain and so they keep checking). Some people even suffer from the phantom vibration syndrome, also called ringxiety or fauxcellarm. It is the perception that a phone is ringing or vibrating when it is not. The fear of missing an important message can lead to an overload of digital information also known as Digital Obesity [177]. People with a communication addiction are very vulnerable to developing an unstable work-life balance and can become chronically distracted, which has a negative impact on their cognitive capacities [362].

The use of cyber technology has some physical effects on humans too. According to studies, our working day has become two hours longer during recent decades [326]. Much of that additional time is spent on the Internet on mobile devices. Apart from

sleep deprivation, this constant use of cyber technology devices can also lead to excess body weight and problems in the hands (De Quervain syndrome – also called texting thumb) from over-use of touch screens.

The Internet and the smartphone do to the brain what using a lift rather than the stairs does to the body. Rather than memorizing information, people constantly refer to the Internet, which can lead to digital amnesia [211], and forgetfulness known as the Busy Lifestyle Syndrome [295]. Skills like mental arithmetic, memorizing numbers (mathematical constants, phone numbers) and driving without a navigation system have largely disappeared in youngsters. More disturbing is that the web is full of texts that fit on just one or two screens, and that many people have lost the ability of ‘deep reading’, that is to say, the ability to focus on a long text for an extended period of time, a skill which is needed to read a book, this HiPEAC Vision or to study [362].

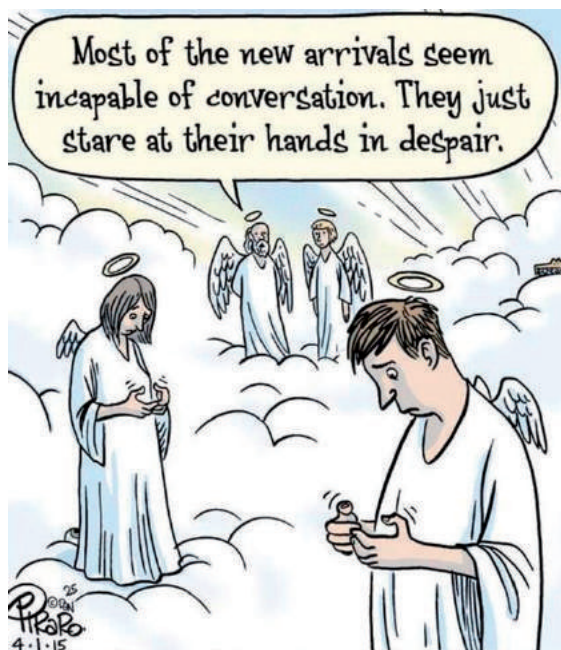
Mobile devices invite users to multitask: to use the device during other activities. Using mobile devices while driving is now forbidden in most countries, but unfortunately it still happens all too often. Using the Internet during meetings is a very common practice even though it reduces the effectiveness of the meeting as people are often mentally absent in the meeting and therefore do not address inefficiency because they are not really part of what has been going on. Many people believe that multitasking increases their productivity, but there is clear scientific evidence that it is detrimental for productivity and for quality of the work, and that it even damages the brain [26]. Instead of interacting socially during meals or at parties, some people prefer to interact with their smartphone, sometimes leading to negative reactions about ‘unsocial’ media by the other people in the room. In some restaurants and pubs, the use of smartphones is forbidden.

Social media punches above its weight when it comes to politics. Many politicians and citizens are eager to immediately react to news items. Those with the most extreme opinions are most noticed and therefore often end up as the ‘trending’ ideas and which eventually dominate a whole discussion. The balanced view of an expert in a newspaper the day after barely matters anymore after the flood of opinions in 140 characters. In the past, discussions about long-term strategies and policies were made in political committees and parliament, and were evaluated at the time of the next election. Nowadays, they are put to the test on social media even before they are discussed in committees making it increasingly difficult to propose unpopular but necessary measures. The Internet makes the job of a politician more difficult than in the past. He or she is being watched by millions of people who have access to very large amounts of information about them, often even from many years ago when they were still in high school or university. Social media is a dream come true for populist politicians however.

RELEVANCE FOR COMPUTING

After many decades of ‘faster is better, more digitalization is better’ mentality, it is now time to accept that there are limits to the growth of computing, and assess in which parts of our lives com-

puting adds value, and in which parts it does not. This is not different from other aspects of life. Fifty years ago, the distance travelled by car per year was considered an indicator of progress; yet it no longer is today for environmental and health reasons. Similarly, consuming lots of energy-rich food and drinks was considered good by people who suffered from lack of food during their youth; it is no longer considered healthy today. It seems likely that we might have to develop a healthy digital lifestyle in the future too.



Source: Dan Piraro

2.3.10. JOB MARKET CHANGES

There have never been more jobs than there are today (which is in line with the fact that there have never been more people than today either), yet employment is not growing for all types of jobs. The numbers of high-skilled and low-skilled jobs are growing while the number of medium-skilled jobs has been shrinking since the financial crisis of 2008 (Figure 17). The medium-skilled jobs are mainly found in manufacturing, in the service industry and in government. The former group is being replaced by cheap labour outside Europe or by automation. The latter group is being replaced by computers, websites, apps, and so on.

This is also illustrated in Figure 18 for the technology-intensive sector. The number of knowledge-intensive jobs is growing while the number of manufacturing jobs is decreasing.

With respect to employment by level of qualifications, the employment of highly-skilled workers is predicted to increase by 26.9% in the period 2013-2025 while employment of medium-skilled and low-skilled workers will decrease (Figure 19). Hence, the best guarantee to find a job is to be trained as a highly-skilled worker.

However, in terms of actual occupations, the picture is different. There is a clear growth in number of high-paid and low-paid occupations, and a strong decrease in the medium-paid occupations (Figure 20). This seems to contradict Figure 19, but it does not. It can be explained by the concept of over-qualification. The number of highly-skilled workers is growing faster than the number of high-paid jobs. The surplus of highly-skilled workers accepts medium-paid jobs, and the resulting surplus of medium-skilled workers is taking the low-paid jobs, leaving large numbers of low-skilled workers jobless. The solution is not to create more low-paid jobs, but to create more high-paid jobs so as to avoid over-qualification of the people in the medium-paid and low-paid jobs.

The current education system is not adequately preparing the next generation to deal with this future because it is still training millions of people for medium-paid routine jobs in professions such as accountancy and administration. Similarly, our political and economic institutions are poorly equipped to handle these radical changes [16].

The limiting factor is not technological advancement, but society's ability to absorb it. It is clear that education, our political systems, unions, markets, economic incentives, welfare systems and legal frameworks are currently too rigid to deal with the speed of technological innovations in a humane way. It is a major challenge for society to cope with this change.

RELEVANCE FOR COMPUTING

There is disagreement among analysts about whether or not technology will eventually create enough jobs to compensate for those lost. Some analysts call technology the biggest job creation machine in history [415]. Based on historical evidence, they claim that technological innovations have always created more (but different) jobs than they have destroyed, and that the productivity gains made products and services cheaper leading to increased demand, and hence to economic growth and more jobs. The proof is that there have never been more jobs than today, and that humanity has never in its history been living in a more prosperous world than today, thanks to technology.

Other analysts predict that this is true for the past, but that this time is different and that, in the long term, 50-75% of people will be unemployed simply because there is too little work left for which the economy needs humans: jobs will either be very specialized or will require hardly any skills at all [363]. They predict that the number of middle-class jobs will keep shrinking and that income inequality will continue to rise. This is not good news for the high-end consumer market. Households with low-paid jobs spend a larger share of their income on basic needs (rent, food, medical costs) and so have less money to spend on consumer electronics. Middle-class households are the cornerstone of the market for consumer electronics; if their number shrinks, so might the consumer electronics market.

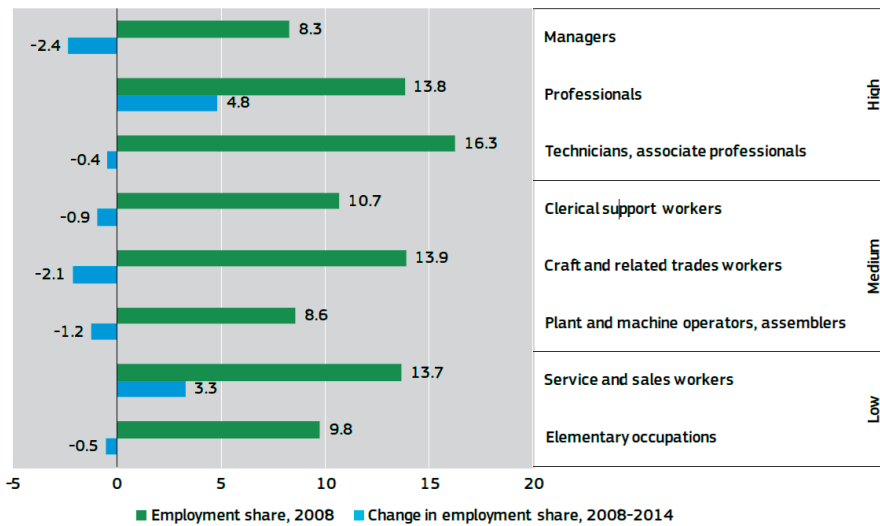


Figure 17: Employment by occupation as % of total employment in the EU (2008) and change in % points between 2008 and 2014
Source: European Commission

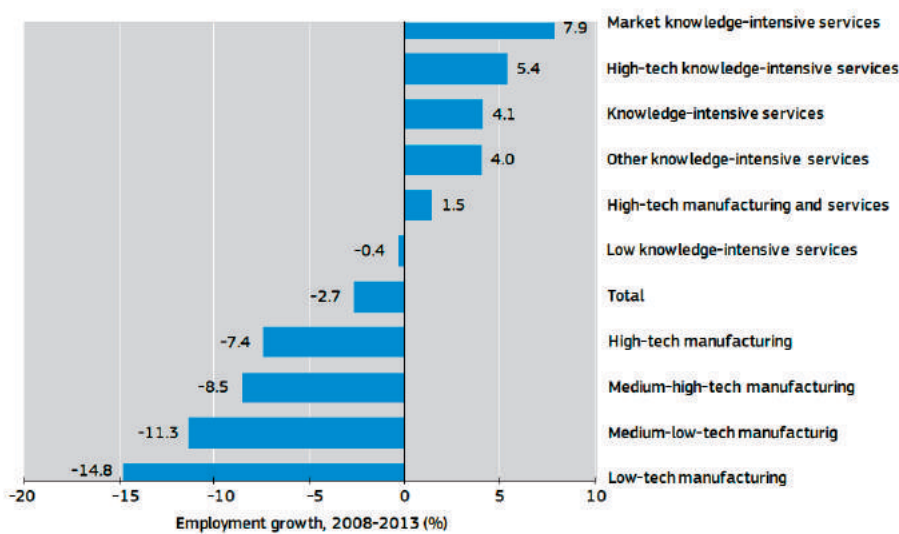


Figure 18: Growth in employment in the EU by technology intensity sector between 2008 and 2013
Source: European Commission

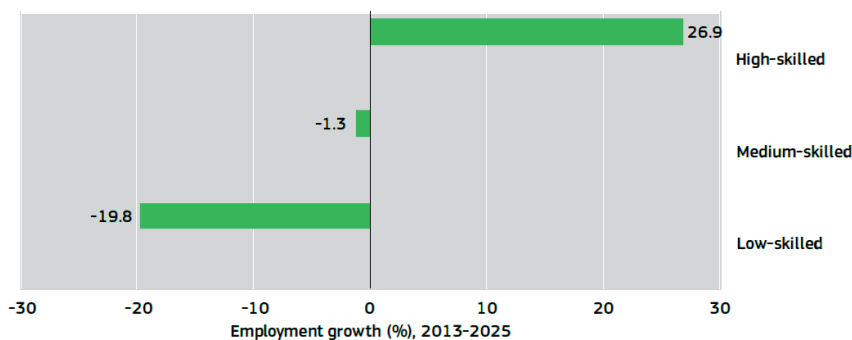


Figure 19: Forecast of employment growth in the EU by qualification 2013-2025
Source: European Commission

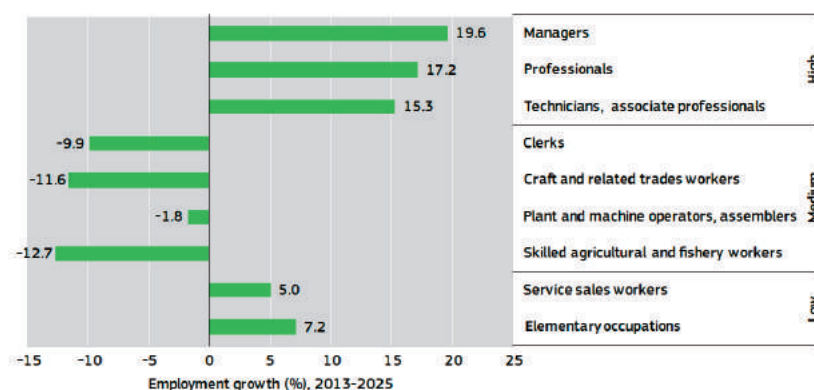


Figure 20: Forecast of employment growth in the EU by occupation 2013-2025
Source: European Commission

2.4. MARKET TRENDS

2.4.1. GENERAL TRENDS

The electronic systems market is very important and it is forecast to be worth US\$1.6 billion by 2020, according to IC Insights.

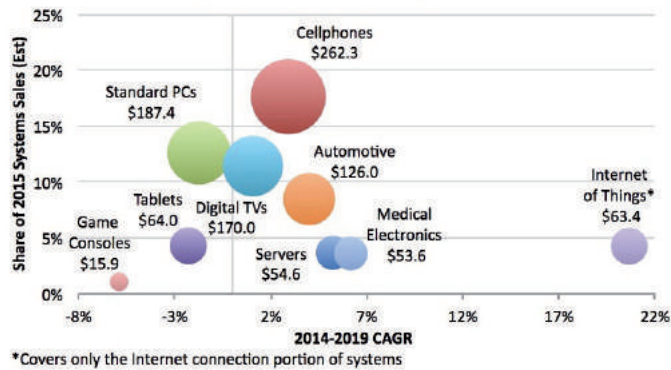


Figure 21: End-use systems market and growth rates
Source: IC Insights

This is a market which evolves rapidly: today's top dog may be gone in a few months' time. This is well illustrated by the hype curve of the emerging technologies proposed by Gartner (see Figure 22).

Even if few technologies are optimistically positioned, we can isolate several trends that seem to continue and are sound for the coming years. Within the market structure, we can observe:

- Cannibalization of mobile digital devices such as cameras, MP3 players and GPS by smartphones;
- Saturation of PC, laptop, tablet and smartphone markets;
- The cloud will remain important, but more and more processing will be done at the 'edge', or 'on-the-fly' before being stored in the cloud servers. This is motivated by cost, privacy and reducing the aggregated bandwidth;
- Evolution of the processor landscape in several respects. Due to the huge success of mobile devices, Intel had to revisit its strategy and ARM is the de-facto standard for smartphones. We also observe that more and more countries want to have their own ICT production. In a sense, this can also be extended to small groups with a 'maker' attitude that develop their own system independently from big companies;
- Facilitation of processor evolution by open source, which is gaining inroads in the hardware market; and open source software very present in some areas (compilers, etc.);
- Cryptography, in all its forms, including blockchain, as a key technology, and not only for ensuring privacy and security.

However, amongst the emerging technologies, it seems that intelligent, smart, cognitive devices are increasingly important thanks to the development of AI. It is not the Artificial General

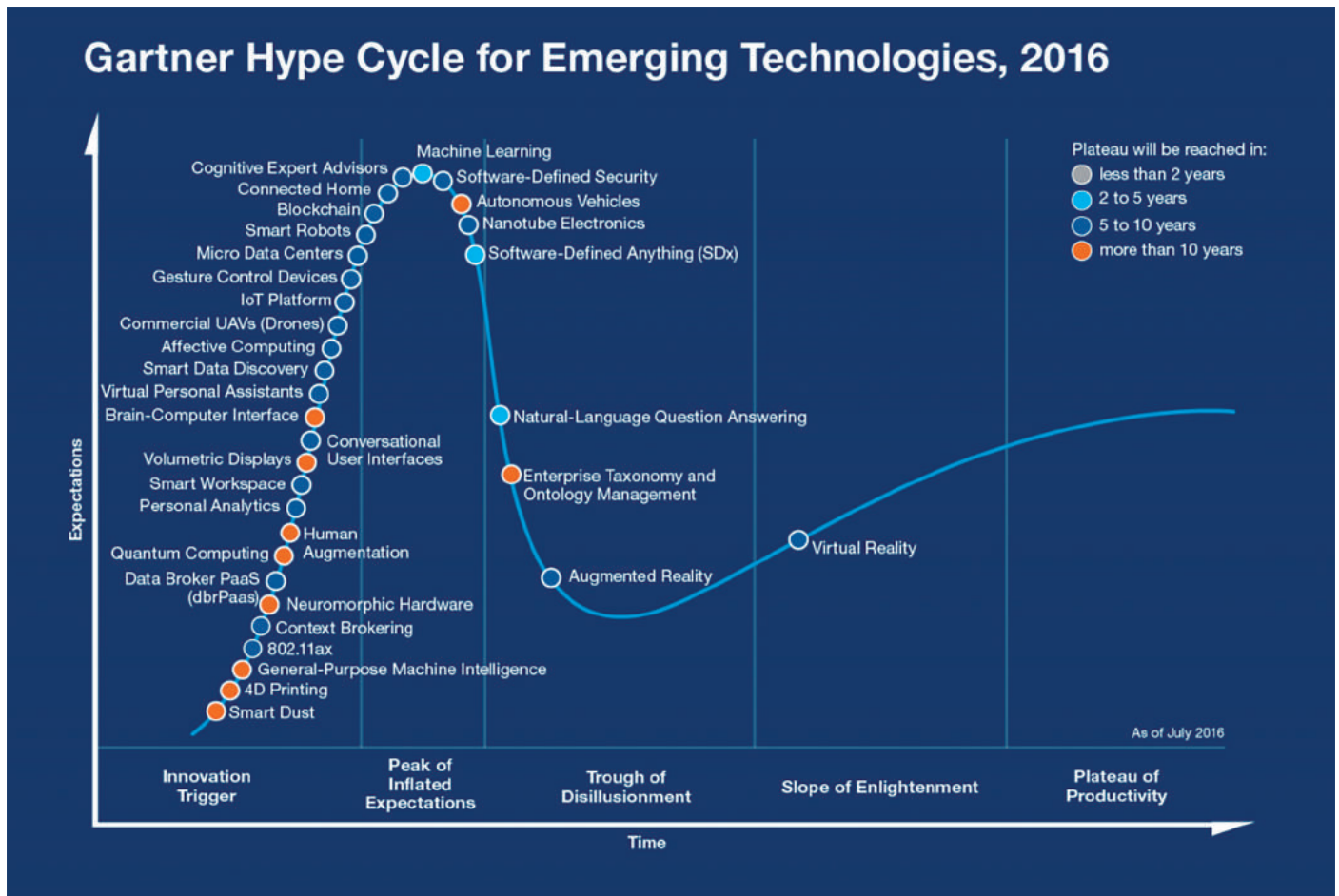


Figure 22: Gartner hype cycle
Source: Gartner Inc.

Intelligence (AGI), but more 'narrow' AI, which is specialized in a specific domain like, for example, oncology or easing the use of services. This has fuelled the development of new personal assistants that communicate with the user with multimodal interfaces, e.g. screen, voice, gesture. On the Gartner 'Hype Cycle', such properties and competencies include *Natural-language question answering, Autonomous vehicle, Machine learning, Cognitive expert advisors, Smart robots, Gesture control devices, Affective computing, Smart data discovery, Virtual personal assistants, Brain-computer interface, Conversational user interfaces, Personal analytics, Neuromorphic hardware, and General-purpose Machine Intelligence*.

Another important domain is the increased realism of computer-generated graphics, either in games or for industrial simulation. Virtual reality will become an important topic, not only in the games industry, but also in 'real' life with augmented reality. According to Gartner, augmented and virtual reality will come after the peak of inflated expectations.

The combination of new sensors and actuators, AI and virtual or augmented reality will mean that computers, as we know them, will disappear and will blend into the environment. After its first announcement, more than ten years ago, we are really entering the era of the 'disappearing computer'. Smart IoT and CPS devices will be omnipresent. The computer will not only blend into the environment, it will also blend into ourselves, and augment us.

Therefore, we consider that the new main market trends that will have an impact on the HiPEAC community could be:

- More and more 'cognitive', intelligent, smart devices everywhere
- Virtual/augmented reality
- The disappearing computer
- The 'augmented' human.

2.4.2. CANNIBALIZATION OF DISCRETE DEVICES

The introduction of the iPhone by Apple on 29 June 2007 was the trigger of a profound revolution in the ICT market. Smartphones became the 'Swiss Army knife' of the beginning of the 21st century. Before, companies used to develop specialized devices for each function: pocket PCs/PDAs, audio recorders, MP3 players, GPS (road or track), cameras, video recorders, compasses, pagers, mini flashlights, voice-recorders, body and health monitoring systems, portable gaming systems, remote controllers, watches, alarm clocks, timers, electronic dictionaries, radios, and much more. Even paper maps, tickets and cheques have now been replaced by map applications, e-ticketing and e-banking on smartphones. The smartphone has become the universal digital information device, at the expense of discrete application devices and their manufacturers [194, 336].

Large screen smartphones even cannibalize sales of tablets, while the PC market continues shrinking [145, 270].

Smartphones have changed habits and trends of the past by using the versatility and the processing power of CPUs. Apple provided a white sheet (in fact, a black screen) of paper to external developers. They provide the programmable processor, numerous sensors and APIs, and it is then up to external application ('App')

developers to customize the hardware. The touch screen didn't impose a pre-defined user interface with buttons and knobs: everything can be customized depending on the application. The device was small enough to be a real 'personal computer' that people can put in a pocket and always have with them. If the local processing power is not enough, the distant servers of the cloud, through a wireless connection, allow the device to perform functions that cannot be done locally.

The smartphone is now a personal Swiss Army knife device that is replacing many discrete devices. As it is small and always with the user, it is more convenient than carrying a number of specialized devices. A lot of traditional companies did not see this coming; as a result, their markets have shrunk and specialized devices are now more and more often found in high-end niche markets (loss-less audio, high quality cameras, professional devices). For the average user, the smartphone is the perfect trade-off between quality and convenience for daily use.

Smartphones and tablets have even cannibalized the PC market for some functions. Young people prefer to watch video on smartphones or tablets instead of on TVs or PCs. They chat and communicate through apps and no longer via phones or PCs. PCs are now predominantly used only for professional or office tasks.

This has had a profound influence on the computing landscape: the market is now dominated by manufacturers of SoCs and processors for smartphones rather than for PCs. The ARM architecture is dominant in this field at Intel's expense.

Despite more than two billion users in 2016 (and more than 2.6 billion predicted by 2019 [434]), market growth is slowing down, in spite of the fact that the lifetime of a smartphone is short (in the range of two years) compared to that of other consumer devices. What will be the next step after the smartphones? Will it be the Internet of Things (IoT)? IoT will be very diverse and specialized with many elements connected via and to smartphones, which will certainly stay the main 'personal' device until the era of 'disappearing computer' arrives.

2.4.3. SATURATION OF THE PC AND MOBILE PHONE MARKETS

Western European mobile phone markets, as well as US and Chinese markets, are soon reaching saturation point [178, 263].

As a result, increased competition on prices is expected, both on the device side and on the phone company side. New popular applications such as Pokémon Go may again increase the sales of new, more powerful devices, but these applications will have to be devised and to be sufficiently innovative to push consumers to upgrade their device in order to use them. A relatively easy path for this is the cannibalization of other, already existing markets, by integrating ever-more external applications into the ever-increasingly central and ubiquitous smartphone. Such applications might encompass GPS (road or track), cameras, body and health monitoring systems, portable gaming systems, remote controllers, and so on. The main role of the smartphone is to capture, process and store information.

We also see in industrialized countries that more and more users have more than one smartphone; for example, one for work and one for private use. This is mainly driven by the fact that the cost of the hardware of smartphone is often supported by carrier providers for keeping their subscribers captive with long duration plans. Therefore, they are not supportive of dual or multi-SIM phones that will unlock users from their services. Initiated by Apple, the market was moving from low, basic phones to multi-usage high-end phones with a high cost (and often a high margin for the manufacturer). Even if more functionalities are added to smartphones (dual camera, RFID, smart payment, etc.), the new features are less and less sufficiently attractive to drive users to buy the latest smartphone, and it might be the start of the replacement market.

The increased screen size of smartphones is cannibalizing tablets, which are also being replaced more and more by hybrid PC/tablets. It is possible that what will subsist in few years' time are smartphones with screens large enough to be comfortable, but small enough to always being carried by the user, and hybrid devices with a large screen and a removable keyboard, that will replace the laptops and even desktop at least for personal activities. A smartphone provides a number of services and functionalities, such as display, local and global connectivity, computing power, and user interface, that can be coupled with a large number of devices in our lives: toys, tools, household appliances, sensors, etc. This market will require appropriate applications for smartphones to be developed, as well as the fitting of machine-to-machine interfaces (via Bluetooth, Zigbee, etc.) within the devices. At the same time, it is likely to entail the use of smartphones as a host platform by appliances that today have dedicated user interfaces and computing power. This could make these appliances cheaper to manufacture [39, 46, 261, 344].

These trends should reinforce the need for scientific and technical solutions to create more compact, less power-hungry and less energy-hungry programs on smartphones and in other embedded systems. Research on IoT and communication protocols is also likely to be stimulated.

2.4.4. FOG AND EDGE COMPUTING

Due to the large amount of data generated by users (including pictures and videos), and the need to be online all the time and to share information (social networks), we have moved from stand-alone desktop computers to mobile devices connected to the cloud. In the cloud, data is stored on remote servers, processed by companies, and can be exchanged and addressed by multiple terminals of various types (from computers to smartphones, tablets and wearables). Current computing and storage clouds, both for private and for business users, are mainly hosted by large companies like Google, Amazon, Microsoft and DropBox. This allows such companies to tune their hardware and software stacks to customer and periodic usage patterns, and overall strategic directions of their own business.

However, a growing awareness that this data is often abused by spy agencies, private companies and malevolent hackers encour-

ages a move towards people keeping their own data closer to them. Thanks to increasing storage and broadband capacities available at reasonable cost to home users, the latter tend to keep their data in Mesh-like distributed environments (IoT) at home, as well as in private personal NAS/SAN devices. Individual data stores can also be mutualized: shared among a limited number of trusted users, in which case they can be called federated, or distributed, clouds. An important question then arises: how do we ensure optimal processing, distribution and safeguarding of distributed data?

Furthermore, new disruptive technologies, like non-volatile storage, can easily change the data storage and distribution landscape. If they could have several Exabytes of storage for a cheap price in 10 cm³, users and (small) companies might prefer to store their data in a device that they own and of which they know the location, leaving the current cloud computing for the fog computing approach [94]. In that case, most computation could also be performed locally, as edge computing [77], while only the more advanced functionality would be remotely distributed, having access to only the (meta)data it does need, not to all raw data. In that case the problem of data confidentiality as it exists today with unified clouds would again arise. Instead, these remote applications should only be provided with the information required to perform the task. This information could moreover be anonymized, or limited to statistical or to metadata, thereby abstracting the real information from the user. Reliable anonymization and anonymizing statistical abstraction of information are thus probably a necessary feature for the concept of federated clouds to really take off.

Another emerging approach is to send encrypted data to the remote application (i.e. homomorphic encryption), that then performs its operations without ever decrypting. As a result, the application never knows the actual data nor the meaning of the results it computes. This is the ultimate solution for keeping data private, but it runs against the current business model of companies such as Facebook and Google that are built on gathering and reselling as much information about their users as possible (*'If the product is for free, you are the product'*).

2.4.5. THE NEW PROCESSOR LANDSCAPE

The landscape of the processor market is slowly moving. Even Intel seems to be reconsidering its strategy, which was previously based on PC and high-performance systems. ARM is the de-facto leader in the architecture of smartphones and bets are still open for the emerging market of processors for IoT devices.

IBM also appears to be slowing down its processor activity and opening its POWER architecture [92]. Both Intel and IBM are making significant staff redundancies.

Hence, the PC market is no longer the dominant market and the smartphone market is getting saturated. As long as the IoT processing engines are still very diversified, it remains very difficult to predict the future leader in the processor domain (if there will be one). A 'killer' application could make a new strong player in just a few years' time.

2.4.5.1. SMARTPHONES DRIVE THE DEVELOPMENT OF PROCESSOR ARCHITECTURE

ARM-based processors are the de-facto standard for the smartphone market (more than 90% of the market share) and Intel is giving up its line of Atom processors for smartphones [205], becoming now also a foundry for ARM processors [73].

The new strategy of Intel is within the core *growth* areas of cloud and data centre, IoT, memory and programmable solutions [128]. In this context, it did not come as a surprise that Intel bought FPGA-vendor Altera in 2015.

The new line of low-power Intel processors (Atom) are now mainly targeted towards IoT, low-end networking processors, and even low-cost personal computers [230]. This shows that the PC market, once the driving force behind Intel's growth, is no longer the main market. IoT and CPS are now seen as the new growth direction. ARM, with its established position in the smartphone markets due to its low power processors, is already well positioned for the IoT market too. In the domain of ultra-low power devices for IoT, although ARM is very well positioned, other microcontrollers will be able to quickly grasp a share of the market if they have a very good energy efficiency. It is yet unclear which companies will become dominant in the IoT market and will eventually determine the (ad hoc) standard.

ARM is not a chip provider, but sells IP blocks (i.e. the blue prints) allowing its customers to design ARM processors. The key asset is the architecture definition (instruction set, interrupt model, debug interface, privilege levels, etc.): all designs must be compliant, and therefore interoperable with the software ecosystem designed for the particular set of ARM processors. The companies that acquire an architecture licence (such as Apple and Broadcom) are allowed to make their own microarchitecture as long as it is compliant with the reference architecture.

2.4.5.2. VERTICALIZATION IS STILL CONTINUING

We continue to observe the verticalization of the market. After Apple, which is making its own processor, Google has revealed that it is also making chips, with its TPU used as an accelerator for Deep Neural Networks (DNN).

	before 1980	1980-2000	2000-2014
MATERIALS		INTEL	ARM
COMPONENTS			
DESIGN	IBM	DEC	GOOGLE
ASSEMBLY		DELL	SAMSUNG
OPERATING SYSTEM		MICROSOFT	APPLE
APPLICATIONS		DEVELOPERS	
SALES/DISTRIBUTION		BEST BUY	
Industry wave:	vertical	horizontal	vertical
Strategic Direction:	benefit leadership	cost leadership	benefit leadership
Innovation Type:	integrative innovation	disruptive innovation	integrative innovation

<http://www.theinnovativemanager.com/innovation-lessons-steve-jobs-apple-story-iphone/>

Figure 23: More and more verticalization

Source: The Innovative Manager

Another illustration of verticalization is that ARM, a fabless semiconductor company, was acquired for €29 billion by Softbank [327], a Japanese company originally specializing in telecoms (it bought Sprint in the US) and which is also active in robotics (it acquired the French company Aldebaran at the origin of the Pepper robot) and in renewable energy.

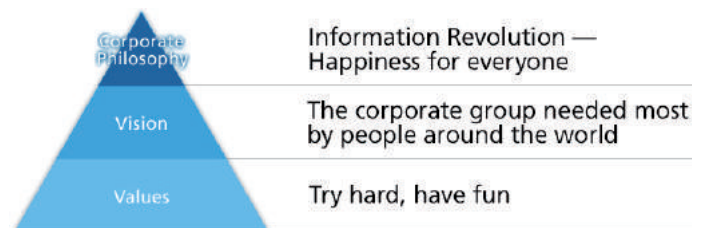


Figure 24: Softbank corporate philosophy, vision and values

Source: Softbank



Source: Slashgear

About 25 years ago, Europe had several vertically integrated companies that had multiple activities and that were involved in many application domains. Companies like Siemens, Philips and Thomson covered domains from basic silicon technology up to the end product. Technologies were shared between application domains. Then, about 15 years ago, more constrained economics, combined with the growing pressure on shareholder value, called for more 'horizontal' companies, i.e. companies focused on the domains in which they excel. This led to the spin-off of semicon-

ductor divisions (NXP from Philips, Infineon from Siemens, STMicroelectronics from Thomson), among other restructurings. Today, Europe is full of many specialized companies, each of which focuses on its own know-how and on its part of the value chain. They are squeezed between their providers and their customers, who in turn also try to maximize their margins and thereby put pressure on other players lower or higher in the value chain.

Since 2000, especially in the domain of consumer electronics, companies that control a large part of the value chain (from technology, to hardware, to software, to devices, to services, to end users' systems) have gained dominance. Thanks to their diversified activities, they also weathered the financial crisis fairly well. They achieved this by creating complete ecosystems and locking in the end users. By controlling the end-user data, they have access to a gold mine of useful information that allows them to further improve their ecosystem.

Google, starting as a search engine, now collects a lot of information about users and their behaviour by tracking their web activities, by analysing their free mail service and by locating them thanks to Android phones, using Google's operating system. Google is building its own data centres and has tablets and phones with their own brand name, even if the design and construction was sub-contracted. It is investing in new generations of devices – such as wearable ones – with smart watches and augmented reality tools. Google is also making chips.

Apple is also enlarging its part in the value chain by designing its own processors. By controlling the hardware and the software, Apple can have optimized solutions without having to pay for extra layers of interoperability. The software is tuned to the hardware and vice versa, allowing for reduction of the amount of memory required, for example, thereby saving costs. Amazon, Facebook and Samsung also have been trying to grow their share in the complete value chain, from basic technology to devices and retail shops.

Two key elements that explain the success of vertically integrated companies are that they do not have to pay the cost of interfaces, compatibility layers or standards to link the various pieces together, and that they can optimize (costs) globally, which is more efficient than trying to optimize all parts independently.

In the domains covered by HiPEAC, it would be interesting to see if one or two strong European leaders could emerge and crystallize a coherent ecosystem around themselves (a sort of 'Airbus' of computing systems). This might be possible given that the consumer market is slowly moving away from traditional PCs, smartphones and tablets towards 'interconnected things', with new constraints of low power, safety, security and the deeply embedded in the physical world.

At least as important as having a couple of tent-pole companies is the ecosystem that they create to operate in. This ecosystem consists of suppliers, service companies, universities, research institutes, amongst others. That same ecosystem also attracts start-up companies that develop technology that might eventually be used by the tent-pole company. The perspective of becoming a supplier, or even of being bought by the large company, at-

tracts venture capitalists interested in investing in such technology start-ups, looking for quick profits. It is far more attractive for venture capitalists to invest in companies that already have a potential exit plan. Many European technology start-ups are eventually acquired by non-European companies as a result of the fact that there are very few European companies interested in buying out such start-ups. There are counterexamples though: Gemalto recently acquired the US-based Safenet. Sysgo was acquired by Thales in 2012. Gaisler Research was acquired by the US-based Aeroflex in 2008, and then bought back in 2014 by the UK company Cobham plc.

Tent-pole companies are not created overnight, but they grow vertically and horizontally over time. Apple started as a desktop computer company and gradually expanded into other markets. Amazon started selling books, and is now a major player in the cloud business. These companies re-invent themselves regularly. Companies like Apple reinvented mobile telephony; Google is re-inventing car-based mobility. European companies could also have expanded into different sectors, but most of them did not. There are success stories though. Nokia successfully transformed from a paper mill and rubber factory to a cable factory, a telephone switch company and a car telephony company, to finally become the global market leader for mobile phones that it was for more than ten years. Unfortunately, it didn't reinvent itself in time to cope with the emergence of smartphones. Europe could use more companies that want to become global market leaders.

2.4.5.3. SEVERAL COUNTRIES DEVELOP THEIR OWN PROCESSORS

As was explained in 2.3.12, several countries are currently developing their own processors, thereby avoiding reliance on US processor vendors such as Intel.

China is developing a full range of different processors. BLX IC Design Corporation (founded in 2002) focuses on designing the 64-bit Loongson general-purpose and embedded processors, together with developing software tools and reference platforms [97]. China's top ranked computer in the TOP500 HPC, the Sunway TaihuLight, is based on a new chip design. It is interesting that the features of this custom-designed ShenWei SW26010 processor are quite different from 'classical' high-end microprocessors: instead of having a very complex architecture, it has fewer caches and a simpler architecture (for example, only 1 thread per core), hence its lower power consumption. There are 260 cores per node (4 clusters of 64 processing elements (called CPE) and one management core per cluster) running at 1.45 GHz [297].

Several other processors are developed in China, we can also note the Phytium Technology (Tianjin, China) FT-2000/64, aimed at 'high throughput and high performance servers'. They are based on ARMv8 cores and they are very competitive in performance and in energy efficiency [182]. Its main characteristics are described in Figure 25.

- Process: Manufacturing with 28nm process
- Core: Integrating sixty-four FTC661 cores
- Frequency: Running at 1.5GHz~2.0GHz
- Cache: Integrating 32MB L2 cache and extending 128MB LLC
- Extension Interface: Integrating eight proprietary extension interfaces, each delivering 19.2GB/s effective r/w bandwidth
- Memory Interface: Extending sixteen DDR3-1600 memory controllers, which can deliver 204.8GB/s memory access bandwidth.
- I/O Interface: Integrating two x16 or four x8 PCIe Gen3 interface
- Power: Max. power 100W
- Package: FCBGA package with 2892 pins



Figure 25: Characteristics of the FT-2000/64
Source: Xinhua

Japan is also developing its own solution for HPC, after the Earth Simulator (a vector machine built by Nec), the K computer (built by Fujitsu) and the PEZY-SC (built by TSMC). The K computer is very efficient at running the new High Performance Conjugate Gradient (HPCG) test that is a follow-on to Linpack to measure HPC performance. HPCG is considered to be more representative of actual HPC loads than Linpack. Fujitsu has been commissioned by Riken to build the next Exaflop machine by 2020. But Fujitsu is switching architecture for this 'post-K' computer: it chooses the ARM architecture instead of its previously developed SPARC architecture [298]. Incidentally, as previously mentioned ARM is soon to be under Japanese ownership, having been purchased by Softbank (see section 2.4.5.2). One side effect could be that Fujitsu will enter the market of ARM-powered servers.

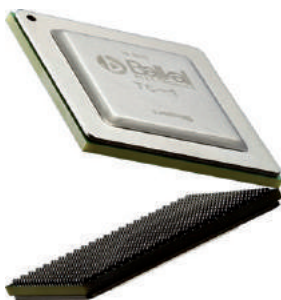


Figure 26: The Baikal-T1 micro processor
Source: Baikal Electronics

Russia is also developing its own processors, e.g. the Baikal-T1, a processor for routers, gateways, industry and embedded applications. It is a SoC with two MIPS 32 bits Warrior P5600 cores 1.2 GHz, with advanced security, virtualization and SIMD units. It has a full set of peripherals: Ethernet (10 Gb/s and Gigabit), PCI Express, USB and SATA. It is made in 28 nm and has a power consumption of less than 5W.

Russia is also developing the Elbrus line of processors (from MCST) for mainframes and servers. The Elbrus-8C will reach 250 Gflops thanks to its 8 VLIW cores running at 1.3 GHz. It is even code compatible with the x86 instruction because of on-the-fly code translation [324].

In the area of operating systems, Russia will develop its own Android-like OS [333] to become independent from the US-originated Android and iOS.

India is also starting development of its own range of processors, this time based on the open source RISC V 64-bit architecture [240].

There are several activities that are important for developing a high performance SoC or microprocessor:

1. Software, compiler and operating system development. This could be a large investment, but mainly consisting of 'brain power' and knowledge. It could be leveraged by adapting open source software;
2. System architecture, SoC or microprocessor architecture. Generally, the investment consists of a small team of very skilled people. Existing architectures can be reused, because they are either open source (e.g. RISC V) or accessible by buying the IP from an external company (ARM being the best example of a company selling such IPs, but also IMG in the case of MIPS). The IPs are generally in readable format, so that they can be inspected and modified to cope with new needs (according to the licence bought);
3. Microarchitecture. This step is not always necessary, but it is if new innovations or better performance is required from the 'standard' licence acquired in phase 2. This requires a highly skilled team. Apple, Broadcom, Fujitsu and others are doing this for their implementation of the ARM architecture (they have what is called an 'architecture licence' allowing them to make whatever changes they want to the core, as long as it complies with the ARM architecture).
4. Design of the SoC or microprocessor, leading to a netlist. This is done with a larger (but easier to find) team than for the architecture, but the key ingredients are the software tools for the design, place and route, synthesis and so on. No real open source tools are available unlike in the domains of compilers and operating systems. The main providers of these tools are US-based companies (including Cadence, Mentor and Synopsys), although Mentor was recently bought by Siemens [419]. From a conceptual viewpoint, these tools 'compile' in space while a classical compiler compiles in time. It should be noted that FPGA tools are in between classical compilers and the ASIC tools;

5. Foundry of the SoC or microprocessor. For advanced nodes, until now, only US, Taiwanese (TSMC) or Korean (Samsung) companies have been prepared to make the investment of working on advanced nodes. China is also consolidating its local foundries [254]. We observe that the new high-performance chips from China or Russia use only technologies above 22nm, so this node, which is recommended for IoT devices (in its FDSOI version [385]) might also be enough for current strategic developments. Recently, Global Foundries in Dresden announced a 12nm version of the FDSOI technology [264] which keeps it in the competition for further denser technology nodes. It is also important to consider that a key bottleneck for deep sub-micron technology is the lithographic equipment, where the leader is still European (ASML [386]);

6. Test and validation: Test equipment is expensive, but accessible;

7. Realization of the system. System integrators are still present in most major countries, even in Europe, and they can integrate the components to make a complete running system. For complete systems, besides the processors, some other key components are required:

- DRAM: The main providers are Korean (Samsung, Hynix) and from the US (Micron)

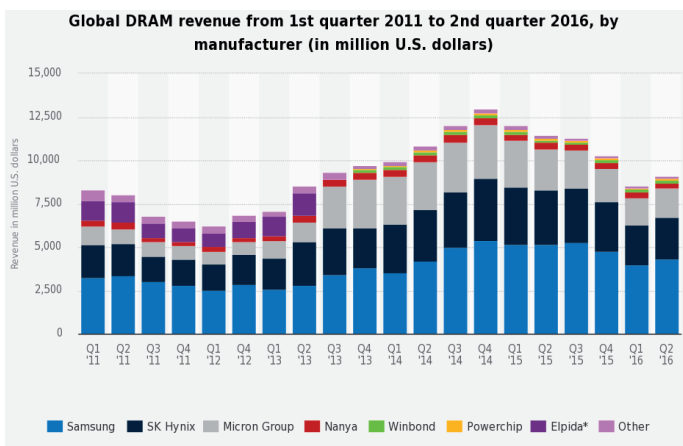


Figure 27: Global DRAM revenue 2011-2016
Source: Statista

- FLASH: the main manufacturers are from Korea (Samsung, Hynix), Japan (Toshiba) and the US (Sandisk, Micron)

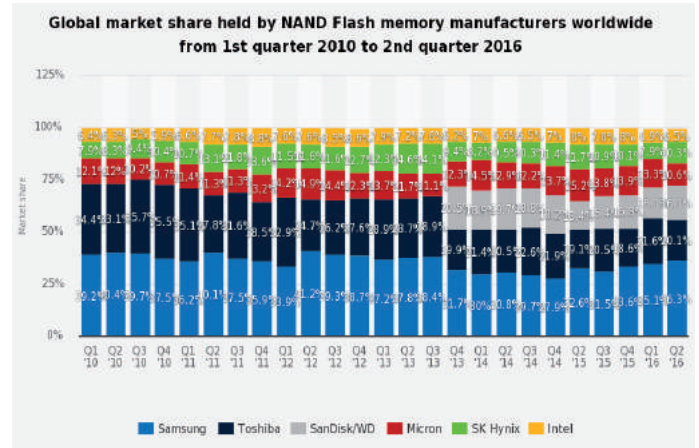


Figure 28: Global market share of NAND flash memory manufacturers 2010-2016
Source: Statista

- Hard drive: only three manufacturers remain from the more than one hundred in operation in previous years: Toshiba (Japan), Seagate and Western Digital (US).
- Interconnect chips: similar to microprocessor development, except that for the phy (physical = analog interface), specific knowledge on analogue design is required. Europe seems still to be strong in that field.
- Interconnect (photonics): also, a strategic domain for supercomputers and servers, allowing fast communication between racks.

Europe still has a good pool of knowledgeable people who could develop efficient architectures, but big European semiconductor companies are focusing more on ‘smaller’ ICT devices like micro-controllers, with application in automotive, network, smartcard, or sensor nodes. As these markets are growing, it seems that this has been a good approach. However, not being present in the ‘high-end’ computing solutions might have drawbacks, besides the dependence on providers outside Europe. What is high-end today will become consumer tomorrow, and it is difficult to catch up because of the increasing complexity of both hardware and software. High-end, low-cost and affordable computing solutions will be required for self-driving cars. For example, the BlueBox autonomous driving platform from NXP delivers 90,000 MIPS at under 40W which would make a very competitive PC – see [387]. ‘Intelligent’ devices will need more and more local intelligence, and a good knowledge of managing highly complex designs and software is an important asset.

Furthermore, as previously explained, Europe needs to address the complete ecosystem including education: if companies don’t work on those topics, they will not recruit staff or use the research results of academia. Therefore, the attractiveness of the field will decrease and fewer and fewer people will train in those topics resulting in it being very difficult to catch up later. Results of projects or new start-ups will be used outside of Europe and specialists will move out of Europe or out of the domain.

2.4.5.4. THE 'MAKER' OR DO-IT-YOURSELF MOVEMENT

New, cheap (less than €50) board computers and new, cheap programmable boards based on microcontrollers and microcomputers like the Arduino [85, 142], Raspberry Pi [106, 170], BeagleBone Black [22, 91], and Intel's Galileo and Edison controllers, are easy to program and make it relatively easy for hobbyists and start-ups to create new devices and artefacts that can be used in daily life. As such, they bring computing and electronics to a large number of interested people, beyond hard-core geeks, well in sync with the growing current trend known as the 'maker movement'.

"The maker movement is primarily the name given to the increasing number of people employing do-it-yourself (DIY) and do-it-with-others (DIWO) techniques and processes to develop unique technology products. Generally, DIY and DIWO enables individuals to create sophisticated devices and gadgets, such as printers, robotics and electronic devices, using diagrammed, textual and or video demonstration. With all the resources now available over the Internet, virtually anyone can create simple devices" [173]

Launched in 2012, the Raspberry Pi alone had sold more than 8 million of its various versions by February 2016 [106].

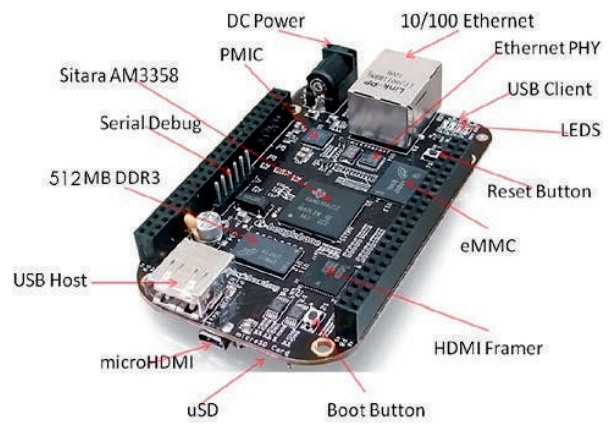


Figure 31: BeagleBone Black

Source: Probotix

Even cheaper boards exist, dubbed as '\$5 microcontrollers' or '\$5 computers' that, on a small board, feature a processor, memory, I/O ports and network connectivity, and run a Linux-based OS.

One example is the self-built MC HCK, a small, versatile ARM Cortex-M4 based microcontroller platform, which is entirely open source and created by the community [121].

Other \$5 computers come fully built, such as the Raspberry Pi Zero [171] or the ESP8266 [52]. These extremely cheap board computers make it possible to experiment in even wider ways, costs not being an issue anymore.

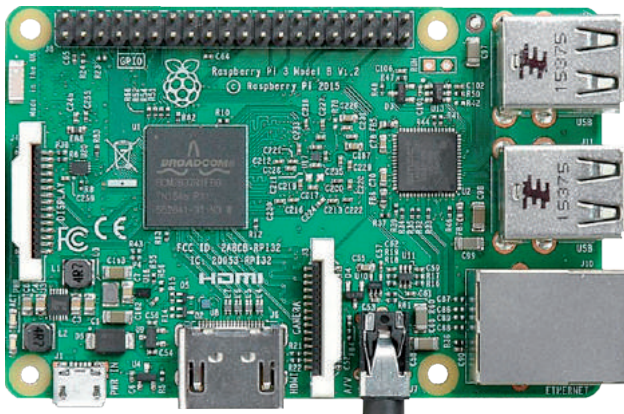


Figure 29: Raspberry Pi 3 Model B

Source: Herbfargus/Wikipedia

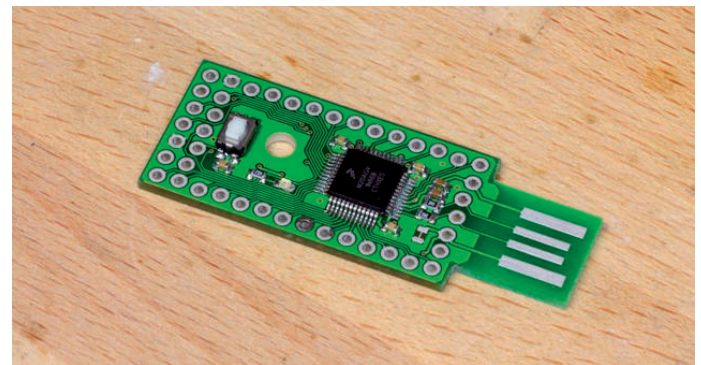


Figure 32: MC HCK board assembled

Source: Simon Schubert and the MC HCK project

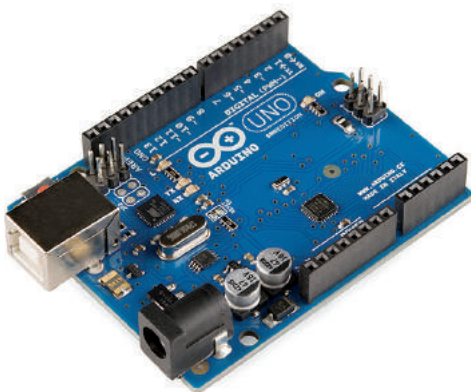


Figure 30: Arduino UNO - R3

Source: SparkFun Electronics/Wikipedia



Figure 33: Raspberry Pi Zero

Source: Gareth Halfacree/Flickr

The big players are also engaging in this direction, because those boards are enablers for IoT developments (and as we don't know yet what is the killer application, a broad exploration of various solutions is interesting).

ARM, after 'personal' initiatives from its employees, internalized the mbed platform, in which all the compilation tools are cloud-based [166]. Mbed aims to simplify and speed up the creation and deployment of IoT devices based on ARM microcontrollers. As stated by ARM: 'The ARM mbed IoT Device Platform provides the operating system, cloud services, tools and developer ecosystem to make the creation and deployment of commercial, standards-based IoT solutions possible at scale.' This is a collaborative project between ARM, partner companies and a community of mbed developers.

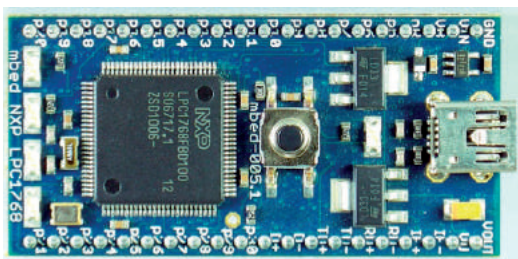


Figure 34: Mbed board with NXP LPC1768 ARM Cortex M3 MCU
Source: Viswesr/Wikipedia

Similarly, Intel, after releasing Edison its 22nm dual-core PC the size of an SD card [152], is now proposing the Joule platform [127]. Intel Joule is a high-performance System-on-Module (SOM) based on Atom processors in a tiny, low-power package. This platform aims at enabling people to rapidly prototype a concept and then take it into production in a fraction of the usual time and development cost. However, the high-end version of this platform was for sale in August 2016 for US\$369, which makes it about 10 times the price of a Raspberry Pi, hence targeting higher-end, professional developers rather than hobbyists. The lower-end version is yet to be released.

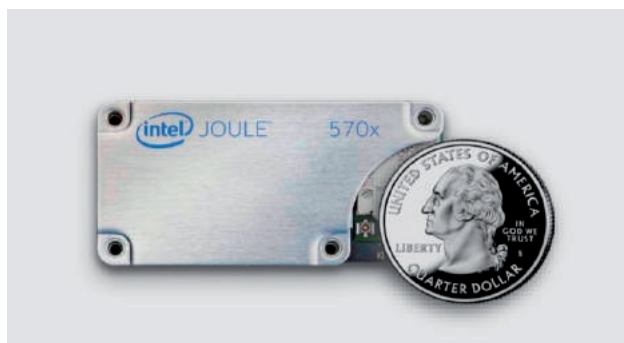


Figure 35: Intel Joule 570x
Source: Intel

On the software side, the maker movement produces and benefits from a large number of online tutorials, courses, helpers, tools and various kinds of resources very much in sync with the open source movement. Efforts are being made to allow for easy programming of these inexpensive devices, to help spread them fur-

ther and to help propagate the literacy they instil in their users [61, 199]. An example of this is the Arduino language, which is a simplified version of C++ [143] with a large number of libraries. The structure of Arduino programs, called 'sketches' in Arduino parlance, have a few specifics to ease development, such as the mandatory `setup()` function for initialization and the `loop()` function that is run continuously [133].

2.4.5.5. IMPACT AND PROPOSED COURSE OF ACTION

Since these 'cheap board computers' of today correspond to high-end configurations of a few years ago, they make use of hardware and software components that have withstood the test of time. They should thus not incur R&D costs, nor call for new solutions. These cheap board computers and the associated maker movement nonetheless have a number of significant implications.

First, they make it possible to have a very large number of small scale experimentations, thus fostering multi-directional and unexpected innovations in technologies and their applications. As such, design-space exploration could be significantly widened by these hobbyists and could lead to new ideas and solutions that could not have been explored without them.

Academia should take note of this and increase its usage of these cheap board computers in curricula. The existence of a large community of makers should also be leveraged, by integrating its members in some large scale, IoT-like experimentations.

This also opens up very interesting paths for (self-)education in electronics and computing by a large number of (potentially young) people, which may to some extent help solve the lack of workforce educated in computer engineering and science that is currently a threat for Europe.

A remarkable initiative in this direction is clearly the BBC Micro Bit (aka Micro Bit, or micro:bit), 'an ARM-based embedded system designed by the BBC for use in computer education in the UK. [...] The Micro Bit was designed to encourage children to get actively involved in writing software for computers and building new things, rather than being consumers of media.' To do this, 'the device has been given away free to every year 7 pupil in the UK, and is also available for purchase by anyone.' [101]. By reaching as many young people as possible, this initiative aims to increase computer and electronics literacy, in a way very similar to the highly successful BBC Micro initiative of the 1980s [90]. The latter can indeed be seen as a strong contributor to Acorn's success, and even - ultimately and in a more distant way - that of ARM. Since ARM now is a world leader, one can only strongly recommend launching across all of Europe educational and computer literacy-oriented initiatives such as these old BBC Micro and recent BBC Micro Bit, but adapted to the present time. Once again, this could help to address the shortage of IT skills in Europe in the long term.

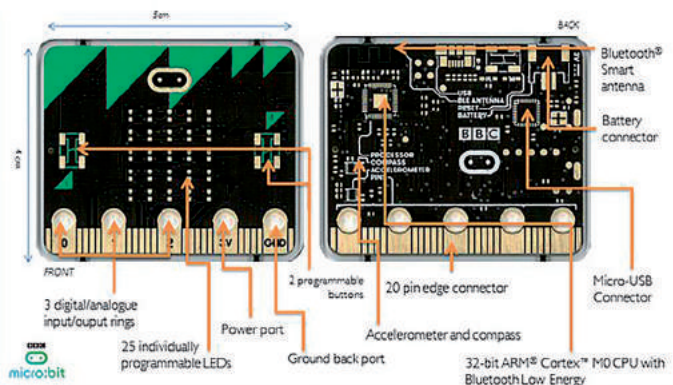


Figure 36: BBC micro:bit
Source: 2015 BBC micro:bit

2.4.6. OPEN SOURCE HARDWARE AND SOFTWARE

2.4.6.1. OPEN SOURCE SOFTWARE

Open source software continues to gain popularity in the business world [165], both in terms of use and contributions. As explained in previous Vision documents, in many cases open source is used to implement basic computing infrastructure, such as operating systems (Linux, Android), security frameworks (OpenSSL/LibreSSL, NSS, various firewalls and intrusion detection packages) and compiler toolchains (GCC, LLVM). Such software is fundamental to the operation of many businesses, but it is not what gives them their competitive edge. In other cases, the main motivator for participating is that a particular open source project enables a large community to use that company's products (e.g. ARM's contributions to various developer tools).

As the definition of 'basic computing infrastructure' evolves, so does the nature of the projects that are open sourced by companies. For example, Facebook [346], Google [174], Baidu [338] and Microsoft [24] have released various AI tools and frameworks as open source following the surge in cognitive computing for everything from advertising to self-driving cars. The commodity these companies deal in is data, and AI is just a tool to extract value from it. As a result, it makes sense for them to try to collaborate with the rest of the world on tools that improve their ability to efficiently process their data, while keeping the data itself (their competitive edge) private.

Even Microsoft, still notorious for its '*Linux is a cancer*' stance in 2001 [332], has completely come around to embracing open source's advantages [219]. For example, in 2016 alone they joined the Eclipse Foundation [68], released an AI environment based on the Minecraft game they acquired earlier [25], and made their PowerShell software open source for macOS and Linux [307]. They also grabbed headlines by surprising friend and foe with support for a fully functional Ubuntu Linux environment running inside Windows 10 [51].

Open source is also increasingly promoted by public administrations. While open source is already preferred for internal ICT development by the European Commission [35], the US went one step further and recently published a draft policy requiring every public agency to publish their custom-built software as Free Soft-

ware or public domain [132]. The reasoning is that since the public paid for its development, they should also get full access to the results. Similarly, Bulgaria recently passed a law requiring all software written for its government to be open source [122]. Currently, the State of New York is even considering a bill that would give individuals a maximum US\$200 tax credit for expenses associated with the development of open source and free software [168].

It is clear that the importance of open source and free software keeps growing, both in commercial and non-commercial settings. As more parts of our world become automated and more common-infrastructure software is developed to support this trend, we expect that more and more of this software will become open. The drivers will be various and will include reducing maintenance costs to growing communities, transparency (e-voting, public administration), fostering confidence (backdoors), and legal and even ethical requirements (algorithm of a self-driving car to decide what to do in no-win situations that may hurt either the driver or another person).

2.4.6.2. OPEN SOURCE HARDWARE

While open source software has become an integral part of the ICT world, the open hardware movement is still much smaller. Nevertheless, open hardware is growing and gaining popularity and some argue that it is at the start of an exponential growth curve much like its software counterpart in the early 1990s [296]. Raspberry Pi boards have become a household name in hobbyist electronics circles over the past several years, with both very low end [171] and higher end [172] models launching over the past year. Apart from main SoC, which includes the (ARM) CPU and graphics chips, it is completely open.

There is a growing trend towards open CPU architectures though. While not the first, a notable example is the OpenSPARC CPU architecture that was launched more than a decade ago [104]. It resulted in the LEON family of chips designed by ESA, which are used on the International Space Station [243]. More recently, a team at Princeton University built a 25-core OpenSPARC-based processor that has been designed to scale up to 8000 processors in a single system [312].

Another example is the RISC-V 2.1 architecture, which was finalized in February 2016 [131]. Similar to OpenSPARC, the RISC-V architecture is completely open and free to manufacture and use by anyone. This means that while IBM's OpenPOWER [92] and ARM architectures can only be licensed and modified by partners, RISC-V CPUs can in principle be made by anyone. Since not everyone has the expertise to design CPU cores, this results in new business models whereby any start-up can design cores that they subsequently specialize according to customers' needs [206].

On the graphics front, things are also starting to move, albeit slowly. Until several years ago, GPUs were even more closed than CPUs, with no documentation available whatsoever regarding the low-level technical details and programmability. Nowadays, engineers from Intel [169] and AMD [207] actively contribute to free software drivers and publish detailed documentation on

their chipsets. Even NVIDIA, which traditionally has been very open-source-unfriendly, occasionally helps the open source community with developing drivers for its products [119]. In early 2016, AMD also launched its GPUOpen [50] initiative for providing more documentation about graphics hardware. Actual open source GPUs are limited to research projects at this time, such as MIAOW [55].

In terms of systems design, we also see small opening steps. The Open Compute Project [303] provides an infrastructure standard with interoperable, efficient data centre components. Facebook released IKEA-style plans to build your own 360-degree surround VR camera [137]. And here in Europe, Fairphone [159] released its Fairphone 2 with a completely modular design that allows for interchangeable components [160].

2.4.6.3. IMPACT AND PROPOSED COURSE OF ACTION

The processor landscape is evolving, and the previous key players are not certain to continue their leadership. The PC era, dominated by Intel, and the smartphone era, dominated by ARM, will reach their peak and all players are looking for the next market. Will it be IoT, with a diversity of solutions? Or something else? All are targeting low power, efficient computing solutions. All possibilities are open to find the killer application or the future 'mainstream' system. We see that small start-ups, or 'markers' are exploring a lot of ideas in this fields, thanks to affordable boards and an easy-to-use development environment. The 'big' companies like ARM or Intel are supporting this 'design space exploration' by hobbyists by providing rather cheap computing solutions (board with processors, memory, sensors and IOs) and development environments.

We also observe that traditional semiconductor companies, that previously only delivered components, are now focusing on reaching a higher position in the value chain, by providing complete computing (sub-)systems with the associated Software Development Kit (SDK) and libraries. The ease of use of these systems, with their large application libraries, will be an important factor for success.

From the hardware point of view, energy efficiency will be the key success factor, both for high performance and for embedded systems. New requirements of IoT and CPS systems, including security, low latency and responsiveness to external events, might also be important differentiating factors.

It is also important for Europe to build up and secure knowledge on high performance systems (high end, FPGA, etc.), fuelling education and research in the complete field (hardware and software), avoiding brain drain and accumulating knowledge for the future lower-end systems.

2.4.7. CRYPTOGRAPHY AND SECURITY

2.4.7.1. CLASSICAL CRYPTOGRAPHY

Cryptography is an important element in our digital world. Banking and data exchanges rely on encryption mechanisms and a lot of services, such as credit cards and online banking, are only fea-

sible because of encryption. It is also a major ingredient for the protection of privacy. To avoid others 'spying' on what one is looking for online, the recommended protocol is now HTTPS (HyperText Transfer Protocol Secure), which is an encrypted version of the initial HTTP (HyperText Transfer Protocol). With HTTPS, the computer and the server agree on a 'key' and then scramble the messages using that key, so any interception of the communication (by a hacker for example, or a 'man in the middle') is not useful anymore without knowing the key.

To ensure the security of your private data on a smartphone or computer, the files can be encrypted on the devices, therefore inaccessible if a user does not know the password of the session. BitLocker from Microsoft encodes disk partitions. Apple introduced encryption in their mobile devices in 2014 with iOS 8 and the majority of Android devices using Marshmallow (Android 6.0) are encrypted by default. The slower adaption of encryption on Android devices is certainly due to the extra computing resources required for encrypting/decrypting the files (make the device slower and reducing its battery life).

The cost of encryption is not negligible, and coprocessors are developed to efficiently encrypt/decrypt with minimum latency and energy use (but it still has an impact on the cost of the device). Apple introduced a 'Secure Enclave' and a dedicated device to store the private information (like the fingerprint data used by TouchID). This required a processor conceived with encryption and security as basic design constraints (using ARM's TrustZone/SecurCore technology) and dedicated hardware to make a segregated and secure area within the processor. Apple claims that it cannot decrypt data on your device [388], which recently lead to the FBI-Apple encryption dispute [389]. Backups stored in the cloud do not have this feature.

Privacy and therefore encryption have become more important since the revelations of Edward Snowden and it is clear that the success of IoT devices (such as IP cameras) will be only possible if they have enough protection against data leaks and hackers [82, 118]. One important topic is to validate software updates, so that hackers cannot completely reprogram the device. This will be even more important when lives are at stake, in the cases of cars or pacemakers, for example. Security should not be an afterthought, but must be introduced from the very conception of the device and at all levels.

It is very important to develop technical solutions for privacy, including support by hardware (privacy hardware enforced), and to return control over privacy to the user. Of course, personal assistants calibrated with the user's preferences can be used to avoid the burden of having to control access to private data for every single transaction.

The massive distributed denial of service (DDoS) attack of October 2016 in which 150 000 IoT devices were involved [83] illustrates that such devices should be secure not only to protect the privacy of their owner, but also to prevent hackers from using them to set up attacks against third parties.

2.4.7.2. HOMOMORPHIC ENCRYPTION

Until now, encrypted data can be used in storage or during communication, but not when computing with it. Data has to be decrypted in order to do this. Homomorphic encryption allows general computations in the encrypted domain. In 2009, Gentry published a theoretical breakthrough on a fully homomorphic encryption (FHE) scheme [7]. Since then, several improvements have been made to make this technique practically possible, which is needed because homomorphic encryption schemes have a huge algorithmic complexity (see the following table for more details). Homomorphic encryption could have a large impact on preserving confidentiality. Sending unencrypted data to a server enables the server to use it for what it is supposed to, but it could also use it for other purposes (advertising, analysing behaviour, etc). If the communication and computation happens using a homomorphic encryption scheme, the server will have access to neither the actual data nor the results it is computing. Only the owner of the data can decrypt the results. This has multiple applications, for example, analysis of medical records and access with biometric data (never stored in clear on the server). As explained before, the main limitation of the homomorphic schemes is the cost of the computation in the homomorphic space, and advances in algorithmic, software and hardware accelerators will be required to make it mainstream.

2.4.7.3. POST-QUANTUM CRYPTOGRAPHY

There is a fear that a quantum computer running Shor’s algorithm [66] can be used to break cryptographic algorithms, mainly public-key algorithms based on integer factorization, discrete logarithms or elliptic-curve discrete algorithms problems. This has triggered research on post-quantum cryptographic algorithms that can resist quantum computer attacks. It should be noted that symmetric cryptographic algorithms (symmetric ciphers and hash functions) are considered to be relatively secure against attacks by quantum computers [27]. Worldwide research is currently exploring several directions in this area. However, it is expected that post-quantum cryptography will require a lot of computations, and dedicated accelerators will certainly be required.

[TABLE 1] CHARACTERIZATION OF A FEW ELEMENTARY ALGORITHMS.

	$b^2 - 4ac$ (8 b)	$b^2 - 4ac$ (16 b)
# ADD	332	1,188
# MUL	302	1,126
DEPTH	43	83
× DEPTH	16	32
AV. //	14.74	27.88
	$\sum_{i=1}^{10} t[i]$ (8 b)	$\sum_{i=1}^{10} t[i]$ (16 b)
# ADD	207	423
# MUL	135	279
DEPTH	24	48
× DEPTH	8	16
AV. //	6.75	14.62
	B. SORT (10 × 4 b)	B. SORT (10 × 8 b)
# ADD	1,620	3,240
# MUL	1,350	2,790
DEPTH	214	350
× DEPTH	68	136
AV. //	13.88	17.23
	FFT (256 × 32 b)	
# ADD	7,291,592	
# MUL	52,96,128	
DEPTH	674	
× DEPTH	166	
AV. //	18,676.10	

Figure 37: Characterization of a few elementary algorithms
Source: [435]

Figure 37 provides characterization data for a number of elementary algorithms obtained using instrumented clear domain bit-level executions. For each algorithm, the number of bit-level additions (# add), number of bit-level multiplications (# mul), depth, multiplicative depth (x depth) as well as the average number of operations per topological equivalence classes of the underlying Boolean circuit (a number which gives an idea of the amount of circuit-level parallelism and is labelled “av. //”) are given.

This shows the large increase of the number of operations required to process in the encrypted domain compared to the clear domain, hence the need for optimization and hardware accelerators.

Figure 38 shows the principle of two use-cases.

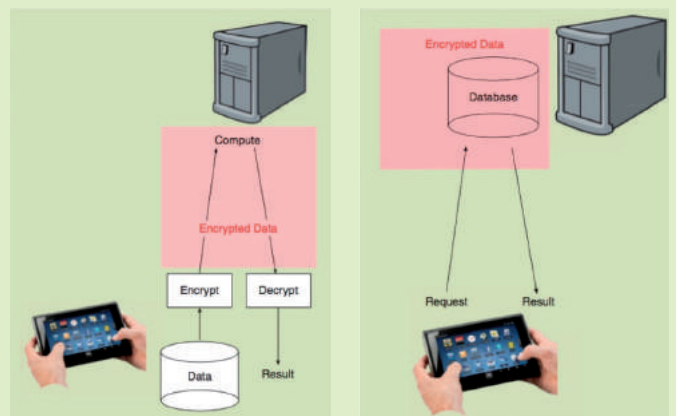


Figure 38: Two use-cases of homomorphic encryption
Source: [435]

2.4.7.4. BLOCKCHAIN AND NEW DIGITAL CURRENCIES

The blockchain concept was first popularized as the infrastructure of the decentralized currency called BitCoin [67]. It relies on a combination of public key cryptography, hashing and proof-of-work-based majority voting to establish a trustable, distributed, public ledger of transactions. This ledger is distributed over a network of peer-to-peer nodes, with each node holding a copy of the complete ledger (much like a Git repository [115]).

BASICS OF THE BLOCKCHAIN

The concept of the blockchain is based on ledgers. An example of a ledger is shown in Figure 39. Every entry in the ledger has a consecutive ID, the corresponding party, a date, the transaction amount, and the total balance. Originally, this made it hard to insert new entries in between when someone wanted to cook the books since they had to ensure consistency and modify all subsequent entries as well.

In the age of computers, such an operation is trivial. We could make this harder by adding an extra column that contains a checksum based on the current transaction and every transaction coming before it, so that inserting or modifying an entry will also require recalculating this checksum for that entry and every entry coming after it. If we would make the checksum very hard to calculate, then changing and modifying entries will become very expensive in terms of time and required computing power.

ID	Description	Corresponding party	Date	In	Out	Balance
1	Starting capital	Bank	20/01/2016	1000		1000
2	Sales	Joe	21/01/2016	100		1100
3	Purchase	Jane	23/01/2016		200	900

Figure 39: Example of a simple ledger

This technique forms one of the foundations of a blockchain: adding transactions to the blockchain is very compute-intensive. Similarly, once a transaction is part of the blockchain, it becomes very compute-intensive to change it, or to insert a new one before it. As illustrated in Figure 40, this is achieved by having the ledger consist of linked transaction blocks, whereby every block contains an extra nonce (a number) that has to be chosen so that the block's hash starts with a predefined number of zeroes. Finding such a nonce is only possible via brute force.

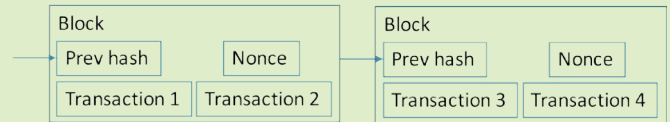


Figure 40: Blocks of transactions in the blockchain

Source: Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", 2008

The second foundation of a blockchain is transaction authentication based on public key cryptography, as illustrated in Figure 41. Since the blockchain contains transactions between any number of entities, every entry must refer to both the source and destination. The source is always the hash of another transaction (Transaction 1), so that you can only spend what you received. The destination is a public key, Owner 2's in this case. The source, Owner 1, signs its outgoing transactions (Transaction 2) with its private key, which must correspond to the recipient's public key in the source transaction (Transaction 1). A transaction can only appear once as source, which prevents double spending.

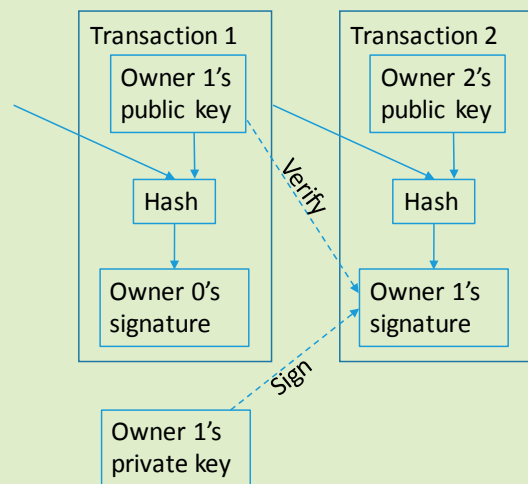


Figure 41: Blockchain transaction (based on [67])

Source: Satoshi Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System", 2008

The third foundation of a blockchain is that it is distributed: the ledger is not just stored on a single, central server, but all systems that want to take part in the maintenance of that ledger keep a full copy. Ledger maintenance consists of verifying the validity of broadcasted transactions and collecting them into blocks, finding valid nonces for such blocks of valid transactions, and adding blocks with valid nonces (found by you or others) to your copy of the ledger after verifying all transactions in them. Participating in the maintenance of the ledger is encouraged by awarding resources to the node that finds the nonce for a transaction

block. If multiple copies of the ledger circulate, the one used by the majority of nodes is considered to be the canonical one. This means that the ledger can only be corrupted if the majority of nodes, and hence computing power, is controlled by an adversary.

BLOCKCHAIN INTEGRITY

An obvious attack on the blockchain would be to remove past transactions from it, in order to be able to reuse —i.e., steal back—, e.g., a BitCoin with which the adversary paid someone before. The adversary will generally have to do this after the block containing the first transaction was verified, since before that point a savvy recipient would not confirm having received the payment. In order to roll back that transaction, the adversary would have to create a new longest chain in the ledger.

This means while all other nodes are working on adding blocks to the current (valid) top of the chain, the adversary's nodes would start working on finding nonces for blocks added to his alternative branch of the chain. In order for that branch to become the new main branch of the blockchain (and hence its transactions getting priority), it would have to become the longest one. This is only possible if the adversary controls the majority of computing power in the peer-to-peer network.

Tampering with transactions is also counteracted by a non-technical barrier in case of virtual currencies: someone with the computing ability to do this, would presumably be able to make more money by participating in the currency's blockchain than by corrupting it and undermining all trust in its operation. This may not hold for state actors or other agents that are not primarily motivated by economic gains though.

Similarly, upgrades or changes to the protocol of an established blockchain are impossible if the majority of computation nodes does not wish to adopt them. At the same time, this can also be seen as a protection against a hostile takeover. The BitCoin protocol is currently in such a situation, whereby an increase of the transaction blocksize is needed to scale up the transaction volume. While some paint this as a conflict between US BitCoin start-ups and the Chinese owners of the majority of the current BitCoin processing power [300], reality appears to be more complicated than that [123].

2.4.7.4.1. Beyond BitCoin

Blockchain technology can be used for many different purposes, since it essentially enables the recording of a sequence of transactions without the need for a central authority to vouch for their integrity. In the first place, many alternate 'Coins' based on different hashing functions came into existence: LiteCoin [120], PeerCoin [129], DogeCoin [31], CureCoin [74].

A second, more generic, application consists of self-executing contracts. In this case the transaction not only contains the re-

ipient's public key, but also a script that tells the nodes to verify that certain conditions are fulfilled before the transaction should be considered as carried out. Existing implementations mostly focus on trading derivatives, futures, swaps and options [331], but possibilities are only limited by the expressiveness of the supported scripting language [274]. Again, parties to the contract would rely on the peer-to-peer network to ensure that a contract is only executed when the encoded conditions are fulfilled, with a transaction fee going to the node that verifies and chains the transaction's block.

Another possible application is using the blockchain as a decentralized, encrypted data store [331], whereby the nodes are again remunerated by transaction fees. The applications vary from secure online voting [113] to internet name registries [23].

Going one step further, IBM sees such blockchain-based decentralized, encrypted storage and communication as indispensable for managing the IoT [390]. Until the early 2000s, most transaction processing was centralized, e.g. in the context of airline booking and phone companies. Today's cloud approach successfully deals with online shopping, social media and stock trading. When we start adding billions of IoT devices, from self-driving cars to toasters and returnable bottles, the centralized approach becomes very hard to scale. In the post-Snowden era, centralized processing in a black box that may be compromised without any way for its users to verify the processing of their data, is also a serious downside. These security, trust and scalability issues are largely addressed by the blockchain by construction.

2.4.7.4.2. The influence of BitCoin on computing

The node that verifies a block of BitCoin transactions and finds a valid nonce for it (see inset) is awarded a fee in BitCoins for performing this work. Obtaining BitCoins in this way is colloquially referred to as mining. BitCoin's blockchain uses a SHA-256 hashing algorithm, and initially its hash constraint requirements were simple enough for mining to be feasible using plain desktop CPUs. As BitCoin grew, the constraints became harder to fulfil and the first GPU-based implementations started to surface. FPGA implementations came next, but with BitCoin's exchange rate steadily climbing it soon became profitable to create ASICs for this purpose [294]. Today, there are several cloud-based platforms that offer hosted ASICs for rent, although this concept is sometimes criticized as it removes the decentralized nature of the blockchain infrastructure [84].

If the use of blockchain is democratized for other applications, as proposed by IBM, dedicated customizable accelerators will certainly be required, even in small IoT nodes, for example.

2.4.7.5. SIDE CHANNEL ATTACKS

Even when using cryptographically secure methods on a system without backdoors and without isolation issues between processes, confidential data may leak through side channels. Side channels are ways to observe properties of code or data by monitoring or inducing side-effects, such as cache misses or electromagnetic radiation.

As such, techniques have been devised recently to spy on data on computers or smartphones through electro-magnetic radiation across walls [148], through a parabolic microphone [54], via USB soundcards [149] or even by touching a metal part of a laptop [225]. It is not only computing devices themselves that are vulnerable: so are storage devices while they are internally transferring data [335], and even 3D-printers [156].

On the defensive side, the design of RFID chips that should be resistant to such attacks [57] is a welcome and necessary development for the security.

For general purpose computing, defences most commonly consist of masking the processed data by ensuring that the computation uses exactly the same amount of time and resources regardless of the input. This can be achieved through code transformations [30], by using specific cryptographic CPU instructions [279], or by keeping both code and data in what Apple terms a 'secure enclave' [134], which is hardened against tampering and external observation.

There is also considerable research into attacks that can be carried out over a network. Several attacks focus on observing and even modifying traffic without having man-in-the-middle abilities, often by finding ways to predict TCP packet sequence numbers [144, 184]. Another type of attack sends certain kinds of traffic to networks managed via software-defined networking (SDN), and discovers internal network topology and other information by observing side-effects of the flow changes that the SDN infrastructure applies as a result [334].

In summary, side channel attacks are an important security aspect that requires a completely different set of protections compared to defending against traditional exploits or preventing cryptographic weaknesses. Defending against them often has a high performance (compiler-based) or hardware resource (co-processor-based) cost, or requires redesigning network protocols or their implementations. It is therefore a distinct issue that needs to be taken into account from the start of the design in all fields of hardware, software and protocol design.

2.4.7.6. IMPACT AND PROPOSED COURSE OF ACTIONS

Security should no longer be an afterthought, but it should be a design requirement from the beginning, for hardware as well as for software. Cryptography in all its forms, from safe software upgrades to user identification, data protection during storage, communication and computation, must be a key ingredient of the future ICT. Hardware and software designers should have at least a basic understanding of security, and of how to defend against attacks. All connected devices should be able to receive security updates in order to protect them against attacks. But all these protection mechanisms have a cost, both in term of energy and computing power and time. Optimized software tool chains and accelerators (or specialized instruction sets) will be required to make it efficient and affordable.

2.4.8. COMMUNICATION/RELATION WITH THE ENVIRONMENT: VIRTUAL AND AUGMENTED REALITY

'Virtual reality' (VR) refers to the creation of a completely artificial visual world, while 'augmented reality' (AR) refers to the addition of some artificial graphical elements to the (picture of) reality. Currently, VR is mainly used in games, simulators, and movies; while augmented reality is found in some games and many industrial applications, including simulators.

Augmented reality and virtual reality will be very important in the coming years: all (Windows 10) PCs will be compatible with this technology from 2017 on [154]. Microsoft is adding its holographic shell to operating systems. It will be compatible with its 'Project Alloy' VR headset, which is a fully autonomous virtual reality headset that does not need a connection to a PC or smartphone [153].

Intel is developing smaller and smaller sensors for sensing the environment in 3D: after the realSense device they announced the Euclid sensor, which integrates sense, compute and connect capabilities in an all-in-one candy bar size form-factor [126]. Sensing the environment in 3D is key for numerous applications, ranging from gesture control, to robotics, games, and security, amongst others. The Kinect [75] was designed for use in games, but was widely used in robotics applications. Low-cost sensors are of paramount importance for many cyber-physical applications; and combining the input of multiple sensors to increase the reliability of environment sensing is mandatory. Low cost LiDARs will lead to development of reliable self-driving cars [305].

2.4.8.1. AUTOMATIC TRANSLATION

Automatic translation has existed for a long time [99]. Its coupling with optical character recognition and cameras on a smartphone as made it possible to provide everybody with automated translation of texts and signs [135], which, by use of augmented reality, provides an easy-to-use and widespread tool:



Figure 42: Automatic translation in action
Source: Google

2.4.8.2. HEAD-UP DISPLAYS IN CARS

The first Head-Up Displays (HUDs) were proposed to buyers of high-end cars in 1988-1989 (GM, Nissan), but very few were actually deployed until 2012. Since 2012, many more brands offer HUDs in a wider range of cars, including Audi, BMW, PSA Peugeot Citroën, Nissan, Mazda, Mercedes, Renault, and Volvo. Today they are available in high-end models [88].

HUDs for cars are also offered as add-ons, but seem to have had a mixed success as such.

Head-up Display



Figure 43: Head-up display from Audi
Source: Audi



Figure 44: Head-up display from BMW
Source: BMW

2.4.8.3. AUGMENTED REALITY GAMES

Various games have taken a shot at Augmented Reality (e.g. Lyte-shot AR FPS [98]), but Pokémon Go [105], released in July 2016, is the first to have achieved widespread use [56, 226].



Source: 2016 Niantic, Inc.

Source: 2016 Pokémon. Source: 1995–2016 Nintendo / Creatures Inc. / GAME FREAK inc.



Figure 45: Crowd of Pokémon Go players
Source: Business Insider UK, Tech News, 2016

It is worth noting however that this game imposes a significant power drain on the smartphone, because power is needed for GPS localization, for image processing and for screen display (especially in generally sunny conditions). As such, it became very common to see Pokémon Go players having their smartphones connected to an external battery while playing. Reports [281] signal that 'during the two weeks following the release of Pokémon Go, the sales of battery packs doubled in the United States. The report shows a year-over-year growth of 101 percent.'

2.4.8.4. SMART GLASSES

Optical Head-Mounted Displays (OHMD), the precursors of 'smart glasses', have existed for a long time (available in 1997, see [391]). At the time, none of these, however, became widespread, despite the involvement of several large players in the consumer market (Sony, Epson...).

The word smart glasses [107] refers to the fact that more computing power is added to them, merging OHMD and smartphones, and is more in fashion nowadays.

The most well-known are probably the infamous Google glasses [95] that were released in 2013. However, these faded from popularity, probably because of cost, battery, 'silly look' and privacy is-

sues [213]. Google nonetheless states that their development continues.



Figure 46: Google glasses
Source: Rijansoo7, Flickr, 2013

Other big players are also currently releasing smart glasses, under various names. Microsoft is providing their HoloLens smart glasses to developers, with various kinds of applications [102, 288]. Sony have their augmented reality SmartEyeglass [76], which includes gyroscope, accelerometer, ambient light sensor, built-in camera and monochrome screen. HTC proposes their Vive [185] VR headset that blends real-world elements into the virtual reality thanks to its front-facing camera. Oculus sells their much-anticipated Rift Virtual Reality headset [140], although this one does not include augmented reality, like Sony with its Playstation VR.



Figure 47: Microsoft HoloLens
Source: Microsoft



Figure 48: Sony SmartEyeglass
Source: Sony

Although these devices have not yet seen widespread adoption, a first significant foray has been made by Google smart glasses, and other large actors are releasing products. This market may thus become mature and spread widely in the next few years.

2.4.8.5. GAMES WILL BECOME INDISTINGUISHABLE FROM REALITY

Electronic games have become a major economic activity: the games industry already surpassed the movie industry [139], growing from US\$91.8 billion in 2015 (mobile and PC, console and handheld games combined) to a predicted US\$118.6 billion in 2019. The biggest category of apps on smartphones is games. Without going as far as Elon Musk in his arguments [337], it is true that the realism of games and the ‘intelligence’ of the computer-driven characters have increased massively. The current development of VR glasses for games further increase the feeling to be ‘in’ the game. This increase of fidelity is due to the increased ease of use of middleware, the performance increase of the software game engines, and also the increase of raw performance of computing devices: both at the processor and the GPU side.

These GPUs were initially developed for 3D games, and now reach the performance levels of the supercomputers of 15 years ago. For example, the consumer-grade Titan-X board from NVIDIA has the following specifications [69]:

- 11 TFLOPS FP32
- 44 TOPS INT8 (new deep learning inferencing instruction)
- 12B transistors
- 3,584 CUDA cores at 1.53GHz
- 12 GB of GDDR5X memory (480 GB/s)

GPUs are now not only ‘pixel pushers’, but can also be used as highly energy-efficient coprocessors for computing. As a result, we now have GPGPU (general-purpose GPU) programming using languages such as OpenCL. The efficiency of GPUs is due to their massive parallelism (originally SIMD-like), and their simplified computing elements compared to general purpose processors. They are also used as co-processors in high-performance computers, and they are often present in high-ranked computers of the Green500 list.

“40 years ago, we had pong, two rectangles and a dot,” Musk said. “That is what games were. Now 40 years later we have photorealistic 3D simulations with millions of people playing simultaneously and it’s getting better every year. And soon we’ll have virtual reality, augmented reality, if you assume any rate of improvement at all, the games will become indistinguishable from reality.” [337]

Elon Musk (SpaceX and Tesla CEO)

It is realistic to think that their performance will continue to increase due to architectural improvements, algorithmic support (e.g. ray-tracing), their parallelism, and technology (even if we are approaching the end of Moore’s law and already passed the one

of Dennard's), and, that in few years from now, 'the games will become indistinguishable from reality'. The difference between virtual reality and augmented reality will fade.

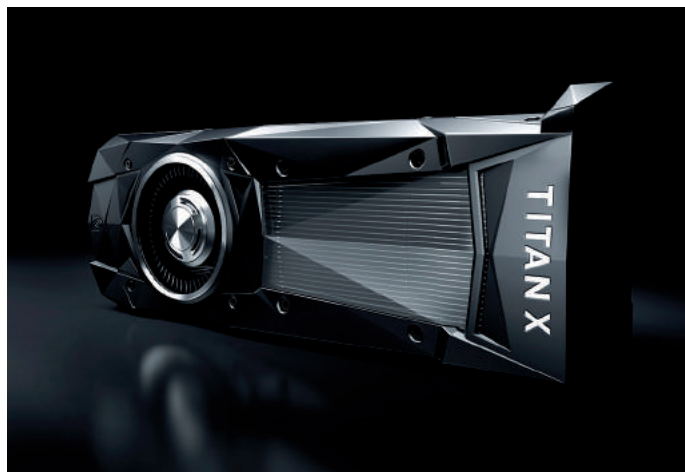


Figure 49: The New Titan-X from NVIDIA
Source: NVIDIA

2.4.8.6. IMPACT AND PROPOSED COURSE OF ACTIONS

Augmented Reality is maturing and to some extent becoming mainstream, mainly in use-cases on the smartphone.

Image processing capabilities are very important for AR, so requirements are likely to increase demand for such capabilities. High-performance embedded image processing systems (algorithms and hardware) will have to be further improved, first in terms of sheer processing power.

Second, since AR mandates quasi-immediate response times to changes in the environment to update the display provided to the user, Real-Time (RT) constraints are paramount in such systems. With the latter becoming ever-more complex, new solutions will have to be found to guarantee that these RT constraints are met.

In addition, since image processing is power hungry and many of the use cases of AR appear in autonomous devices (such as smartphones and smart glasses), low-power and low-energy issues will have to be significantly addressed to avoid the need for quasi-permanent connection to an external battery. Energy harvesting may also provide (elements of) solutions, as well as improved, i.e. larger and/or higher energy density batteries.

However, in the longer term, for a number of use cases of AR, devices closer to the body than smartphones may be found easier to use (smart glasses, smart earplugs, or even in-body systems) and be preferred. Such systems pose a wide range of research and technical challenges on their own, including low-power and low-energy, heat dissipation, miniaturization, bio-tolerance, and so on.

2.4.9. FROM IOT TO CPS TO 'DISAPPEARING COMPUTER'

2.4.9.1. IOT AND CPS

'The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it.'

Mark Weiser, Chief scientist at Xerox PARC. He is the father of ubiquitous computing, a term he coined in 1988.

The last decade we have experienced a transition away from computing systems that were easily identifiable as 'computer', generally with keyboards and screens as human interfaces. They were bulky and not so easy to transport, as even the 'portable computers' are heavy and you cannot have them with you all the time. The mobile phone became the new computing device of choice, with a touch screen as its main interface, and its size and weight enable keeping it nearly all the time with you. Both markets are saturated, with the PC market already declining and the smartphone one soon to follow down the same road.

Computers and, until recently, smartphones mainly operated in the cyber world: they had few interfaces between them and their users, little integration with the physical world, and speed/timing is completely defined by processing and network. Smartphones are now starting to have sensors that make them more aware of their environment: localization, magnetic field, acceleration, and even humidity and temperature. This evolution will continue and culminate in computing systems that offer context-sensitive services, but that are not recognisable as a 'computer'.

Currently, we already have smart sensors, integrated in the environment, that communicate (often wirelessly) with gateways (i.e. specialized computing devices), which in turn are connected to big, remote servers (the cloud). The enormous amount of data generated by these sensors, called big data, is analysed (data analytics) to extract information that will allow for offering new and better services. In other words, the resulting information is mainly used by computing systems, rather than directly applied to or influencing the physical world.

This combination of physical-world interaction and computing systems is the Internet of Things (see Figure 51). The *main concerns are keeping the systems protected from hackers (malevolent actions from outsiders), security, and privacy (data use for purposes not authorised by the data subject, and unauthorised data accesses)*.

The main characteristic of an IoT system is that it is composed of several physically separated, communicating devices whereby, most of the time, the communication does not need a human in the loop (machine-to-machine communication). It is a distributed system. Machine-to-machine communication constitutes only 5% of 2016 global mobile data traffic, totalling 508,022 terabytes per month [221]. However, it is expected to grow exponentially.

2016 What happens in an INTERNET MINUTE?

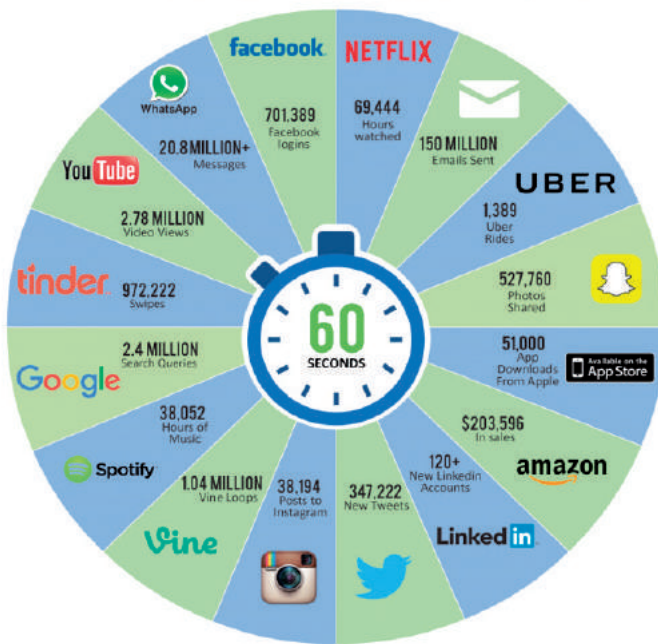


Figure 50: Samples of human-generated data in 60 seconds of Internet. This is only human generated data, machine to machine traffic is expected to rise exponentially in the coming years. Source: Exelacom Inc., 2016

Cyber-physical systems take the integration with the physical world one step further by directly interacting with the physical world based on the results of the data analytics, like steering a car (self-driving car), running a factory or simply switching on a light. CPS are *active*. In addition to the requirements of IoT devices, they also need *safety (the system should not harm its environment)*.

IOT SYSTEMS VERSUS CYBER-PHYSICAL SYSTEMS

There are many definitions of *Internet of Things* and *cyber-physical systems*, and a lot of controversy. We choose to define a *CPS* system as being characterized by having an actuator that directly impacts the physical world (a screen is not considered as an actuator in this definition), while an *IoT* system is distributed and composed of elements that communicate typically via the Internet. With our definition, CPS and IoT are not exclusive. For example, a self-driving car that is not connected and makes all its decision locally is a CPS device, but not an IoT device. It only becomes an IoT device (still being a CPS) if it is connected, e.g. to get maps from a server. A smart sensor transmitting the local temperature to a smartphone is an IoT device, but it is not part of a CPS. If it is connected to a thermostat that controls heating, the combination (i.e. the system composed of the sensor, the various servers and the thermostat) becomes a CPS (and the sensor is still a IoT device).

IoT systems, and a fortiori CPS, add to the computing systems the requirement that they have to cope with non-functional properties imposed by the physical world, such as time. In the 'old cyber world', timing is completely under control of the computer: the user has to wait until the system is ready, or has processed the data, in order to continue. We consider that a keyboard, mouse and screens are part of the cyber world, as they are purely limited by the timing and speed of the computer. In CPS (and some IoT systems), timing is imposed by the external environment: it the system is not fast enough or if it is busy, it will lose data and cannot ask the environment to resend the data. In a CPS, reactions are governed by the laws of physics: if the response time of the computer is not adequate, this could lead to, e.g. an accident with a self-driving car.

In IoT devices, communication is an inherent part of the system. This system is distributed and heterogeneous. Therefore, any optimization efforts should be global and take into account the cost of communication, storage and computation.

There is a compromise between storage, communication and computation that will be further developed in section 2.5.6.3.

The cost of communication is, in most cases, higher than the cost of computation, so it is usually beneficial to compute locally instead of transmitting raw data. This is what happens in edge computing, fog computing, local intelligence and streaming analytics, where data is processed as early as possible instead of in the cloud. It is an answer to the latency constraints for CPS and reduces energy consumption at the same time. Privacy also benefits due to data staying locally on the device. Extracting meaningful information from data often requires storing context locally, which increases the cost of the local system. It also requires more powerful processing resources. As a result, in spite of advantages in terms of privacy and global energy efficiency, local computing may face hurdles in its adoption. After all, the extra cost of the device due to these requirements are often more visible to the end-user than remote energy consumption or privacy issues.

Wearables and consumer IoT seem to be slow to emerge, because the users do not really see the benefits (except for fitness devices) and battery life is also problematic: people do not like to recharge their devices every couple of days, even if they are forced to do so with their smartphones. Some people also become aware that their personal health data may be used for unintended purposes beyond their control.

Interoperability is also mandatory for users' acceptance: unless you buy all your devices from the same company, generally you will have to download a special App to control your new device. Applications like IFTTT (see [393]) are successful in automating situations using different devices and services, but it does not cover everything. There are also a lot of interoperability platforms (see some of the European ones in Figure 53), but until they are exchangeable from the user's point of view, they will be a bottleneck for acceptance.

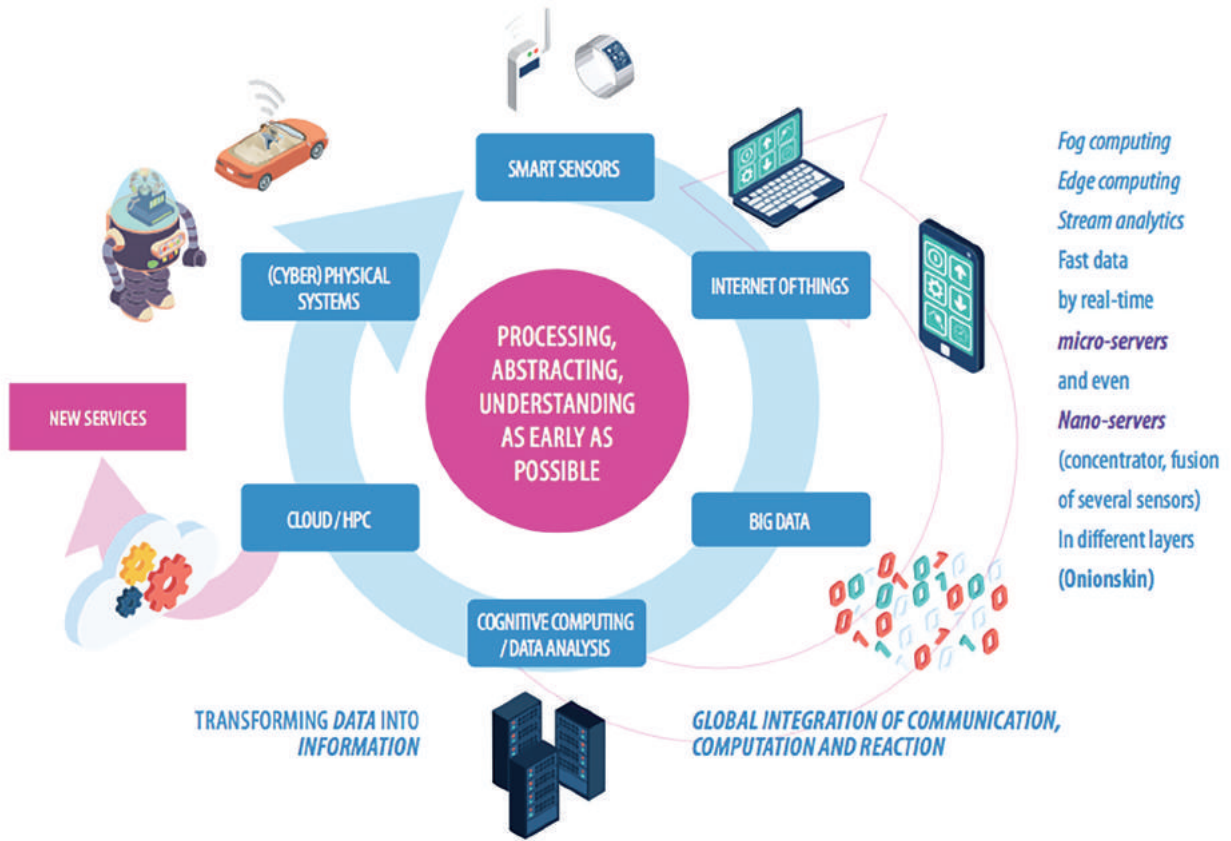


Figure 51: Interconnected systems
 Source: HiPEAC, 2015 and Artemis SRA, 2016

Figure 52 shows the complexity of the IoT ecosystem, which is still clustered by application domain. Users, however, generally want cross-domain applications, e.g. if an electric car should connect to the home grid in order to reduce peak hour consumption, that

should be indicated by the smart-grid ecosystem. The car should also be connected with the city to find recharging stations and parking places, to the weather services, and to the private and work agenda of the user, and so on.

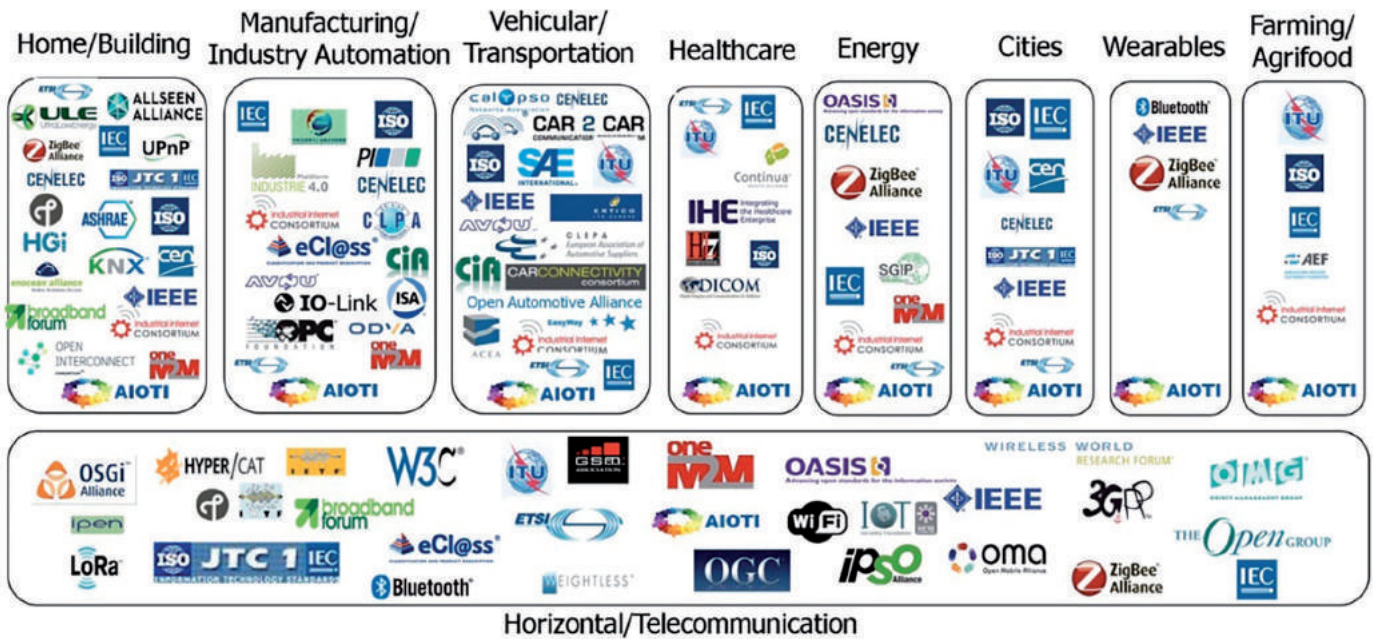


Figure 52: IoT SDO/Alliance landscape
 Source: AIOTI WG3

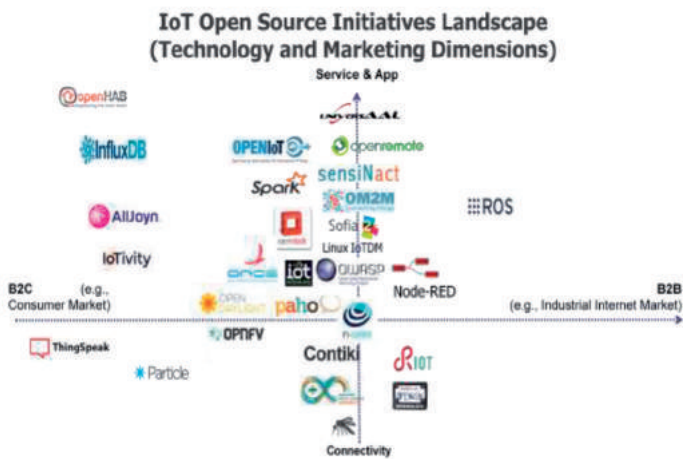


Figure 53: European Open IoT platform landscape
Source: AIOTI WG3

Industrial IoT, i.e. using interconnected smart sensors to capture data in an industrial process and analytics that provide global process improvements, is gaining more traction as it offers direct cost reductions and efficiency improvements ('the power of the 1%').

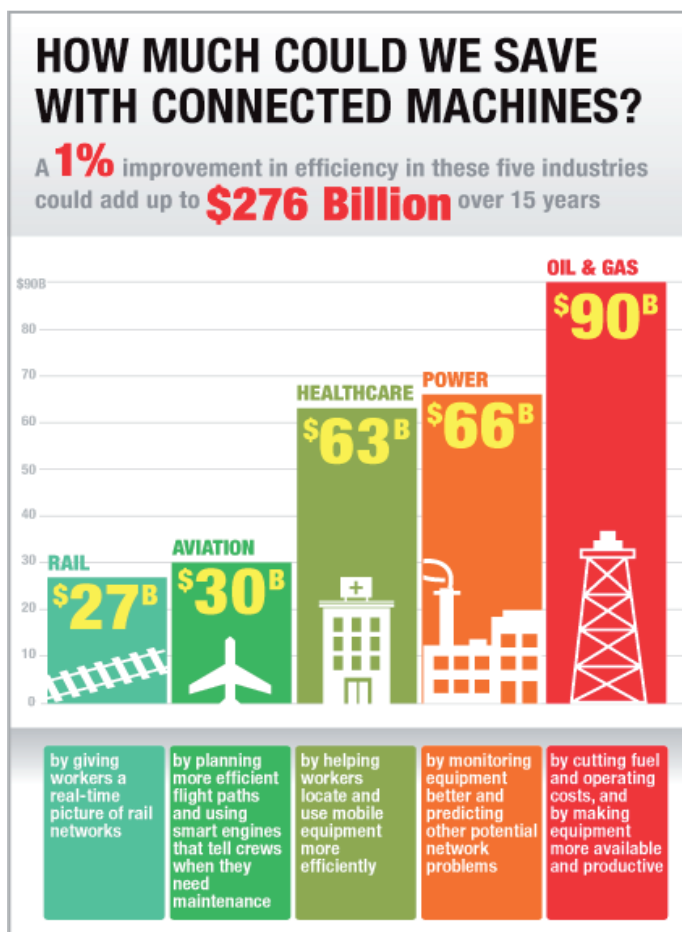


Figure 54: Potential savings offered by connection of devices and systems

Source: The Economist, GE Look ahead, 2013

We forecast that we are entering the era of the 'disappearing computer' in which computers are no longer represented by a screen and a keyboard: they are everywhere and they are invisible. They will need to be made transparent for users so that these devices can enhance quality of life. We will talk to the computers, they will see what is going on; this might be the next step after the era of the keyboard/screen and the touchscreen era. This is an old notion (1988) that is now becoming reality. Similar ideas were developed 15 years ago (*vanishing computer, pervasive computing, ubiquitous computing, invisible computer, everywhere, intelligent environments, ambient intelligence*, etc). Ambient Intelligence (Aml) is a concept introduced by the European Commission's IST Advisory Group ISTAG (ISTAG, 2001) (ISTAG, 2002) [203]. The disappearing computer was also an ICT Call for projects back in 2003: 'The Disappearing Computer (DC) is a EU-funded proactive initiative of the Future and Emerging Technologies (FET) activity of the Information Society Technologies (IST) research program' [231].

The most exciting evolution of computers is in how they interface between the real world and people, as well as the increasing ability of machines to 'understand' the environment and the context – not just numbers. Voice recognition has been integrated into all smartphones and is appearing on home devices (see section 2.4.11.6.1), while smart cameras have started to interpret what is going on. New algorithms are making machines smarter, taking applications such as self-driving cars out of the realms of science fiction and making them a reality. 'Deep learning' and cognitive computing will change how we use computers and will open up a whole new range of applications. We will interact with machines in a natural manner, and they will seem to us more and more like 'beings' than 'things'.

2.4.9.2. HIGH PERFORMANCE COMPUTING: AN ENabler FOR NEW CPS APPLICATIONS

High Performance Computing is a key technology for simulation of a vast range of things including airplanes and cars, crashes and collisions, protein folding, fluid dynamics, computational chemistry and physics, and cosmology. It is more and more often used to replace experiments because it is cheaper or because the experiments are no longer allowed (testing cosmetics on animals, nuclear explosions, etc.).

The European Technology Platform for High Performance Computing (ETP4HPC) is publishing a Strategic Research Agenda that lists applications and future challenges of the domain [245]. The main goal of the community is to reach the Exaflop (10^{18} floating point operations per second) as soon as possible in order to meet the application requirements. One of the main challenge is to reduce the energy consumption to about 20 MW (Figure 55 gives an indication of the required energy per floating point operation). Reducing energy dissipation will also decrease the cost of the cooling systems and the electricity bill.

÷ 4
every
2 years



Figure 55: Evolution of energy per floating point operation
Source: www.top500.org

HPC systems are being used for new applications that require more near real-time capabilities. Weather simulation has always included a notion of time (the forecast must be ready on time), but new High Performance Data Analysis (HPDA) applications are emerging that require the HPC machine to be in a (loose) loop, e.g. for the optimization of a process. We can even imagine going further, with a HPC machine simulating a phenomenon in advance so that it can be regulated in real-time. A real-time situation of the physics of an engine or a reactor could allow for tuning its parameter in advance in order to gain efficiency.

2.4.9.3. SELF-DRIVING VEHICLES AND DRONES

Drones and self-driving cars are natural test-beds for cyber-physical computing and sensing devices.

2.4.9.3.1. Self-driving vehicles

Although autonomous cars, also named self-driving cars, driverless cars or robotic cars, were prototyped as early as the 1980s [89], they have only recently become mature enough to begin spreading into our daily lives. The most famous and common ones are probably the Google Car [163] and the Tesla Model S in autopilot mode (*computer on wheels*) [175]. However, most car makers, including the largest ones, have working prototypes of self-driving cars these days (Mercedes-Benz, General Motors, Continental Automotive Systems, IAV, Autoliv Inc., Bosch, Nissan, Renault, Toyota, Audi, Hyundai Motor Company, Volvo, Tesla Motors, Peugeot, Local Motors, AKKA Technologies and so on) and seem to be close to beginning to sell them.

As a consequence, self-driving cars are expected to appear in the streets very soon, bringing about a revolution. Uber announced in 2016 it will deploy self-driving Volvo cars in Pittsburgh [212]. BI Intelligence is forecasting the first widespread sales of fully autonomous cars in 2019, with 10 million self-driving cars on the roads in 2020 [196]. Ford promises mass-production of self-driving cars in 2021 [306].

The first sector in the economy that might embrace autonomous vehicles is long haul transportation. The highway environment is much easier to model (no pedestrians, no bikes), private toll highways might be willing to equip their infrastructure for autonomous trucks in return for a higher toll, and the return on investment for transportation companies is huge (trucks that can run 24h/day, fewer drivers, fewer accidents, lower fuel consumption). Tomorrow's driver is likely to be a 'logistics manager' who over-

SAE international has published a classification system for automated vehicles. It distinguishes six levels [89].

Level 0: Automated system has no vehicle control, but may issue warnings.

Level 1: Driver must be ready to take control at any time. Automated system may include features such as Adaptive Cruise Control (ACC), Parking Assistance with automated steering, and Lane Keeping Assistance (LKA) Type II in any combination.

Level 2: The driver is obliged to detect objects and events and respond if the automated system fails to respond properly. The automated system executes accelerating, braking, and steering. The automated system can deactivate immediately upon takeover by the driver.

Level 3: Within known, limited environments (such as freeways), the driver can safely turn their attention away from driving tasks.

Level 4: The automated system can control the vehicle in all but a few environments such as severe weather. The driver must enable the automated system only when it is safe to do so. When enabled, driver attention is not required.

Level 5: Other than setting the destination and starting the system, no human intervention is required. The automatic system can drive to any location where it is legal to drive.

sees the automated systems, talks with dispatchers, and drives the truck in places like cities. A first and intermediate step could be the introduction of platooning, and automatic driving between two highway parking lots.

At a press conference in Nevada on 06 May 2015, the US's largest heavy-duty truck manufacturer, Freightliner, an affiliate of Daimler, unveiled a prototype 18-wheeler called the Inspiration Truck [259]. This was the world's first self-driving truck licensed for road tests, and the first to operate on an open public highway in the US, after prototype tests in Europe.




Figure 56: Daimler Freightliner Inspiration Truck
Source: Daimler Trucks North America LLC.

The Freightliner Inspiration Truck is a Level 3 autonomous vehicle, as defined by the National Highway Traffic Safety Administration, which means a driver must remain at the wheel at all times while the truck is in motion and be able to take over driving 'with a sufficiently comfortable transition time'.

What it can do

- ▶ Read lane markings
- ▶ Detect vehicles in front
- ▶ Steer through curves and turns
- ▶ Let drivers text, or talk on the phone, or watch YouTube
- ▶ Coordinate "platoons" of trucks for better fuel economy
- ▶ Start a countdown when the driver needs to take the wheel

Detachable tablet in the dash can be used by truckers while self-driving is engaged



What it can't do

- ▶ Overtake slower cars
- ▶ Change lanes
- ▶ Exit highways
- ▶ Park
- ▶ Work on roads with insufficient markings
- ▶ Work in cities

Figure 57: Daimler Freightliner Inspiration Truck capabilities
Source: Daimler Trucks North America LLC

On 20 October 2016, Otto, a San Francisco-based self-driving truck company acquired by Uber in August, made the first commercial delivery (50,000 cans of beer) with a self-driving truck [188, 197]. The truck left Fort Collins, Colorado, at 01:00am and drove itself 120 miles on I-25 to Colorado Springs at an average speed of 55 mph. It is to be noted that the driver took care of the trickier parts of the journey and of city driving at both ends of the journey. Instead of building its own trucks, Otto designs hardware kits to retrofit existing trucks. The Volvo truck that completed this delivery was equipped with US\$30,000 worth of hardware and software: two cameras for lane detection, a LIDAR sensor to create a 3D environment, two front-facing radar sensors to detect obstacles and other vehicles, and a GPS sensor to pinpoint the truck's location. However, this commercial delivery is still a prototypal one. A Colorado state patrol vehicle followed the truck from a distance to monitor the journey. To prepare for it, a few weeks before, Otto had sent some of its trucks and team to Colorado to map the roads, gradually adding dummy trailers, first empty ones then loaded ones, studying the traffic patterns and landscape before deciding that 01:00am was the best time to run the shipment. Otto are currently still focusing on the basics, like smoothing out acceleration and braking, and improving lane control. The next set of goals include predicting other drivers' behaviour, dealing with hazards like construction zones or sudden bad weather, while the longer-term goal would be driving in cities.



Figure 58: Otto's self-driving truck
Source: Otto and Budweiser

All these self-driving cars and trucks will have tremendous communication requirements. Vehicle-to-vehicle communication will increase exponentially with the number of such vehicle, as this is required to better cooperate and use the road.

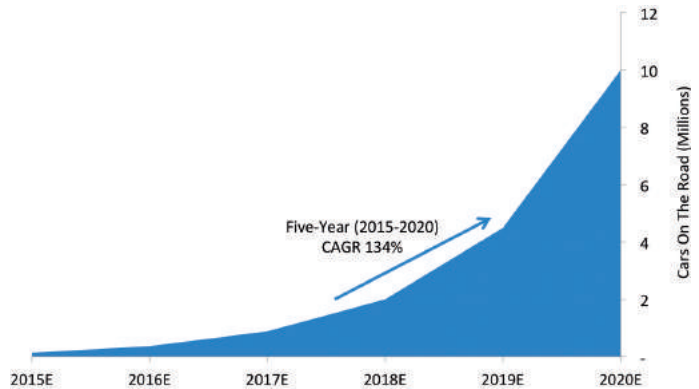


Figure 59: Estimated global installed base of cars with self-driving features
Source: BI Intelligence Estimates, 2015

Vehicle-to-road infrastructure communication will likely appear for the same reason, and also as a way to communicate information to cars in more reliable ways than through optical scanning. These two kinds of communication will result in calls for standardization for the sake of interoperability. In addition, self-driving vehicles will continue to communicate, as they already do, with their base station at their manufacturer's premises, both for maintenance and forensic/legal reasons. These communications will pose new challenges in terms of determining what to communicate to whom at which point in time, but also to ensure security. Indeed, these vehicles will have to be protected from 'rogue vehicles' or people injecting wrong data into the system, for unfair advantages on the road or for purposes of hijacking or breaching privacy. These self-driving cars and trucks and their communication will generate huge amounts of data, bringing new challenges to the big-data community [53]. In order to make autonomous vehicles a reality, there are still a lot of challenges to be tackled (technical, communications and infrastructural, as well as ethical and liability issues). There is, however, progress in all these domains. Recently, Volvo announced that it

will take liability for its self-driving cars when they are running in autonomous mode [253].

Computer vision and data fusion are key elements for autonomous systems and deep learning techniques are today among the best performers in the field. Several companies are developing embedded vision solutions, e.g. NVIDIA (see above), Synopsys with its CNN IP and MobileEye, which was the provider of the vision solution for Tesla.

CHARACTERISTICS OF THE MOBILEYE EYEQ4:

- More than 2.5 Teraflops
- Power budget of approximately 3W
- Vector processors give the EyeQ4 the ability to use computer vision algorithms such as deep-layered neural networks and visual modelling
- Can process information from eight cameras simultaneously at 36 frames per second.

2.4.9.3.2. Drones

Once the realm of hobbyists, unmanned aircraft, popularly called drones, are progressively pervading everyday life. Although the US armed forces already used drones in the 1960s, the increasing technological performance made possible a wider range of uses of drones. The Federal Aviation Administration expects some 600,000 drones to be used commercially within a year, up by a factor of 30 from the 20,000 registered for commercial use in August 2016 [299]. It is mainly the advent of materials research that made the construction of light yet powerful electro motors possible. The development of lithium polymer batteries resulted in high-energy-density power sources. Sensor technology development resulted in compact, integrated accelerometers, gyroscopes and GPS receivers. Most of these advances were initially mainly targeted at the consumer market for smartphones, resulting in their prices gradually going down to readily affordable levels around 2010.

Drones are used for many applications: for views from inaccessible or dangerous locations such as for inspecting power lines, pipe lines and oil rigs. Companies such as Amazon [141] and even Domino Pizza [33] have experimented with deliveries of orders by drone. Consumer-like applications for drones are starting to appear as well: see Kimon, the drone for taking selfies [32].



Figure 60: Amazon air delivery drone

Source: Amazon

As always, with the advent of new, unthought-of possibilities come new problems. With drones, these problems are situated in the area of privacy, security and safety. A drone makes it possible to view your neighbour's garden from almost any viewpoint at any time of day. New legislation has to be drawn up to regulate where and when a drone is allowed to fly. And drones can also reach places where their presence poses a threat to security. Drones can fly autonomously, and that creates the same type of legal problems as self-driving cars, but with that added third dimension of motion: a self-driving car that detects a power problem can try to park itself at the side of the road. A drone, high up in the air, may not be able to land safely once a power problem is detected.

Authorities are slow to come up with rules for drones, and that certainly weighs down on their economic impact. Amazon, after its successful experiments, has not been given permission to use drones for order delivery. Only recently, the US Federal Aviation Administration (FAA) has enacted new rules for drones, relaxing at least the requirements for drone pilots (until August 2016, commercial drone pilots were required to have a traditional pilot's license). Other FAA rules (flight within line of sight requirement, flying only allowed during daylight hours) are expected to be relaxed in the near future. The situation in Europe is slightly more complicated: although basic safety rules apply, they differ from nation to nation, and rules are not written in a coherent way. This situation may be remedied in the near future as the European Commission has drawn up a 'Prototype Commission Regulation on Unmanned Aircraft Operations' [150].

Many technologies, not just information technology, converge and interact in drones, making them an attractive testbed for cyber-physical computing and sensing devices. This is clearly shown by the interest demonstrated by Intel, who introduced a flying drone to promote its RealSense camera technology [155].

2.4.9.4. IMPACT AND PROPOSED COURSE OF ACTIONS

CPS and IoT systems are the new market hope for ICT. As nothing is fixed yet, there are a lot of opportunities for Europe to stake its presence in these fields.

Besides the classical ICT requirement of low energy and low cost, new challenging requirements in terms of *keeping the systems protected from hackers* (malevolent actions from outsiders) or *security, privacy* (use of data unauthorized from the data owner) and *safety* (especially for CPS systems) are mandatory.

IoT systems will require to be *interoperable*, and to simplify the current jungle of software stacks and APIs, while guaranteeing a correctness at the system level. This will require new research in the domain of composing heterogeneous 'grey' systems (where only interfaces are known, not the inner operations) or even black systems.

CPS systems will need to cope with the 'world' requirement, in term of *response time, predictability and safety*. In contrast with the past, where time was abstracted away as far as possible, a paradigm shift will be needed to make it a first-class citizen, both in software than in hardware.

2.4.10. HEALTH AND AUGMENTED HUMAN: TOWARDS 'CYBORGS'?

2.4.10.1. ON-BODY CYBER-PHYSICAL SYSTEMS

With the IoT, computing devices tend to be everywhere and in ever increasing numbers. The next step that can logically be expected would be to have them on-body, then in-body.

On-body wearable computer systems have existed for quite some time and are becoming more widespread. Smartwatches are the most visible recent iteration of such systems. They first appeared in the 1990s, but remained uncommon, and were reintroduced as wearable smartphones or smartphone extensions in 2013 [108].



Figure 61: Apple watch
Source: Justin14/Wikipedia

A number of other ad-hoc connected objects sporting body sensors are also publicly available. Finally, thanks to MIT and Microsoft, connected tattoos called DuoSkin are even starting to appear as prototypes [394].

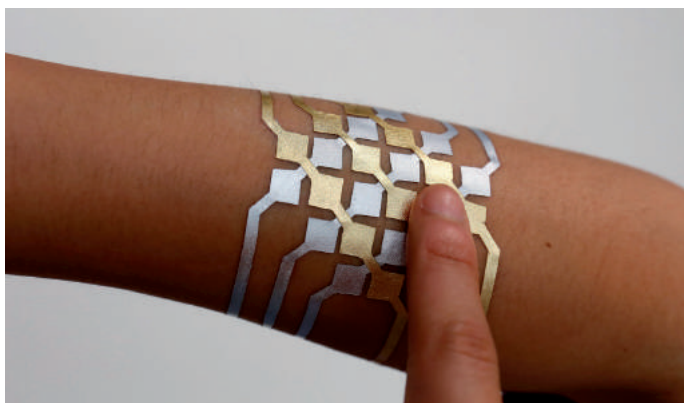


Figure 62: DuoSkin connected tattoo
Source: 2015 MIT Media Lab

These and other wearable systems are sometimes called 'On-Body Cyber Physical Systems' (OBCPS).

2.4.10.2. IN-BODY CYBER-PHYSICAL SYSTEMS

Another trend that is beginning to pick up consists of 'In-Body Cyber Physical Systems' (IBCPS), where the CPS are not above the skin anymore, but can be implanted —possibly deeply— below the skin.

The possible applications of such OBCPS or IBCPS are many. They include helping to cure or live with disease, either by monitoring body parameters or by taking actions (e.g. injecting medicines, sending electrical stimuli). Today's 'neural dust' prototypes by UC Berkeley have already been reduced to 3x1x1 mm in size, but are expected to further shrink to the width of a human hair [125].

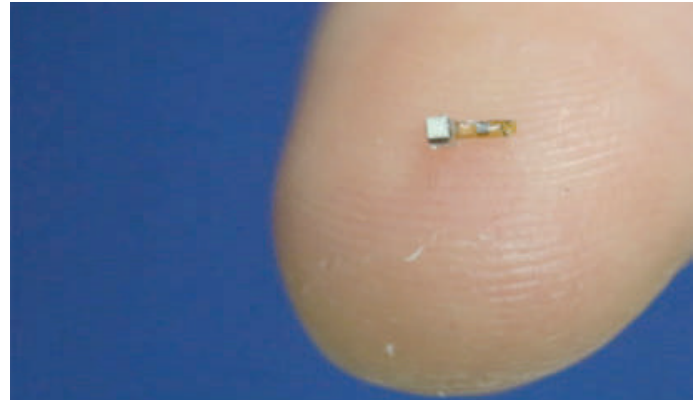


Figure 63: 'Neural dust' prototypes at UC Berkeley
Source: UC Berkeley, Roxanne Makasdjian and Stephen McNally

Similarly, in its ElectRx program, DARPA 'seeks to create ultraminiaturized devices, approximately the same size as individual nerve fibers, which would require only minimally invasive insertion procedures such as injectable delivery through a needle.' Such devices would 'continually assess conditions and provide stimulus patterns tailored to help maintain healthy organ function, helping patients get healthy and stay healthy using their body's own systems.' [227, 368]. This kind of IBCPS could represent quite a disruption in healthcare, allowing for much more fine-grained monitoring and prescribing than ever before, in an automated, hence potentially affordable, way.

In 2016, DARPA also launched a call for bi-directional brain-computer interfaces, able to read more than one million neurons, and stimulating more than 100 000 [329].

OBCPS and IBCPS could also be used to enhance or improve human capabilities for 'fun' or performance reasons, with no compelling medical need. The result would be 'augmented humans' or 'humans++' à la Cyberpunk [114].

Finally, OBCPS and IBCPS could be used 'simply' because they are the most convenient ways to carry CPS and to interact/interface with them. Indeed, today the biggest market opportunities are in the consumer segments, i.e. technology for people's entertainment. Technology thus has to be more human-inclusive and easy-to-use, just like paper does not require any intermediaries in order to use it. One way to better put the human in the loop and to get rid of large cumbersome devices (mainly because of the screens) might be to rely on OBCPS or even IBCPS to suppress intermediaries. This would be attractive to at least some people (e.g. the 'most connected man in the world' [215]).

2.4.10.3. IMPACT AND PROPOSED COURSE OF ACTIONS

OBCPS and IBCPS pose a number of challenges. First, there is a challenge of size. However, miniaturization is progressing significantly. On the electronics and computing side, circuits are still shrinking. Intel announced plans to deploy its 14 nm process in 2016 [278], its 10nm process in 2017 [200] and its 7 nm process in in 2020 [201].

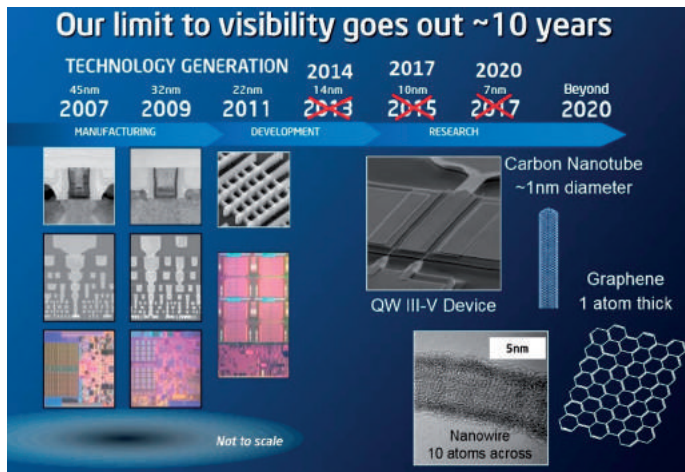


Figure 63: Intel's miniaturization plans
Source: Intel

TSMC, the world-leading Taiwanese founder, forecast that 2017 will bring its 10 nm process, 2018 its 7nm process and 2020 its 5nm process [248].

A second challenge is that, as for augmented reality, for OBCPS and even more for IBCPS, low-power and low-energy consumption will be paramount. Energy harvesting will likely be required, since changing batteries inside the human body is rather inconvenient. On a similar note, heat dissipation will also have to be addressed. On the hardware side, FPGA accelerators or custom ASICs should be more efficient than general-purpose processors. The software side will also have to be highly optimized.

A third challenge will be that of safety and reliability. Since serviceability and reparability are going to be very limited in IBCPS, it will be crucial to produce devices that are correct (bug-free) from the beginning. Techniques such as verification, proving and testing will have to be further developed to this end.

A fourth challenge will be the security and privacy issues. With devices ever closer to the body, or even inside it, and always accompanying their owner, security and privacy must be guaranteed, since breaches would have significant, possibly deadly, personal impact. There again, research on techniques such as verification, proving and testing can help provide the answer.

Finally, a number of issues unrelated to computing arise, such as the connection to the human body (possibly the brain) and biocompatibility (medical challenges), as well as the ethical issues (societal and political challenges).

All of these challenges will have to be tackled.

2.4.11. ARTIFICIAL INTELLIGENCE AND COGNITIVE, SMART DEVICES

The next big challenge in ICT will be to create intelligent, 'cognitive' systems. Artificial Intelligence was several times at the top of the hype curve (see Figure 22), but never really succeeded in practice following its introduction at a conference at Dartmouth College in 1956. In the 1980s, expert systems and LISP and Prolog machines appeared, and Japan launched its fifth generation computer program, but again expectations were not met. Since then, the term 'Artificial Intelligence' has had negative connotations, due to its inability to meet expectations.

In the 1990s, significant improvements made to Artificial Neural Networks, inspired by the work of several people including Warren McCulloch and Walter Pitts (1943), Frank Rosenblatt (1958), David E. Rumelhart and James McClelland (1986), led to the development of hardware accelerators and real-world applications such as handwritten character recognition for zip code machines. Artificial Neural Networks regained interest in 2012 when the 'supervised' Deep (i.e. with 9 layers) Neural Network from Hinton et al. reduced the error rate of image classification on the ImageNet data set to about 15%, while the best classical approaches had a rate of 26% (today, the best Deep Neural Network's error rate is below 5%).

The AlphaGo program developed by Google beat Lee Sedol (a 9-dan professional in the game of Go) in March 2016, generating lots of publicity for deep learning and AI techniques. Previously, IBM's Watson system, which is able to answer questions asked via natural speech won the first prize in the Jeopardy! game in 2011. Since then, it has been further developed by IBM to serve as consultant in different domains, including medicine.

Machine learning, cognitive expert advisors and the like are at the peak of the inflated expectations on the Gartner Hype Cycle, but as all major companies (Google, Apple, Facebook, Baidu, Microsoft, are so on) are investing heavily in it, it might be the 'right' time this time; even the term AI has returned. There is also a boom in new start-ups related to the domain, with associated equity deals increasing nearly six-fold, from roughly US\$70 million in 2011 to nearly US\$400 million in 2015 [146].

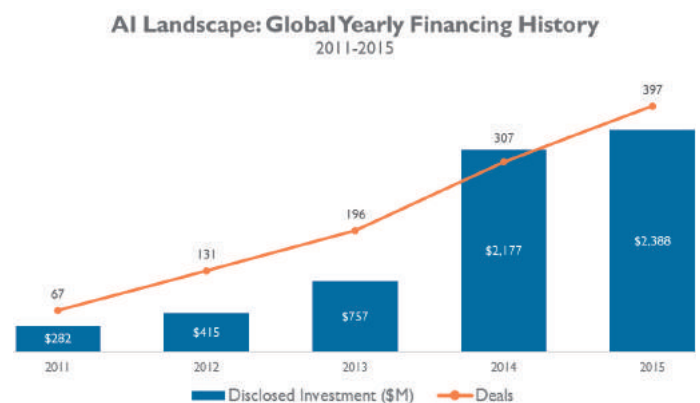


Figure 64: Start-up in AI financing history
Source: [146]

Investments related to artificial intelligence are forecasted to reach \$11.1 billion by 2025.

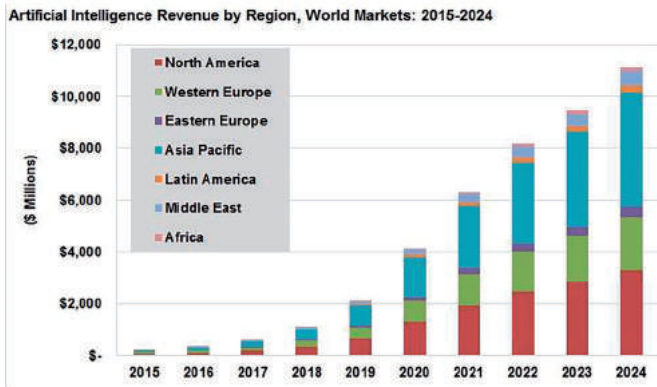


Figure 65: Forecast of revenue of AI
Source: [436]

Many start-ups working in the AI domain have recently been acquired by larger companies. For example, in 2014 Google bought the UK-based DeepMind (the company that created AlphaGo), while in 2016 Intel bought the Ireland and US-based Movidius (which makes the deep learning accelerator Fathom and vision chips for drones) [339] and Nervana for its Nervana engine [510]. In all, more than 30 companies working on AI have been acquired over the last five years by Google, IBM, Yahoo, Intel, Apple and Salesforce [147].

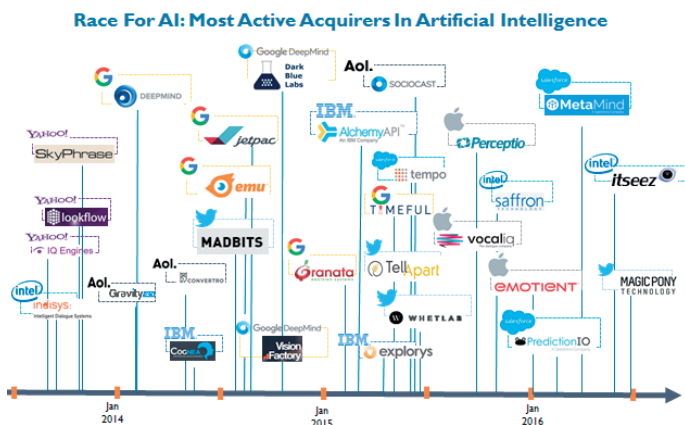


Figure 66: Acquisition of AI-related companies
Source: [147]

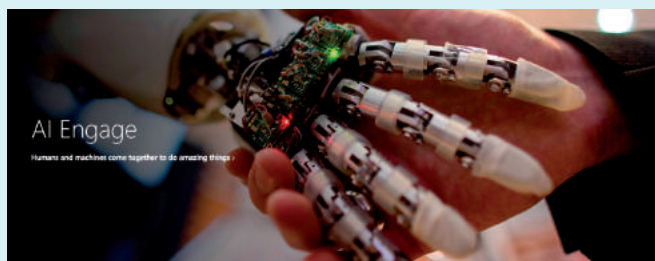


Figure 67: AI at Microsoft
Source: Microsoft, 2016

Microsoft is introducing “Open Mind Studio”, described as Visual Studio (their development environment) for machine learning [247].

DEFINITION OF ARTIFICIAL INTELLIGENCE

An ideal ‘intelligent’ machine is a flexible rational agent that perceives its environment and takes actions that maximize its chance of success at some goal [86].

Weak AI (also known as **narrow AI**) is artificial intelligence that is focused on one narrow task. Weak AI is defined in contrast to either **strong AI** (a machine with consciousness, sentience and mind) or **artificial general intelligence – AGI** – (a machine with the ability to apply intelligence to any problem, rather than just one specific problem). All currently existing systems considered artificial intelligence of any sort are weak AI at most [111].

Neil Jacobstein analyses that there are 7 factors that explains **the rebirth of AI** [395]:

1. **Capital:** Capital is ‘rushing’ into the AI space to the tune of \$2.4 billion in 2015. In the first half of 2016 alone, there were 200+ AI start-ups that raised over \$1.5 billion. It’s an understatement to even say deal activity is fast and furious;
2. **Algorithms:** Jacobstein pointed to algorithms like deep learning and its hierarchical pattern recognition as a major force driving the adoption of AI. With software like RStudio and Sentient, companies who would never have thought about getting into AI suddenly can;
3. **Hardware:** Whether it’s Alphabet’s recently announced tensor processing unit (TPU), Qualcomm’s new neural processing unit (NPU), NVIDIA’s deep learning chip, or IBM’s TrueNorth neuromorphic computing platform, more chips are being developed to enable faster and more powerful AI;
4. **Data:** As important as hardware is to AI, large data sets are where machine learning algorithms really learn by refining hypotheses iteratively. From real-time information discovery to the integration of algorithms and data with TensorFlow, more tools for working with data are enabling analysis of an increasing number of publicly available datasets;
5. **Talent:** For all the focus on software and hardware, Jacobstein’s thesis is that humans are as critical to the AI equation as machines. He points to the number of AI start-ups (Turi, Nervana, and DeepMind, to name a few) in recent years in which talent was the primary driver for acquisition;
6. **Applications:** Today, we experience AI through the applications we use. It delivers value by augmenting human skills and extends our capabilities. But this isn’t an overnight thing. When first released, Siri had plenty of issues, but over time, it’s become increasingly useful and more companies have released virtual assistants, such as Microsoft’s Cortana and Amazon Echo;
7. **Responsibility:** The final driver of AI adoption that Jacobstein highlighted was responsibility, especially noting that AI comes with tradeoffs that often boil down to trust. While applauding efforts like OpenAI to democratize access to AI, he emphasized the need for AI to demonstrate core human values. In the business world, this means that ultimately the winning formula is humans-plus-AI processes.

More than 8,000 international personalities, including Stephen Hawking, Elon Musk (the CEO of Tesla), Steve Wozniak (co-founder of Apple), Jen-Hsun Huang (CEO of NVIDIA), have signed a letter [48] warning about the potential pitfalls of AI.

AN OPEN LETTER RESEARCH PRIORITIES FOR ROBUST AND BENEFICIAL ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) research has explored a variety of problems and approaches since its inception, but for the last 20 years or so has been focused on the problems surrounding the construction of intelligent agents – systems that perceive and act in some environment. In this context, ‘intelligence’ is related to statistical and economic notions of rationality – colloquially, the ability to make good decisions, plans, or inferences. The adoption of probabilistic and decision-theoretic representations and statistical learning methods has led to a large degree of integration and cross-fertilization among AI, machine learning, statistics, control theory, neuroscience, and other fields. The establishment of shared theoretical frameworks, combined with the availability of data and processing power, has yielded remarkable successes in various component tasks such as speech recognition, image classification, autonomous vehicles, machine translation, legged locomotion, and question-answering systems.

As capabilities in these areas and others cross the threshold from laboratory research to economically valuable technologies, a virtuous cycle takes hold whereby even small improvements in performance are worth large sums of money, prompting greater investments in research. There is now a broad consensus that AI research is progressing steadily, and that its impact on society is likely to increase. The potential benefits are huge, since everything that civilization has to offer is a product of human intelligence; we cannot predict what we might achieve when this intelligence is magnified by the tools AI may provide, but the eradication of disease and poverty are not unfathomable. Because of the great potential of AI, it is important to research how to reap its benefits while avoiding potential pitfalls.

The progress in AI research makes it timely to focus research not only on making AI more capable, but also on maximizing the societal benefit of AI. Such considerations motivated the AAAI 2008-09 Presidential Panel on Long-Term AI Futures and other projects on AI impacts, and constitute a significant expansion of the field of AI itself, which up to now has focused largely on techniques that are neutral with respect to purpose. We recommend expanded research aimed at ensuring that increasingly capable AI systems are robust and beneficial: our AI systems must do what we want them to do. The attached research priorities document gives many examples of such research directions that can help maximize the societal

benefit of AI. This research is by necessity interdisciplinary, because it involves both society and AI. It ranges from economics, law and philosophy to computer security, formal methods and, of course, various branches of AI itself.

In summary, we believe that research on how to make AI systems robust and beneficial is both important and timely, and that there are concrete research directions that can be pursued today.

Furthermore, Elon Musk helped to create the non-profit AI research company OpenAI with various donations totalling US\$1 billion [251]: ‘*Our goal is to advance digital intelligence in the way that is most likely to benefit humanity as a whole, unconstrained by a need to generate financial return.*’ [17].

The fact that well-known scientists and large companies are investing heavily in AI and that countries like the US [2, 327], China, Japan are launching large AI projects offers confidence that new breakthroughs will happen, and that these will certainly have a profound impact on our society in the coming years. President Obama says that “My Successor Will Govern a Country Being Transformed by AI” [357], showing the impact that could have AI in the future.

In fact, Artificial Intelligence is not a single technology; several complementary approaches are used, including statistical analysis, expert system and neural networks (now called deep learning approach).

2.4.11.1. DEEP LEARNING

Deep learning, or the use of Deep Neural Networks is a machine learning approach based on learning representation of data, using structures of Artificial Neural Networks. Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results on various tasks. The first DNNs were introduced by Kunihiko Fukushima in 1980 with his Neocognitron, and the task of training networks with multiple layers was partly solved in 1989 by Yann LeCun et al. by applying the back-propagation algorithm. But the real explosion began in October 2012, when a DNN called ‘Supervision’ by Alex Krizhevsky in the team of Geoff Hinton won the large-scale ImageNet competition by a significant margin over classical machine learning methods. Supervision is composed of 650,000 artificial neurons connected by 630,000,000 connections (synapses) (as the synapses are shared, it has only 60,000,000 parameters). Since then, DNNs provide the best results in classification on the ImageNet competition (see Figures 68, 69 and 70), now very comparable with the results obtained by humans. Thanks to the fact that a DNN is trained, and not explicitly programmed, it is applied in numerous applications, from image recognition, to speech understanding, lip reading and playing various games. A large ‘labelled’ database

showing what has to be done is all that is required; these are often available from the big Internet players (Google, Baidu, Facebook, etc.), explaining why they lead in the research and uses of deep learning.

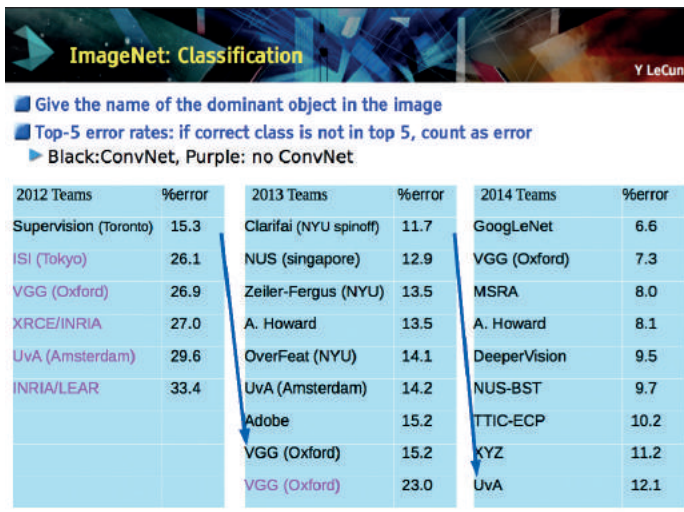


Figure 68: Progress in the ImageNet classification benchmark
Source: Yann LeCun

Team/algorithm	Date	Test error
Supervision	2012	15.3%
Clarifai	2013	11.7%
GoogLeNet	2014	6.66%
Microsoft	05/02/2015	4.94%
Google	02/03/2015	4.82%
Baidu/ Deep Image	10/05/2015	4.58%
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences	10/12/2015 (the CNN has 152 layers)	3.57%
	Now	?

Figure 69: Further progress on ImageNet

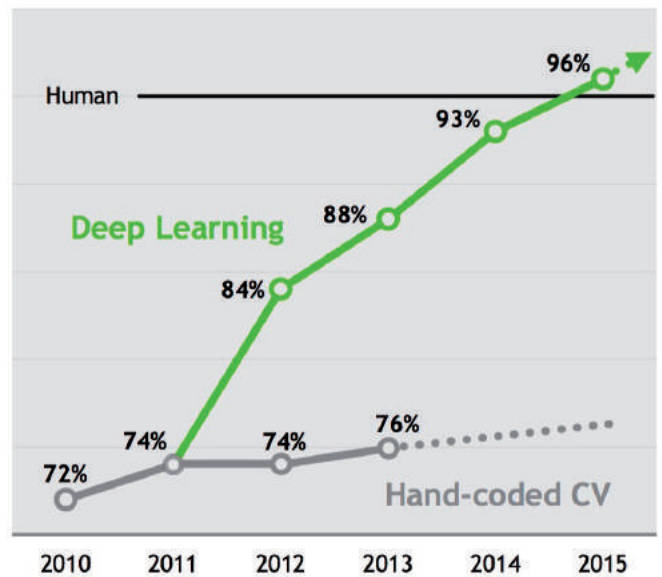


Figure 70: Progress of DNNs on ImageNet
Source: Imagenet: accuracy in %; from NVIDIA: 'Supercharge Deep Learning with DGX-1', Markus Weber, SC16, November 2016

Generally, there are two phases in the use of DNN: the learning phase, in which the parameters of the network (topology and the weight of the connections (synapses)) are determined by the learning rule, and the inference phase in which the DNN is used to classify data. The learning phase is the most demanding, with millions or billions of presentations of examples and modifications of the parameters of the DNN. It is now generally done on GPUs with simple precision floating points of event 16 bit floating points. The NVIDIA DGX-1 [416] is a system designed for accelerating the learning phase. The inference phase is less demanding and can be done on less precision (integer, even scaled down to 8 bits). It is generally this phase which is implemented in embedded devices for image recognition, etc.. The synaptic weights are downloaded after learning and can be updated after a new learning phase extending the number of recognized objects or doing a different function.

There are a large number of approaches for the learning phase, but they can be categorized in 3 classes:

- Supervised learning: presentation during learning of the inputs AND desired output corresponding to the particular class of the presented input;
- Unsupervised learning: the DNN determines its output from various inputs. It tries to discriminate the inputs into different classes automatically;
- Reinforcement learning: focuses on the prediction of a reward. It is this kind of learning that was used to train the AlphaGo program.

Deep learning techniques perform very well in recognition and classification of natural data, such as images, sound and signals. Classical approaches often deliver worse results and are, moreover, more difficult to program. Their inherent parallelism and tolerance to less accurate computations enable very efficient hard-

ware implementations, such as the Tensor Processing Unit (TPU) from Google [256] and IBM's TrueNorth chip [316].

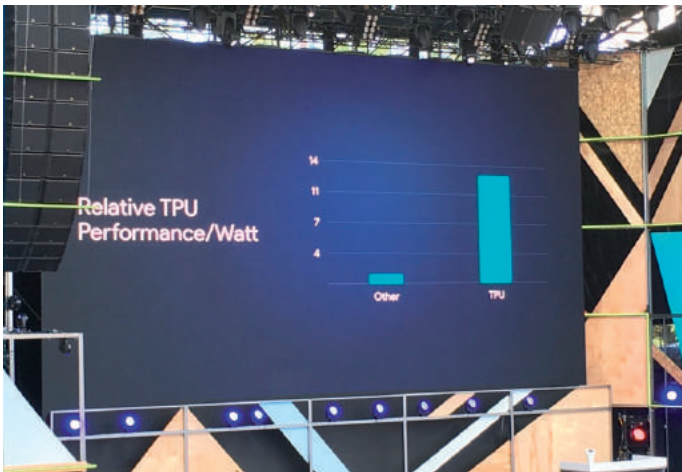


Figure 71: Google stresses the power efficiency as an important characteristic for its TPU Deep Learning accelerator
Source: Patrick Moorhead from Google I/O

Baidu made a machine dedicated to deep learning: Minwa. It consists of 36 server nodes, each with an Intel Xeon E5-2620, FDR Infiniband (56Gb/s) and 4 NVIDIA Tesla K40m GPUs with a total of 8.6 TB of memory.

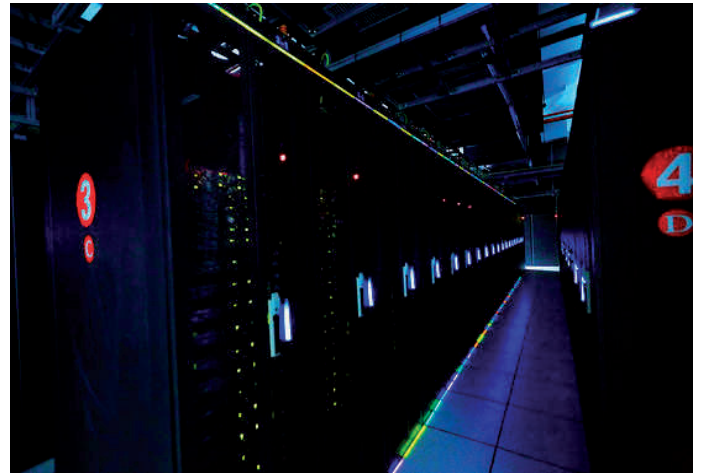


Figure 74: The Minwa machine
Source: Baidu



Figure 72: IBM TrueNorth system, able to simulate 48 millions of artificial Neurons
Source: IBM

NVIDIA is now focusing part of its strategy on deep learning, both for the learning phase with its DGX-1, and for use in vision applications such as self-driving cars. 'The DRIVE PX is an auto-pilot computing platform that can process video from up to 12 onboard cameras to run capabilities providing Surround-Vision, for a seamless 360-degree view around the car, and Auto-Valet, for true self-parking [...]. It will detect objects, tell you what they are, and even stop your vehicle if it has to. This means the car is not only sensing what's around you, it's also interpreting what's taking place around it.' From [396].



Figure 75: The NVIDIA DGX-1
Source: NVIDIA

f **Deep Learning is Everywhere (ConvNets are Everywhere)** Y LeCun

- **Lots of applications at Facebook, Google, Microsoft, Baidu, Twitter, IBM...**
 - ▶ Image recognition for photo collection search
 - ▶ Image/Video Content filtering: spam, nudity, violence.
 - ▶ Search, Newsfeed ranking
- **People upload 800 million photos on Facebook every day**
 - ▶ (2 billion photos per day if we count Instagram, Messenger and Whatsapp)
- **Each photo on Facebook goes through two ConvNets within 2 seconds**
 - ▶ One for image recognition/tagging
 - ▶ One for face recognition (not activated in Europe).
- **Soon ConvNets will really be everywhere:**
 - ▶ self-driving cars, medical imaging, augmented reality, mobile devices, smart cameras, robots, toys.....

Figure 73: The large number of photos uploaded per day, and they systematic process by Deep Neural Networks explain the need for optimized accelerators, like the TPU from Google.
Source: Yann LeCun

Facebook is proposing open-source AI hardware design [72]. There is a broad range of implementations, ranging from CMOS-based accelerators using classical binary coding up to 3D stacks of spike-encoded data processing chips using emerging memory technologies. Energy efficiency is the main driver for most implementations: a 'classical' CMOS realization of a DNN accelerator IP in FDSOI 28nm reaches 1.8 Tops/W on less than 0.5 mm² for a

quad core configuration (the architecture is scalable to several hundreds of cores), while a mixed analogue-digital implementation using spike coding can realize signal processing function using only 2.3% of the energy of an Atom-like core (Figure 76) [1].

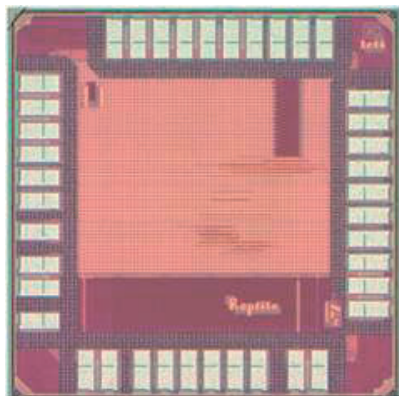


Figure 76: Spiking Neural Network (SNN) hardware accelerator for DSP functions (contains 3 tiles of 12 neurons)
Source: [5]

To reach the next step, synapses built using the same kind of technology as non-volatile memory (PCM – Phase Change Material - or CBRAM – Conductive Bridging RAM -) are under development. These ‘memristors’ can locally change their value according to the history of the current flow, thereby implementing something like an STDP (Spike Timing Dependent Plasticity), which is a biologically plausible local learning rule.

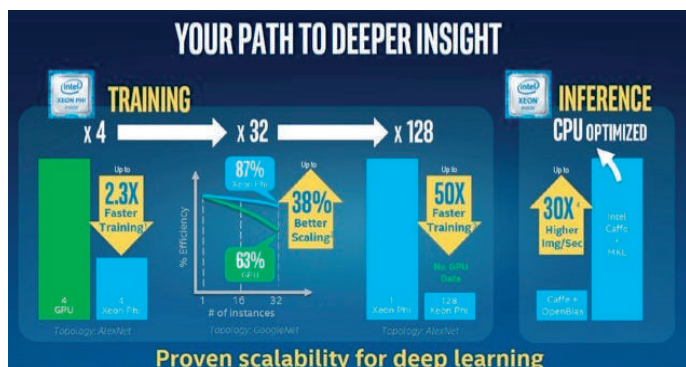


Figure 77: Intel claiming its solution is more efficient for deep learning tasks
Source: Intel

System	Hours to Train	Speed-up vs Xeon Phi
PC with 4x NVIDIA TITAN X (Maxwell) <small>* based on NVIDIA Caffe implementation as of March 2015</small>	25 Hours	-
Four Xeon Phi servers	10.5 Hours	-
PC with 4 NVIDIA TITAN X (Maxwell) <small>* based on publicly available Caffe as of August 2016, cuDNN5</small>	8.2 Hours	1.3x faster than Xeon Phi
PC with 4 NVIDIA TITAN X (Pascal) <small>* based on publicly available Caffe as of August 2016, cuDNN5</small>	5.5 Hours	1.9x faster than Xeon Phi
NVIDIA DGX-1	2 Hours	5.3x faster than Xeon Phi

Figure 78: And the answer from NVIDIA...
Source: NVIDIA

Big players are also delivering their AI software development tools in open source, like TensorFlow (Google), CNTK (Microsoft), DSSTNE (Amazon [198]), Theano, Caffe (Berkeley), Torch (Facebook contribute with open-source deep-learning modules Torchnet [130], OpenAI Gym (from Open AI), etc. [397]). In fact, software is a non-crucial element to make an effective deep learning system. A large database and the topology of the neural networks are the main ingredients: the value resides in the topology of the neural network and its weights, determined after learning on a particular database.

2.4.11.2. DATA ANALYTICS

The universe of IoT systems will create ZetaBytes of data, but most of it will be ‘dark data’: data that is written once and never read again, possibly because it gets overwritten or because access is lost. The current ‘gold rush’ consists of trying to extract meaningful information from this data using various data analytics approaches, mostly running in the cloud. Big data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights. It is the ingredient to transform the ‘data deluge’ into meaningful information. It is currently being developed by many companies in order to optimize processes or improve their business.

Two remarks can be made: now, data analytics is mainly carried out with ‘classical’ processors which may not have the best energy efficiency. Dedicated hardware or reconfigurable systems could lead to further improvement. Secondly, moving to a hierarchical approach, in which first analyses are performed at or near the capture location, can save energy. *Streaming analytics*, where data is analysed on the fly and not after being stored in a data server (in the cloud), follows a similar approach to reduce energy costs (and also provides faster results, more suitable to be immediately interpreted and used - data analytics in the loop).

2.4.11.3. IBM’S WATSON AND NATURAL MAN-MACHINE INTERACTION

The IBM Watson computer system idea started in 2004 during a dinner at a restaurant where all guests were looking at the Jeopardy! game on TV. As IBM was looking for a new ‘grand challenge’ to demonstrate its mastery of computing technology, the idea of a computer playing this game slowly gained momentum until the 2011 competition, in which Watson beat former human champions Brad Rutter and Ken Jennings and received the first place prize of US\$1 million. This was a major milestone in the human versus computer competition after IBM’s Deep Blue’s victory over Garry Kasparov in chess in 1997.

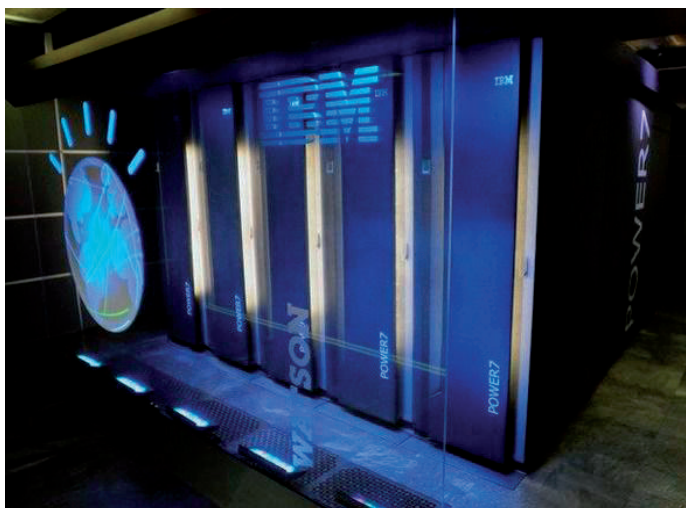


Figure 79: The original Watson machine
Source: IBM

The main characteristic of Watson is to answer questions posed in natural language. It combines various techniques and runs IBM's DeepQA software, which generates hypotheses, gathers massive evidence, and analyses data [317]. 'More than 100 different techniques are used to analyse natural language, identify sources, find and generate hypotheses, find and score evidence, and merge and rank hypotheses' [398]. The original machine had a processing power equivalent to 80 Tflops and 16 TBytes of RAM (where all the data is stored).

IBM has since started a completely new business leveraging Watson's capabilities [269], employing 2,000 people. It has invested US\$1 billion to get the division going and has investigated uses for medical diagnosis, [399] e.g. for helping take decisions on lung cancer treatment, and for various other applications including creating cooking recipes! IBM CEO Virginia Rometty said she wants Watson to generate US\$10 billion in annual revenue within ten years [348]. More details about Watson can be found at [400] and [401].

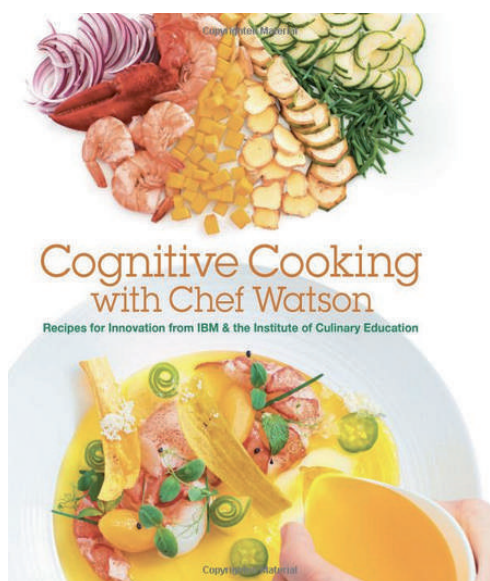


Figure 80: IBM Watson is also used to create new cooking recipes
Source: IBM

For IBM 'Watson is a technology that understands all forms of data and reasons and learns at scale [...] The goal is to have computers start to interact in natural human terms across a range of applications and processes, understanding the questions that humans ask and providing answers that humans can understand and justify.' [293].

We took the example of Watson because it demonstrates natural interaction with humans, using natural language. Additionally, the end-user does not have to 'program' the machine, but instead cooperates with the machine during various iterations of dialogue between man and machine. Initially, in 2011, the hardware requirements to run Watson were those of an HPC machine, costing about US\$1M. Today, it runs on cloud machines and the required capabilities have decreased by 90%. In the future, 80 Tflops and 16 Tbyte of (non-volatile) RAM could be on a cm³ device, so the kind of interaction we have now with Watson could be embedded in a mobile device or robot.

It is also of interest that it originated from the idea of a grand challenge, took seven years to be achieved, and now is expected to generate US\$10 billion in annual revenue.

2.4.11.4. THE 5TH RESEARCH PARADIGM?

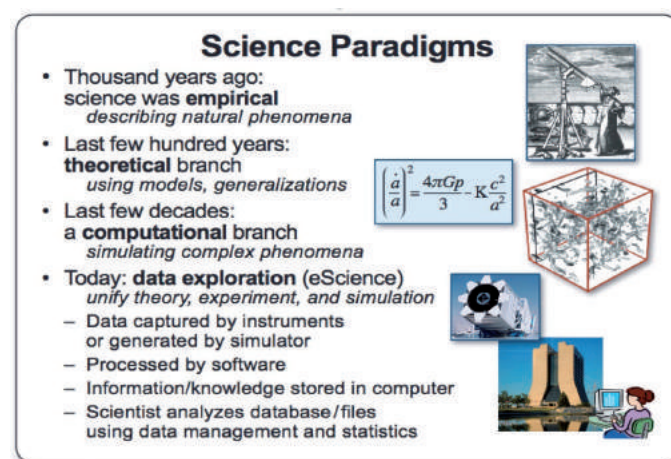


Figure 81: Evolution of science paradigms
Source: T. Hey, S. Tansley, K. Tolle, "The Fourth Paradigm: Data-Intensive Scientific Discovery"

Microsoft states that we are currently using the fourth paradigm of scientific discovery. The first three paradigms were experimental (empirical description of phenomena), theoretical (discovery of laws, models, etc. able to predict results) and, more recently, computational science (computer simulations). The fourth paradigm of scientific discovery is the analysis of massive data sets, enabled, e.g. by data capture, curation, mining and analytics techniques and thus permitting new scientific discoveries.

In the fourth paradigm, computers are used to extract information from raw data, but it is still humans who perform analyses of the information and make the scientific discovery. We believe that within the next decade there will be a **fifth paradigm**, in which computers will be not only extracting information from data, but will also formalize hypothesis, invent new simulations

or make new formal proof and finally make themselves scientific discoveries without human intervention. We already have examples of this with formal provers, data analytics systems, and approaches like IBM's Watson. Potentially, the Ultra-Intelligent machine could solve problems that are beyond the reach of human intelligence.

2.4.11.5. THE DECLARATIVE OR PARENTING SYSTEMS

Machine learning and artificial intelligence are solutions to the complexity problem: when the problem is too complex to break it down into algorithms, or too ill-defined, or when the problem can only be 'defined' using examples (such as image recognition). They are also examples of reuse: the 'core' software or hardware can be similar for very different applications, only the 'learning' or the adaptation to the problem must be customized. Dialogue and interaction with the system allows human and machine to converge towards the solution of the problem: **we will be rather teaching and parenting than explicitly programming** the systems.

From a software point of view, we can see two levels: the 'basic' software required to run the machine learning or the neural network. This part is still traditional HiPEAC technology based on languages and compilers. But the result will be an 'engine' (the 2nd level) that will be 'tuned' to the application by learning. From the hardware point of view, while the engine could run on a classical architecture, more specialized and therefore optimized architecture could be more interesting.

We can see this evolution already taking shape in the generative design approach where the user only states desired goals and constraints. The computer then generates entire designs, iterations and solution sets [186].



Figure 82: An early, bone-like, prototype of the Lightning Motorcycle swingarm. Autodesk's Dreamcatcher software can craft a swingarm – the piece that hinges the rear wheel to the bike's frame – by starting with weight and strength as the beginning design elements, rather than a predetermined shape of metal.

Source: Autodesk

Similar approaches can be used for the design of the architecture of a multicore processor (Rapid Technology Aware Design Space

Exploration For Embedded Heterogeneous Multiprocessors [402]) or the selection of the best optimization flags for a compiler.

2.4.11.6. OUR NEW ASSISTANTS: VIRTUAL OR ROBOTS

Another instantiation of the results of natural man-machine interaction, AI and self-learning systems are the new personal assistants, either in cyberspace (voice-activated like Siri, Alexa, Cortana or Google Now) or with virtual images like Gatebox from Vinclu [49]), or in the real world, like robot companions.

2.4.11.6.1. Voice controlled personal assistants

Intelligent Personal Assistants (IPA) such as Apple Siri, Google Now, Microsoft Cortana and Amazon Alexa (Echo) are now widely available on mobile devices, computers and custom devices. They use speech and natural language processing techniques, are connected to the cloud to offer different kinds of services, and use machine learning and neural network techniques to perform voice recognition. As a result, natural language might supplant the keyboard and touch screen as the new standard user interface to drive applications.

These services are provided for free on smartphones and computers. Amazon proposes a custom device, Echo, which does not have other interfaces besides an array of microphones and a speaker: 'Amazon Echo is a hands-free speaker you control with your voice. Echo connects to the Alexa Voice Service to play music, provide information, news, sports scores, weather, and more—instantly. All you have to do is ask.'

Google offers a similar device to the Amazon Echo with its 'Google home' device.



Figure 83: The echo from Amazon, only microphones and a speaker... and Alexa

Source: Amazon



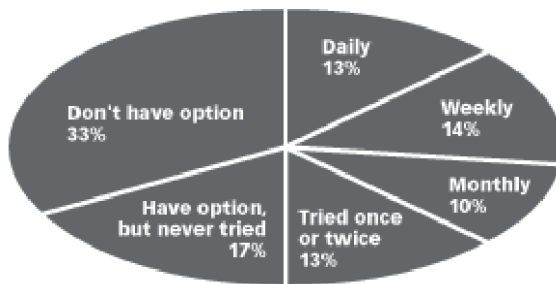
Figure 84: The google Home device, powered by Google assistant
Source: Google

'What's becoming loud and clear is that a machine's ability to recognize and process speech will be integral to the 'Internet of Things' universe, from wearables to connected cars, to home automation and appliances.' [343]

E.C. Baig

Frequency with Which US Mobile Phone Owners Use a Voice-Controlled Personal Assistant on Their Device, June 2015

% of respondents



Note: n=15,209

Figure 85: Use of voice controlled personal assistant in US
Source: www.emarketer.com

Figure 85 shows that in the US, more than 37% of smartphone users regularly use a voice-controlled personal assistant. As mentioned above, it might become the user interface for IoT devices and the disappearing computer era [80].

"Personable or not, intelligent voice agents are poised to alter the nature of computing as we know it. Since the dawn of the PC, humans have been forced to learn an arcane, unnatural language. But in a voice-first world, mediated by artificial intelligence and machine learning ... the next 50 years will see computers learning to be more like us."

Brian Roemmele [322]

2.4.11.6.2. Companion robots

'Personal' robots could be an instantiation of the disappearing computer, and might become the 'cyber-physical personal assistant'. Many 'home robots' are currently under development, often supported by crowdfunding (Indiegogo or Kickstarter). One of the most well-known is the Pepper robot, designed by Aldebaran, bought since by Softbank. It was announced on 5 June 2014 and demonstrated in Softbank shops. In the meantime, it has been introduced in about 700 Japanese and US businesses [321]. Since June 2015, 1000 units of the robot have gone on sale every month, each time selling out in a matter of seconds. Pepper was introduced as the first 'emotional robot', with some arguing that *'the notion of a companion robot pushes the boundaries of AI by not just figuring out who we are emotionally, but by simulating near-human levels of empathy and compassion. Pepper is just the first iteration'* [220].

Masayoshi Son, the founder of Softbank, is teaming up with Honda to bring robotics to cars, so that they can communicate with the driver, offering them company on long trips, reading their emotions and reacting accordingly [319]. Relatedly, Toyota and Honda are establishing research facilities on AI. Sony, who was well known for its Aibo robot, has also re-committed itself to robotics [349].



Figure 86: Pepper (originally developed by Aldebaran, now part of SoftBank) arguing its point on the impact of companion robots to the HiPEAC Vision Coordinator.

Studies forecast that robots will reside in more than 1 in 10 US households by 2020 [282]. Most of them will be utilitarian robots, like vacuum cleaners or lawn mowers, but robots like Pepper will also see increased use.



Figure 87: Zenbo from ASUS
Source: ASUSTeK Computer Inc.



Figure 88: Buddy
Source: 2016 Blue Frog Robotics [405]

There are tens of start-ups that propose robot companions on crowdfunding sites, with generally the same functionalities as a 'tablet on wheels': they generally embed a tablet as face or on the chest, are controlled by Apps available on the Apple or Google App stores, and can control home IoT devices like intelligent lighting and thermostats. They support voice recognition and synthesis (like the virtual personal assistant, but on a physically mobile platform), image and face recognition, surveillance functionality, and can navigate a house without bumping into the furniture. One cute example is Buddy from the French start-up Blue Frog Robotics.

LINK TOWARDS WEB SITES PRESENTING ROBOT COMPANIONS

<http://www.softbank.jp/en/robot/>
<https://zenbo.asus.com/>
<https://www.indiegogo.com/projects/buddy-your-family-social-companion-robot-family-social#/>
<https://www.indiegogo.com/projects/xibot-your-new-robotic-family-member--2#/>
<http://wowwee.com/chip>
<http://www.ubtrobot.com/product/aboutinfo1.html#matterIn1>
<https://www.indiegogo.com/projects/jibo-the-world-s-first-social-robot-for-the-home#/>
<https://www.indiegogo.com/projects/aido-next-gen-home-robot--2#/>
<http://www.aidorobot.com/press.html>
<https://www.indiegogo.com/projects/tapia-ai-robot-companion-learning-your-lifestyle-home-japan#/>
<http://mjirobotics.co.jp/en/>
<http://www.rob.it.io/>
<https://www.kickstarter.com/projects/403524037/personal-robot>
<https://www.autonomous.ai/deep-learning-robot>
<https://anki.com/en-us/cozmo>

The market of robot companions, assistants and humanoids is just emerging in 2017, with a few thousand to be sold by the end of the year [275]. It nevertheless represents the next evolution of the personal assistant, after smartphones and personal voice assistants. Studies show [320] that even if their practical use might be limited, they will contribute to fulfil a sociability need and will be accepted as a member of the family in a non-negligible proportion of households. They will also contribute to the intellectual well-being of elderly people and thereby enable them to stay in their homes for longer.

Contrarily to the personal voice assistants, these robots are not pure IoT devices, but real cyber-physical systems. This means they have to deal with all related constraints in terms of latency, real-time control of the motors, safety, image and sound analysis and recognition, and low power, all the while embedding more and more artificial intelligence functionalities.

2.4.11.7. IMPACT AND PROPOSED COURSE OF ACTIONS

Cognitive systems and narrow AI will have a drastic impact on the market and applications. They will have also a profound impact in the ICT domain. These systems will appear everywhere, and are posing new challenges in terms of processing power, storage and interaction with the world (a Watson in a cm³). *New architectures*, including *dedicated coprocessors*, will be needed to meet those challenges, and the safety and security requirement will become more stringent.

Even if a system uses the cloud rather than fully process information locally (which we think is only a stopgap approach), a dedi-

cated chip may be required to safely trigger the cloud transfer. For example, a personal voice assistant cannot transmit all of the sound it captures to the cloud, as this is unacceptable in terms of both privacy and energy requirements. Therefore, a local system should be able to recognize when to “call the cloud” (in the case of speech recognition, it must be able to detect the trigger sentence).

We also see that *tools will help, advise on and finally produce the code* instead of a programmer, which will become more of an adviser that formulates goals rather than a list of commands to carry out. A first step towards this approach could be to assemble libraries, followed by a next step where the complete code is optimized according to the characteristics of the target. This does not only apply to software, it could also lead to automatic generation of FPGA netlists in a first stage, and later on to ASICs.

The requirements of the interaction with the real world could also trigger innovative architectures such as event driven processors and systems with guaranteed response time. In all cases, security will have to be an important aspect of the design constraints.

Finally, we will have more and more *dynamic systems*, or systems that are not explicitly programmed, so *how do we ensure they will reach the assigned goal* and will operate within (safe) boundaries? New models will certainly need to be developed, evaluated and tested. We are not yet at the level of the ‘3 laws’³ of robotics of Isaac Asimov, but a safety guard approach will be soon required.

THE 3 LAWS OF ROBOTICS BY ISAAC ASIMOV

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

³ In fact, at the end, Isaac Asimov added a 4th law, called law “zero”, reading: ‘A robot may not harm humanity, or, by inaction, allow humanity to come to harm’, and adapted the initial 3 laws accordingly.

2.5. TECHNOLOGICAL TRENDS

2.5.1. TIME TO REVISIT THE BASICS: VON NEUMANN, NEURAL NETWORKS AND QUANTUM COMPUTING

In the previous HiPEAC Vision, we already showed that we are slowly approaching the limits of what technology could bring in term of scaling. Even if Moore's law is still continuing, and allowing the increase of transistor density, cost reduction (the cost of an individual transistor) is not so obvious. Dennard's scaling has already been over for years, meaning that the increase in transistor density is no longer linked with a similar increase in frequency at which the transistor could operate, nor is it linked anymore to a similar decrease in energy [417].

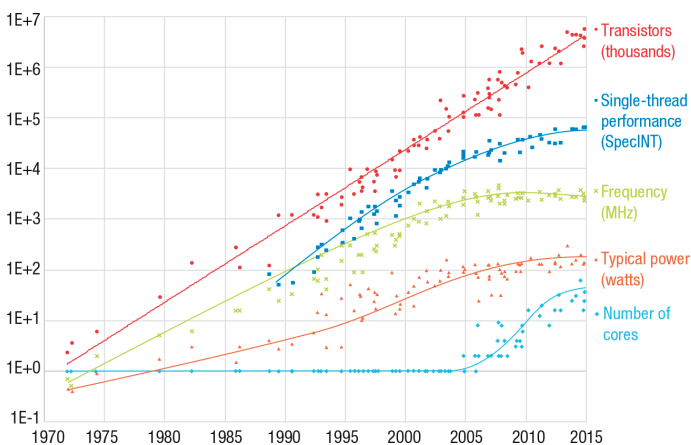


Figure 89: Evolution of microprocessor performance over time. From 2000-2005, we seen the end of “Dennard’s scaling”, leading to stagnation of the frequency of the cores and the rise of multi then manycore era.

Source: [418]

As we have been reaching the limits of the silicon technology that fuelled the ICT evolution for years, it is perhaps time to sit back and revisit the basics and all the hidden assumptions that drove the evolution of computing systems, both hardware and software. Of course, this should take legacy into account, and the new ideas that will emerge for this re-thinking of ICT should be progressively introduced, for example implemented as accelerators for specific tasks, slowly replacing more and more the low efficiency parts of classical systems. This re-thinking should be holistic, and should cover not only new architectures and new ways to make good software, but should also encompass new computation models and the use of new technologies, even non-silicon based.

As an illustration, if we revisit all the work of John Van Neumann, we see that the Von Neumann architecture is only the consequence of the technological limits of his time, and not really what Von Neumann pursued as the main direction of his research. When looking more closely at his work, we can see that it covers concepts that are again hot research topics today, such as quantum computing.

THE VARIOUS CONTRIBUTIONS OF JOHN VON NEUMANN

Stochastic computing was first introduced in a pioneering paper by von Neumann in 1953. However, the theory could not be implemented until advances in computing of the 1960s.

He also created the field of cellular automata without the aid of computers, constructing the first self-replicating automata with pencil and graph paper.

Beginning in 1949, von Neumann's design for a self-reproducing computer program is considered the world's first computer virus, and he is considered to be the theoretical father of computer virology.

In a famous paper of 1936 with Garrett Birkhoff, the first work ever to introduce quantum logic, von Neumann and Birkhoff first proved that quantum mechanics requires a propositional calculus substantially different from all classical logic and rigorously designed a new algebraic structure for quantum logics.

Von Neumann founded the field of game theory as a mathematical discipline.

The following June, in 1945, von Neumann penned what would become a historic document titled “First Draft of a Report on the EDVAC,” the first published description of a stored-program binary computing machine—the modern computer. The EDVAC's predecessor, the ENIAC, which took up 1,800 square feet of space in Philadelphia, was more like a giant electronic calculator than a computer. It was possible to reprogram the thing, but it took several operators several weeks to reroute all the wires and switches to do it. Von Neumann realized that it might not be necessary to rewire the machine every time you wanted it to perform a new function. If you could take each configuration of the switches and wires, abstract them, and encode them symbolically as pure information, you could feed them into the computer the same way you would feed it data, only now the data would include the very programs that manipulate the data. Without having to rewire a thing, you'd have a universal Turing machine.

To accomplish this, von Neumann suggested modelling the computer after Pitts and McCulloch's neural networks. In place of neurons, he suggested vacuum tubes, which would serve as logic gates, and by stringing them together exactly as Pitts and McCulloch had discovered, you could carry out any computation. To store the programs as data, the computer would need something new: a memory. That's where Pitts' loops came into play. “An element which stimulates itself will hold a stimulus indefinitely,” von Neumann wrote in his report, echoing Pitts and employing his modulo mathematics. He detailed every aspect of this new computational architecture. In the entire report, he cited only a single paper: “A Logical Calculus” by McCulloch and Pitts.

Source: [406]

We can therefore say that Von Neumann foresaw the technologies that are currently at a research stage, and that are currently high on the hype curve (see Figure 22), but he was not able to implement them due to the technology available at that time. We can thus argue that the *true* Von Neumann era is the one of neural networks and quantum computing, into which we are just entering *now*.

2.5.2. THE SILICON ROADMAP

2.5.2.1. CURRENT STATUS

The semiconductor market as a whole is now a mature one that feeds into the global electronics production industry. The material and equipment supply is not negligible, reaching \$1,457B in 2016 (Figure 90).

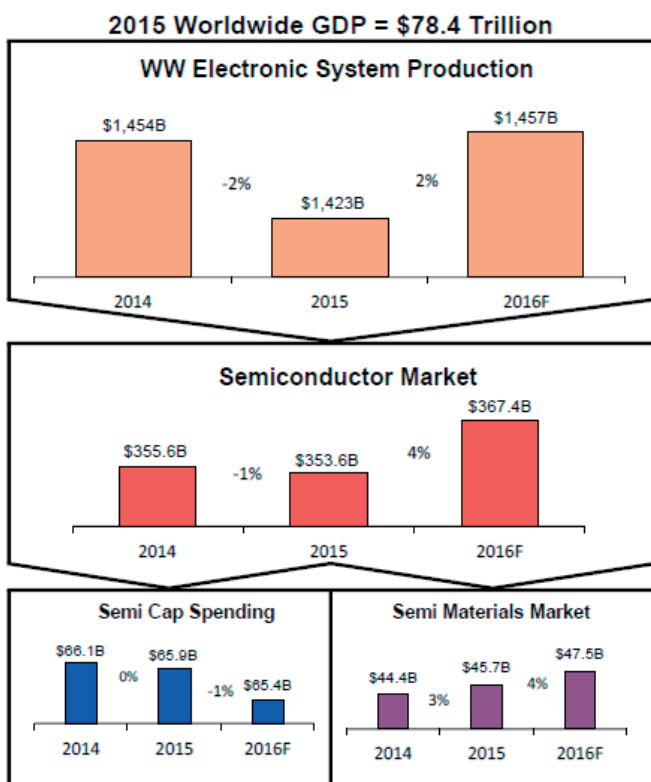
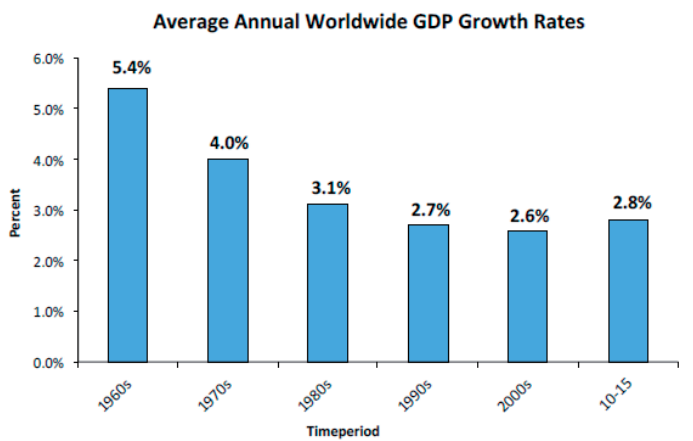
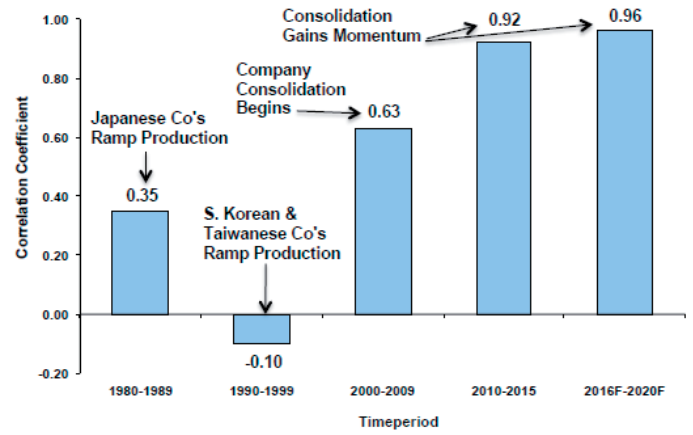


Figure 90: Interdependence of the electronic industry
Source: IC Insights

Its maturity is proven by the lower growth rate and by the increased correlation with GDP growth; the market will continue to grow in both volume and value but it will be profoundly different from the past. The consequences that can already be seen are a consolidation and specialization of the actors and a reduction of the overall R&D spending.



Figures 91 and 92: Worldwide GDP growth
Source: World Bank and IC Insights; IC Insights

Recently and in quite a short time, a lot of mergers and acquisitions have taken place: Intel bought the FPGA company Altera, Avago bought Broadcom, and NXP bought Freescale and was then bought by Qualcomm (for US\$47 billion).

We also observe that in the pure-play foundries companies (i.e. companies in which the main business is foundry, unlike Intel, Samsung and ST Microelectronics), there is one company which covers more than half of the global market.

In term of devices, MPUs (Micro Processor Units) represent only 10% of units but make up 78% of revenue. In particular, PCs and servers are 2% of units but 50% of revenue. MCUs (Micro Controller Units) make up most of the units (88%) but they are not the drivers for development as the revenue is too low and they are lagging in the technology node (mostly >40nm). The MPUs are the driver of the CMOS race and, even if the segmentation changes (we observe a shift from PC and servers to mobile devices), the growth is strong and the revenue still high. We can also observe that the x86 architecture loses ground everywhere but in servers and mainframes.

Top 10 Worldwide Semiconductor Sales Leaders* (\$B)

Rank	1990		1995		2000		2006		2015		2015 With Mergers	
1	NEC	4.8	Intel	13.6	Intel	29.7	Intel	31.6	Intel	50.5	Intel/Altera	52.1
2	Toshiba	4.8	NEC	12.2	Toshiba	11.0	Samsung	19.7	Samsung	41.6	Samsung	41.6
3	Hitachi	3.9	Toshiba	10.6	NEC	10.9	TI	13.7	SK Hynix	16.9	SK Hynix	16.9
4	Intel	3.7	Hitachi	9.8	Samsung	10.6	Toshiba	10.0	Qualcomm**	16.0	Qualcomm**	16.0
5	Motorola	3.0	Motorola	8.6	TI	9.6	ST	9.9	Micron	14.5	Avago/Broadcom**	15.3
6	Fujitsu	2.8	Samsung	8.4	Motorola	7.9	Renesas	8.2	TI	12.1	Micron	14.5
7	Mitsubishi	2.6	TI	7.9	ST	7.9	Hynix	7.4	Toshiba	9.7	TI	12.1
8	TI	2.5	IBM	5.7	Hitachi	7.4	Freescale	6.1	Broadcom**	8.4	NXP/Freescale	10.2
9	Philips	1.9	Mitsubishi	5.1	Infineon	6.8	NXP	5.9	Avago**	6.9	Toshiba	9.7
10	Matsushita	1.8	Hyundai	4.4	Philips	6.3	NEC	5.7	Infineon	6.9	Infineon	6.9
Top 10 Total (\$B)		31.8	86.3		108.1		118.2		183.6		195.4	
Semi Market (\$B)		54.3	154		218.6		265.5		353.6		353.6	
Top 10 % of Total Sem		59%	56%		49%		45%		52%		55%	

*Not including foundries **Fabless

Figure 93: Consolidation in the semiconductor industry

Source: IC Insights

Top 20 Pure-Play Foundry Companies

2016F Rank	2015 Rank	Company (Headquarters)	2014 Sales (\$M)	14/13 % Change	2014 Share of Total	2015 Sales (\$M)	15/14 % Change	2015 Share of Total	2016F Sales (\$M)	16/15 % Change	2016 Share of Total
1	1	TSMC (Taiwan)	24,975	25%	59%	26,439	6%	59%	28,570	8%	58%
2	2	GlobalFoundries (U.S.) ¹	4,355	6%	10%	5,019	15%	11%	5,645	12%	11%
3	3	UMC Group (Taiwan)	4,331	9%	10%	4,464	3%	10%	4,490	1%	9%
4	4	SMIC (China) ²	1,970	0%	5%	2,236	14%	5%	2,850	27%	6%
5	6	TowerJazz (Israel)	828	64%	2%	961	16%	2%	1,245	30%	3%
6	5	Powerchip (Taiwan)	1,291	9%	3%	1,268	-2%	3%	1,240	-2%	3%
7	7	Vanguard (Taiwan)	790	11%	2%	736	-7%	2%	780	6%	2%
8	8	Hua Hong Semi (China)	665	14%	2%	650	-2%	1%	700	8%	1%
9	9	Dongbu HiTek (S. Korea) ³	541	20%	1%	593	10%	1%	640	8%	1%
10	10	SSMC (Singapore)	480	-3%	1%	474	-1%	1%	470	-1%	1%
11	11	X-Fab (Europe)	330	14%	1%	331	0%	1%	460	39%	1%
12	12	WIN (Taiwan)	327	-8%	1%	379	16%	1%	440	16%	1%
13	14	LFoundry (Europe) ⁴	220	38%	1%	242	10%	1%	260	7%	<1%
14	13	Altis (Europe)	255	6%	1%	252	-1%	1%	255	1%	<1%
15	15	TSI Semi (U.S.)	245	7%	1%	240	-2%	1%	245	2%	<1%
16	17	XMC (China) ⁵	165	10%	0%	175	6%	0%	195	11%	<1%
17	16	Silterra (Malaysia)	180	0%	0%	180	0%	0%	180	0%	<1%
18	19	Shanghai Huali (China)	105	24%	0%	115	10%	0%	130	13%	<1%
19	18	ASMC (China)	130	11%	0%	119	-8%	0%	120	1%	<1%
20	20	AWSC (Taiwan)	86	146%	0%	136	58%	0%	100	-26%	<1%
—	—	Others	87	3%	0%	93	3%	0%	100	3%	<1%
—	—	Total	42,356	17%	100%	45,102	6%	100%	49,115	9%	100%

1. Includes \$740 million in 2H15 sales from IBM purchase. 2. Partially owned by TSMC.

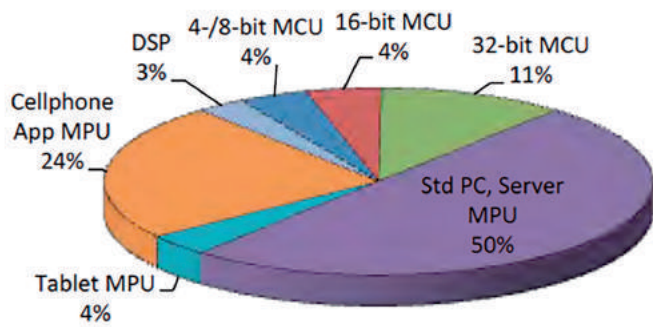
3. Currently up for sale. 4. To be 70% owned by SMIC.

5. Reported to have been acquired by Tsinghua Unigroup in 3Q16.

Figure 94: market share of pure-play foundries

Source: IC Insights

2016F Microcomponent Marketshare (\$83.8B)



2016F Microcomponent Unit Marketshare (26.6B)

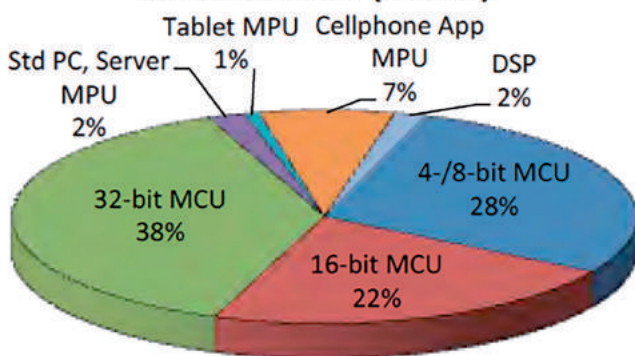


Figure 95: MCU market evolution
Source: IC Insights

The separation between these two products explains the different focus of the two roadmaps (MPUs and MCUs) and the difference of the introductions of advanced technologies between pure play foundries and Intel:

- High performance/low volume and single user for Intel (with the latest technology node, driving it for performance)
- High yield/high volume and multiple users (flexible platform) for pure play foundries (using the most cost effective technology node for the product).

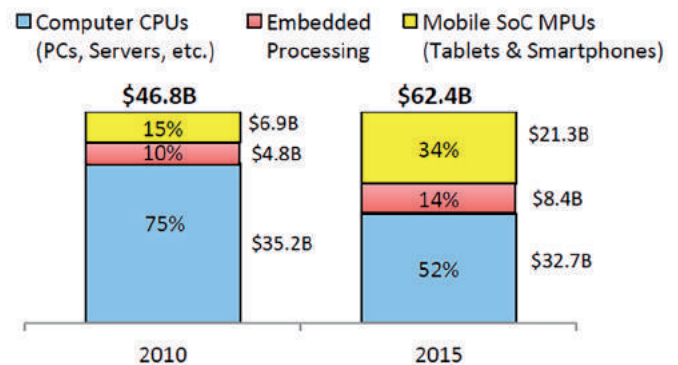
2.5.2.2. WHAT IS COMING NEXT?

In order to continuously increase the performance of devices, the race for ever-smaller transistors is still ongoing. While it has been claimed several times in recent decades that Moore's law had ended, it still holds today and the number of transistors per square millimetre is increasing with each new technology node. What *has* changed in the last decade is the 'law' of Dennard [43]. In the early technology nodes, going from one node to the next allowed for a nearly doubling of the transistor frequency, and, by reducing the voltage, power density remained nearly constant. With the end of Dennard's scaling, going from one node to the next still increases the density of transistors, but their maximum frequency is roughly the same and the voltage does not decrease accordingly. As a result, the power density increases now with every new technology node. The biggest challenge therefore now

consists of reducing power consumption and energy dissipation per mm².

Furthermore, due to reductions of on-chip feature sizes, power leakage increases. The reason for this is that transistors cannot be completely 'switched off' anymore, which significantly increases their standby current. Hence, the ratio of leakage versus active current will increase, further degrading the systems' energy efficiency. The pattern sizes will become so small that signal loss in on-chip interconnects due to high wire resistance will no longer be negligible, increasing inefficiency yet further.

Shifting Microprocessor Sales



2016 MPU Sales by Application (Fcst, \$65.0B)

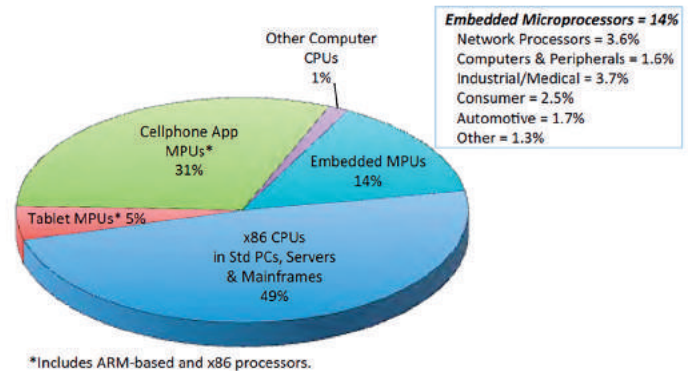


Figure 96: MCU market evolution
Source: IC Insights

The traditional technology roadmaps (like the one on technology nodes and also ITRS) are therefore running into power and cost limits. To continue scaling with power limits will require disruptive technologies, and scaling within cost limits will also require disruptive architectures and integration technologies.

The cost of development of new technology nodes is so high that, for the first time, the cost per transistor could increase, and the number of companies able to invest in lower size nodes is decreasing.

In April 2015, IC insight predicted that only three companies will offer nodes below 14nm, but FDSOI (with its planar structure which is easier to realize than the 3D structures of FinFet) allowed Global Foundries to continue to be in the race with a 12 nm FDSOI offer [264].

Cost of technology increasing after 28nm

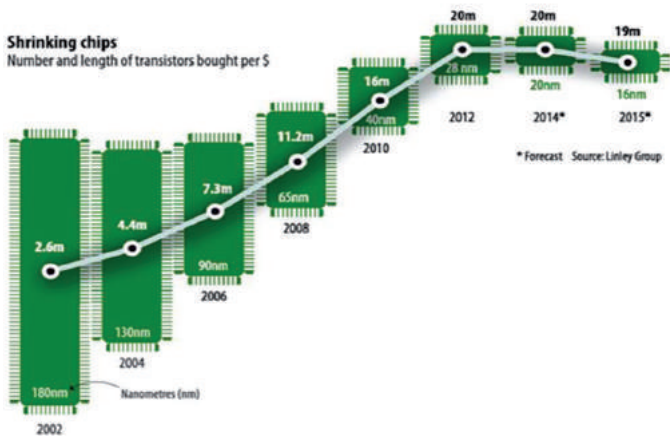


Figure 97: Increasing transistor cost
Source: SAMSUNG

Lesser number of players for leading edge nodes

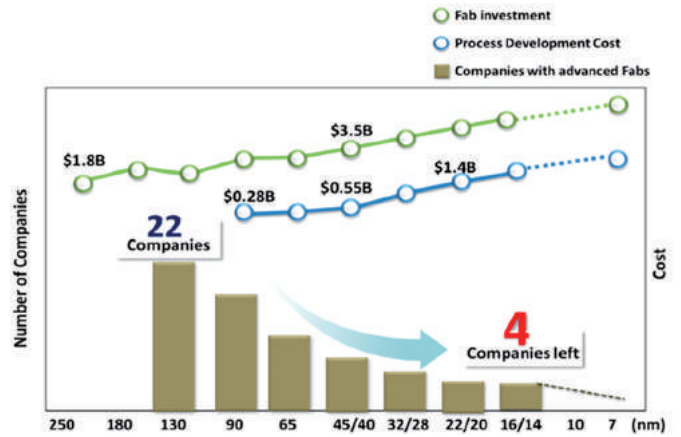


Figure 98: Decrease of the number of players
Source: SAMSUNG

FDSOI is now a credible solution for energy-efficient systems, for example for IoT developments, while FinFet is offering the fastest designs. Currently the solution for computing for the next few years is clearly mapped out by the current leaders:

- FinFet down to 7nm node;
- More open path for 5-3nm node;
- No clear idea below 3nm node.

There is also the idea of the splitting of the world into two paths: below 22nm for the high-end devices and above 22 nm for the bulk of designs [404]. The 22nm and above become the commodity market

The key limiting factors are lithography, properties of the material, device physics and manufacturing (variability).

For lithography, the state-of-the-art is multiple patterning with immersion lithography (using different flavours). It has generated a big market for lithography/etching/metrology solutions (from ASML, LAM, AMAT). The technique has its drawbacks but it has been shown to be capable in terms of both imaging and throughput. The major issues are the capital required and the NRE (design, masks) costs which impact upon the final product costs.

Technology Roadmap

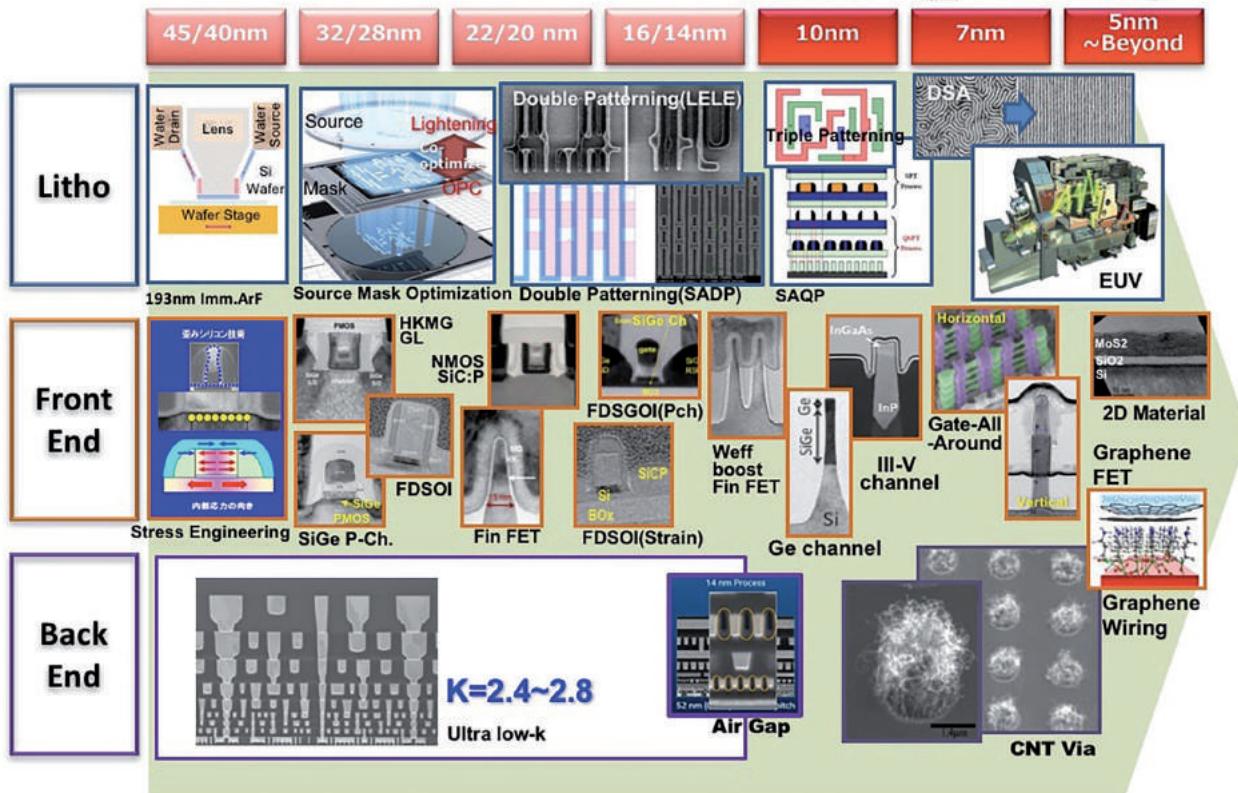


Figure 99: Various technology for going to denser nodes
Source: Yuzo Fukuzaki

Our fourth-gen NXT is rapidly ramping into production
Better overlay, focus, productivity in support of leading-edge nodes

ASML
Page 14
28 April 2015

- NXT:1980i starting shipping late 2015
 - 1.2 nm dedicated chuck overlay
 - Better than 10 nm focus uniformity
 - 10% throughput increase to 275 wafers/hour
- Designed for mix-and-match use with EUV
 - About 2 nm matched-machine overlay
- Flexible product configuration provides extension for logic, DRAM and NAND



Figure 100: ASML is a European company leader in lithography, enabler for the all semiconductor industry
Source: ASML

The next step is using EUV (Extreme UV lithography), but there will be a delay in this happening due to its technological complexity. Yet all the pieces are finally coming together, and the main and sizeable benefits help retain the faith of the actors involved. It could be introduced at a later point at 7nm (to reduce costs) or have an earlier initial introduction at 5nm. It remains the mainstream choice and further developments are to be announced.

In the current CMOS development, the limiting material factors are the driving current (simplified wrongly with the need for higher mobility materials), the supply voltage (hence the fashion for steep subthreshold slope devices), the internal node capacitances, the total access resistance (contacts, etc.) and the RC of the interconnects (higher R – resistivity – and higher C – capacitance – with shrinking feature sizes).

The old pure material or pure device approach has shown its limit as the solution has to come from a closely coupled analysis of the interactions. The problem of the reducing drive current with scaling has plagued the CMOS since the 65nm node and a solution has been the introduction of strain and SiGe. For new materials, 2D and topological insulators are still in infancy and at more than 10 years from any possible introduction in pre-industrial development due to deposition issues, quality, repeatability and fundamental understanding.

Concerning the transistor architectures, two main issues are to be considered:

- how to maintain a good electrostatic control of the device when shrinking (Ion/Ioff, DIBL – Drain Induced Barrier Lowering – a measure of the electrostatic control of the channel by the gate) in order to have a good window of operation for the inverter and low power operation;
- how to choose the structure to maximize the drive current per unit area and reduce the cell size.

Today the roadmap until 5nm node (2021) is pretty much fixed, with FinFet as the leader and giving first introduction of the nodes and FDSOI as an alternative, with 2-3 years lag time (some nodes may be skipped – e.g. 14/16 nm).

At around 7/5 nm node, the FinFet reaches its physical and manufacturing limitations (electrostatic confinement, aspect ratio, uniformity), and there is a nearly unanimous response with a move towards stacked nanowires.

Table 2 | Summary of the opportunities and challenges for 2DM-based electronics in different applications.

	Opportunities	Challenges
Production	Uniform single-crystal films Inexpensive fabrication technology High mobility Atomically thin films 2DM-inks with on-demand electronic properties	Reduction of surface states Growth of large (>1 cm ²) single crystals Doping control (less than the solubility limit) Metal/semiconductor interface control On-demand control of morphological (lateral sizes and thickness) and rheological (surface tension, viscosity, density) properties of 2DM-based functional inks
High performance	Reduced short-channel effects Good performance in terms of τ , I_{on}/I_{off} (>10 ⁴)	Fabrication of ultrashort channel devices (channel length smaller than 10 nm) Fabrication of devices based on new principles to reduce V_{DD} and SS Good ohmic contacts with low source-drain parasitic resistance
Low power	Low-power performance as compared with Si (power supply <0.5 V) Good control of the gate over the tunnel barrier, with SS << 60 mV dec ⁻¹ Large I_{on} currents (>10 ³ μ A μ m ⁻¹)	Fabrication of doped tunnel junctions Low interface states to reduce SS Design of new device architectures
Radiofrequency	High μ and saturation velocity (graphene) Development of 2DM heterostructures for terahertz operations	Reduce contact resistance (<100 Ω μ m) Obtain f_{max} in the THz range
Flexible electronics	Ultrathin bendable material Mobility larger than in materials already available (>40 cm ² V ⁻¹ s ⁻¹) Enabling technology for wearable electronics	Improve material mobility Development of roll-to-roll fabrication processes Development of device/circuit fabrication technologies

Figure 101: Opportunities and challenges for 2DM-based electronics
Source: [407]

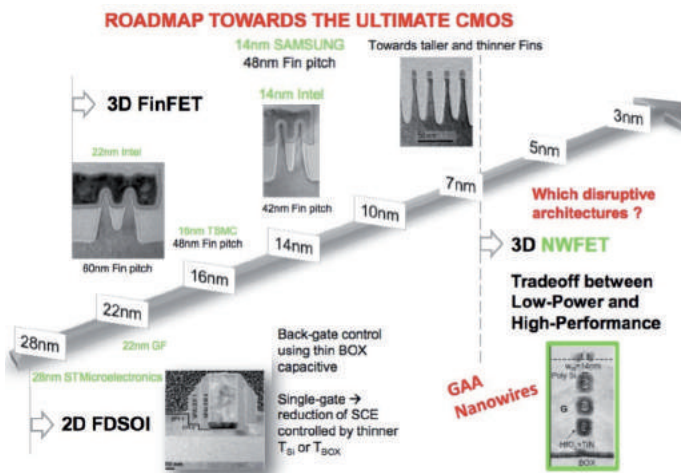


Figure 102: From FinFet to Nanowires
Source: [437]

The nanowires are a logical evolution from FDSOI, with its ultra-thin Box separating the active transistor gate from the substrate, and the FinFet, which are evolving into taller and thinner fins. Very small structures like fins separated by structure like FDSOI's BOX are in fact making nanowires that can be stacked to increase the current they can drive.

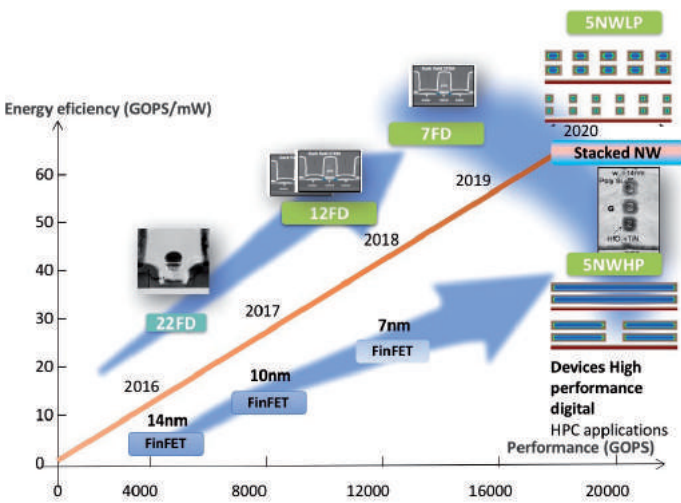


Figure 103: From FinFet and FDSOI to nanowires
Source: [437]

It should be noted that stacked nanowires can represent a step towards making quantum dots in silicon (especially Si-28 at ultra-low temperature of a few mKelvins). This could leverage the entire infrastructure of the semiconductor industry to build quantum computers [377].

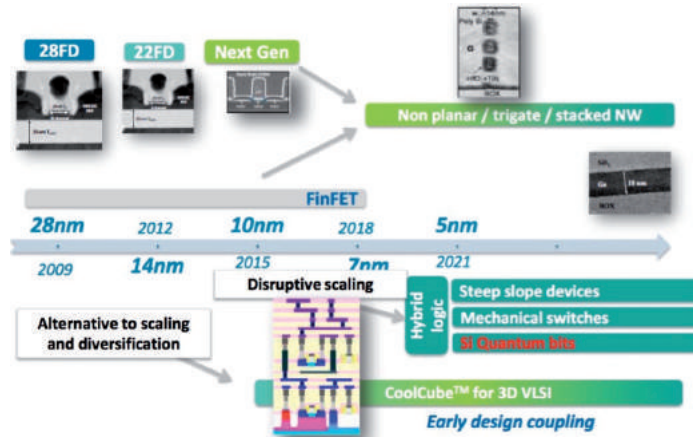
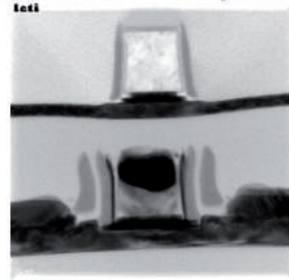


Figure 104: Summary of the possible evolutions of semiconductors
Source: [437]

Besides decreasing the device size, other solutions can also be used to further increase the device density. For example, 3D stacking, and especially its advanced forms like monolithic 3D that allow a very fine partitioning between layers (so low that an inverter can be done with its two transistors in two different layers).

Monolithic 3D principle



CMOS/CMOS: 14nm vs 2D:
Area gain=55%
Perf gain = 23%
Power gain = 12%

LETI, DAC 2014 QUALCOMM

Figure 105: Principle of 3D monolithic (we can see two transistors one above the other)
Source: LETI

More 'classical' 3D stacking, e.g. using TSVs (Through Silicon Vias) or Copper-to-Copper interconnect, and 2.5D (small silicon dies connected on a silicon 'interposer', acting as a miniature Printed Circuit Board) are a solution to the development cost and diversity of monolithic SoCs. It is also illustrated by the CHIPS initiative from DARPA [228], which promotes the idea of chiplets and interposers.

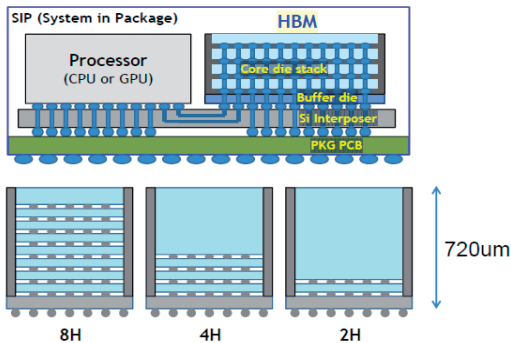
This was already illustrated in the HiPEAC Vision 2015 ('Entering the third dimension', pp. 33-34). Die stacking creates several opportunities:

- It enables the building of new composite architectures by physically placing separately manufactured components very close together through stacking. For example, we can place memories and processors (for many-core architectures) or

sensors and processing (example: intelligent retinas) on top of each other. Directly stacking the memory above the processor increases energy efficiency by reducing the interconnect losses and, due to the higher number of interconnect points, can increase the bandwidth between the memory and the processor. The new memory systems, Hybrid Memory Cube (HMC) and High Bandwidth Memory (HBM) are new standards for 3D stacked memories;

HBM Architecture

- SiP using HBM and 2H/4H/8H intersection



Chip Implementation

- Process: 20nm DRAM
- Capacity: 9Gb/core die
- Supply voltage: 1.2V/1.2V/2.5V
- Chip size: 12mm x 8mm (buffer die)

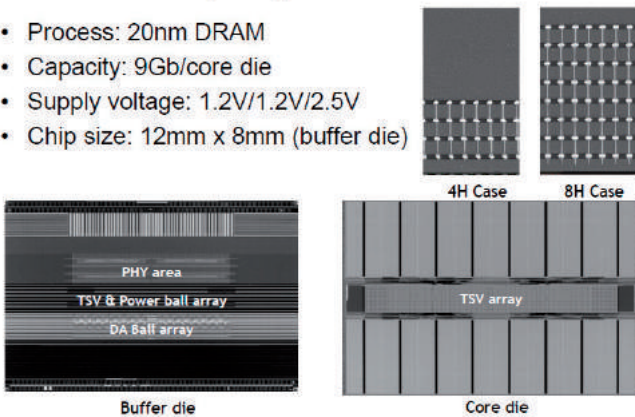
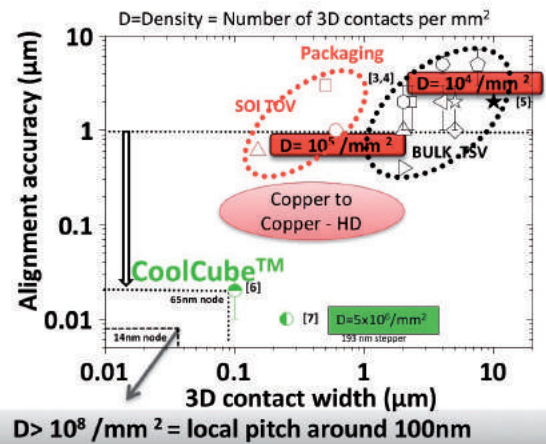


Figure 106: The Memory cube is using 3D stacking technology
Source: IEEE

- It allows the combining of different technologies in one package, meaning that not all the dies in a package need to be produced in the same technology node. This can extend the lifetime of existing fabs by, for example, producing a sensor chip in an older technology and mounting it on top of a modern processor die;
- Through different combinations of dies, it is possible to regain the chip diversity lost due to the increasing costs of chip design and particularly due to the cost of semiconductor chip fabrication plants, which doubles every four years (Rock's law). By reusing dies in different packages, the die volume of one design will increase, thereby lowering the cost, while simultaneously providing for more differentiation through different stacking combinations;

- Silicon interposers are also promising for the integration of silicon photonics, thereby enabling optical interconnects between systems or dies, with the potential advantages of lower energy for communication and higher bandwidth.



[3,4]: P. Garrou et al., Handbook of 3D integration, Vol 1,2 (Wiley ed) / [5]: B. Banijamal, ECTC2011
[6]: S-M. Jung et al., VLSI 2005 pp220 / [7]: P. Batude et al, ECS journal 2008, VO16,pp47

Figure 107: Various density of contacts for 3D interconnect
Source: see legend embedded in image

Some other devices are also being investigated for computing 'beyond CMOS'. However, it is clear that the new devices will be very different from and will not replace CMOS devices, but will augment and complement CMOS. Like for computer architectures, accelerators of different forms (and using different technologies) will complement the classic CMOS-based microprocessors. Anyway, there are and will continue to be opportunities for active and interesting research. However, the unknowns are many: benchmarking activity shows that there are intrinsic limitations in many approaches and we are at the beginning of a >15 year cycle before arriving at pre-industrialization [13].

Hierarchical level	CMOS	Beyond CMOS
Materials	Silicon	III-V, Correlated oxides, High-Z metals
Device	MOSFET	Tunneling-FET, MESO (Magneto-Electric / Spin Orbit Torque)
Interconnect	Electronic	Electronic, Photonic
Circuits	CMOS	Electronic, Spintronic
Architecture	Von Neumann	Von-Neumann, Non Von-Neumann
Memory	SRAM/DRAM	Electronic, Spintronic, NVM

Figure 108: Example of "beyond CMOS" research directions

The Future is Full of Opportunity

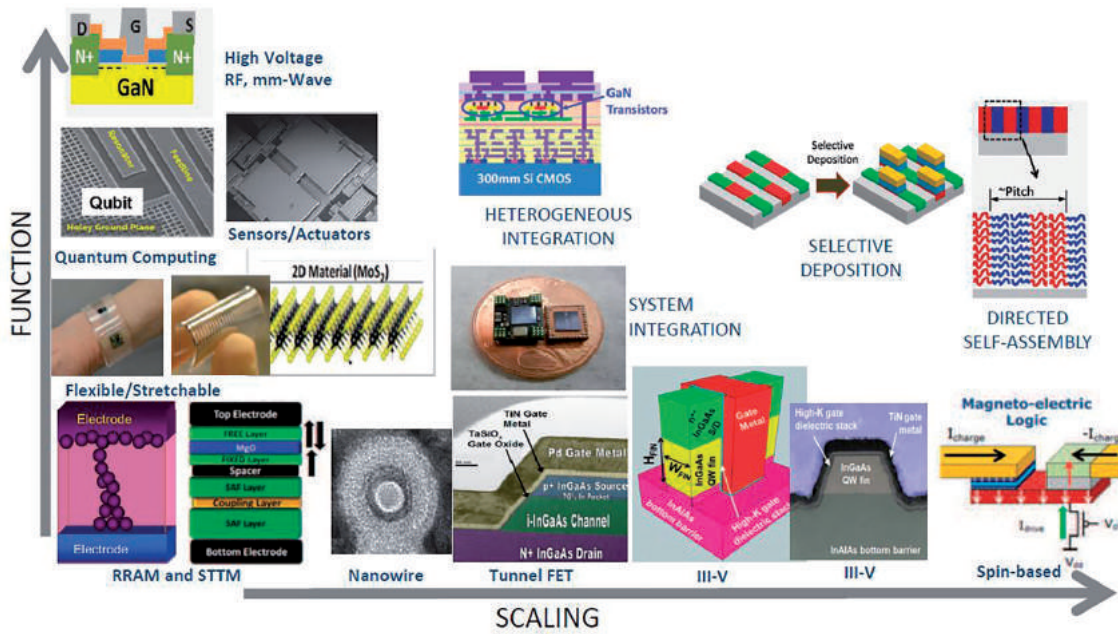


Figure 109: Beyond CMOS devices

Source: William M. Holt, Moore's Law: A Path Forward, Keynote at International Solid-State Circuits Conference, February 2017, http://isscc.org/videos/2016_plenary.html



Figure 110: Organic and Printed Electronics Roadmap

Source: OE-A

2.5.3. THE NON-SILICON ROADMAP

A number of technologies have been studied in an effort to look beyond the CMOS (Silicon) roadmap. In the following sections, we will show few of them (of course non-exclusively), like printed electronics, carbon-based technologies, memories based on new physical phenomenon, quantum computing.

2.5.3.1. PRINTED ELECTRONICS

2.5.3.1.1. The roadmap of the Organic and Printed Electronics Association

The Organic and Printed Electronics Association (OE-A) [301] presents a roadmap (Figure 110) for a printed and organic electronics market consisting of five key application areas: 1) OLED lighting, 2) Organic Photovoltaics, 3) OLED Displays, 4) Electronics and Components and 5) Integrated Smart Systems. The roadmap predicts the evolution of these application areas from short term to medium term and then to long term. 'Integrated Smart Systems' (ISS) focus on the manufacturing of smart ubiquitous objects consisting of the integration of printed sensors, transistors, memory, batteries and displays onto one substrate. The roadmap predicts that ISS will appear in the medium term (2019-22) when batteries, memory, logic and sensors can be integrated. More complicated smart standalone monitoring systems are predicted beyond 2023. The roadmap states that truly smart objects are possible if more and more on-board functionality is added to the printed system.

2.5.3.1.2. Market of Printed Electronics

According to market research by MarketsandMarkets [286], the PE market size is forecasted to be worth € 35 billion by 2020 with a CAGR of 34% from 2014 to 2020. IDCTechEx [273] predicts that the total market for printed, flexible and organic electronics will grow from € 24 billion in 2016 to € 62 billion in 2026. Although the majority of the market share is to be OLED displays, as shown in Figure 111, it is expected that logic, memory and thin film sensors will have huge growth potential by 2026.

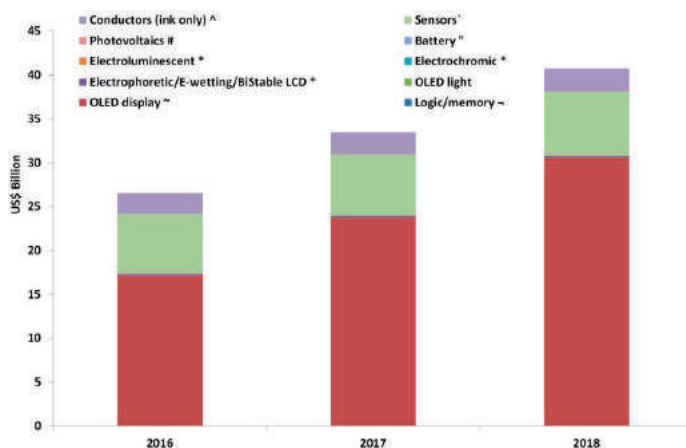


Figure 111: Growth prediction of printed, flexible and organic electronics by IDTechEx

Source: IDTechEx

2.5.3.1.3. Printed sensors and logic

Printed Electronics (PE) [370] offers cost-effective fabrication of electronics with low cost substrates (e.g. plastic, paper), materials and simpler processing/patterning steps. PE has found applications in sensors, RFIDs, solar cells, batteries and displays in the medical, automotive, human-machine interfaces, mobile computing platforms and embedded systems fields [3].

Sensors such as pressure [369], temperature [1], image [280] and biosensors [361] have been successfully manufactured using PE technology. Printed sensors are generally integrated into systems using 'conventional' (i.e. Si-based) interfaces (i.e. analog and digital electronics for amplification, signal conditioning and readout). However, these complex Si-based interfaces will be unable to deliver the volume and cost reduction demand envisaged in the Trillion Sensors Vision [193].

There has not been a strong presence of printed processors because PE technology today does not allow high speed and high-density digital circuits similar to Si technology. Recently, a research prototype of a thin-film 8-bit digital microcontroller on a plastic substrate has been implemented to demonstrate the viability of integrating a printed processor into an IoT device [355]. However, PragmatIC [311] is today able to routinely manufacture flexible ICs with complexity of >1,000 gates and is now prototyping circuits with up to half a million individual devices. PragmatIC's current technology node (critical dimension/CD) is 1µm and is expected to reduce to <0.5µm by 2018. In parallel, p-type oxide materials are being developed to support full flexible CMOS integration. Circuits operate at high frequency (>10MHz) with production yields greater than 90% within an onset voltage spread of +/-0.2V. Devices are stable for months under storage at 60C/90% rh. Improvements enabled by a new fully integrated production line are expected to bring significant improvements in reducing defects due to handling and the use of dedicated process equipment.

2.5.3.1.4. Printed Integrated Smart Systems (ISS)

Although there are many hybrid integrated (i.e. printed + Si) smart systems in the literature [358, 360], there are not many printed ISS. Thin Film Electronics ASA built the first proof-of-concept of an integrated electronic tag using a printed rewritable memory, logic and temperature sensor that can detect temperatures and record them in the memory [191]. Building on this printed integrated system, Thin Film Electronics ASA later demonstrated another integrated system combining printed temperature sensor, comparator and NFC in a smart label [192]. The smart label is attached to a surface measuring the temperature, and the temperature data can be read through the NFC tag using a smartphone.

Another example of smart packaging is a smart label on beer bottle which lights up when pressed [313]. The smart package contains a flexible pressure switch, LEDs, thin film batteries and printed integrated circuits. The intelligent circuit was integrated within the label substrate where the thumb naturally falls while holding a beer bottle. Once pressed, the LED lights begin to shine

through the eyes of the mask on the front of the bottle. The University of Tokyo [373] built a printed alarm armband that monitors the vital signs of hospital patients. The flexible armband contains a solar panel, piezoelectric speaker, temperature sensor and power supply circuit all of which are organic components in a wearable form factor. It is self-powered by a solar panel and the speaker sounds an alarm when the temperature sensor measures a temperature between 36.5C to 38.5C.

2.5.3.1.5. Impact and proposed course of actions

Today, PE manufacturing uses moderately expensive and large equipment, and we predict that PE technology, particularly for integrated circuits (ICs), will track a similar trend to that of optical disc manufacturing which has evolved from manual, batch-based cleanroom production to fully-automated production in a self-contained module as shown in Figure 112. These modules cost only a few million dollars, which is three orders of magnitude lower than a Si fab. Only a few giant semiconductor companies can afford to own Si fabs, and most of them are geographically located in Asia to reduce operating costs.

For PE, self-contained modules will be small and cheap enough to be owned by smaller companies, research institutes and even consortia of universities. These modules can be located in Europe due to the fully automated nature that reduces the operating costs. The projected production times of future printed ICs will be <1hr, in contrast to 7 days at present and to 8-12 weeks for Si (excluding design effort).

PE technology will enable low-cost customization per specific application even in low volumes. This is because PE uses largely additive process that has fabrication process and material costs advantages, and therefore can provide cost efficient customized products in low volumes [21]. The combination of affordable PE fabs with customization toolflow will eventually enable printing rapid and low-cost prototypes for flexible smart electronic devices by SMEs and even universities. In the most conservative case, SMEs/universities could buy a consumable (e.g. plastic sheet with embedded PE) and then print custom parameters digitally in order to configure it for their target product/prototype like for FPGA. In a more optimistic case, PE allows configuration to be changed even per device (thus allowing an optimization of the printed system to the particular task) and even for the purposes of adaptation to its own variability (e.g. adaptation to the exact characteristics of a sensor system).

In the next ten years, users will personalize the sensing and intelligence of wearable/IoT devices by selecting sensors and their interfaces, and by attaching them directly to customized processing (e.g. by neural networks) adapted to intelligent applications. The entire electronic system, including battery, will be printed on a low-cost and flexible substrate in full automation. This will have unprecedented effects on the electronic/computing industry and research communities such as rapid prototyping of custom printed products, faster time-to-market for SMEs, low-cost research



*Figure 112: Future projection of PE from manual cleanroom production to self-contained full-automated module
Source: LETI*

prototyping and testing of novel ideas. This could have a similar impact on society to that of 'Fab Labs' and 3D printers, allowing small groups or even individuals to express their creativity in realizing real and usable systems composed of printable elements [234].

2.5.3.2. NON-VOLATILE MEMORY TECHNOLOGIES

Non-volatile memories offer a number of very interesting properties that can be used in future computing systems. These novel memory technologies can be classified based on the physical processes implementing the data storage and recall process. Figure 113 shows a high-level taxonomy.

Of all the technologies shown in Figure 113, MRAM and PRAM devices are the most advanced from an industrial point of view. MRAM was introduced in 2004 by Freescale and, since then, the

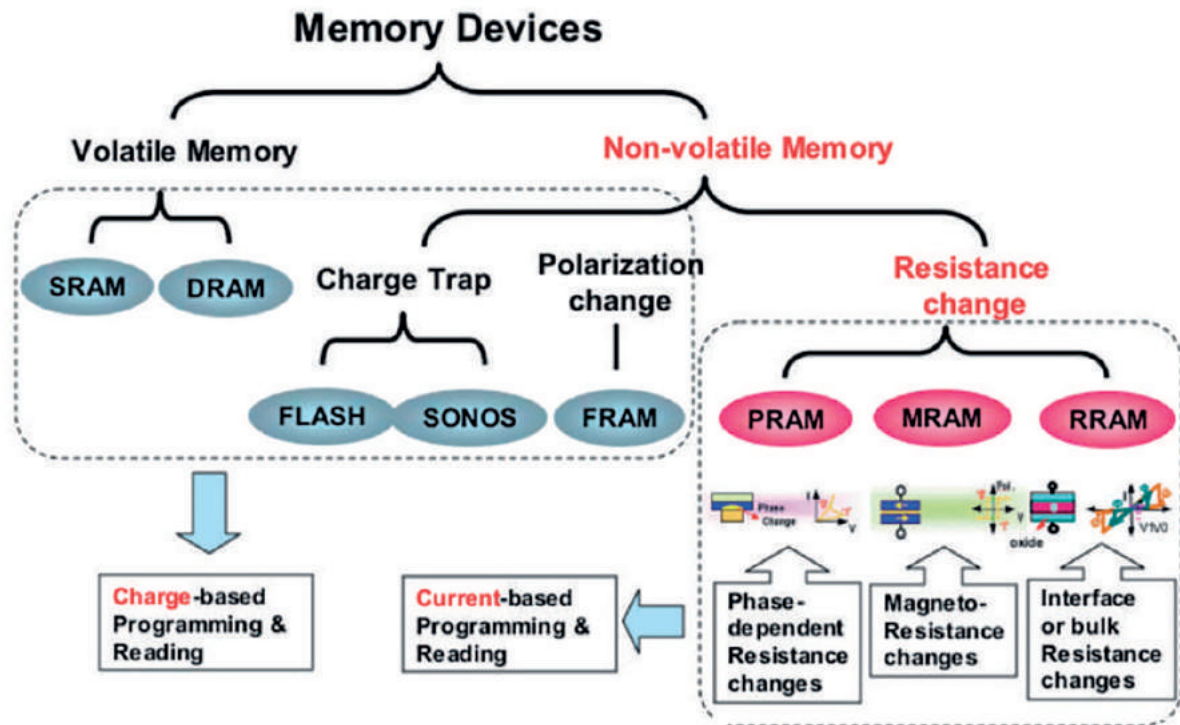


Figure 113: Various kinds of non-volatile memories
 Source: Y. Pershin and M. Di Ventra

technology has been refined to the point at which Everspin was able to introduce a 64 Mb ST-MRAM chip. PRAM is currently in active development and industrial products are on the way with Micron just announcing a 1Gb PRAM module for mobile applications. Resistive RAM (RRAM) comes in many technological flavours such as Conductive-Bridge or Oxide Ram; they are actively developed by industrial companies such as Adesto technologies and Hynix, and may prove to be the best option in the future for very dense arrays of non-volatile memory.

Since some novel resistive devices are dipoles (i.e. they are two-terminal devices) for which resistance can be controlled by applying voltage levels, the idea to organize them in crossbars comes naturally with the expectation of getting ultra-dense arrays of non-volatile memory bits. Although using the crossbar scheme with devices of nano-metric sizes can provide dramatic memory densities, doing so requires many problems to be solved: select logic, sneak paths, and process variability, to name but a few. However, a convincing demonstration of a 2Mb resistive device crossbar without selection logic has been unveiled recently by Hynix and HP [20]. This work shows that putting those devices to practical use in digital design would be possible at the price of rethinking the whole architecture of the memory plane.

Up to now, practical demonstrations have targeted either reconfigurable logic or full arrays of memories. Another very promising application of non-volatile memories could be to use them in the design of processor register files with the aim of building instant-on/instant-off CPUs. HP announced an architecture concept, called 'The Machine' [204], which uses memristors for storage

and photonics for the interconnect. The architecture of 'The Machine' has not been disclosed, but a new consortium has been formed to design and publish its redesigned memory hierarchy [381].

Non-volatile memories will also have an impact on software, and more particularly on the OS. HP calls for the participation of academia in the development of a new OS for its architecture using non-volatile memories. According to the HP roadmap, memristor memories (in the form of DIMMs) shall be launched in 2016. Similarly, Toshiba plans to industrialize 3D ReRAM in 2020: 'But in and after 2020, we will need a new memory having a different mechanism. ReRAM (resistive random-access memory) and ion (magnetic wall displacement type) memory are candidates. We are also considering the manufacture of those new memories by stacking layers three-dimensionally, and they can possibly be combined with scaling beyond 10nm' [189].

For the reader interested in this burgeoning topic, a valuable review on emerging memory technologies can be found in [380].

2.5.3.3. CARBON-BASED TECHNOLOGIES

Carbon-based materials have long been proposed as a candidate technology for electronics beyond CMOS. Starting at the turn of the millennium with works on fullerene [365] and carbon nanotubes [379] transistors, carbon based technologies for electronics have so far not yet really emerged as feasible solutions. However, two potentially disruptive technologies in this context are graphene and graphyne [60] (Nobel Prize in Physics, 2010) transistors, which seem capable of increasing their clock frequency beyond the capabilities of silicon transistors (in the 100 GHz to THz

range). Currently such transistors are significantly bigger than silicon transistors and only limited circuits have been implemented. Their current application scope is mainly fast and low-power analogue signal processing, but research on graphene and graphene transistors is still in its infancy. The computing industry may shift back to the frequency race from the parallelism race if complex devices running with low power at 100 GHz become possible [266].

2.5.3.4. QUANTUM COMPUTING (QC)

The idea of quantum computing has been put forward some time ago. The concept dates back to Feynman [15] with great progress at the end of the 20th century by Di Vincenzo, Shor, Bennett and many others [366]. However, the topic has become very trendy since recent results seem to turn what looked like a nice but hypothetical idea into something that might be feasible after all.

2.5.3.4.1. Unitary QC vs. simulated QC

We can roughly classify QC architectures in two categories:

1. Unitary QC in which each QuBit is controlled through operators. This looks like a quantum version of a digital ISA-based computer. Such a concept requires a long coherence time (at least long enough to perform useful computations) and needs quantum error correction to mitigate the inevitable decoherence;
2. Simulated QC in which a collection of QuBits is implemented through a physical substrate, and whereby we are only interested in the collective (hopefully quantum-assisted) behaviour. A good example of such a concept is a quantum annealer such as that proposed by the D-Wave company. This concepts of QC seems readily feasible but the potential acceleration factor is not yet very clear.

2.5.3.4.2. Recent breakthroughs

The first circuits based on Si-28 implementing a two-qubit logic gate were demonstrated by [375] at the University of South Wales. The compatibility of such circuits with conventional CMOS process has been demonstrated by LETI.

It has been recently demonstrated a small programmable QC based on atomic qubits [11].

2.5.3.4.3. Post-quantum cryptography and machine learning

One of the main drivers of the recent interest in quantum computing comes from the application side and in particular from the computing requirements of two classes of applications: cryptography and machine learning. These application fields are indeed at the heart of many computing challenges ranging from security and privacy to artificial intelligence and autonomous systems.

Quantum cryptography was proposed in 1970 [378] and developed in the 1980s [4]. It proposes a secure way for the problem of exchanging keys: quantum key distribution. A method for secure

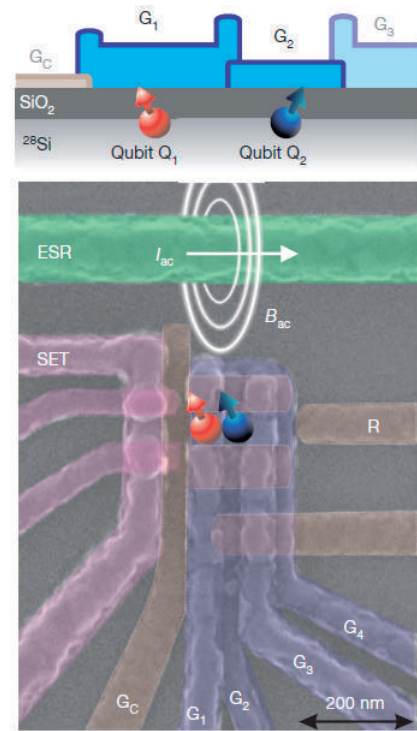


Figure 114 - A two qubit logic gate implemented in Silicon
Source: [375]

communication, called BB84, has been developed from this work and is now used in commercial products (ID Quantique, MagiQ). The announcement in 2015 by the US National Security Agency (NSA) that it 'will initiate a transition to quantum resistant algorithms in a not too distant future' [367] has boosted a number of research studies into what is known as post-quantum cryptography.

Machine learning and in particular deep learning have seen rapid progress in the last two years. However, to tackle bigger and more realistic cases, deep-learning algorithms face a hard computing constraint on the learning side. High-performance computers with the help of GPU accelerators often run for days or weeks with very deep networks and extremely large learning sets. Since learning algorithms are basically highly dimensional optimization problems requiring careful descent in very complex phase spaces, quantum computing could theoretically solve this much more efficiently.

2.5.3.4.4. Programming the quantum

In the event that QC hardware becomes available, it is clear that the resulting machine will be hybrid. It will combine a quantum engine with a classical digital computer. The program that would run on such a machine will need to combine at least two computing models: a classical part, to prepare data and process results and a quantum part. This will require a tight connection between the two programming models.

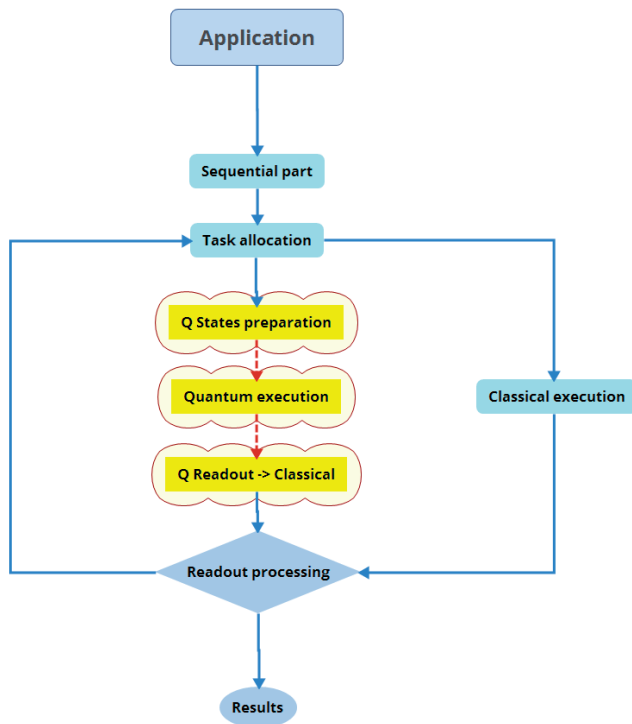


Figure 115: Programming the quantum
Source: Valiron

Some preliminary ideas have been put forward to tackle this problem [374], but there is still a lot of research and development to be done.

2.5.3.4.5. The European quantum computing initiative

A European team has been working on a ‘Quantum Manifesto’ to formulate a common strategy for Europe to stay at the forefront of the ‘Second Quantum Revolution’. This document, released in May 2016 [408], is the foundation for the € 1 billion Flagship initiative. For the reader interested in the implication of quantum computing in Europe and its implication on policies, see the report “Quantum technologies: implication for European policy” [62].

2.5.3.5. IMPACT AND PROPOSED COURSE OF ACTIONS

The ITRS 2015 document on emerging research devices [352] gives a taxonomy for emerging information-processing devices and is a very complete document on new technologies that can be used for building new computing devices. We encourage the reader who wants to have more details on emerging technologies to read it. Note that ITRS 2015 will be the last issue of ITRS as we all knew it. The successor of ITRS, the International Roadmap for Devices and Systems or IRDS, will focus on functions and systems rather than on processes and technologies like the old ITRS.

2.5.4. THE ARCHITECTURE ROADMAP

2.5.4.1. ARCHITECTURES FOR PREDICTABLE COMPUTING

Taking into account all the necessary aspects to overcome the software crisis and solve the various related challenges (see section

2.5.7.2) will mandate a fresh view on ICT systems development, at both the conceptual and tooling levels. Holistic approaches are required (see section 2.5.7.2.5), as well as transparency across all levels, and most likely new paradigms (see section 2.5.1).

As a consequence, new hardware and software architectures will have to be designed both for ICT systems and for ICT systems development tools. The need for predictable computing, for various criteria (time, energy, etc.), pushes for ‘re-engineering computing’ for formal methods.

An example of this is the DeepSpec [229] Expedition in Computing project financed by NSF (US\$10 million over five years) that focuses on the specification and verification of full functional correctness of both software and hardware. Their ambitious and holistic goal is to connect, at specification interfaces, various conceptual and technical bricks (specifications, compilers, model checkers, certification tools) to provide development environments to build whole ICT systems while being able to prove end-to-end correctness, across all the software and hardware layers. These development environments should make it possible to work ‘with specifications that are both precise, that is, integrated with the code they describe, and live, that is, continuously and automatically checked, and that can express rich descriptions of correct system behaviours’.

Another foray in this direction is the PRET project [70] and Precision Timed (PRET) machines. This project, focusing on predictable timing for embedded systems, relied on microarchitecture and memory system design to achieve precise and repeatable timing of software with no loss of aggregate performance. The project showed that that timing predictability and repeatability were not at odds with performance, and provided architectural optimization techniques that were carefully selected to deliver performance enhancements without sacrificing timing predictability and repeatability. The project’s approach included extending the instruction-set architectures (ISA) with control over execution time, a way to reach a more transparent and holistic control of the ICT system.

Another example of an effort for predictable computing that is closer to industry is European company Kalray’s MPPA family of manycore processors [283]. MPPA evolved from the accelerator domain to a standalone processor aiming to fulfil the goals of predictable computing while offering the performance of a manycore processor.

The Kalray MPPA processors, implementing a massively parallel architecture on a single 28nm CMOS chip, are designed to ensure predictable and certifiable response times, by combining a three-level architecture based on 5-issue VLIW cores, cluster of cores with a shared local memory, and the association of compute clusters with I/O clusters (see Figure 116).

Kalray MPPA-256 Bostan processor integrates 16 compute clusters with 17 VLIW cores and 2 I/O clusters with two quad-cores, for a total of 288 VLIW cores. The 18 clusters are interconnected through a 2D torus network-on-chip which is RDMA capable, and also terminates 8 Ethernet links at 10 Gbps each. According to

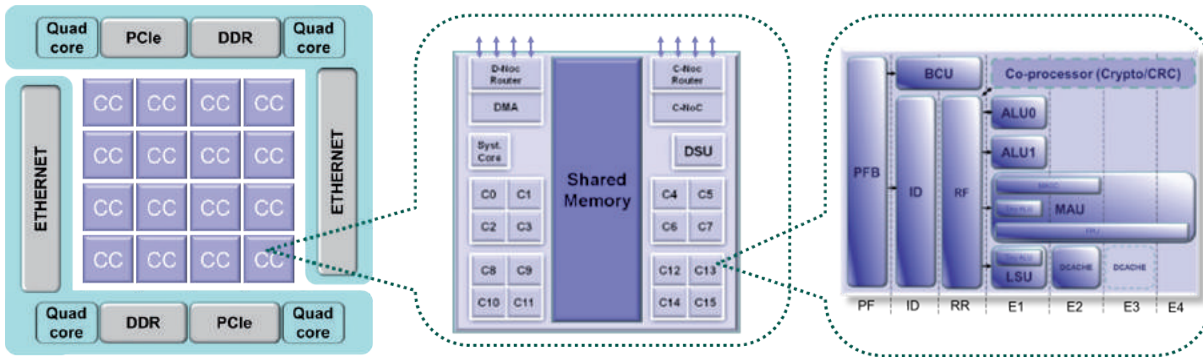


Figure 116: Kalray MPPA manycore architecture
Source: Kalray

Kalray, the processor operates at 600 MHz while consuming less than 20W, with peak floating-point performance over 900 GFLOPS for single-precision (450 GFLOPS for double precision).

Kalray claims ‘the VLIW core implementation is fully timing compositional, while its cache and write-buffer feature a strict LRU eviction policy’, which enables ‘static timing analysis tools to report accurate results. On the compute clusters, the local memory is implemented as 16 independent banks connected to the 20 bus masters using dedicated busses and round-robin arbitration. Address mapping can be configured as interleaved (every 64 bytes) or blocked (128 KB). The former mode is used for high performance, while the latter is for high integrity as software may allocate one bank to each core and control interferences through a model of computation. The network-on-chip supports guaranteed services through the configuration of rate and burstiness parameters at ingress. These configurations are obtained by the application of network calculus in a way similar to AFDX (Avionics Ethernet). Finally, the DDR controllers have a flexible address mapping scheme and other tuning parameters that enable to exploit DDR memory in a composable way with guaranteed throughput.’

Kalray targets its MPPA platform and the associated software platform towards both number crunching applications and control-command applications.

We recommend that research efforts be supported in the direction of predictable computing, both at the formal and at the development tool level.

2.5.4.2. ACCELERATORS

As a way to help cope with the ever-more diverse and numerous applications that run in today’s and tomorrow’s computing systems, especially in the CPS world (see section 2.4.9), and to help cope with the data deluge (see section 2.5.7.2.4), systems that are more and more heterogeneous will appear, featuring a wide range of dedicated sub-systems, especially accelerators.

2.5.4.2.1. Reconfigurable accelerators

Reconfigurable computing provides several interesting means of designing accelerators.

The most known and widespread is probably the FPGA (Field-programmable Gate Array). ‘FPGAs contain an array of programmable logic blocks, and a hierarchy of reconfigurable interconnects that allow the blocks to be ‘wired together’, like many logic gates that

can be inter-wired in different configurations. Logic blocks can be configured to perform complex combinational functions, or merely simple logic gates like AND and XOR. In most FPGAs, logic blocks also include memory elements, which may be simple flip-flops or more complete blocks of memory’ [93]. FPGAs offer far more flexibility than ASICs, since there are reprogrammable, but there are slower and more energy-hungry (see Figure 117).

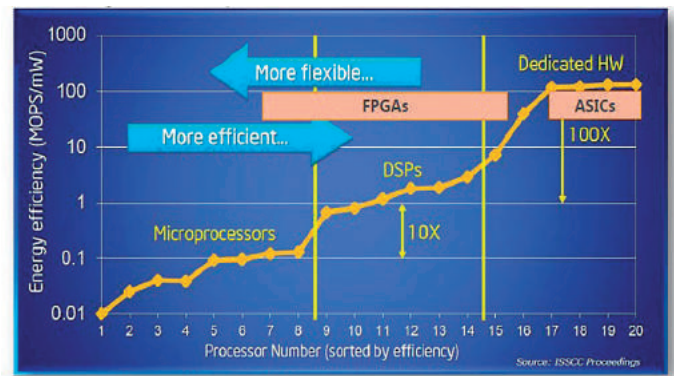


Figure 117: Compared energy efficiency of processors categories
Source: Bob Broderson, Berkeley Wireless group

An example of using FPGAs for acceleration is Microsoft’s Catapult project to speed up its Bing search engine [242]. There, FPGAs were used to (hard-)code part of the search algorithms (see Figure 118), leading to very significant improvements (see Figure 119), while retaining the ability to change and improve the algorithm over time within the system.

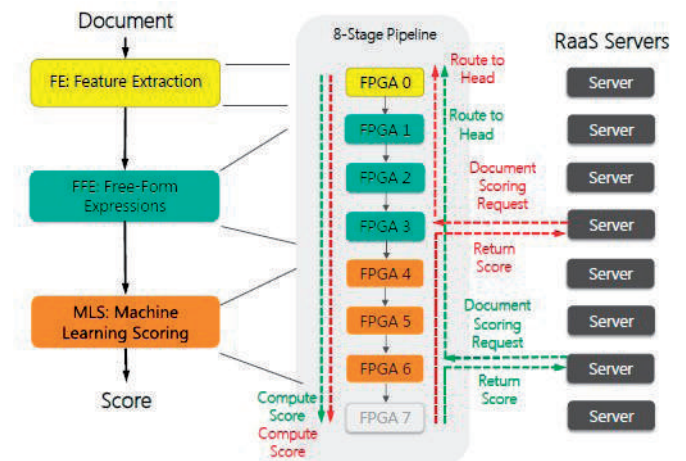


Figure 118: Architecture of Microsoft Catapult
Source: 2016 EnterpriseTech.

1,632 Servers with FPGAs Running Bing Page Ranking Service (~30,000 lines of C++)

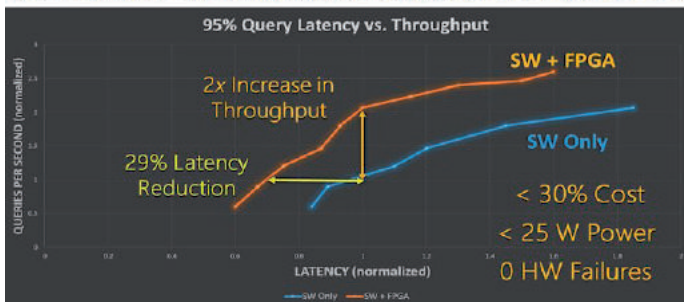


Figure 119: Performance improvements brought by Microsoft Catapult

Source: 2016 EnterpriseTech.

An alternative evolution to FPGAs is CGRAs (Coarse Grain Reconfigurable Arrays) [224], which fill a gap between ASICs and FPGAs (see Figure 120).

- Coarse-Grained Reconfigurable Array (CGRA)
 - ASIC-like performance
 - Higher flexibility than ASIC

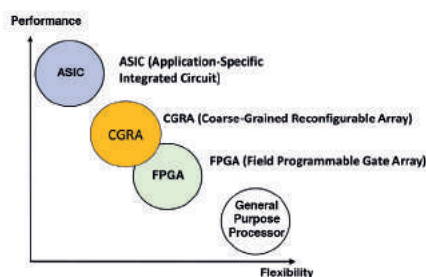


Figure 120: Compared performance and flexibility for FPGAs, CGRAs and ASICs

Source: Jeremy Lee, University of New South Wales

A CGRA is basically a reconfigurable array of processing elements interconnected by a 2 mesh-like network. Compared to FPGAs, CGRAs typically are reconfigured at a coarser, operation or functional level, while FPGAs tend to be reconfigured at low, bit or operation level. This makes it possible to avoid the two main typical weaknesses of FPGA (see Figure 121): slow performance due to the fact a lot of interconnect has to be passed through before getting to the appropriate logic; and the fact that only a small fraction of the die actually does processing (most is interconnect and memory). In a CGRA, interconnect is much smaller, having to serve coarser blocks, which saves area and time.

What is Holding FPGAs Back?

Honeywell

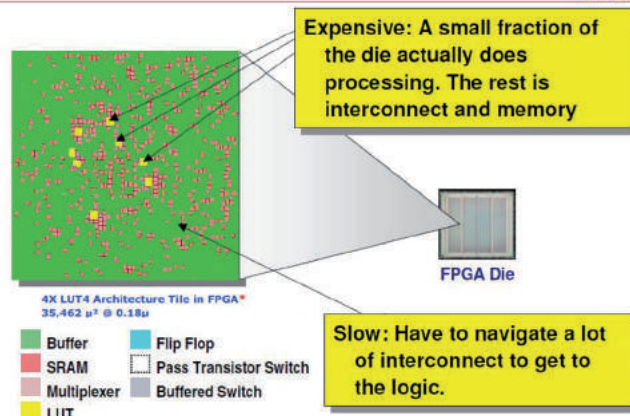


Figure 121: Honeywell "Radiation Hardened Field Programmable Object Array (FPOA) for Space Processing," MAFA 2007
Source: Jeremy Lee, University of New South Wales

Compared to ASICs, CGRAs are a bit slower, but much more flexible, while ASICs are completely predefined and non-reconfigurable (see Figure 122). As a consequence, the relative pros and cons of ASICs, FPGAs and CGRAs can be summarized as in Figure 123:

CGRA Operation Example

- CGRA example
 - System-on-Chip Platform for MPEG encoder

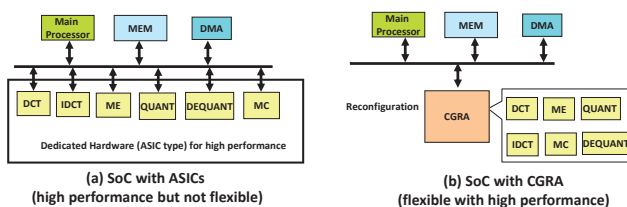


Figure 122: Compared architectures of a SoC with ASICs and a SoC with CGRAs

Source: Jeremy Lee, University of New South Wales

Technology	Example suppliers	Development costs	Time to market	Die size	Performance	Reconfiguration
ASIC	Qualcomm, Samsung	High	Slow	Low	High	Low
FPGA	Xilinx, Altera	Low	Fast	High	Low	High
CGRA	PACT, MathStar	Low	Fast	Medium	High	Medium

Figure 123: Relative pros and cons of ASICs, FPGAs and CGRAs
Source: Jeremy Lee, University of New South Wales

CGRAs have shown very promising results, especially in terms of low-power properties, and are, for example, used by Samsung in a flexible video processing platform for 8K UHD TV [268], or in the ULP-SRP: Ultra Low-Power Samsung Reconfigurable Processor for wireless sensor networks that Samsung is pushing for biomedical applications [8].

The market of accelerators is changing quickly these days. Intel acquired the FPGA maker Altera in December 2015, in a move to boost its data centre and IoT capacities [347].

On the high end of the market, Intel first plans in 2016 to begin selling products with a Xeon chip and an Altera FPGA in a single package, hoping to bring a 30% to 50% speed improvement over using processors and FPGAs separately. Intel then intends to provide a single chip packing both traditional processor and FPGA circuitry, hoping to double the performance. This technology should help support *'rapid virtualization of multiple VMs through one CPU while simultaneously running Network Function Virtualization (NFV) which is touted as the future of network security and stability. Put simply, the dynamic capabilities that FPGA technology contains could help reinvent the processor and, subsequently, the data centre as we know it.'* [217].

On the embedded and IoT market, Intel plans to integrate its small Atom chips with FPGAs, targeting new areas such as automobile electronic systems.

Reconfigurable accelerators, especially CGRAs, are thus a very active and promising area. We recommend that research efforts be supported in this area, and stress the need for cooperation with EU industry.

2.5.4.2.2. Accelerators for neural networks

Another booming area lies in accelerators specifically designed for neural networks. These address the requirements brought by the deep learning push in artificial intelligence (see section 2.4.11). Historically, neural networks were run on general purpose CPUs. Then the GPUs, with their increase performance and massively parallel architecture, came into use for neural network computation, regardless of the GPUs' initial targeting of graphics. Now, new architectures specifically designed for neural network computation are appearing. One kind is what are now starting to be referred to as Tensor Processing Units (TPUs) [109]. *'Tensor processing units are application-specific integrated circuits developed specifically for machine learning. Compared to graphics processing units-they are designed explicitly for a higher volume of reduced precision computation (e.g. as little as 8-bit precision), and lack hardware for rasterization/texture mapping'.*

The name-giver of tensor processing units is Google's TPU (Tensor Processing Unit), an ASIC specifically built for speed for machine learning (see Figure 124). It is customized to give high performance and power efficiency when running Google's TensorFlow widespread Open Source framework for machine learning [174]. Google announced the TPU in May 2016. Little details are known about this TPU [71, 330]. It has an instruction set, is tolerant to reduced computational precision, delivers an order of magnitude better-optimized performance per Watt for ma-

chine learning than all commercially available GPUs and FPGAs. It is now used in Google's data centres, for map creation, navigation, voice recognition and picture processing, and it powered the AlphaGo computer that defeated the world Go champion. Some elements (lack of apparent large memory) points to the TPU doing the inference (i.e. exploitation of a trained neural network), with the initial training itself being performed outside and off-line, probably by GPUs, then implemented in the TPU.

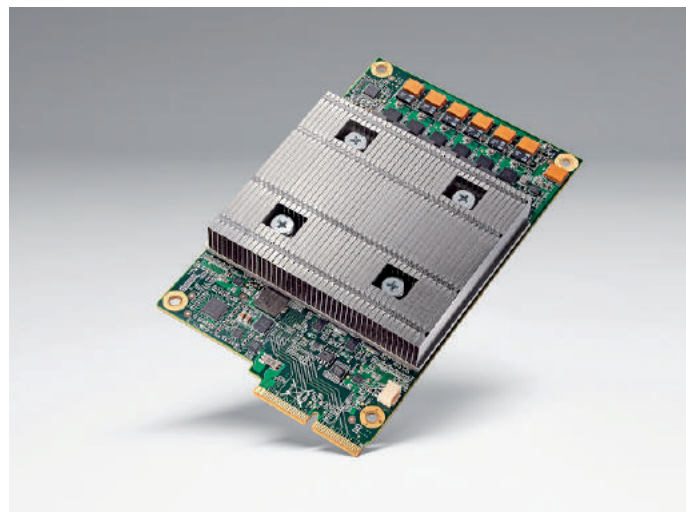


Figure 124: Google's TPU (Tensor Processing Unit) deep learning accelerator

Source: Google

IBM provides more information about its own TrueNorth brain-inspired neural net chip [110, 316]. TrueNorth features 4096 processors, each of which has 256 integrate-and-fire spiking neurons each with 256 inputs. That represents a total number of programmable synapses over 268 million (2^{28}). Each processor can compute all 256 neurons 1000 times per second. Overall peak performance reaches 46 GSops (billion synaptic operations per second). Power consumption is around 70mW. IBM have designed an end-to-end ecosystem complete with simulator, programming language, integrated programming environment, libraries, and teaching curriculum (called "SyNAPSE University"). The architecture aims to solve a wide class of problems from vision and audition to multi-sensory fusion, since it should be able to efficiently process high-dimensional, noisy sensory data in real time, while consuming orders of magnitude less power than conventional computer architectures. Over time, IBM hopes that SyNAPSE will become an integral component of the IBM Watson group offerings.

Intel, too, has been entering this field, in particular through acquisitions. They bought Nervana Systems for its Nervana Engine [510], and also Movidius [289], which specializes in computer vision systems and provides solutions based on holistic hardware-software cooperation. Movidius has designed the Myriad 2 hardware Vision Processing Unit family (see Figure 125).

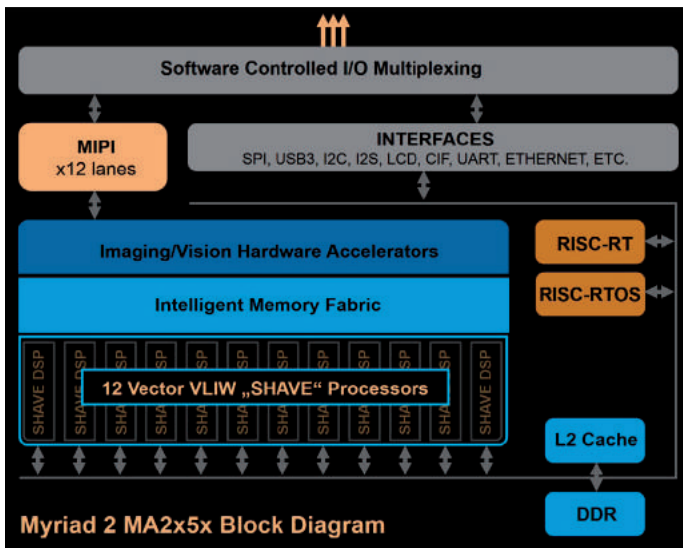


Figure 125: Intel Movidius Myriad 2 architecture
Source: Movidius Inc.

The design principles for Myriad 2 relied on an increase in the number of programmable vector-processors and additional dedicated hardware accelerators. As a Vision Processing Unit (VPU) System-on-Chip (SoC), Myriad 2 has a software-configurable multi-ported memory subsystem and caches (2MB of on-chip memory and 256 MB of L2 cache), providing data and instructions to twelve SHAVE processors, two 32-bit RISC processors and video hardware accelerators, at 400 GB/sec of sustained internal memory bandwidth, according to Movidius. Power management features include 20 power islands and low-power states, with nominal 600 MHz operation at 0.9 V. The SHAVE proprietary processor itself is a hybrid VLIW 128-bit vector stream processor with 8/16/32 bit integer and 16/32 bit floating point arithmetic, and hardware support for sparse data structures.

This platform has a number of features that make it quite appropriate for neural network implementation, such as performance, on-chip RAM and flexible precision. Movidius thus recently extended the reach of this platform to neural network acceleration, by providing the Fathom machine learning software framework. Fathom parses existing trained offline neural networks from TensorFlow or Caffe and converts them in an optimized way for the Myriad 2 VPU. In an additional interesting move aiming at putting neural network accelerations to the masses, Movidius announced [290] the Fathom Neural Compute Stick, that is an USB stick containing an embedded neural network accelerator based on the Myriad 2 platform. It can be used as a prototyping tool for neural network profiling and evaluation. It also makes it possible to easily connect it to a wide range of devices and enhance significantly their neural computing capabilities, giving them, for example, computer vision capabilities.

We believe the move of major ICT companies (Google, IBM, Intel etc.) towards accelerators for neural networks is an important shift in computing, and is likely to very significantly boost neural network, deep learning and AI usages in the next few years (see section 2.4.11).

We thus recommend supporting research in the design of neural network accelerators, that may be a key strategic asset in the future, stressing the need for cooperation with EU industry in this domain too.

2.5.4.2.3. Quantum accelerators

Finally, the high hopes brought by quantum computing and the first quantum computer such as those of D-wave Systems [92, 233] may lead to significant acceleration of at least some specific class of problems. Although this horizon is still in the future, the first quantum machines are beginning to be used, e.g. at NASA and Google.

2.5.4.3. IN-MEMORY COMPUTING

In-memory computing, or in-memory processing [96], is another simple way of accelerating computation. It consists of keeping data in the live memory (e.g. RAM) of a server rather than on external, much slower disks. In-memory computing has become possible since the price of RAM memory has significantly decreased, and 64bit-OSes, that allow addressing much larger memory sets, have become common.

In-memory computing benefits from much lower latencies and higher transfer speeds, hence helping cope with the data challenge (see section 2.5.7.2.4). In-memory computing is especially useful where large amounts of data are to be processed, such as in many big data, business intelligence applications.

Developing in-memory computing, through specific applications, frameworks and strategies can thus be efficient and cost-effective way of improving the performance of ICT systems, since most of the infrastructure investment has already been made. We thus recommend supporting research in these domains.

2.5.4.4. IMPACT AND PROPOSED COURSE OF ACTIONS

Predictable computing is becoming necessary to overcome the software crisis and solve the various related challenges. Holistic approaches are required (see section 2.5.7.2.5), as well as transparency across all levels. We thus consider it important to support research efforts in the direction of predictable computing, both at the conceptual and tooling levels, including architectures.

Accelerators are becoming a major and booming area and the field is likely to expand in the near future. Indeed, they provide an effective way of answering the ever-increasing demand for more computing power, especially with regards to AI, as well as a way to somewhat limit the increase in complexity. We believe accelerators are going to be a key strategic asset in the future. We thus recommend supporting research, with strong EU industrial involvement, on accelerators, including CGRAs and neural network accelerators, as well as research on how these can be smoothly integrated in software and applications development processes.

We believe that developing in-memory computing, through specific applications, frameworks and strategies can also be an efficient and cost-effective way of improving the performance of ICT systems, especially those related to big data. We thus recommend supporting research in these domains.

2.5.5. EVOLUTION OF MASS STORAGE

2.5.5.1. MAGNETIC DISKS

Bits on a hard disk are stored in sets of magnetic grains. A magnetic grain is about 8nm, and it cannot be made much smaller because super-paramagnetism will cause random flips of the magnetic grains under the influence of temperature. One stored bit consists of 20-30 grains and has a bit width of 75nm and a bit length of 14nm. The number of grains cannot be reduced much if we want to keep a sufficient signal-to-noise ratio. Therefore, the maximal density of perpendicular recording is about 1 Tb/in². Today, hard disks with a density of 1 Tb/in² are commercially available.

The bit density can be further increased by reducing the bit (track) width. The idea is that a track is written full-width, but the next track partially overwrites the previously written track (just like shingles on a roof, hence the name 'shingled magnetic recording'). The remaining strip of the track is wide enough to be read, but it cannot be written anymore without destroying the data in the neighbouring tracks. This leads to disks where data must be stored in bands. These hard disks have to be used like solid-state disks; bands must be written sequentially and cannot be changed, they can only be overwritten. The fact that changes require more work turns out to be not that problematic because many contemporary data is write-once (like images, movies, audio files). Shingled magnetic recording increases areal density about 25% [252]. In 2016, major hard disk vendors introduced helium-filled hard drives. Helium is seven times lighter than air, and creates less friction and less turbulence inside the hard disk, and hence less heat. This allows for higher rotational speeds (10,000 rpm) and 50% more platters in the same volume, increasing both the bandwidth and the capacity of the hard disk.

Narrower tracks lead to more interference from adjacent tracks when reading. Two-dimensional magnetic recording improves the signal-to-noise ratio by using multiple read heads: one to read the central track, and two heads to measure the interference from neighbouring tracks. By combining the three signals, the signal-to-noise ratio can be improved, and the track density can be further increased. This technology might appear in the market in 2017.

Beyond shingled magnetic recording, other approaches are needed. One approach is energy-assisted magnetic recording, of which heat-assisted magnetic recording is the best known. It uses heat in combination with a magnetic field to record the bits. This, however, requires that a heat spot must be localised on a single track and that the rise and fall times must be in the sub-nanosecond range. Designing such a head is challenging. Heat-assisted magnetic recording has been demonstrated in the lab and could lead to an areal density of 5 TB/in². Such hard disks won't be generally available in the market before 2018.

The next approach is to make use of patterned media. In patterned media, each bit is recorded on a small island of magnetic material, surrounded by a non-magnetic material. In this case, a bit can be made as small as a single magnetic grain (instead of

20-30 grains for perpendicular recording). In order to reach 1 Tb/in² in patterned media, we need to etch islands of 12 nm, which is beyond the resolution of current lithographic systems. That means that patterned media will have to rely on self-ordering. Densities of up to 10 Tb/in² by 2025 seems to be theoretically possible with patterned media, if combined with heat-assisted magnetic recording. However, today, bit patterned media is not yet ready for the market. The first prototypes might become commercially available in 2018 [258].

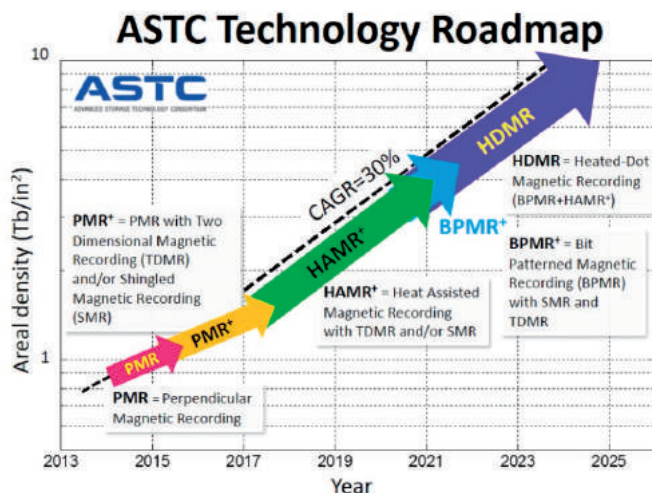


Figure 126: Increase in Areal Density of magnetic recording
Source: IDEMA

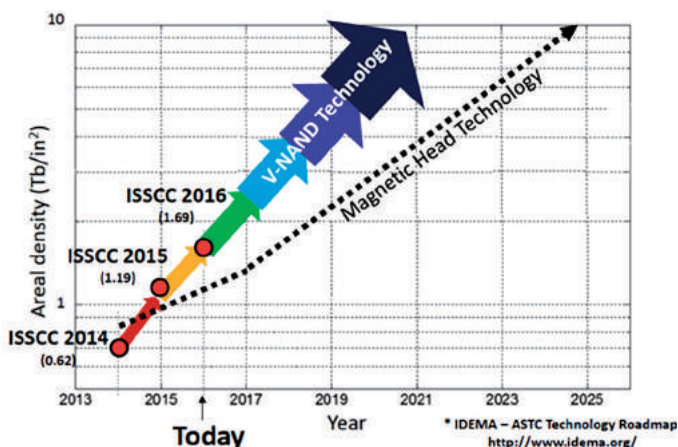


Figure 127: Increase in Areal Density of solid state storage
Source: IDEMA

Today's bits on a hard disk are of the same order of size as a transistor on a chip, but they do not need to be interconnected. This leads to a higher areal density for magnetic recording compared to single layer flash. By storing multiple bits per flash cell, and by stacking multiple flash layers, the effective areal density of flash surpassed the density of magnetic recording in 2016 [222].

2.5.5.2. SOLID STATE DISKS (SSD)

Solid state storage is standard in mobile devices like tablets, navigation devices, smartphones, cameras and so on. Solid state drives are also gradually replacing hard disks. They are available in a similar capacity as hard disks (up to 10-13 TB), but an SSD is still five to ten times more expensive than hard disk storage. However, the cost of an SSD is decreasing fast, and one third of laptops sold already have an SSD. As long as hard disk manufacturers are able to increase the capacity of their disks, it is unlikely that SSDs will reach parity with hard disks. Hence, until then, hard disks will remain the cheapest storage technology for large capacities.

Flash endurance is however much lower than for hard disks. It used to be 100,000 cycles for SLC (cells storing one bit), and 10,000 cycles for MLC (cells storing multiple bits), but that was for 50 nm technology. In today's technology, write endurance for MLC is between 1000 and 3000 cycles, and in the next technology node, it is expected to be reduced even further.

Finally, data retention for flash is 10-20 years for new flash cells. For used cells (5 000-10 000 write cycles) data retention at 50°C is 6-12 months. Even reading stresses the cells in the same block, which means that SSDs that are written only once but read many times, will eventually start losing their data as well.

That means that Flash is at this point in time neither affordable nor reliable enough for (long-term) data storage, and further shrinking beyond 20nm will make it even worse. At best, flash could be used as a cache, not as stable storage in an enterprise storage setting [350].

Possible alternatives for flash are the non-volatile memory technologies mentioned in the section on non-volatile memories. Of all these alternatives, Phase Change Memory seems to be the most promising candidate (Figure 128).

In 2015, Intel and Micro announced 3D XPoint technology. It did not disclose information about the materials or physics, but it is known not to be based on electrons, phase change or memristor technology. Instead, it uses a change in resistance of the bulk material. The latency is claimed to be 10 times lower than flash, and the throughput and durability are claimed to be 1000x better than flash. SSD devices will be commercialized under the name Optane (Intel) and QuantX (Micron). In 2016, IBM announced a triple level cell (TLC) Phase Change Memory challenging 3D XPoint [340].

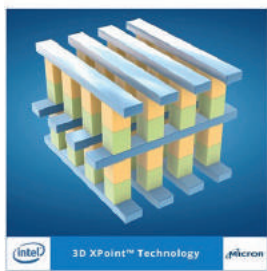
The major problem for a new technology to enter the market is that the current market is a commodity market with huge volumes, very low margins and dominated by less than a handful of large players. This makes it very hard for new ideas to scale up quickly enough to recoup the research costs. An extra difficulty is that existing technologies are still evolving and improving at an exponential rate. So, unless a new technology has some very significant advantage, it will have a very hard time to disrupt the existing storage market. In the near future, the cost of SSD storage will not match that of hard disk storage, but neither will hard disks become as fast as SSDs in the same period.

	Memristor	PCM	STT-RAM	DRAM	Flash	HD
Chip area per bit (F ²)	4	8-16	14-64	6-8	4-8	n/a
Energy per bit (pJ) ²	0.1-3	2-100	0.1-1	2-4	10 ² -10 ⁴	10 ⁶ -10 ⁷
Read time (ns)	<10	20-70	10-30	10-50	25,000	5-8x10 ⁴
Write time (ns)	20-30	50-500	13-95	10-50	200,000	5-8x10 ⁴
Retention	>10 years	<10 years	Weeks	<Second	~10 years	~10 years
Endurance (cycles)	~10 ¹²	10 ⁷ -10 ⁸	10 ¹⁵	>10 ¹⁷	10 ³ -10 ⁴	10 ¹³ ?
3D capability	Yes	No	No	No	Yes	n/a

Figure 128: Phase Change Memory versus other NVM

Source: Aimee Chanthadavong/2016 CBS Interactive

Intel and Micron begin production on new class of non-volatile memory, creating the first new memory category in more than 25 years.



- New 3D XPoint™ technology brings non-volatile memory speeds up to **1,000 times faster than NAND**, the most popular non-volatile memory in the marketplace today.
- The companies invented unique material compounds and a cross point architecture for a memory technology that is **10 times denser than conventional memory**.
- New technology makes new innovations possible in applications ranging from machine learning to real-time tracking of diseases and immersive 8K gaming.

From: http://newsroom.intel.com/community/intel_newsroom/blog/2015/07/28/intel-and-micron-produce-breakthrough-memory-technology

Figure 129: Intel and Micron's non-volatile memory

Source: see link embedded in image

2.5.5.3. IMPACT AND PROPOSED COURSE OF ACTION

Today, there is no clear alternative for magnetic storage, and since any good alternative will require years to reach the market and become mainstream, we will probably experience a slowdown of Kryder's law for storage (on top of a similar slowdown for Moore's law for semiconductors). We are slowed down by the challenges of reducing the physical size of stored bits (due to the limits of lithography), by the limits of reliability at small feature sizes, and by the recording power needed by alternative technologies.

At the same time, non-volatile memory will definitely find its way into existing computing systems and, there, it will lead to the need to revisit software stacks. Likely areas of interest include virtual memory, distributed shared memory, and relaxed memory models without coherence. Increasingly cheaper persistence should push for dramatic changes in I/O design and interfaces, hybrid memory architectures, data-centric execution models, and tying data and computations.

2.5.6. EVOLUTION OF COMMUNICATION

Networks unleash the power of computers. Where unconnected systems can only process locally stored and produced data, computer networks enable the processing of data from any connected storage on any connected computer. Computer networks have shaped the connected society of today, and are key for the future success of the IoT.

Networks come in many different forms, but broadly speaking two general forms can be distinguished: wired and wireless. Wired networks are point to point connections. In terms of energy they can be made very efficient. Tampering with wired connections requires physical contact and is, in principle, but not always in practice, detectable. The construction cost of wired connections directly depends on the distance between the endpoints. Emphasis in wired connections is first on speed, and then on security and energy.

Wireless communications are often used to broadcast information over an area, to an a priori unknown number of receivers. In terms of energy consumption, wireless connections are inefficient: the amount of energy used for transmitting data is independent of the number of receivers, which can be anything from zero up. Tampering with wireless connections is more difficult to detect, eavesdropping is undetectable. The construction cost of wireless connections depends on the distance covered. Emphasis in wireless connections is on all three aspects: speed, security and energy.

Wireless connections are also used for point-to-point connections, e.g. to span physically difficult to access, or even inaccessible, trajectories. In those cases, construction costs are less dependent on distance.

Both wired and wireless connections are used for (terrestrial) network communication distances ranging from millimetres (on one chip) to thousands of kilometres (between continents).

Because of the importance of intersystem communications, the next paragraphs describe their future development from a high-level point of view.

2.5.6.1. WIRELESS COMMUNICATIONS.

Wireless communications are key to the further development of the self-driving car and the IoT.

Wireless systems can be divided based on their point to point range:

- the mobile phone system covering in the order of several square kilometres
- Wireless local area networks: Wifi
- Personal area networks: Bluetooth, Zigbee

The currently rolled-out mobile phone system is widely known as the 4G, or LTE Advanced network. Since 1981, a new mobile phone system has appeared roughly every 10 years. The 4G system, which came into full operation around 2009, has data communication as one of its primary functions. But its data speeds are not up to par for the IoT, being limited to a 1 Gbit/s peak rate.

The 5G system under development at this moment promises 1 Gbit/s to simultaneous users, allowing for higher data rates. 5G

will offer something other than just higher speed: an increased number of connections, higher spectral efficiency (data volume per area), lower energy consumption, lower latencies and lower infrastructure cost. Higher baseband frequencies will result in smaller baseband station area coverage, requiring more wire-connected (or point to point wireless-connected) base stations per unit area, which is only practical in urbanized areas.

5G does not yet exist as a standard: the FCC has opened up frequency bands to allow for development in the US [161]. Europe also recognises the importance of 5G and the EU has presented an action plan to take a leading role in 5G technology development, and will start deploying 5G networks by 2020 [1039]. Several technology development projects have already been started, or even concluded [124].

The communications industry has drawn up a manifesto warning the EU for legislation and rules that may negatively impact the development of 5G [190].

But in all, 5G will put a higher demand on data processing, both because of the higher data rate, but also because of the increased complexity of the required infrastructure. 5G stations have a limited range, requiring a denser network of access points. That denser network requires more intelligence to distinguish between the large set of devices connecting to it.

Due to the existing legacy of 2G (mostly in machine to machine communication: M2M) - an estimated 160 million devices at the end of 2014(!) — providers need to keep older network generations in the air even with a declining number of users. M2M devices have a lifetime exceeding consumer devices, and some providers will even kill 3G before 2G! Note that successive generations of mobile networks are not backward compatible [315].

Security is always an issue in wireless communication and much research is undertaken to constantly develop new fundamental and applied techniques. One fundamentally new technique is quantum communication, applying the principle of entanglement of quantum states to create secure communication. In August 2016, China launched a quantum communications satellite, thereby taking the lead in this technological area. In Europe, fundamental research into quantum communications is ongoing. Europe needs to be at the forefront of this new key technology, otherwise it will become dependent on China and/or the US.

Wireless local area networks (WLAN) are mostly implanted with the technology known as WiFi, based on the IEEE 802.11 standards. Most new developments in WLAN are in extending the available frequencies, to lower frequencies (larger wavelengths) for wider area networks (several kilometres range), and to higher frequencies (smaller wavelengths) for higher data rates. Managing WLAN-based communication systems with a broad spectrum of frequencies, data rates and distances, requires sophisticated network management.

There are a lot of wireless personal area network (WPAN) protocols, for communication between devices such as phones, computers and headsets, etc. Some of these protocols allow for wired communications, e.g. using in-house power lines. Others are carried over infrared links. The best-known WPANs are Bluetooth and

ZigBee for connecting devices such as smartphones and cameras to computers or speakers. These networks will play a pivotal role in the IoT (for example Bluetooth LE - Low Energy-) as they will often form the first link from the device to a local Internet access point. As with WLANs, managing WPAN based communication systems with a broad spectrum of standards, data rates and distances, requires sophisticated network management.

Mobile communications: from 1G to 5G

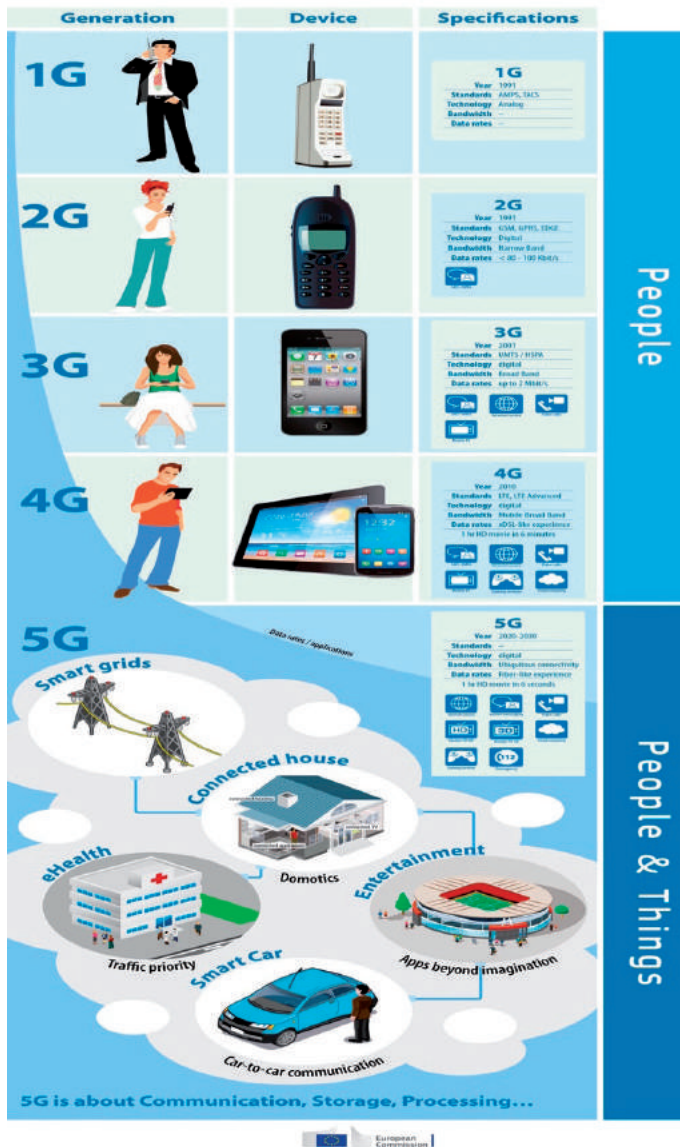


Figure 130: Evolution of mobile communications
Source: European Commission

2.5.6.2. WIRED COMMUNICATIONS

Wired communication links carry the brunt of the data, between data centres and within data centres. Wireless devices are connected to access points that themselves are generally connected through wired (including optical) connections. Overall, the connection speed for consumers has grown on average by about 50% year on year, which is less than Moore’s law, but still an exponential growth. The speed of wired communication technology increased by a factor of ten every ten years. The reason that consumer connection speeds increased much faster is

due to upgrades of to-the-home communication infrastructure (telephone lines, CATV, glass fibre). But just like silicon technology, communication technology is pushing against physical boundaries, and the next ten years will see a slower rate of speed increase [244]. Much of the increase will come from improved data processing techniques, and is therefore closely coupled to silicon technology. Improvements in throughput will come from sophisticated communication link management, such as FlexEthernet [241] In both cases, Europe is well-positioned, through its strength in theoretical computer science.

Technological developments will also aid in increasing the capacity of communication links. One such development is Space Division Multiplexing, the embedding of a number of optical channels in one optical fibre [304].

2.5.6.3. THE TRIANGLE: COMPUTATION - COMMUNICATION - STORAGE

Communication unleashes the power of computers. Isolated computers can only process locally available data, but interconnected computers can process geographically distant data and they can present data anywhere. The LAN, the WAN and especially the internet have made the rapid growth of data exchange possible, boosted by rapid network technology advances. Since 2008, cloud communication dominates peer-to-peer communication on the Internet, and by 2019, over 80% of data traffic will be cloud data traffic. The total amount of traffic, inside and between data centres, is estimated to be 5.6 ZB in 2016, growing to 10.4 ZB by 2019, and, based on a CAGR of ~25%, 13 ZB in 2020. The numbers include intra centre data rates, which reach about 2 ZB in the Internet in 2016 [216]. The expected data production rate is 44 ZB by 2020: that is almost 4x the amount of Internet traffic, meaning that 75% of the produced data is either destroyed or not moved.

The mobile phone network has seen a comparable transition: from voice-only to data-dominated communication. The mobile network’s next generation standard, 5G, expected to become available in the first half of the 2020s, is a data communication network, with voice telephony as one of the many applications. Connecting all the devices on the IoT critically depends on communication, wired and wireless. The 5G mobile network is expected to play an important role, but new wireless standards, targeted specifically at the IoT are appearing as well (Z-Wave, HomePlug, MoCA, SigFox, LORA.)

Just as computation takes energy, so does communication: a high-speed fibre interconnect switch can easily consume several kW. In some cases, it is more efficient to bring the processing to the data than the other way around. In other cases, especially in mobile IoT-connected devices, it pays to offload the communication stack running on a general purpose processor to specialized processors/accelerators to save energy.

The internet enabled cloud computing, whereby data plus processing is moved to some available storage and processing resource. Where that resource is located is unknown in advance. Therefore, it may be in another country with its own legislation.

The Internet has made computing borderless, requiring new laws and treaties.

Even with the growth of available data storage, it will become impossible to store all the data produced by the devices connected to the IoT in data centres. The reason is that the internet was designed hierarchically and that intermediate nodes cache popular content in order to save bandwidth on the backbones for servers and data centres. Caching works well in the downstream direction, but not in the upstream direction because there are no two identical streams. That means that the traffic generated by billions of IoT devices cannot be cached on its way to the data centres and that every single data stream has to be transported over the backbones.

2.5.6.4. IMPACT AND PROPOSED COURSE OF ACTIONS

The IOT cannot exist without yet-to-be-designed communication technology. Although physical network technology is not HiPEAC's area of expertise, network characteristics determine many aspects of IoT system architecture. The HiPEAC community must link up to existing network technology projects (such as the METIS-II project) defining and developing projects to shape the future IoT.

2.5.7. A NEW SOFTWARE CRISIS AND COURSE OF ACTION

The major cause of the software crisis is that the machines have become several orders of magnitude more powerful! To put it quite bluntly: as long as there were no machines, programming was no problem at all; when we had a few weak computers, programming became a mild problem, and now that we have gigantic computers, programming has become an equally gigantic problem. [44]

Edsger Dijkstra, The Humble Programmer

Controlling complexity is the essence of computer programming.

Brian Kernighan

Simplicity is prerequisite for reliability.

Edsger W. Dijkstra

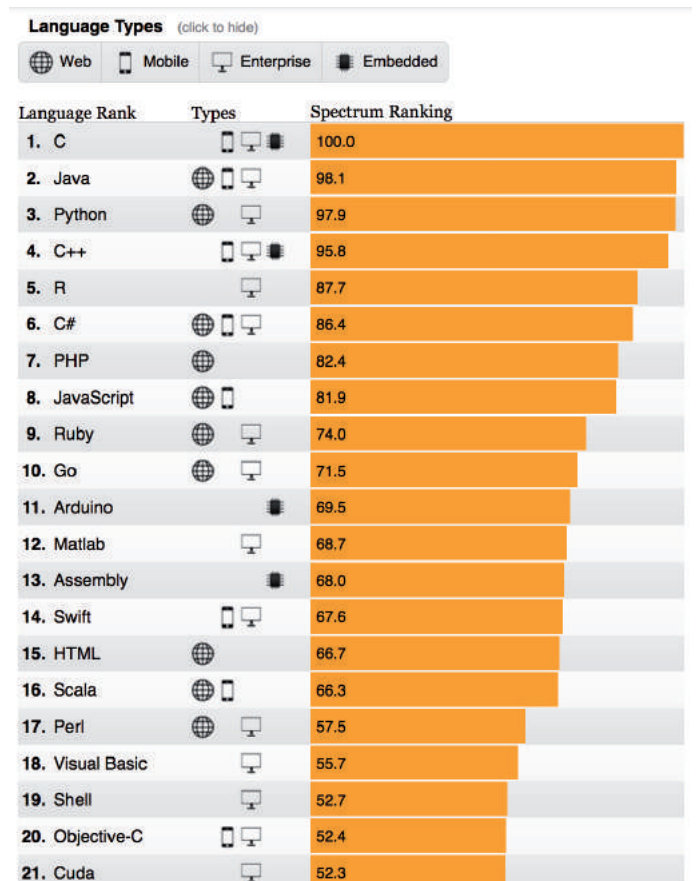


Figure 13: List of the most used language in 2016

Source: IEEE

2.5.7.1. A CHANGING LANDSCAPE

The next decade will see a change in landscape, from both a hardware and a software point of view. The one dominating characteristic is increasing complexity. It poses a number of significant challenges to the software development community.

2.5.7.1.1. A changing devices landscape

There is no doubt that technological stumbling blocks on the CMOS technology road ahead will change the system development landscape. More hardware details will have to be exposed to the software developer. That necessity comes from both disciplines: the software needs to be aware of more dimensions of the hardware (timing, energy, other resources), but the software may also have to take more hardware events into account (ageing, error conditions).

Developers can take at least two approaches: predictive and corrective. Corrective measures monitor the system's behaviour and correct detected errors. Predictive measures do the same but take measures in advance to prevent events from happening. Both approaches need information about the physical state of the system in order to be able to act. What information is needed, and how that information is sensed and presented to the system requires (again) a multi-disciplinary approach [164].

November Headline: Is Haskell finally going to hit the top 20?

Some people say that Haskell is the most mature purely functional programming language available nowadays. It has quite a long history, dating back from 1990 and its community is growing, although slowly. This month Haskell is only 0.255% away from the top 20 at position 23. Let's see what happens the next few months. Other interesting moves this month are MQL4 (from #52 to #41), Hack (from #76 to #63) and Elixir (from #86 to #64).

The TIOBE Programming Community index is an indicator of the popularity of programming languages. The index is updated once a month. The ratings are based on the number of skilled engineers world-wide, courses and third party vendors. Popular search engines such as Google, Bing, Yahoo!, Wikipedia, Amazon, YouTube and Baidu are used to calculate the ratings. It is important to note that the TIOBE index is not about the *best* programming language or the language in which *most lines of code* have been written.

The index can be used to check whether your programming skills are still up to date or to make a strategic decision about what programming language should be adopted when starting to build a new software system. The definition of the TIOBE index can be found [here](#).

Nov 2016	Nov 2015	Change	Programming Language	Ratings	Change
1	1		Java	18.755%	-1.65%
2	2		C	9.203%	-7.94%
3	3		C++	5.415%	-0.78%
4	4		C#	3.659%	-0.66%
5	5		Python	3.567%	-0.20%
6	8	▲	Visual Basic .NET	3.167%	+0.94%
7	6	▼	PHP	3.125%	-0.12%
8	7	▼	JavaScript	2.705%	+0.23%
9	11	▲	Assembly language	2.441%	+0.56%
10	10		Perl	2.361%	+0.33%
11	14	▲	Objective-C	2.246%	+0.82%
12	15	▲	Swift	2.039%	+0.80%
13	48	▲▲	Go	2.001%	+1.80%
14	9	▼▼	Ruby	1.978%	-0.06%
15	16	▲	MATLAB	1.967%	+0.78%
16	12	▼▼	Delphi/Object Pascal	1.950%	+0.27%
17	13	▼▼	Visual Basic	1.923%	+0.24%
18	37	▲▲	Groovy	1.811%	+1.48%
19	19		R	1.715%	+0.70%
20	18	▼	PL/SQL	1.512%	+0.48%

Figure 132: TIOBE Index November 2016

Source: TIOBE

2.5.7.1.2. A changing systems landscape

Once, computers were *stand-alone*, single processor systems. Nowadays, computers systems consist of *interconnected*, *many-processor* systems. Both transitions come with their respective difficulties. Both have the notion of parallelism in common: things happen at the same time and systems need to communicate at all levels. At the system level, in multiprocessor systems, this is a big challenge. Mainstream programming languages (C/

C++/Java) have no language abstractions to express parallelism, or in other words: parallelism is *not* part of the programming model. As the systems landscape changes rapidly from increasing processor speed to increasing the number of processors with approximately constant speed [43], increasing performance and efficiently using computing resources (avoiding dark silicon [10]) requires parallel programming. Parallel systems are inherently unpredictable, so replaying a faulty condition is very hard, and

probing a parallel system potentially changes its behaviour (Heisenbugs). Debugging parallel systems is thus hard, or better, very hard. Parallel systems should therefore be correct by design. This not a new problem, but its urgency becomes stronger with increasing system complexity induced by parallelism.

2.5.7.1.3. A changing users landscape

Users of next-generation computing systems hardly realize that they are using computers. Computing devices are now hidden in smartphones, smart watches and wearables. This has changed the landscape from stationary, centralized, mostly homogeneous devices into a heterogeneous, decentralised, dynamic system of devices. The problems these systems are supposed to solve have changed into computations triggered by a particular situation ('looking for a restaurant') at a particular location ('in this street'). And the result is to be presented through a variety of human interface devices, not just a simple screen with text.

Next-generation computing systems work with data from multiple sensors. The systems have to be able to deal with sensors contradicting each other, by taking the situation into account (smart, situation-aware sensors and interpretation). The amount of data generated by sensors is potentially enormous, so computing devices must make decisions on aggregating and filtering data.

2.5.7.1.4. A changing development tools landscape

In the early days of computers, specialists who intimately knew the hardware developed the software for it. They had to, because writing correct programs hinged on many system details that were exposed to the programmer.

With the advent of reliable electronics with an abundance of features, computing power, and memory space, those details could be hidden in the system software. The operating system's task was, and still is, to present the system in terms of everyday concepts: disk sectors became files in folders, serial communication channels became devices such as keyboards, screens and mouse pointers.

Software development evolution has reached a point where writing programs has been turned into connecting the right components to get from raw data to results. This development has significantly emancipated programming: the user tends to become the programmer, and almost everyone can develop a data processing system. With this development comes a large gain in productivity for those tasks for which libraries of functions are available.

Developing high performance libraries is still a specialists' task though, because ultimately mastering a system's complexity and squeezing out the last bit of performance or memory usage (which not all libraries do) requires intimate knowledge of said system. And of course, this only works if the performance of a system is dominated by general compute tasks, where the communication between tasks is almost free and can easily be standardized. As an example, several frameworks for processing large data sets (big data) have been developed (Unicore/X, Kepler, KNIME, ...).

These changing landscapes create a new software crisis with new challenges.

2.5.7.2. THE PRODUCTIVITY CHALLENGE

The software crisis is back, if it ever went away. In the late 1960s and early 1970s, systems became so complex that software developers could not keep up. New programming models, languages, methodologies and tools solved many of the problems of this classical software crisis. However, during that time basic computer architecture generally consisted of isolated single-core systems.

When Dennard scaling stopped, systems became multi-core, closely connected, and accelerators were introduced. A new software crisis loomed. With the end of Moore's law on the horizon, together with the expanding application of computer systems, the systems community is faced with additional system complexity factors – energy, time constraints, elaborate human-computer interfaces, cyber physical systems, continuous adaptation – that further deepen the new software crisis.

Since that first software crisis, programming languages have introduced abstractions for data and control, but areas that matter in contemporary and even more in future information systems lag behind. Parallelism at the task level often has to be expressed explicitly, that is at the code level, since programming languages do not have abstractions to model parallelism. Similarly, non-functional requirements such as power, energy or time currently cannot be expressed in mainstream programming languages such as Java, C++ and Python.

To keep productivity high, languages must evolve to treat these concepts as *first-class* citizens, fully supported by the underlying tools. This emphasizes the continued need for low-level software stack development in support of higher abstractions, continuously taking into account the penalties these abstractions bring. In the context of designing and developing systems, programming languages and tools are but one small part in a stack or even a mesh of tools (see Figure 133). Tooling for modelling systems and design space exploration, which allow system designers to study high-level trade-offs in design alternatives, are other necessary parts of this stack. The same goes for tools for debugging and testing, for performance analysis, and, where required, for certification. The tools that make up these stacks must be fully integrated, allowing system developers to focus on system design and development.

Possibly apart from niche markets such as VLSI design, development environments should be applicable to an as wide as possible range of systems. Creating and evolving solid environments requires a considerable effort that is not sustainable by SMEs unless backed by a large user community. The resulting development environments themselves are not the IP of the future, but the developed models in all their levels of abstractions are. Compare with compilers: compilers for a programming language are often not an asset for a company, but the programs are. For tools to be widely adopted, standardization is a must.

We thus recommend stimulating research in required abstractions to keep up with the intra-system, inter-system, and extra-system requirements of tomorrow's information systems. Tool stack development must be supported, as well as tool stack



Designed by DevNetwork.com - May 2014

Figure 133: (A small part of) The developer's technology landscape
Source: <http://www.devnetwork.com/devnetwork-ecosystem/>

maintenance. Finally, standardization efforts have to be stimulated.

Apart from raw productivity for expert developers as described above, there is another dimension to productivity. With the accelerating build-up of online data, building information processing and presentation systems is becoming an everyday task for the non-programmer. High-level data processing oriented abstractions in programming languages are too far away from the specialist areas of other disciplines. One way of dealing with this gap is to provide libraries of common data processing and presentation tasks, and rely on the specialists' ability to glue together these tasks with a programming language (Python) or with visual composition tools. Another way is to design languages specifically targeting particular common tasks, such as data processing. An example of such a language is R.

One other solution, requiring more effort from the programmers' side, is to design and implement domain-specific languages (DSLs), allowing the specialist to express her problem in terms of a specific domain of knowledge. From experience, this appears to be an effective way to enhance the non-programmer's productivity (~50%) [6]. A compiler for such a language transforms the pro-

grams into executables, tapping into the functionality of available libraries for a specific domain. To leverage the advantages of this approach frameworks must be available to aid in the rapid design of DSLs and rapid implementation of DSL compiling systems.

The Open Source software development process seems to be at least as productive, or even more productive, than closed software development (in terms of KLOC/time). However, getting to the first release takes open source software projects more time [45].

The longer a system stays in use, the larger the maintenance effort will be. This effort is related to a very hard-to-measure, non-functional aspect of code, namely maintainability. This is an issue still to be addressed, and tools are certainly required for this subject [6]. Indeed, as systems grow bigger, maintenance complexity often increases at the same rate. Maintainability is linked to availability of the appropriate documentation, to help developers master the complexity of legacy software they did not write.

Having helpful and above all up-to-date documentation, at both the development and maintenance stages, is still very much an issue in practice. Languages have for a long time embedded com-

port timing analysis of code, and optimization for, e.g. worst case (instead of the common case).

CPS' continuous contact with external systems places a heavy obligation for safety and security on the programming languages and tools used to implement them. Part of the system must be able to run in partial or full isolation (not taking input from the outside world, not depending on outside events, etc.). The system must be able to detect intrusion (attempts), and to take countermeasures to guarantee safety and security of the system. This hinges on the hardly investigated semantic properties of programming language constructs in terms of safety and security. There again, abstractions of security and safety must be integral parts of the languages and design methods.

MILS (Multiple Independent Levels of Security/Safety) is an interesting example of a design method for secure and safe systems [392]. Indeed, MILS aims to provide *'a high-assurance security architecture based on the concepts of separation and controlled information flow; implemented by separation mechanisms that support both untrusted and trustworthy components; ensuring that the total security solution is non-bypassable, evaluable, always invoked and tamperproof'* [103].

MILS OVERVIEW

Multiple Independent Levels of Security (MILS) is a high-assurance security architecture based on the concepts of separation and controlled information flow. The cornerstone of the architecture is a separation mechanism that encapsulates trusted and untrusted applications in compartments. The separation mechanism ensures that these applications can only communicate over channels explicitly defined by policies. This key component has to be non-bypassable, evaluable, always invoked and tamperproof (NEAT) [103].

The EURO-MILS project [422] has specified an architectural template and a set of standard components to design and build MILS systems [423]. These were abstracted from the use-cases of an automotive infotainment system (Figure 135) and an avionics network gateway. Further work on more diverse systems, such as industrial control, will probably lead to additions and refinements [424].

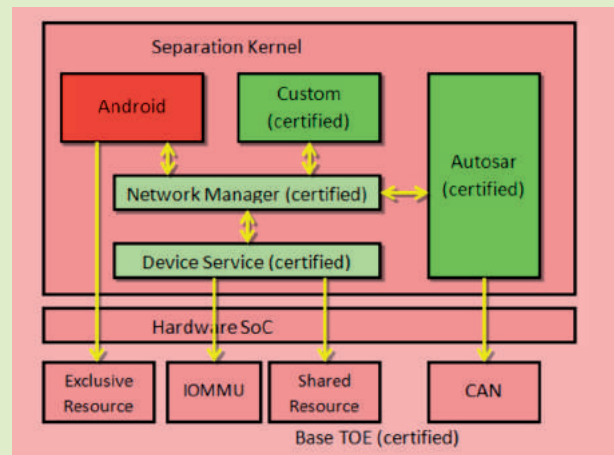


Figure 135: Non-interfering composed infotainment system
Source: [426]

ASSURANCE BY CERTIFICATION: COMMON CRITERIA

Certification is a way to build up trust by assigning the evaluation of a product to a trusted party. MILS addresses both security and safety, although safety standards are often domain-specific. Conversely, security can be handled based on the Common Criteria for Information Technology Security (CC) framework [425].

The Common Criteria provide a large requirements catalogue, which tries to cover a large group of products. For each concrete product to be certified, the developer has to identify a set of meaningful assurance and functional requirements in a document called /security target/. It is also possible to identify a set of meaningful certification requirements for a kind of products sharing similar functionality; this document is called a /protection profile/ (PP). Naturally, a PP benefits from getting input from developers of different systems, so a PP often is a community effort.

In this context, MILS enables compositional certification. This means that composed systems can be certified by re-using certification or evaluation results of individual (MILS) components. Three formal compositional certification methods exist. T-Composition certifies a complete stack of hardware, operating system and all applications running on top of the MILS platform. I-composition limits the certification to the MILS platform, and has been standardised in the form of a PP. Finally, /puzzle composition/ enables replacing an application either statically (at re-design and re-deploy time) or dynamically (secure update in the field) in an existing certified system [426].

RESEARCH CHALLENGES

Europe has a rich infrastructure for security certification, where it is clearly leading in the number of Common Criteria security certifications world-wide: 65% of the certifications published by the Common Criteria are from the EU [429]. This particularly holds for high-assurance certifications: all 39

listed EAL6 products, and four of the five listed EAL7 products, are from the EU.

The MILS methodology enables us to further cement this expertise. One of the two existing PP for a MILS separation kernel (SKPP [427]) has already been invalidated. However, the one by EURO-MILS has not yet been fully evaluated [428]. Further group efforts and involvement of MILS separation kernel stakeholders (both developers and users) will be required here. This could happen in the form of a Common Criteria Users Forum (CCUF) activity and/or a /community Protection Profile/ (cPP).

Other MILS components where work on PPs has begun – specifically in the Open Group – are network systems [430], a MILS platform, MILS Network Systems (MNS), MILS File Systems (MFS), and MILS Integration. So far the PPs are not yet published and need wider community review. Certifying these PPs at high assurance CC levels requires the application of formal methods [425]. EURO-MILS made a first attempt at unifying different national interpretations, but the work so far is limited to France and Germany. It needs to be extended to more national domains as well as possibly to a larger diversity of formal methods tools. Another area so far largely unexplored is to generate a set of formal specifications for MILS components that can be easily adapted to formal tools.

Computer-aided assurance can both aid with design and verification aspects of high assurance. The D-MILS project published a MILS-specific extension [432] of the Architecture Analysis and Design Language (AADL [431]), but further research is needed to develop patterns for MILS requirements, extensions for separation properties, and to apply it to the MILS Architecture Template Specification [423].

In terms of analysis, of note are methods based on formal proofs. These can model individual MILS components, security policies, and even integrated systems [420]. More research is needed to determine how to derive general properties from the properties of single systems, to perform abstract interpretation of content-dependent information flow policies, and to prove the security properties of components.

Sometimes, in numerical accuracy, *good is good enough*. It is not important to know the location of a self-driving car down to the millimetre. Calculating the position to that level of accuracy might cost a disproportionate amount of time, energy or both. Several techniques are under investigation to bring the accuracy of computations down to an acceptable level (we call it *adequate computing*, instead of *approximate computing*), both at the software level (tailored optimizations), and at the hardware level (mixed mode calculations). More can be gained if software developers can leverage all levels in the software stack to achieve a balance between the required precision and available resources, including time and energy.

Energy is another physical dimension that must be visible as a first-class citizen in programming languages to allow the programmer to design energy-efficient systems. Indeed, CPS have power, energy and even for some of them thermal constraints, as part of their requirements. The energy awareness of such systems is crucial, and most of the time has to be dynamic, so that the system can react and adapt to the changing environment, in all its dimensions. Many CPS are autonomous, and hence rely on batteries with limited autonomy and peak power. Some are able to harvest some energy from their environment, to extend their lifespan. Correct modelling of all these aspects is necessary to ensure appropriate operation, i.e. correctness, of such systems.

All the requirements mentioned above must possibly be open to certification by independent authorities. That may create the requirement for certain patterns in programs to allow for tool support, comparable to coding rules for existing languages.

We thus recommend that research on new programming concepts for so-called non-functional properties, such as time, power, energy, security, be stimulated, so as to find accurate yet simple ways to express these notions as first-class citizens when developing tomorrow's systems.

2.5.7.4. THE PERFORMANCE CHALLENGE

Performance is crucial in many ICT systems. Performance is often considered to be synonymous with timing and speed only. One should not forget however, that many other performance criteria exist: energy usage, peak power, memory usage during runtime, program footprint in the device, required computing power, other resource usage, etc.

Because performance is mostly not portable, developers have to take into account the target platform and its execution environment. They have always done that, and this was highlighted by Dijkstra as one of the initial causes of the software crisis [12]. What we did to address the initial software crisis was to start caring more about humans and less about machines. Powerful abstractions were designed to make developers more productive, relying on automation (compilers, runtime execution environments and operating systems) to bridge the gap with the low-level hardware interfaces. Methodologies, languages and tools to tackle complexity have been hugely successful, leading to the thriving software-dominated technological world we live in today. Most programming abstractions assume a Von Neumann architecture, possibly threaded to some degree.

These abstractions map relatively well to general purpose processors, but their suitability for advanced modern (e.g. heterogeneous, reconfigurable, distributed) platforms is being challenged. For example, object-oriented programming principles thrive in a threaded Von Neumann environment, but are almost entirely absent from hardware accelerated, massively parallel or specialized hardware. The lack of static typing, binding and referential transparency in some object oriented methodologies can be a no-go for compilers-based vectorization and for restricted, specialized hardware. Furthermore, the best-established software engineering practices can even be counter-productive when performance

is a non-functional requirement: cross-component optimization, inlining, and specialization break portability, modularity and code reuse, if not automated and made transparent to the application developer.

Higher-level frameworks and libraries always appear, for all languages, providing even higher abstractions or more accurate, larger sets of ready-to-use functionalities, thus boosting developer productivity and time-to-market. They also, to some extent, address the increasing shortage of skilled workforce in ICT that is becoming, at least in Europe, rather prominent. In a very similar way, languages that have been designed for ease of use, or for ease of deployment, spread more and more widely, reaching adoption levels of older, more (time-) performance oriented languages such as C (see Figure 131).

However, these frameworks, libraries and languages have rarely been designed with the impact on other performance criteria in mind, such as security or low-energy. Even when they have, they often address only one criterion, generally speed, without considering other criteria. As such, they provide large gains at development time, and large losses in performance at runtime. Since these frameworks, libraries and non-performance-oriented languages will continue to exist, and because developers tend to be less flexible than software, solutions have to be found to cope with this problem, preferably in an automated way so as avoid shifting the burden of complexity back to the developers. We thus recommend stimulating research aiming to improve the multi-criteria performance of high productivity languages in an automated way.

Furthermore, decades of progress in programming languages, software engineering and education threaten to go largely to waste because of hardware-software interface disruptions. As a result, the crisis also looms in the interaction between development teams with different expertise and procedures. It equally emerges from the interaction between diverse programming languages and runtime systems (e.g., MPI + OpenMP + CUDA, task parallel runtimes + virtualization). The clever automated techniques implemented in compilers and execution environments are not designed to deal with disruptive changes in the hardware-software interface.

Therefore, we recommend designing abstractions and optimizer tools able to cope with heterogeneous (non-) Von Neumann architectures and to reconcile differing development methods.

We strive for automated code generation techniques that provide abstraction without performance penalty. This is a well-known programming language dilemma that constantly needs to be revisited following hardware evolutions. In practice, the kind of modularized, well-defined components that may be good for software engineers are often very different from the components needed for parallelization and efficient resource usage. A real challenge is to allow developers to continue to modularize programs with a code reuse and productivity mind-set, without making it impossible for automated tools to implement cross-module and cross-layer optimizations, and to thoroughly reparti-

tion the application for efficient execution on a heterogeneous, parallel target.

Decoupling programmer abstractions from platform-dependent mapping is exactly what Model-Driven Engineering (MDE) research aimed to demonstrate for many years. The lack of precise semantics for general-purpose modelling frameworks unfortunately made this impossible. We could address this by learning from rigorous formalisms and tools, such as synchronous languages used in the correct-by-construction design of safety-critical systems [359]. We thus recommend reviving model driven design research, but basing it on precise modelling abstraction semantics.

Another path worth exploring is the possibility of alleviating hardware-software disruptions by making it possible to propagate pieces of information that specifically help making performance-related decision across the various abstraction levels, even at runtime. Such information can be related to timing, power, energy, and so on. In other words, we have to exploit all kinds of so-called non-functional properties that should be made first-class citizens in programming and modelling languages. Currently, developers and tools lack that information, which prevents the making of educated decisions. We therefore recommend stimulating research and development for vertically integrated tools allowing the free flow of information on various performance criteria from top to bottom and vice versa.

Other efforts exist in the form of automated, iterative (i.e. trial-and-error), compilation and optimization of programs, often coupled with machine-learning techniques. We indeed believe this human-machine cooperation has the potential to provide a practical way to deal and cope with the huge complexity a developer would face when trying to take into account all the elements that affect the targeted performance.

Currently, the performance challenge is becoming more and more prevalent, especially when all its facets are considered (speed, energy, power, etc.). The holistic view becomes indispensable.

2.5.7.5. THE DATA CHALLENGE

The IT world is becoming increasingly data-centric, with IoT, CPS, ‘smart everything everywhere’, and the immensely connected internet with its various social networks, which collectively fuel the ‘big data wave’ or even the ‘data deluge’.

Meanwhile, memory, storage and interconnect technologies are reaching scaling limits. This paradox will induce massive changes in the software stack and consolidation in the industry [364]. At the same time, non-volatile memory will find its way in existing computing systems. It will occupy a spot between fast, volatile (primary) storage, and slow, permanent secondary storage. This evolution, too, will require the revisiting of software stacks. Likely areas of interest include virtual memory, distributed shared memory, relaxed memory models without coherence, and data-centric execution models tying data and computations. Increasingly cheaper persistence should push for dramatic changes in I/O design and interfaces, and in hybrid memory architectures.



Figure 136: Coping with the data deluge (figurative)
 Source: <https://itmonitor.zenoss.com/dont-try-this-at-home-because-event-storms-are-inhumane-zenoss-forrester-webinar/>

On the other hand, the uncertain industrial maturity of new memory and communication technologies will delay their concrete impact. Researchers and engineers will have to live through a world of fragmented and trial-and-error adoption of these technologies. This will create scientific and business opportunities, but under a tight energy cap, because of the increased market focus on IoT and CPS with their stringent energy performance requirements. We thus recommend the development of memory models that take the new massive, non-volatile memory technology into account.

Research and innovation should break out of the incremental refinement and tuning of existing system architectures and layers. Technology reaching a scaling plateau will push for more efficient, leaner and more specialized solutions all over the software stack. The latter has indeed reached a level of complexity that is difficult to manage and calls for revisiting established assumptions, going back to the Von Neumann basics.

In a data-dominated computing landscape, one goal is to enable high performance for scale-out, heterogeneous parallel and distributed systems without sacrificing programmer productivity. Applications must be made aware of and adaptable to run-time environment changes, and to evolutions in computing system architectures. This ambitious goal depends on the ability to make applications portable and elastic. This is especially important in environments such as mobile devices where power and energy constraints can force the application to migrate to low-power cores or cloud services, where the amount of resources fluctuates depending on the workload. Another scenario is computational tasks that move closer to high-bandwidth sensors to reduce the communication cost and to enable upstream data integration

(e.g. cognitive cameras, augmented reality). These evolutions motivate research and innovation in the area of process and system virtualization, just-in-time compilation, binary-level optimization, dynamic orchestration, fault tolerance and monitoring, and programming models with a global address space for elastic computing.

The data challenge is also characterized by the emergence of new application domains and compute-intensive problems, pushing the limits of existing tools and abstractions for high-performance computing and real-time computing. There is already a need for high-performance libraries and domain-specific languages to support these new applications and their key algorithms. In these fields, security and data integrity are crosscutting concerns, interacting with all aspects of hardware, system, and development methodologies.

Such complex applications will require the collaboration of domain experts. For example, the design of advanced user interfaces can benefit from a close interaction with people having backgrounds in ergonomics and in behavioural and medical sciences. The safe interaction with the physical world through sensors and actuators requires a good knowledge of system theory and signal processing. Applications for health monitoring will naturally require the help of medical professionals, etc. We thus recommend fostering research into multidisciplinary software development teams to identify and solve potential roadblocks.

“I think part of what made the Macintosh great was that the people working on it were musicians and poets and artists and zoologists and historians who also happened to be the best computer scientists in the world” [285]

Steve Jobs

Cloud computing has become mainstream: the self-driving cars (will) rely on the cloud to get improved routing data, personal devices tap into the cloud to get localized, situational data. Therefore, software development, testing and verification tools must incorporate (abstract) connections to dynamic cloud services and devices, also for devices that are part of production and business IT systems. Again it is Europe’s strength in theoretical computer science that gives it an edge in setting the stage for such developments.

Finally, to tackle the ‘data deluge’, novel methodologies and tools will be needed that are data-centric, instead of computation- or algorithm-centric. Management of large amounts of data will be key, during the three stages of data collection, data storage, and data processing. This will have implications on research and technology for the sensors and the networks that connect them to the rest of the world, on storage solutions for large amount of data, and on algorithms and hardware to process large amounts of (potentially simple, in the case of IoT) data in parallel in a relatively inexpensive way, including energy-wise. Some specific domains, such as financial trading, also have very demanding requirements regarding collection and processing of huge volume of data in real-time.

The need for experts will thus continue to be strong, with data science being one of the most sought-after competencies.

One path to limit the data deluge is to minimize sensory data, through sensory information fusion. To this end, smart sensors able to process (part of) the data have to be designed, as well as the appropriate sensor system architectures. We recommend stimulating research in these areas.

Another aspect of the data deluge is the fact that data, and certainly huge collections of data, are a goldmine of information. Next to hardware IP, modelling IP, and advanced data processing algorithms, such collections of data form a company's assets, and require adequate protection. The content itself undoubtedly contains private information, making these data collections subject to privacy laws, and as such possibly subject to inspection by law enforcement agencies.

2.5.7.6. THE HOLISTIC CHALLENGE

To help solve the productivity and the correctness challenges (see sections 2.5.7.2.1, 2.5.7.2.2), system development has historically emphasized modularization and componentization, with significant success. As a result, the past decades have seen growing, but fragmented, development ecosystems with standardized interfaces to connect system parts. As noted in the performance challenge (see section 2.5.7.2.3), this tends to result in only a local view, while a holistic one is often necessary to produce optimized systems.

- Optimizing systems locally is insufficient: individual systems in general will not be sufficiently powerful to perform all necessary calculations (e.g. sensor networks). On the other hand, some form of local pre-processing will always be required as, otherwise, the amount of data that needs to be transmitted will overwhelm the communication links, or at least consume inordinate amounts of bandwidth and energy;
- Similarly, focusing on a limited set of optimizations, hence on a limited set of optimization criteria (and often only one single criterion) is less and less viable. The reason is that optimizing for, e.g. speed, can adversely affect other criteria, such as energy autonomy. System optimization must take into account the various non-functional optimization criteria in a holistic way, not individually;
- Global optimizations similar to the integration performed by current cloud providers will also be infeasible due to the fragmented nature of the systems in terms of ownership, control, security and privacy concerns, and the proprietary architecture of the devices. Virtualization has a lot of nice properties, but it also hides too much of how the system works in view of global optimization.

Within single systems, we thus need APIs and models that enable cross-layer optimizations. Only looking at the hardware or software is not enough. For example, software, including the cross-system optimization layer, may need to know the relative power usage of a particular kind of processing versus transmitting data in order to determine the most efficient way to proceed. In particular, this means that the software layer needs access to de-

tailed probes at the hardware level that provide it with information about power usage. This information cannot be statically encoded in the software, not only because of portability concerns but also in the face of increased hardware ageing effects that change its properties, and because energy consumption is difficult to statically predict. This also means that the software layer needs access to hardware 'knobs': mechanisms to act on the hardware and give it orders, or at the very least hints, on how to adapt. Indeed, only the software may make decisions, based on information both about the hardware and the environment, such as increasing or decreasing Quality of Service (QoS), taking alternate execution paths depending on the favoured optimization criteria (e.g. speed, energy), etc. These requirements imply that optimization criteria, or so-called non-functional properties, must become explicit, first-class citizens in programming and modelling languages. That will make them exploitable holistically across all layers of the software and software development stacks.

We will also have to come up with a new holistic approach that deals with all of these concerns in order to be able to improve the efficiency of large-scale distributed systems. A basic requirement for a holistic approach at this level is a large degree of interoperability between all systems, so that the optimizations can enlist the cooperation of as many involved systems as possible, to an as large extent as possible. This calls for an increased standardization effort. The increased amount of cooperative coordination-related communication among systems results in extra security concerns: systems have to protect themselves against both attacks on their own integrity and being induced into unwittingly attacking other systems (DoS). They also have to balance the optimal functioning of the network as a whole with their own QoS requirements and resource capabilities.

As a consequence, holistic approaches are necessary in several dimensions:

- In height, encompassing all layers of the system, from the top of the software stack to the bottom of the hardware stack;
- In width, encompassing various program and library components and even across distributed systems;
- In depth, encompassing several optimization criteria (speed, responsiveness, energy, power, QoS on various aspects...).

These three dimensions to the holistic challenge require frameworks to leverage their potential for productivity improvements that could give Europe a competitive edge.

The commercial market for software tools is small, but tool development effort is high, certainly in comparison with the market for, e.g. apps. That makes tool development unattractive to SMEs. From this perspective, free and open source software hurts SMEs' economic interests. Another model exists where companies offer paid support and services for open source and free tools, which they or someone else develop. This works for SMEs as they can keep overhead costs to a minimum and it allows them to build up expertise in diverse markets. However, it is a development path that needs stimulation in order to reduce the initial risk for SMEs.

2.5.7.3. IMPACT AND PROPOSED COURSE OF ACTIONS

As the classical Von Neumann model of computing loses its validity due to hardware developments in diverse areas such as processor architectures, memories, accelerators, and communication, the accompanying software models lose their validity as well. The complexity of developing ICT systems becomes the source of another software crisis.

We thus recommend rethinking the ‘traditional’ software stack, from hardware driver, through operating system to application. We believe it is time to revisit, re-assess its fundamental assumptions in the light of all the elements described in this section.

Indeed, the new ICT landscape, with the end of Dennard scaling, the complexity challenge, the trustability challenge, and the strong rise of CPS that entangle two worlds, requires us to revisit the basic concepts. This will offer a strong and unique opportu-

nity for disruptive changes where Europe can (re)take the lead. The creation and development of system development methods and tools that take into account a wide range of constraints, including non-functional ones (time, power, energy, trustability) as first-class citizens, in a holistic yet transparent and manageable manner, across all layers, is key. Stimulating research on software engineering and the appropriate related supporting tools is crucial. In this context, AI and generative design are potential paths to be explored.

Europe should also play a leading role in standardization, with a strong emphasis on industrially realistic solutions, hence with significant involvement of the computing industry in Europe. A strong backing from (European) early adopters is very important. It is worth remembering this quote heard from a member of US industry: *‘The US make solutions, the EU makes standards’*.

2.6. THE POSITION OF EUROPE IN THE WORLD

2.6.1. OTHER ROADMAPS

In recent years, we have witnessed a proliferation of roadmaps, vision documents, strategic research agendas, and the like. Many of them have been produced in the context of FP7 and H2020 projects, technology platforms, joint undertakings, and so on.

2.6.1.1. ETP4HPC

ETP4HPC is an industry-led think tank and advisory group of companies and research centres involved in High Performance Computing technology research in Europe. It was formed in 2011 with the aim of building a world-class HPC technology supply chain in Europe, increase the global share of European HPC and HPC technology vendors as well as maximize the benefit of HPC technology for the European HPC user community. ETP4HPC is also the EC's partner in the HPC contractual Public-Private Partnership (cPPP) which monitors and manages the European HPC research investment programme supported by a € 700 million investment by the EC within the Horizon 2020 programme

The current research priorities of the 2015 update of the Strategic Research Agenda are [246]:

1. HPC System architecture and components
2. System Software and management
3. Programming environment
4. Energy and resiliency
5. Balance compute, I/O and storage performance
6. Big data and HPC usage models
7. Mathematics and algorithms for extreme scale HPC systems

2.6.1.2. PRACE SCIENTIFIC CASE

PRACE, in a document, published in 2012, made seven recommendations [310]:

1. Europe should continue to provide a world-leading HPC infrastructure to scientists in academia and industry, for research that cannot be done any other way, through peer review based solely on excellence;
2. Leadership and management of HPC infrastructure at the European level should be a partnership between users and providers;
3. A commitment to Europe-level HPC infrastructure over several decades is required to provide researchers with a planning horizon of 10-20 years and a rolling 5-year specific technology upgrade roadmap;
4. There is an urgent need for algorithm and software development to be able to continue to exploit high-end architectures efficiently to meet the needs of science, industry and society;
5. European level HPC infrastructure should attach equal importance to compute and data, provide an integrated environment across Tiers 0 and 1, and support efficient end-to-end data movement between all levels. Its operation must be increasingly responsive to user needs and data security issues;

6. Europe's long-term competitiveness depends on people with skills to exploit its HPC infrastructure. It must provide on-going training programmes, to keep pace with the rapid evolution of the science, methods and technologies, and must put in place more attractive career structures for software developers to retain their skills in universities and associated institutions;
7. Thematic Centres should be established to support large long-term research programmes and cross-cutting technologies, to preserve and share expertise, to support training, and to maintain software and data.

2.6.1.3. EUROPEAN EXASCALE SOFTWARE INITIATIVE

The main goals of the second European Exascale Software Initiative (EESI2) were to elaborate an evolutive European vision and roadmap and to propose recommendations to address the challenges of Extreme Data and Extreme Computing on the new generation of Exascale computers expected in 2020. This FP7 project ended in March 2015.

The principles underlying EESI2 vision and recommendations are grouped in three pillars [239]:

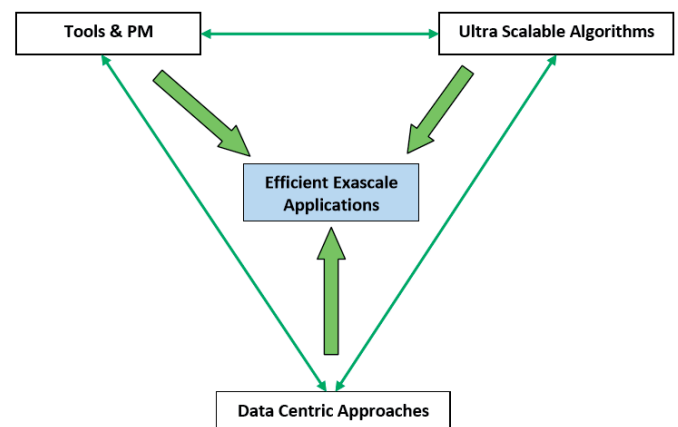


Figure 138: Principles underlying EESI2 vision and recommendations

Source: European Exascale Software Initiative 2 Final Report

In the **Tools & Programming Models** pillar, recommendations concern programming models and methods, heterogeneity management, software engineering and cross-cutting issues like resilience, validation and uncertainty quantification with a strong focus on the specificity of Exascale in these domains.

The following recommendations are proposed for funding by the European Commission:

- High productivity programming models for extreme computing
- Holistic approach for extreme heterogeneity management of Exascale supercomputers
- Software engineering methods for high performance computing
- Holistic approach to resilience
- Verification validation and uncertainties quantification tools evolution for a for better exploitation of Exascale capacities

In the **Ultra Scalable Algorithms** pillar recommendations concern specific and disruptive algorithms for Exascale computing, taking a step-change beyond 'traditional' HPC. It will lead to the design

and implementation of extremely efficient scalable solvers for a wide range of applications.

The following recommendations are proposed for funding by the European Commission:

- Algorithms for communication and data-movement avoidance
- Parallel-in-Time: a fundamental step forward in Exascale Simulations (disruptive approach)

The **Data Centric** pillar links extreme computing and extreme data. For the transition to Exascale, current data life cycle management techniques must be fully rethought, as described in the document ‘Software for Data Centric Approaches to Extreme Computing’ which is more a vision than a concrete recommendation.

This pillar gathers together key strategic issues for Exascale applications which have not been sufficiently addressed in Europe up to this point.

Ensuing from the EESI holistic vision of ‘Software for Data Centric Approaches to Extreme Computing’, the following recommendations, all new at European level, should be supported and funded by the European Commission:

- Towards flexible and efficient Exascale software couplers (direct or not, exchange of big data)
- In-situ extreme data processing and better science through I/O avoidance in high performance computing systems
- Declarative processing frameworks for big data analytics, extreme data fusion e.g. identification of turbulent flow features from massively parallel Exaflops and Exabytes simulations

Not all of these recommendations are at the same level of generalization but they are complementary and linked to each other by their global common objective: enabling the emergence of a new generation of data intensive and extreme computing applications. Some of them are fully disruptive; all need to go beyond known HPC technologies and methods.

2.6.1.4. RETHINK BIG

The objective of the RETHINK big Project is to bring together the key European hardware, networking and system architects with the key producers and consumers of big data to identify the industry coordination points that will maximize European competitiveness in the processing and analysis of big data over the next ten years. Specifically, RETHINK has delivered a strategic roadmap for how technology advancements in hardware and networking can be exploited for the purpose of data analytics while also taking into consideration advancements in applications, algorithms and systems [318].

Key findings in the roadmap are:

- 1 Industry is still focused on finding how to extract value from their data, and they are also still looking for the right business model to turn this value into profit. Consequently, they are not focused on processing (and storage) bottlenecks, let alone on the underlying hardware;
- 2 European companies are not convinced of the Return on Investment of using novel architectures;

- 3 Europe is at a strong disadvantage with respect to hardware / software co-design;
- 4 Dominance of non-European companies in the server market complicates the possibility of new European entrants in the area of specialized architectures.

A number of actions are derived from these key findings:

- Promote adoption of current and upcoming networking standards
- Prepare for the next generation of hardware and take advantage of the convergence of HPC and big data interests
- Anticipate the changes in data centre design for 400Gb Ethernet networks (and beyond)
- Reduce risk and cost of using accelerators
- Encourage system co-design for new technologies
- Improve programmability of FPGAs
- Pioneer markets for neuromorphic computing and increase collaboration
- Create a sustainable business environment including access to training data
- Establish standard benchmarks
- Identify and build accelerated building blocks
- Investigate intelligent use of heterogeneous resources
- Continue to ask the question: do companies think that hardware and networking optimizations for big data can solve the majority of their problems?

2.6.1.5. ECSEL

ECSEL (Electronic Components and Systems for European Leadership) is a public-private partnership set up to keep Europe at the forefront of technology development.

The objectives of the ECSEL Joint Undertaking are:

- Contribute to the development of a strong and globally competitive electronics components and systems industry in the European Union;
- Ensure the availability of electronic components and systems for key markets and for addressing societal challenges, keeping Europe at the forefront of technology development, bridging the gap between research and exploitation, strengthening innovation capabilities and creating economic and employment growth in the Union;
- Align strategies with Member States to attract private investment;
- Maintain and grow semiconductor and smart system manufacturing capability in Europe;
- Secure and strengthen a commanding position in design and systems engineering;
- Provide access for all stakeholders to a world-class infrastructure for design and manufacturing;
- Build a dynamic ecosystem involving Small and Medium-Sized Enterprises (SMEs), strengthening existing clusters and creating new clusters.

ECSEL JU produces a Multi-Annual Strategic Research and Innovation Agenda (MASRIA), which is an input into the Multi-Annual Strategic Plan (MASP).

The MASRIA describes the Vision, Mission and Strategy of the ECSEL JU as well as the strategic research and innovation activities to be undertaken through the ECSEL Calls in order to enable the ECSEL JU to fulfil its objectives. The MASRIA identifies and explores specific Electronic Components and Systems (ECS) technology solutions for smart applications relevant for societal challenges and industrial leadership in Europe.

The 2015 MASRIA focusses on four essential capabilities and ECS development tools, and on five key ECS applications [325]. Rather than formulating recommendations like the HiPEAC Vision, it proposes a work plan with a concrete timeline.

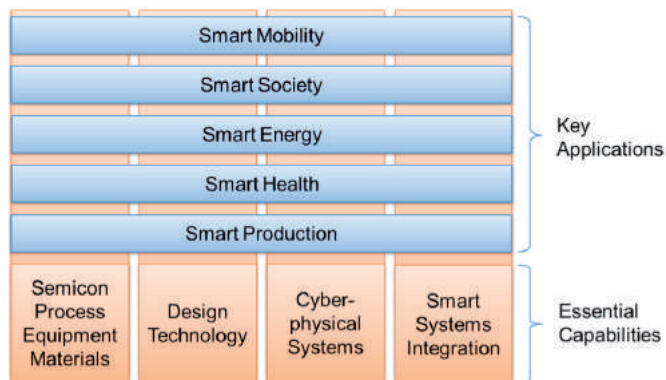


Figure 139: Structure of ECSEL applications/capabilities domain area
Source: 2016 Multi Annual Strategic Research and Innovation Agenda for ECSEL Joint Undertaking

2.6.1.6. ARTEMIS SRA

Cyber-physical systems technology, as a basis for embedded intelligence, is a major enabler of the digital transformation and for the digitization of industry as well as for every business in Europe. It is increasing its innovation potential by boosting its ability to bring to the market a new and larger variety of smarter products and services that are reshaping their future and creating new and unprecedented opportunities.

The ARTEMIS SRA [65] makes a distinction between six application priorities, i.e. new application domains where investments are needed, and six technological domains that are cross-cutting the different application domains.

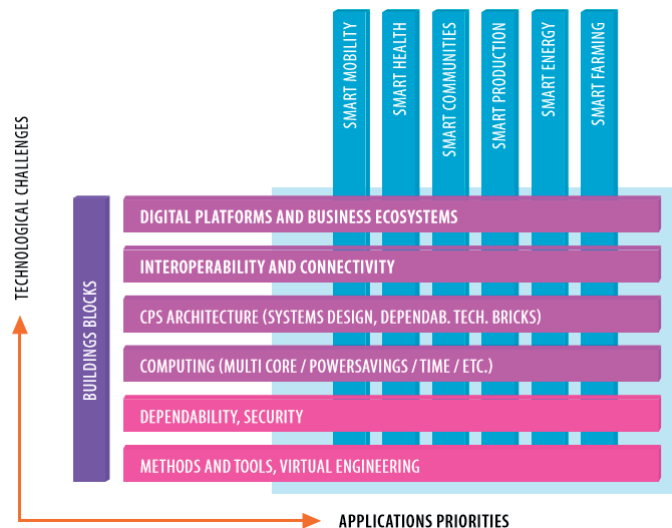


Figure 140: ARTEMIS application priorities and technological challenges

Source: ARTEMIS Strategic Research Agenda 2016

2.6.1.7. NEXT GENERATION COMPUTING ROADMAP

This roadmap was ordered by the European Commission in 2012. It presents a vision on next generation computing for the next 10-15 years. It does this by developing a number of visionary scenarios covering key areas of everyday life.

The scenarios are:

Scenario	Focus
It's all about me	Empowering the individual citizen
It's all about us	Communities and how they collaborate
Trains and other vehicles with brains	Make transport more efficient
Connected Brains	Research, education and knowledge sharing
Health & happiness	Health and social well being
Renewtopia	Sustainability, Energy and resource management
At a factory near you	Manufacturing in the future

Starting from these scenarios, it presents a series of technology roadmaps, associated research / development / innovation challenges and recommendations for Europe to exploit the opportunities offered by the next generation of computing [78].

Its key messages are:

1. Parallel hardware is now mainstream, but parallel software is not;
2. High-performance computing meets cyber-physical systems;
3. Internet of Everything is developing fast.

Areas of opportunity are:

1. Cyber-physical systems
2. Software
3. Energy
4. Computer interfaces

Mobile Computing and Internet of Everything

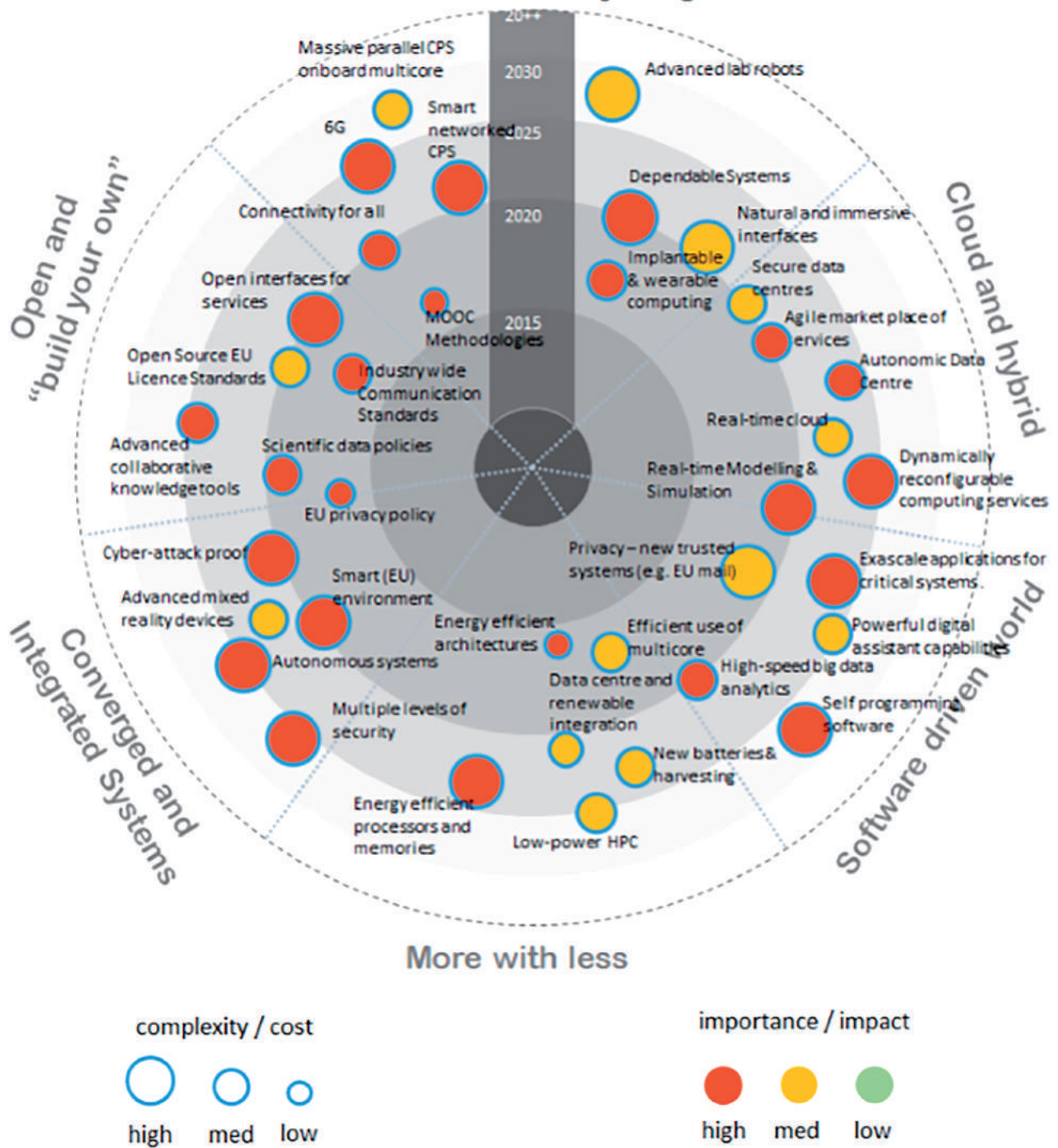


Figure 141: Summary of Next Generation Computing Roadmap

Source: Next Generation Computing Roadmap, study carried out for the European Commission by eutema GmbH in cooperation with Optimat, EPCC and 451 Research

2.6.1.8. EUROLAB-4-HPC

EuroLab-4-HPC [157] is a two-year Horizon 2020 funded project with the commitment to build the foundation for a European Research Center of Excellence in High-Performance Computing (HPC) Systems. Its roadmap targets a long-term roadmap from 2022 to 2030 for High-Performance Computing (HPC). Because of

the long-term perspective and its speculative nature, it starts with an assessment of future computing technologies that could influence HPC hardware and software. The current version of the roadmap is an intermediate draft version without recommendations.

2.6.1.9. CPSOS RESEARCH AND INNOVATION AGENDA

CPSoS – Towards a European Roadmap on Research and Innovation in Engineering and Management of Cyber-physical Systems of Systems – was a 30-month Support Action supported by the European Commission under the FP7 programme. It aimed to build constituencies for a European R&I agenda on SoS. CPSoS provided a forum and an exchange platform for systems of systems related communities and ongoing projects, focusing on the challenges posed by the engineering and the operation of technical systems in which computing and communication systems interact with large complex physical systems. Its approach was simultaneously integrative, aiming to bring together knowledge from different communities, and applications driven. The project ended in June 2016.

CPSoS produced a roadmap [223] in which three challenges were identified:

1. Distributed, reliable and efficient management of cyber-physical systems of systems;
2. Engineering support for the design-operation continuum of cyber-physical systems of systems;
3. Towards cognitive cyber-physical systems of systems

The document furthermore identified 11 medium-term research and innovation priorities that must be addressed to solve the core challenges.

1. System integration and reconfiguration: research and innovation is needed in open platforms, easy-to-test interfaces for semantic integration, and methods for describing and handling couplings between elements to enable the fast deployment of new technologies;
2. Resiliency in large systems: resiliency is a key issue in cyber-physical systems of systems in which faults are the norm;
3. Distributed robust system-wide optimization: cyber-physical systems of systems are too complex for centralized optimization methods and require novel approaches for distributed optimization;
4. Data-based system operation: cyber-physical systems of systems produce huge amounts of data that, for the most part, is not yet used to optimize and monitor the system. There is a need for advances in large-scale, real-time data analytics;
5. Predictive maintenance for improved asset management: maintenance depends on advances in sensors and novel tools for analysis, visualization, and decision support to provide the right information to the right person at all times;
6. Overcoming the modelling bottleneck: model-based methods for CPSoS engineering and management provide large benefits, but the effort needed to build such models often prevents the use of these techniques. New approaches for model adaptation, maintenance, and data-based modelling are needed;
7. Humans in the loop: CPSoS depend on humans, and novel HMI concepts are required to enable human operators to digest and react to large amounts of data and information quickly and effectively;

8. Integration of control, scheduling, planning, and demand-side management for industrial production systems will enable improved efficiency and reduce the carbon footprint;
9. New ICT infrastructures for adaptable, resilient, and reconfigurable manufacturing processes are required to adapt to the trend of product personalization, short time-scales, and fast changing customer demands;
10. Multi-disciplinary, multi-objective optimization of operations in complex, dynamic, 24/7 systems is needed to improve capacity and efficiency, and to reduce the cost of transportation and logistics systems;
11. Safe, secure and trusted autonomous operations in transportation and logistics: The increased degree of autonomy in transportation and logistics systems requires new approaches to guarantee safety, security, and trust.

2.6.1.10. ICT-ENERGY

ICT-Energy is a coordination action among consortia involved in the ICT-energy field with specific reference to bringing together the existing “Toward Zero-Power ICT” community organized within the ZEROPOWER project and the novel “MINECC”. The coordination activity is aimed at assessing the impact of the research efforts developed in the groups involved in the different consortia and proposing measures to increase the visibility of ICT-Energy related initiatives to the scientific community, targeted industries and to the public at large through exchange of information, dedicated networking events and media campaigns.

The coordination action has also produced a strategic research agenda [271]. The main recommendations are:

1. Energy consumption could be minimised if smart communications minimising the amount of data being transported is used over other techniques. Finally, the development and fast deployment of new communications system with low energy consumption per bit to replace legacy systems is essential to circumvent the enormous increase in data volume from cloud and especially high definition video use;
2. Research to find traceable and transparent energy usage throughout the system stack of ICT systems is required. Only once such research has been successfully completed can compilers and software be written to minimise energy consumption;
3. Transistors are approaching the minimum amount of energy per switch and the access resistance of electrical interconnects is a key issue for future scaling of energy in all ICT hardware. Radical new devices, interconnect solutions and system architectures are required if reductions in power per device are to be achieved in the future;
4. Radical solutions are required to remove the dependency on diesel generators and high carbon sourced electricity to maintain robust and reliable deployment of computing and communication services;
5. For autonomous systems, significant improvements in energy harvesting and energy storage at the small scale would also provide disruptive solutions to the use of smart sensors for a

host of applications in personalised healthcare, environmental monitoring, industrial monitoring, security and transportation.

2.6.1.11. ITRS

The ITRS 2015 document on emerging research devices [352] gives a taxonomy for emerging information-processing devices and is a very complete document on new technologies that can be used for building computing devices. ITRS 2015 will be the last issue of ITRS as we knew it. The successor of ITRS, the International Roadmap for Devices and Systems (IRDS), will focus on functions and systems rather than on processes and technologies instead.

2.6.2. ACTIONS IN OTHER COUNTRIES

2.6.2.1. US - SEMICONDUCTOR RESEARCH CORPORATION (SRC)

Semiconductor Research Corporation (SRC) is an American technology research consortium. It is a non-profit organization founded in 1982 and based in North Carolina, US.

SRC comprises four programmes:

- Global Research Collaboration (GRC). The SRC GRC's mission is to provide for innovative, strategic, pre-competitive research guided by the ITRS, conducted in universities worldwide, with a time frame of 7-14 years. GRC provides for a global forum for pre-competitive collaboration among all segments of the semiconductor industry, universities and government agencies. GRC is an advocate to various government and other funding agencies for support of university semiconductor research;
- Semiconductor Technology Advanced Research Network (STARnet). STARnet is a US based university research program that is guided strategically by industry and the US government, but managed by the US university community. It provides a multi-university, multi-disciplinary, collaborative research environment that is highly leveraged by both industry and US Department of Defense funding. STARnet focuses on beyond CMOS technology options and systems integration and discovery to enable both CMOS and beyond CMOS components, with a time frame of 14-20 year. The program also provides access to highly trained university graduate students;
- Nanoelectronics Research Initiative (NRI). The NRI Mission is to demonstrate non-conventional, low-energy technologies which can outperform CMOS on critical applications in ten years and beyond;
- SRC Education Alliance (SRCEA). The SRCEA mission is to attract and educate students of science and engineering through use-inspired research and industry connections, and promote their transition into careers that make a difference. It is a private foundation supporting science and engineering students and encouraging them to pursue careers in the semiconductor industry.

In this context, STARnet is most relevant. The STARnet programme is administered by the Semiconductor Research Corporation and has over 142 researchers from 38 different universities trying to push the limits of chips. STARnet is a collaboration between the

US Department of Defense, the Semiconductor Industry Association lobbying group, various chip and chip-making equipment manufacturers, and universities that do research in semiconductors. STARnet members are: GLOBALFOUNDRIES, IBM Corporation, Intel Corporation, Micron Technology Inc., Raytheon Company, Texas Instruments Incorporated, United Technologies Corporation and Semiconductor Industry Association.

STARnet research helps to keep the US and its industries at the forefront of technology. Member companies, in collaboration with US universities and the federal government, gain access to discoveries that keep them a step ahead of the competition. The benefits of STARnet membership are:

- Direct involvement with, and access to, relevantly educated university graduate students;
- Early access to the results of technologically critical research;
- Participation in a research program that is highly leveraged through combined industry and Department of Defense funding;
- Ability to shape early stage research that directly addresses industry needs;
- Contribution to long-term, innovative research supported through the unique STARnet model: multi-university, multi-disciplinary, collaborative research efforts focused on high-level intractable problems for the industry.

STARnet has a budget of US\$40 million per year, coming from DARPA and participating companies. It runs 6 centres:

1. Center for Spintronic Materials, Interfaces and Novel Architectures (C-SPAN)
2. Center for Low Energy Systems Technology (LEAST)
3. Center for Future Architectures Research (C-FAR) [410, 411]
4. Systems On Nanoscale Information fabricCs (SONIC)
5. TerraSwarm Research Center (TerraSwarm)
6. Function Accelerated nanoMaterial Engineering Center (FAME)

2.6.2.2. US - NATIONAL STRATEGIC COMPUTING INITIATIVE

The NSCI wants to ensure US leadership in high-performance computing. The federal government, industry and academia are working together to maximize the benefits of HPC for the economy and for scientific discovery.

This initiative is supported by several federal agencies [409].

- The Department of Energy (DoE) investigates the potential of neuromorphic computing, quantum-based sensors, and machine learning. It also supports predictive oncology and precision-medicine initiatives. It also continues working on the Exascale Computing Project;
- The National Science Foundation (NSF) launched a programme on "Energy-Efficient Computing: from Devices to Architectures" to reduce the energy consumption of future computing systems and created extra HPC resources. It will establish two Scientific Software Innovation Institutes;
- The Department of Defense (DoD) initiated a "Quantum Science and Engineering Program" and created secure access to high performance computers.

2.6.2.3. US - A NANOTECHNOLOGY-INSPIRED GRAND CHALLENGE

This initiative wants to develop transformational computing capabilities by combining innovations in multiple scientific disciplines. The Grand Challenge addresses three administration priorities – the National Nanotechnology Initiative (NNI), the National Strategic Computing Initiative (NSCI), and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative to: *create a new type of computer that can proactively interpret and learn from data, solve unfamiliar problems using what it has learned, and operate with the energy efficiency of the human brain.*

The initiative has seven focus areas:

1. Materials
2. Devices and Interconnects
3. Computing Architectures
4. Brain-Inspired Approaches
5. Fabrication/Manufacturing
6. Software, Modelling, and Simulation
7. Applications

The goals for computing architectures are as follows [291]:

- 5-year goal: Enable large-scale design, modelling, characterization, and verification of future computing architectures in both digital and analogue domains. Leverage advances in high-performance computing platforms to enable parallel, high-concurrency, and large-scale simulations beyond Exascale performance. This will enable the hybridization and interfacing of current digital computing with quantum- or biology-inspired computing approaches that require analogue and other novel interfaces;
- 10-year goal: Be able to predict the performance of new architectures incorporating new material systems and physical nonlinear phenomena;
- 15-year goal: Be able to predict the design and characterization of computing architectures based on user application needs. These results should enable ready-to-fabricate designs and specifications.

2.6.2.4. US - NATIONAL SCIENCE FOUNDATION (NSF)

The National Science Foundation gets a yearly budget of around US\$7 billion from Congress. About US\$1 billion is spent by the directorate for Computer and Information Science and Engineering (CISE). CISE’s mission is to promote the progress of computer and information science and engineering research and education, and advance the development and use of cyberinfrastructure; to promote understanding of the principles and uses of advanced computer, communications, and information systems in service to society; and to contribute to universal, transparent, and affordable participation in a knowledge-based society. CISE supports ambitious long-term research and research infrastructure projects within and across the many sub-fields of computing, as well as cyberinfrastructure for all areas of science and engineering; contributes to the education and training of computing profes-

sionals; and, more broadly to the preparation of a U.S. workforce with computing and computational competencies essential to success in an increasingly competitive global market [167].

Programme	Budget (million \$) (2016)
Advanced cyberinfrastructure (ACI)	227.29
Computing and communication foundations (CCF)	198.59
Computer and Network Systems (CNS)	236.32
Information and Intelligent systems (IIS)	198.94
Information Technology Research (ITR)	93.27
Total	954.41

Figure 142: Summary of US NSF programme budgets 2016

Progress in foundational research and education in the above areas is considered vital to address key national challenges, spur innovation, increase productivity, secure critical infrastructure, improve data analysis and sharing, and develop the next generation of computing and computational scientist.

First IEEE International Conference on Rebooting Computing (ICRC 2016), 17-19 October 2016 at the Hilton San Diego/Del Mar, San Diego, CA.

IEEE Rebooting Computing [412], an initiative dedicated to reinventing computer technology for the next generation, previously sponsored a series of four invitation-only Rebooting Computing Summits [413] in 2013-2015.

ICRC 2016 seeks contributions describing original research in the broad area of novel computing approaches, covering the entire computing stack from device hardware to applications software. Examples of topics may include the following:

- Neuromorphic, or “brain inspired”, computing
- Approximate and stochastic computing
- Optical computing
- Quantum computation
- Reversible and adiabatic computing
- Cellular Neural/Nonlinear Networks (CNN) and Cellular Automata
- Nonlinear Dynamical Systems and Edge of Chaos
- Superconducting or cryogenic computing
- Error-tolerant logic and circuits
- In-memory processing
- Extending Moore’s law and augmenting CMOS
- Novel device physics and materials

2.6.2.5. JAPAN

In Japan, the Abe administration created a five-year plan to realize a GDP of 600 trillion Yen with a primary budget surplus by 2020 (GDP in 2014 was 491 trillion Yen). This is needed in order to cope with the huge government debt, and the seriously ageing (and soon shrinking) population. The ‘Japan Revitalization Strate-

Strengths		Weaknesses
Science & Technology	High quality education Large number of PhDs Largest publication and citation count of the world Research and Technology Organizations (RTOs)	Weak academia-industry link Strong in research, but not in commercialization
Market & Industry	Largest market in the world Large embedded market	EU ICT contributes less to the GDP than in other advanced countries Europe lacks advanced foundries
Policy & Measurements	Common market Variety of research funding instruments Decent level of public funding of R&D	Europe lacks Venture Capital culture Lack of ICT workers Fragmentation of funding
Opportunities		Threats
Science & Technology	The end of Moore's law	Financial crisis
Market & Industry	Embedded systems, IoT, CPS Cybersecurity	Saturating markets Computing initiatives in China, Russia, Japan, etc. China is building a huge patent portfolio
Policy & Measurements	Solutions for societal challenges	Political instability (Brexit, immigration crisis, etc.)

Figure 143: SWOT analysis of Europe's position

gy' wants to promote a 'revolution in productivity' by launching ten strategic public-private joint projects.

The first one is the fourth industrial revolution (IoT, big data, AI, CPS): businesses have to innovate faster, SMEs must be digitized and robotized, and more ICT workers have to be trained. Other projects are healthcare (smart drug design, personalized healthcare services, the use of robots in nursing), renewable energy, creation of a sports industry (in preparation of the Olympics in 2020), revitalizing the housing market, improving the productivity of the service industry, more support for growing companies, promotion of Japanese agriculture, forestry and fishery, promotion of tourism and stimulating domestic consumer confidence. Even with an annual economic growth of 2%, the GDP in 2020 will only be 582 trillion Yen. The latest realistic projections are 551 trillion Yen assuming the current pace of growth.

2.6.3. EUROPEAN POSITION (SWOT)

In this section, we discuss the SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis of the European computing systems community. We make a distinction between three stakeholders: (i) the publicly funded universities and research institutions (Science & Technology), (ii) the computing industry and its market (Market & Industry), and (iii) the local and European governments responsible for creating an environment in which research, innovation and commercialization can take place (Policy & Measurement).

2.6.3.1. STRENGTHS

2.6.3.1.1. High quality education

Europe has a **good educational system**. Higher education is more affordable than in the US, and of the top one hundred best universities worldwide (2016 Times Higher Education Ranking), Europe has 42 institutions (North America has 43, and Asia 15) [180].

2.6.3.1.2. Large number of PhDs

European universities produce significantly more PhD degrees per 1000 of the population than those of the US, South Korea or Japan (Figure 144). Even better, the majority of individual European countries produce considerably more PhD degrees than the US. The countries producing less are either small, or joined the European Union recently [414].

2.6.3.1.3. Largest publication and citation count of the world

With respect to scientific output, Europe belongs to the strongest regions in the world. One third of all scientific publications in 2000 were generated in Europe. US was second with 28.2% of publications. By 2013, China had caught up and took second place. Europe remained first with 27.3% (Figure 145). The US had lost 9 percentage points and fallen to third place. This shows that research in Europe is of excellent quality and can compete globally.

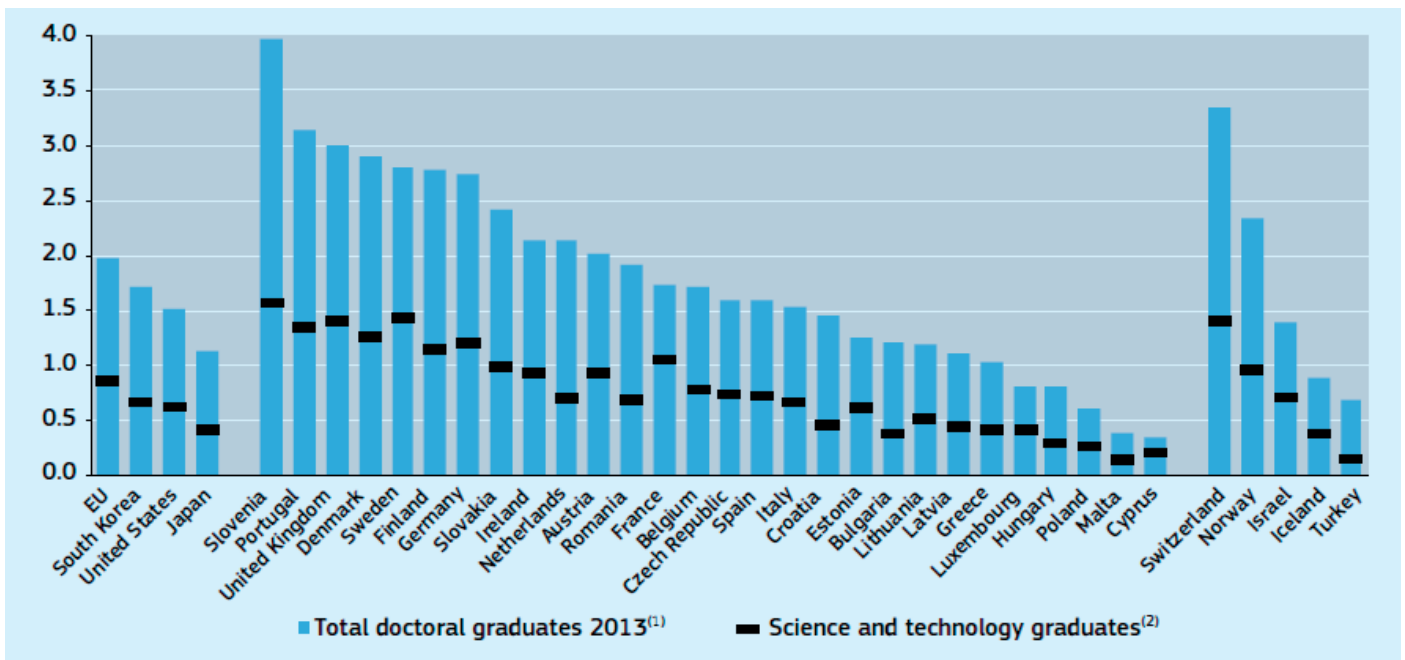


Figure 144: New doctoral graduates 2013 per 1000 of the population
Source: European Commission

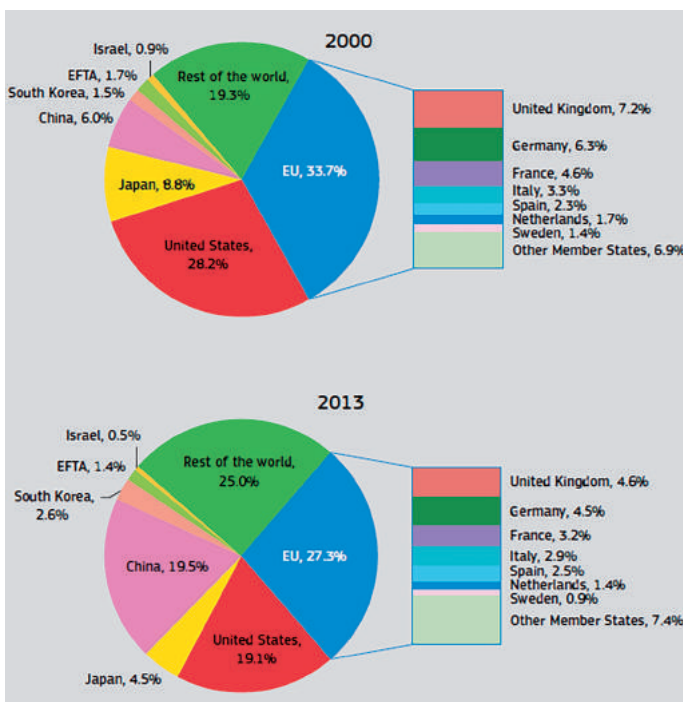


Figure 145: World share of scientific publications, 2000 and 2013
Source: European Commission

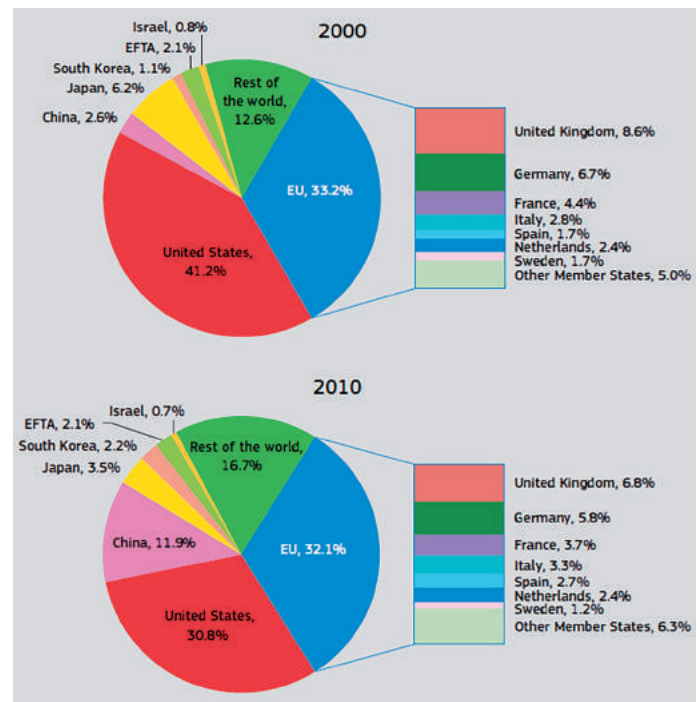


Figure 146: World share of highly cited scientific publications, 2000 and 2010
Source: European Commission

With respect to the number of citations, Europe was second after the US in 2010. By 2013, this situation has changed too. Europe now has more citations than the US (Figure 146). Per publication, US papers are cited more often which might suggest that US researchers publish more high-impact papers. In comparison with 2010, the impact of papers from both Europe and the US has increased in terms of number of citations. The many papers originating from China are cited less than the European and American papers, but the evolution is still spectacular.

Compared by sector, the US has more highly-cited publications in most domains. In ICT, Europe is second after the US. Surprisingly, China is leading security research with regards to the number of cited papers (Figure 147).

2.6.3.1.4. Research and Technology Organizations (RTOs)

Europe has several research institutes and companies that are key players in technology development (including CEA, imec and ASML). They are Europe's biggest asset when it comes to the further development of CMOS-technology, and their expertise

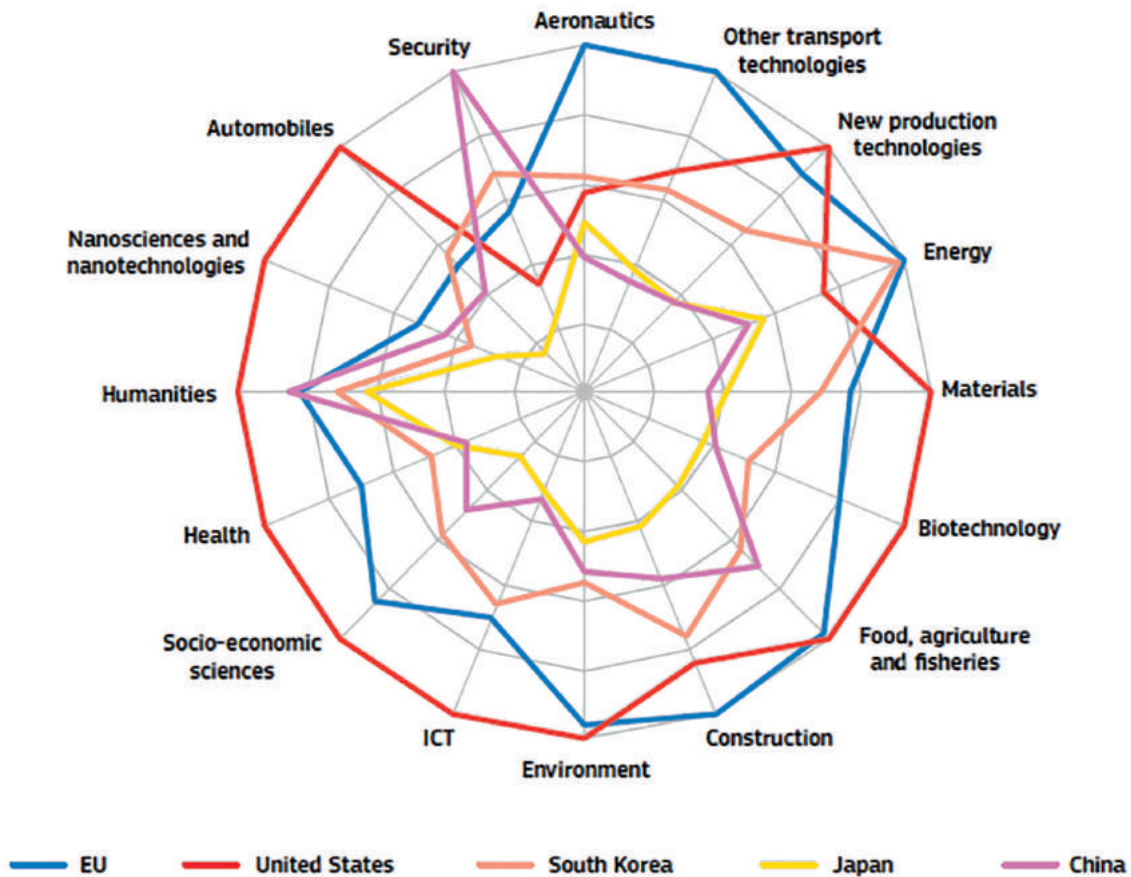


Figure 147: Highly cited scientific publications by sector, 2010
Source: European Commission

might also be crucial to the development of post-CMOS technology. With the recently approved quantum computing flagship, Europe wants to take the lead in quantum computing too.

2.6.3.1.5. Largest market in the world

Europe (EU-28 + Norway, Switzerland, Iceland) is the largest market of the world, based on purchasing power parity [195]. Europe would no longer be the biggest market of the world if the UK was left out of the statistics.

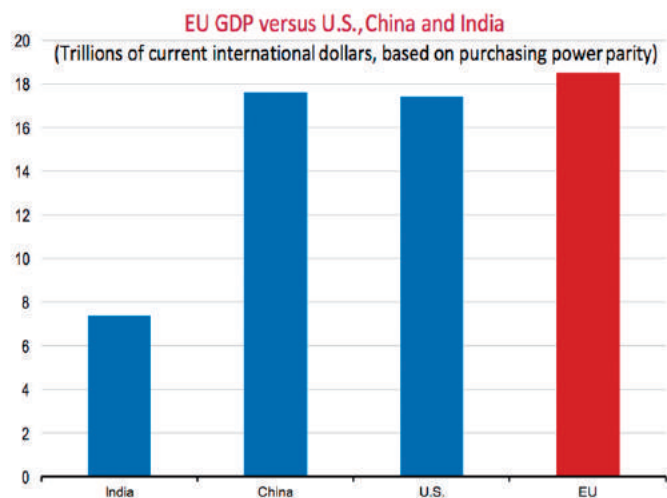


Figure 148: EU GDP versus US, China and India (2014)
Source: [195]

Household consumption expenditure in Europe is considerably higher than in the US, but it took Europe six years to overcome the effects of the financial crisis of 2008. The way in which the EU has dealt with the aftermath of that crisis has clearly slowed down growth in Europe.

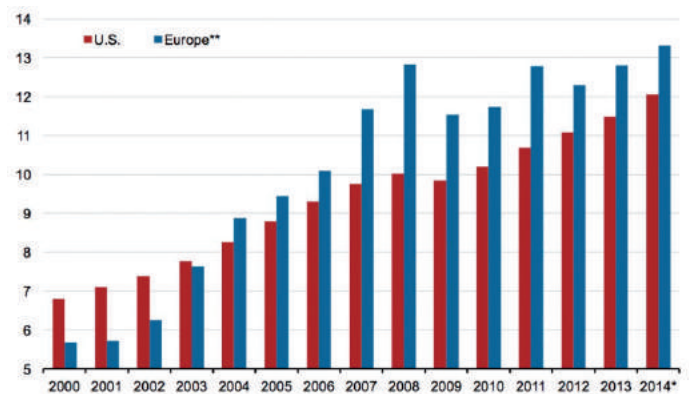


Figure 149: Household consumption expenditures in US and Europe, 2000-2014 (trillion US dollars)
Source: [195]

European businesses have access to a huge domestic market with a large potential still for growth in new member states and in the states recovering from financial crises like Greece. According to Global Industry Analysts [328], Europe also has a larger consumer electronics market than the US.

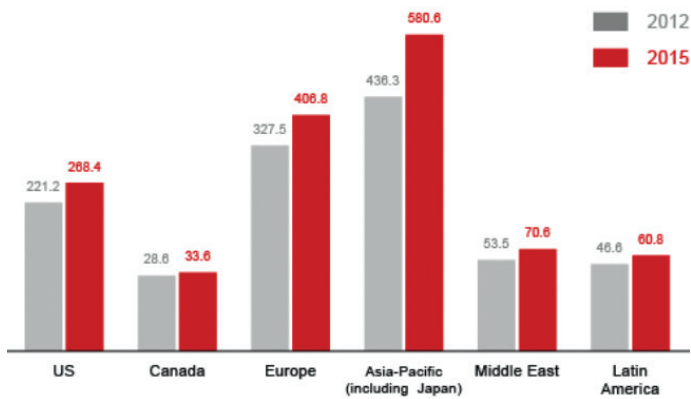


Figure 150: Global consumer electronic market, 2012 and 2015 (billion US dollars)

Source: [328]

2.6.3.1.6. Large embedded market

According to Global Markets Insight [162], the embedded systems market will reach a total size of US\$258 billion in 2023 at an average annual growth rate of 5.6%.

The European embedded systems market is the third largest in the world after North America and Asia, and will have an estimated size of US\$ 62 billion in 2023 (North America will attain US\$84 billion, and Asia US\$81 billion in the same year). The biggest embedded systems sectors in Europe are automotive, followed by health care and military and aerospace.

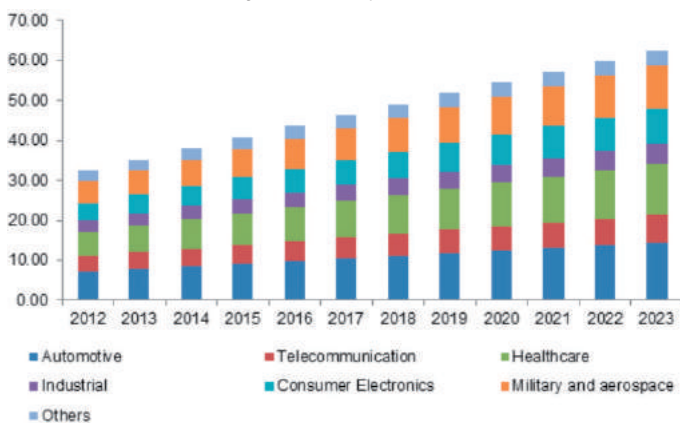


Figure 151: European embedded system market size, 2012-2023

Source: [162]

With an annual growth of 5.3% per year, the potential of the embedded systems industry to create added value, employment and growth cannot be underestimated. On the other hand, the fact that the embedded systems market in Europe is not the largest (in contrast to Europe's leading position in total market size and consumer electronics market) might suggest that the embedded market is weaker in Europe than in the US. According to [162], the embedded hardware market will grow to US\$144 billion, while the embedded software market will only grow to US\$18 billion. In order to grow, Europe's focus should be on embedded hardware, not software. The good news is that Europe has some important key players in this area: Infineon Technologies, STMicroelectronics, NXP Semiconductor. Non-European players are Renesas Electronics, Texas Instruments, and Microchip.

2.6.3.1.7. Common market

At a policy level, one of the strengths is the common market, and the fact that Europe can act as one economic block in global trade negotiations. However, there is still a long way to go before Europe will become a fully integrated market with one set of laws, one currency and one tax system.

2.6.3.1.8. Variety of research funding instruments

Europe has a variety of research funding instruments, complementing the national funding instruments. The research and innovation programmes of the European Commission help to stimulate research collaboration. The ERC instruments support research excellence, the flagship programs want to create critical mass in key research areas, the European Institute of Technology wants to stimulate research and innovation, and the joint undertakings like ECSEL aim to pool local and European funding to encourage research and innovation.

2.6.3.1.9. Decent level of public funding of R&D

The total public funding efforts make Europe a good place to carry out R&D (at 0.7% of GDP). Europe is in third place after South Korea and Japan. The public R&D intensity in the US is declining (Figure 152).

The relatively high public funding does however not compensate for the low R&D investments by industry (see weaknesses).

Combined, Europe is lagging behind the other geographies, and was recently taken over by China. The aim for Europe is to spend 3% of GDP, but it is still far away from that target (Figure 153).

The number of researchers hired is in line with the R&D intensity. The number of researchers hired by the business sector is clearly lower in Europe than in other developed countries (Figure 154).

2.6.3.2. WEAKNESSES

2.6.3.2.1. Weak academia-industry link

According to OECD [9], European universities and research institutes produce more spin-offs per public research dollar than their US counterparts. Although the study does not distinguish between ICT and non-ICT spin-offs, there is no reason to assume that the ICT sector would underperform in spin-off creation. The fact that the US has fewer spin-offs than Europe might be surprising, but most US-based start-ups are not created by universities and research centres, but by their graduates, without using the IPR of the university.

The collaboration between academia and industry (quantified as the number of joint scientific publications) is weak in Europe, and only increased very slowly between 2007 and 2012. Only four (small) European countries have more joint publications than the US (Denmark, Sweden, The Netherlands and Belgium), but not necessarily in computing (Figure 156).

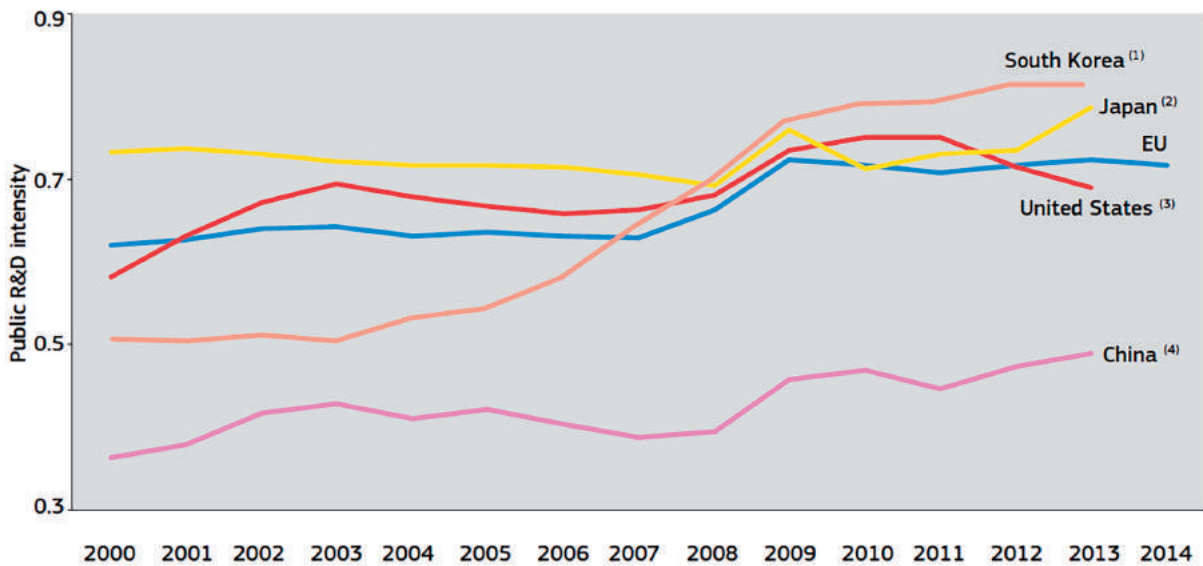


Figure 152: Evolution of public R&D intensity, 2000-2014
Source: European Commission

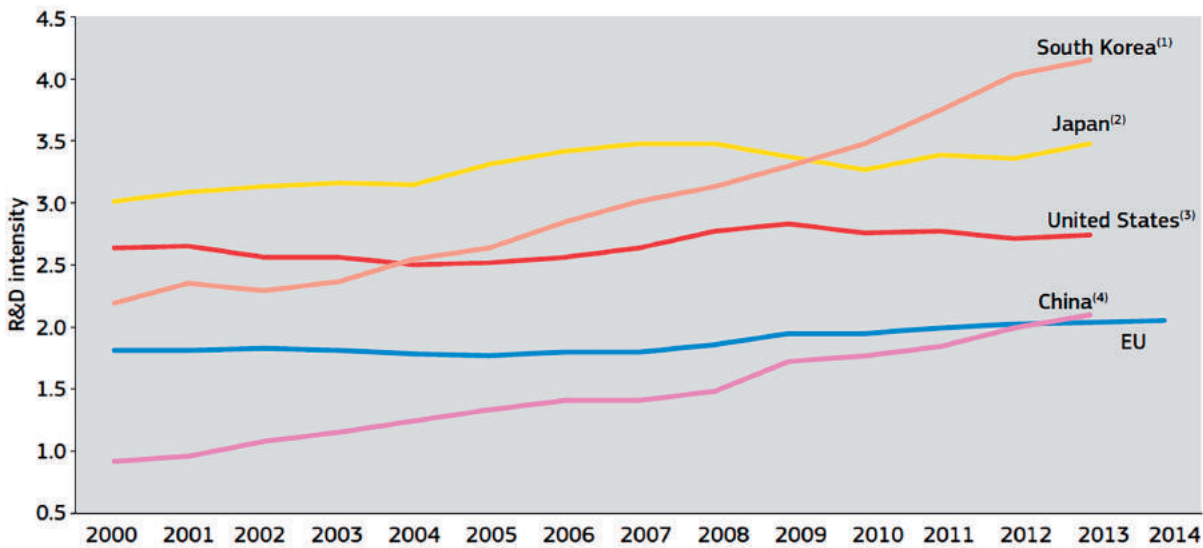


Figure 153: Evolution of R&D intensity, 2000-2014
Source: European Commission

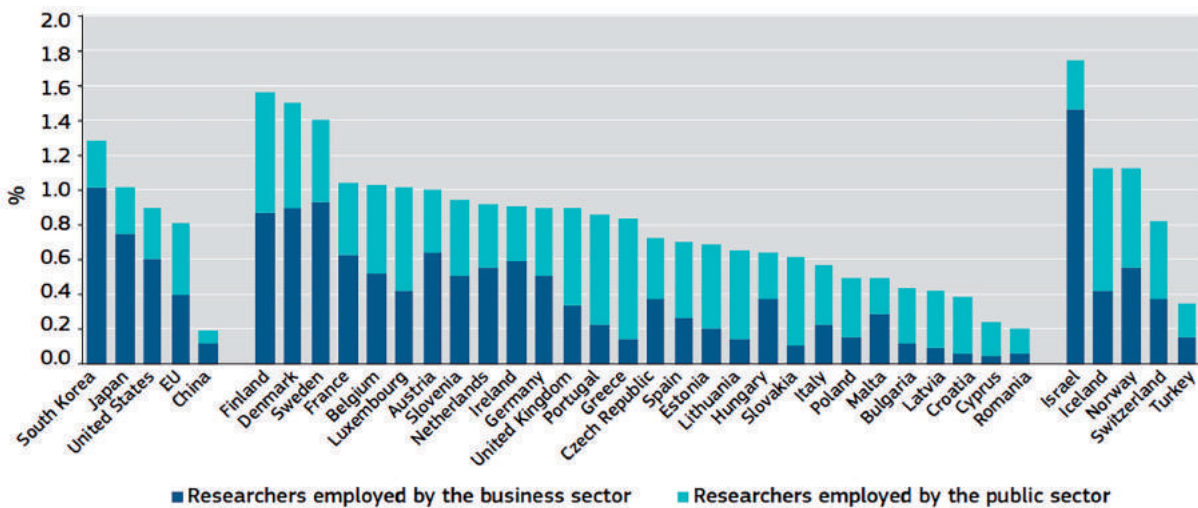


Figure 154: Researcher FTE as % of total employment, 2014
Source: European Commission

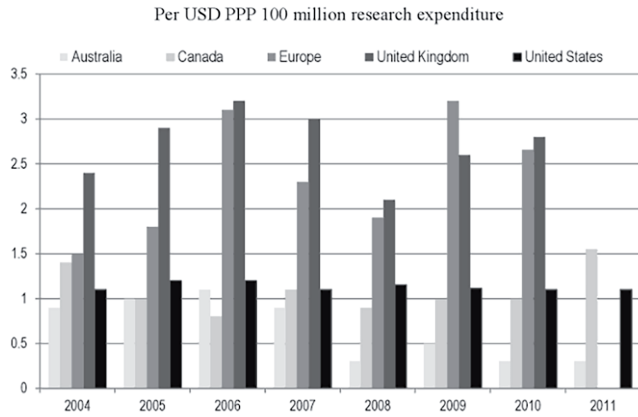


Figure 155: Creation of public research spin-offs, 2004-2011

Source: Organization for Economic Co-operation and Development

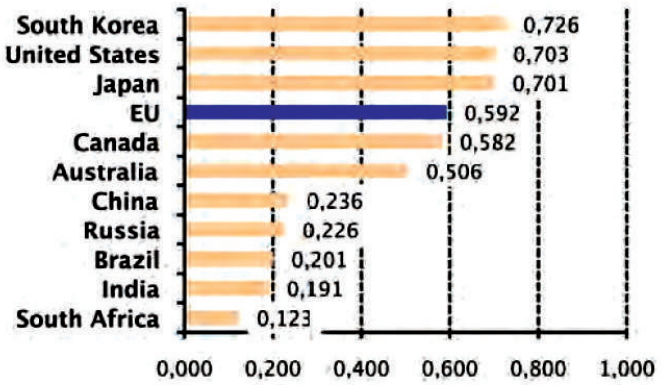
2.6.3.2.2. Strong in research, but not in commercialization

Europe has the highest world share of publications and has one third of the world share of highly cited publications, but this does not make Europe the most innovative region. Europe is only the fourth innovator in the world after South Korea, United States and Japan and on a par with Canada [36]. Whereas the gap between the EU and the United States and between the EU and Japan has been narrowing (US: 29% in 2008 down to 19% in 2015; Japan: 23% in 2008 down to 18% in 2015), the gap with South Korea is widening (5% in 2008 up to 23% in 2015). This trend is expected to continue in the next two years (Figure 157).

There are large differences in innovation performance between member states. Only the two strongest innovation leaders in Europe (Sweden and Denmark) are at the level of Japan and the United States (Figure 158).

At European level, there is a clear innovation gap between the North-Western part and the South-Eastern part (Figure 159).

Global innovation performance



Global innovation growth rates

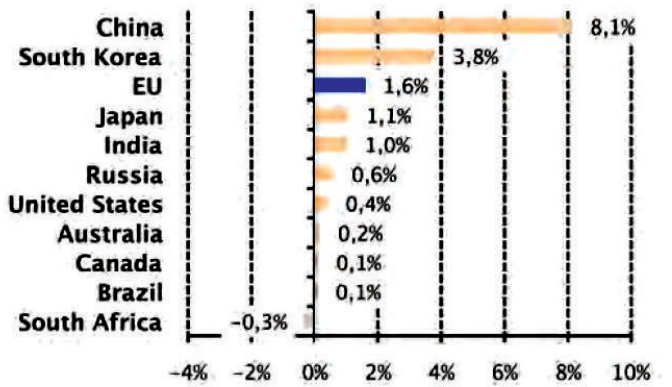


Figure 157: Global innovation performance and growth rates

Source: European Innovation Scoreboard 2016, European Commission

However, the good thing is that the innovation index of Europe is increasing, and is predicted to keep increasing in the coming years (Figure 160).

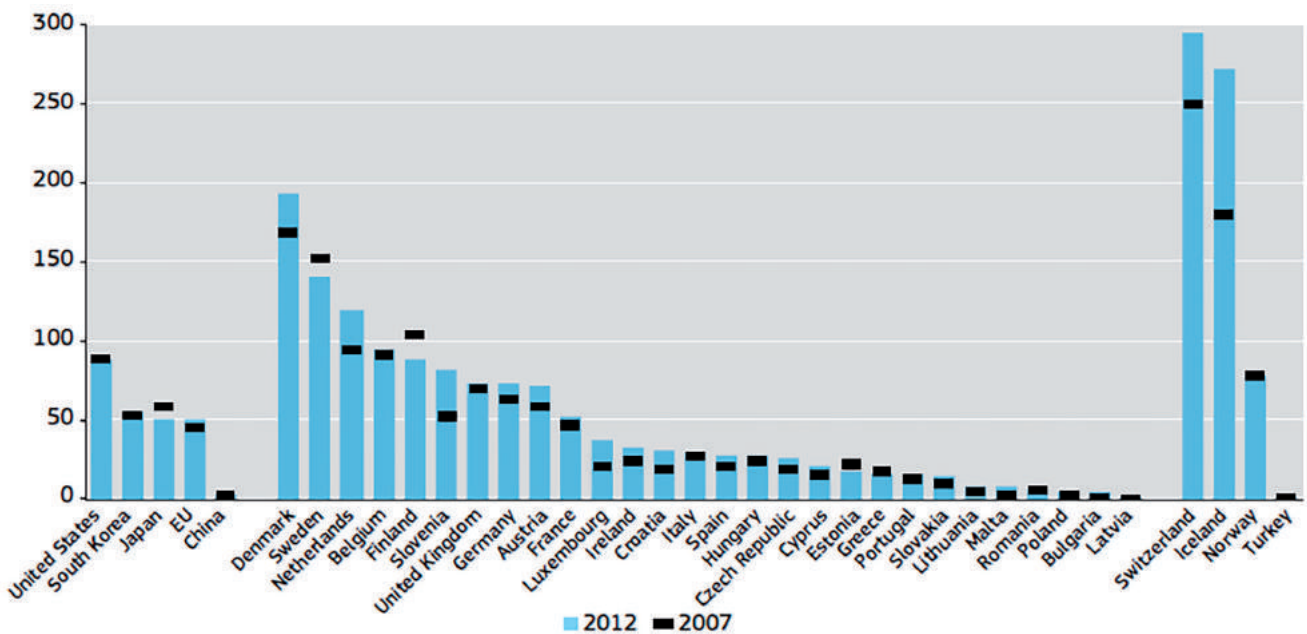


Figure 156: Public-private co-publications, 2007 and 2012

Source: European Commission

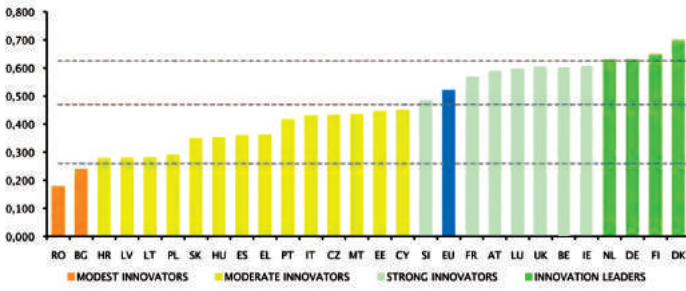


Figure 158: EU Member States' innovation performance
Source: European Innovation Scoreboard 2016, European Commission

2016 EUROPEAN INNOVATION SCOREBOARD EU MEMBER STATES' INNOVATION PERFORMANCE



Figure 159: EU Member States' innovation performance, simple form
Source: European Innovation Scoreboard 2016, European Commission



Figure 160: EU innovation performance
Source: European Innovation Scoreboard 2016, European Commission

2.6.3.2.3. EU ICT contributes less to the GDP than in other advanced countries

The European ICT industry contributes less than 4% to GDP while it is more than 5% in the US and Japan (Figure 161). One explanation is that Europe lacks GAFA (Google, Apple, Facebook, Amazon), and other major ICT companies like Microsoft, HP, Dell, IBM, and

the ecosystem supporting them. This is a structural weakness which also limits the innovation potential for the ICT sector (the smaller the sector, the smaller the resources to invest in research and development).

This is illustrated in the business R&D intensity; this is the fraction of GDP that is invested by businesses in R&D. Of all major geographies, European companies invest the least in R&D (Figure 162).

The business expenditure on R&D in Europe is less than 6% of the value added, compared to more than 10% in the US and in Japan (Figure 163).

For the computing and electronics market, the share of the value added is decreasing for Europe, and it is increasing in the US. However, the good news is that the business expenditure for computing is growing in Europe (while it is slightly decreasing in the US).

Europe is lagging far behind the US in R&D investments in the ICT-related sectors, and this gap has become wider over the last decade (Figures 164 & 165).

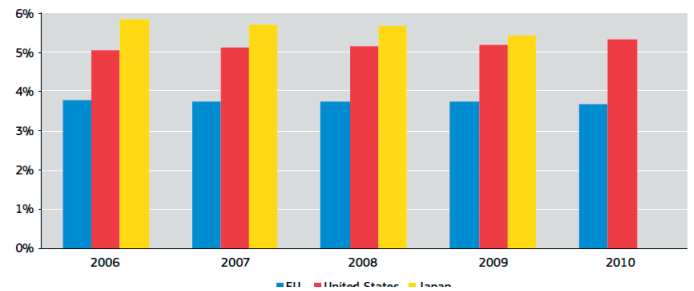


Figure 161: ICT share of GDP, 2006-2010
Source: European Commission

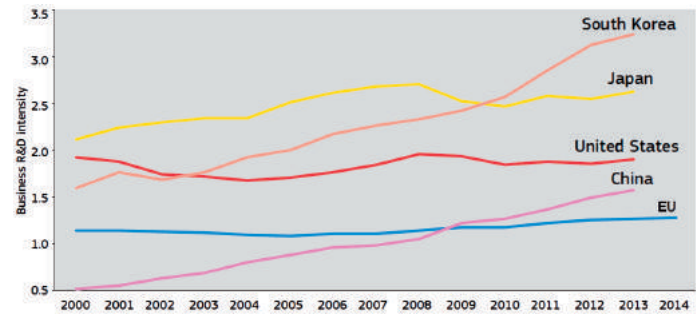


Figure 162: Evolution of business R&D intensity, 2000-2014
Source: European Commission

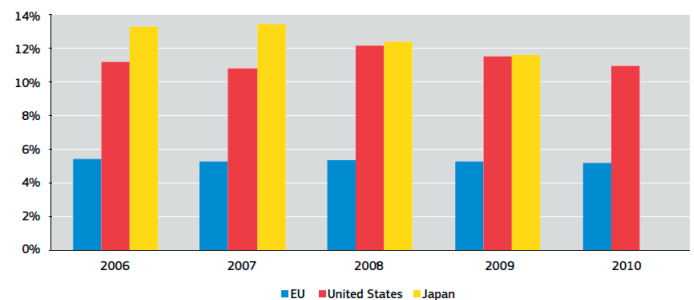


Figure 163: ICT R&D intensity, 2006-2010
Source: European Commission

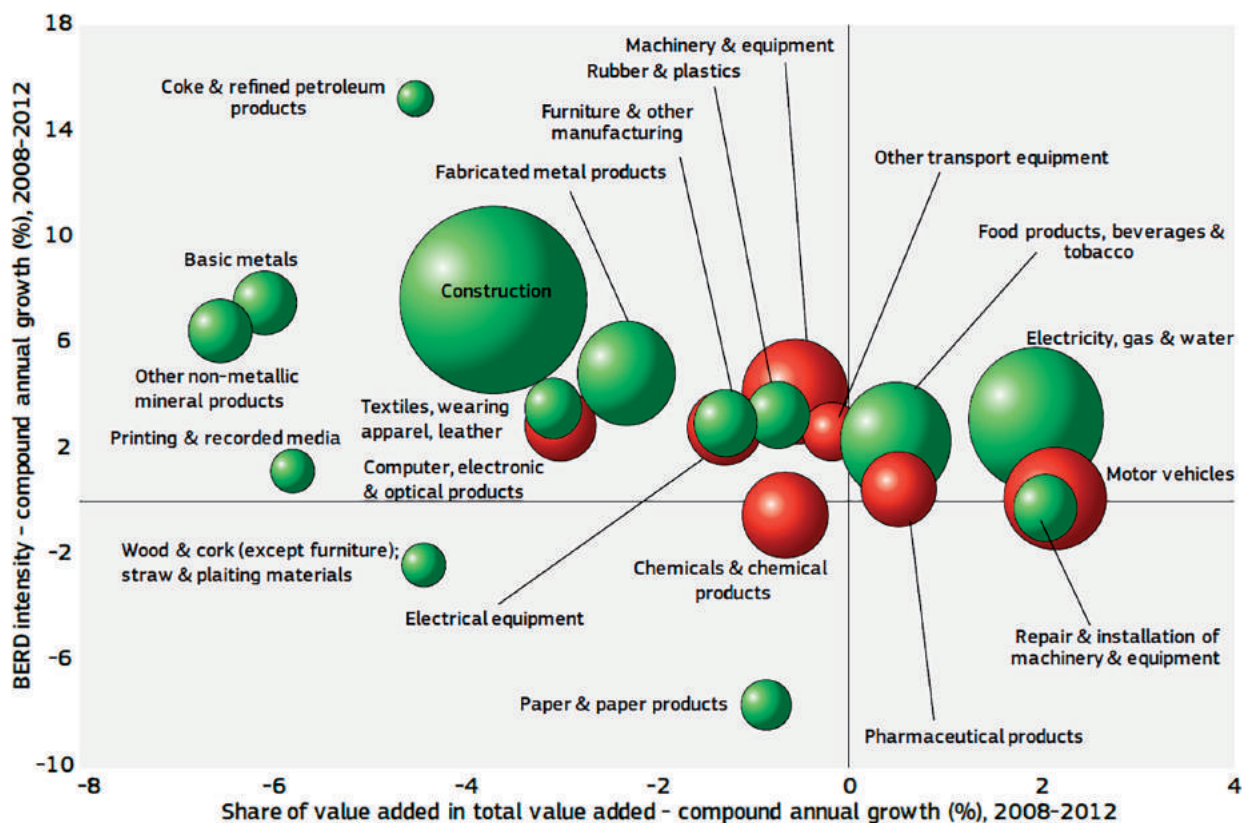


Figure 164: Evolution of R&D intensity and industrial structure in the EU, 2008-2012
Source: European Commission

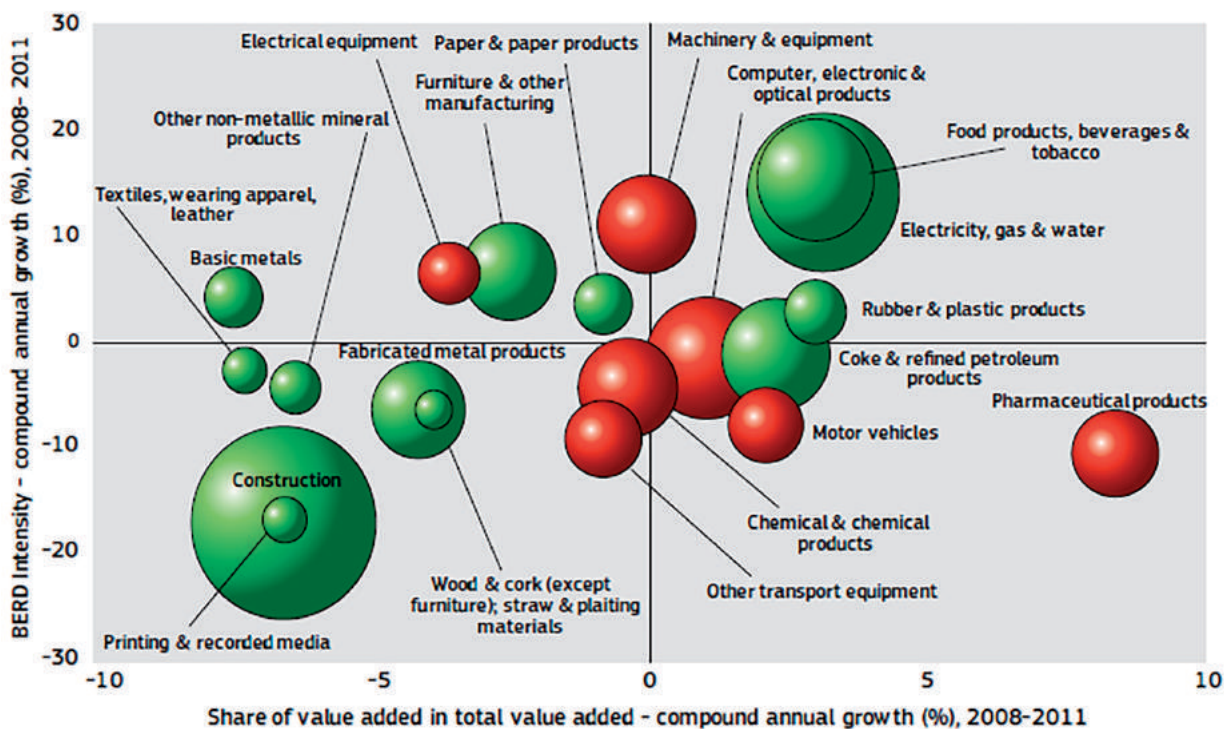


Figure 165: Evolution of R&D intensity and industrial structure in the US, 2008-2011
Source: European Commission

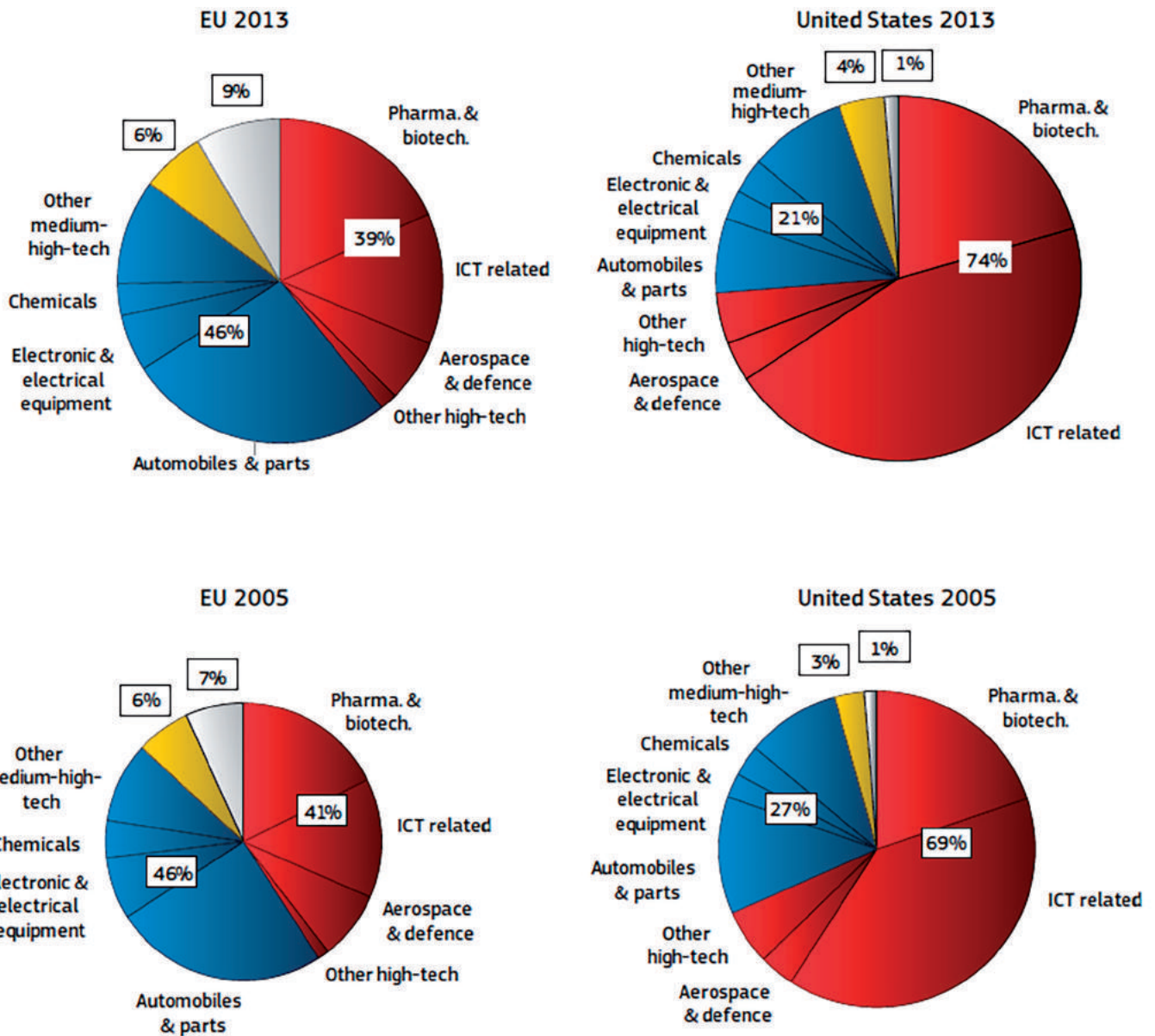


Figure 166: Sectoral composition of R&D intensive enterprises in the EU and the US, 2005 and 2013
 Source: DG Research and Innovation – Unit for Analysis and Monitoring of National Research Policies, European Commission

The fact that Europe lacks major ICT companies has far-reaching consequences: it also means that venture capitalists are less eager to invest in European start-ups and scale-ups because there are fewer companies that might be able to acquire them. The companies that grow big are often acquired by non-European companies: Nokia was acquired by Microsoft, ARM by Softbank, Movidius by Intel. There are a few counterexamples like Sysgo, which was acquired by Thales.

Non-European business leaders like Elon Musk, Tim Cook, the Google founders, and Masayoshi Son seem to have a clearer vision on the future than their European counterparts, and they actively promote their vision in the media. Very few people know the CEO of major European computing companies like Infineon, Ericson, Atos, ARM or STMicroelectronics. They are lacking a rock star status.

2.6.3.2.4. Europe lacks advanced foundries

There used to be foundries in Europe, but they were acquired by non-European companies, and disappeared. The fact that Europe depends on foreign foundries means that it has to import most of its semiconductors. Since the embedded hardware market is many times bigger than the embedded software market, this is a lost market opportunity. The leading foundries are not located in low-wage countries, meaning that they did not leave Europe due to labour costs. Given the fact that Europe is a world leader in the development of the technology used in foundries (CEA, imec, ASML and so on), it is surprising that no large foundries are left in Europe. One explanation is that European countries did not aggressively invest in new foundries (while this was the case in South Korea and in Taiwan), and that European venture capitalists are not interested in foundries (while they are in the US).

2.6.3.2.5. Europe lacks Venture Capital culture

More generally, Europe lacks a VC culture, and in this metric, the gap between the US and Europe could not be bigger. Not a single European country can match the average for the whole of the US, and the top five are all small countries.

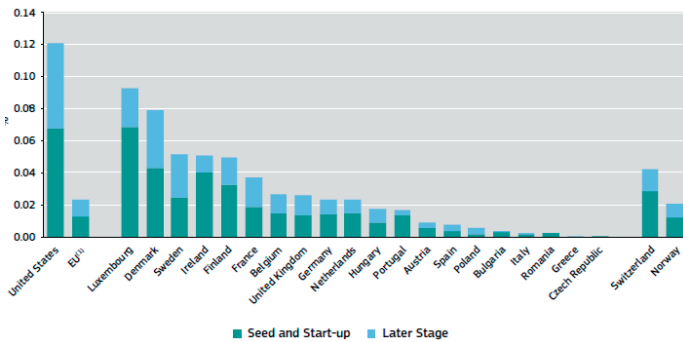


Figure 167: Venture capital as % of GDP, 2013
Source: European Commission

2.6.3.2.6. Lack of ICT-workers

Europe will have an estimated 825,000 vacancies for ICT professionals by 2020. Unfortunately, the number of European graduates in computing has been growing only by about 0.5% per year in the period 2007-2013, which is too low to fill all the vacancies. It is alarming that major countries like France, UK and Poland have a declining number of graduates (Figure 168).

Bringing a large number of well-trained foreign workers to Europe to help mitigate the shortage is not an effective solution. First of all, Europe needs more than one million ICT workers in the

next decade. Secondly, the countries of origin try hard to keep the local talent in their countries. Finally, Europe has recently become less inviting to foreigners. On top of that, the foreign ICT workers will be attracted by well-paid jobs in major cities, and it will be more difficult to convince them to accept a job in smaller cities, or in poorer countries. The only long-term and sustainable solution is to make maximum investment in the technical education of local people.

2.6.3.2.7. Fragmentation of funding

The public funding system in Europe is highly fragmented. There are national funds, regional funds and European funds. There are funding instruments for applied research, for innovation, and for fundamental research. There are individual grants and collaborative research grants. A particular research proposal could fit multiple funding instruments and calls. Sometimes, a research proposal can only be funded if different agencies agree to each fund part of the proposal. On top of this, the success rate for research proposals is sometimes lower than 10%.

Within a funding agency, different committees deal with particular topics, which makes multidisciplinary project proposals very hard to get funded because committees tend to give priority to the proposals that belong to the core of a domain and this leads to lower acceptance rates for interdisciplinary projects. The organizational structure of the funding agency is thus leading to constraints in the research work that can be proposed in one single project. The design of a novel, secure cloud based IoT solution will cut across the topics of at least three units of DG Connect. The fact that European Regional Development Funds are now also starting to be used to fund research only adds to the complexity.

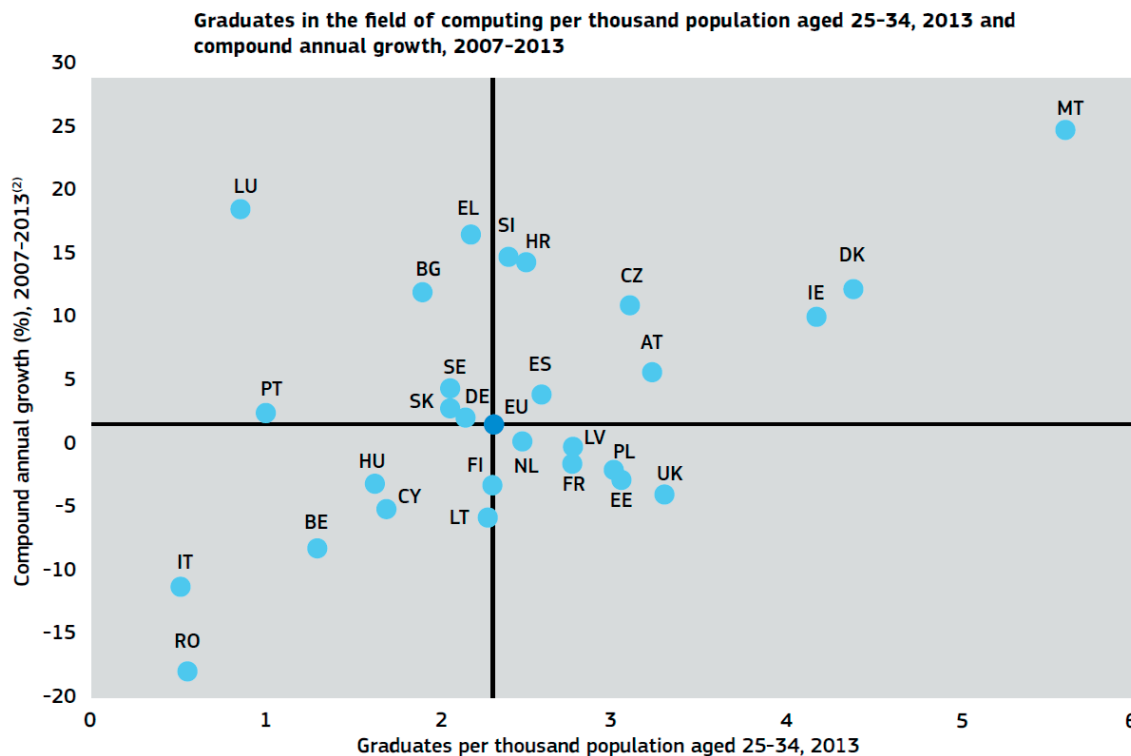


Figure 168: Graduates in computing
Source: DG Research and Innovation – Unit for Analysis and Monitoring of National Research Policies, European Commission

2.6.3.3. OPPORTUNITIES

2.6.3.3.1. The end of Moore's law

With respect to opportunities, the end of Moore's law is a clear opportunity for research. The increase of sequential performance at the pace of Moore's law already ended a decade ago; parallelism kicked in to keep the performance increasing in lockstep with number of transistors and cores, but now power consumption starts limiting the number of active cores. This means that the computing systems community has to start thinking outside the box, and come up with clever solutions to make the best use of the computing resources offered by the computing substrate and the available power envelope. Today, specialized accelerators seem to be the preferred solution. There is however room (and also a need) for more disruptive solutions, possibly replacing the (rather inefficient) von Neumann architecture by another computing paradigm.

2.6.3.3.2. Embedded systems, IoT, CPS

The number-one market opportunity in computing systems is the strongly growing market of embedded systems (including the IoT, CPS, and the digitization of European industry). Europe has the largest economy in the world, it has a larger household expenditure than the US, it has a number of world-class players producing the key enabling technology for advanced embedded systems, and it has strong automotive, health and aerospace industries. Furthermore, there are no dominating companies like Google, Apple, Facebook or Amazon (GAFA) in this space yet. The stars of the IoT era will probably not be the same as the ones of the Internet era (which are different from those in the mainframe era). Could the company dominating computing in 2030 be European?

2.6.3.3.3. Cybersecurity

Cybersecurity is a growing challenge, and it will become even bigger in the coming years. According to cybersecurity ventures [28], the cybersecurity market grew from US\$4 billion in 2004, to US\$75 billion by 2015, and it is forecasted to grow to US\$170 billion by 2020. This is comparable to the size of the global embedded systems market of a couple of years ago. The annual growth rate will be twice the growth rate of the embedded systems market, which makes it one of the fastest growing markets in computing.

On 20 June 2016, the European industry created ECSO (European Cyber Security Organization) with the objective of supporting all types of initiatives to develop, promote and encourage European Cybersecurity [236]. According to ECSO [237], the European cybersecurity market is about 25% of the global market while the North American market is 43%. The share of the global market secured by companies originating in Europe is only 8.5% (or 35% of the European market) and representing around 100,000 jobs. Given the importance of cybersecurity for the future, Europe needs to catch up. In July 2015, the European Commission signed a public private partnership with ECSO and will invest € 450 million in research and innovation via Horizon 2020. The objective is

to raise three times more investments from industry, leading to a total investment of € 1.8 billion by 2020.

2.6.3.3.4. Solutions for societal challenges

The societal challenges form a huge opportunity for the European computing industry. Europe is the region with the highest number of people aged 60 or older [342]. Only Japan has an older population. That means that Europe and Japan will have to search for solutions for the ageing population first. Since the rest of the world will face the same challenges in the future, Europe has an opportunity to develop and commercialize services and products for the elderly first and to sell them to the rest of the world.

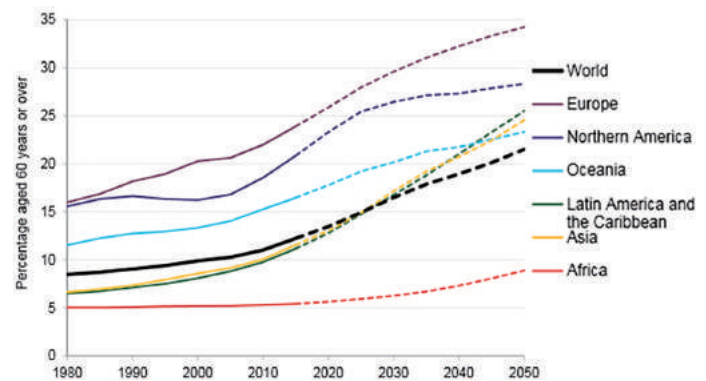


Figure 169: Percentage of world population aged 60 or over, 1980-2050

Source: World Population Prospects: The 2015 Review

The same reasoning holds for the environment. The European population (together with the US) has one of the largest ecological footprints of the world. Solutions for reducing our footprint may also work on other continents, and thus may create opportunities for European businesses.

2.6.3.4. THREATS

2.6.3.4.1. Financial crisis

Europe seems so far unable to find effective solutions to end the financial crisis and the economic stagnation. The lack of economic growth, decline of the middle class, and the growth of income inequality [302] put stress on the businesses and the governments. Current approaches have to be reassessed and replaced by more adequate solutions. If this stagnation keeps mainly affecting Europe, the continent could quickly lose its leading position in the global market.

The weakness of research is its dependence on investments by industry or governments. Low or no economic growth easily leads to cuts in R&D budgets, especially when these budgets are requested to fund long-term research that might not lead to short term results and new market opportunities.

2.6.3.4.2. Saturating markets

The market of desktop computers and laptops is shrinking, and the market of smart phones is bound to be shrinking too (after

having cannibalized the market of other devices like navigation systems, cameras, music and video players). This puts pressure on the companies to cut costs and jobs, and to focus on short-term results instead of mid-term innovations or long-term research.

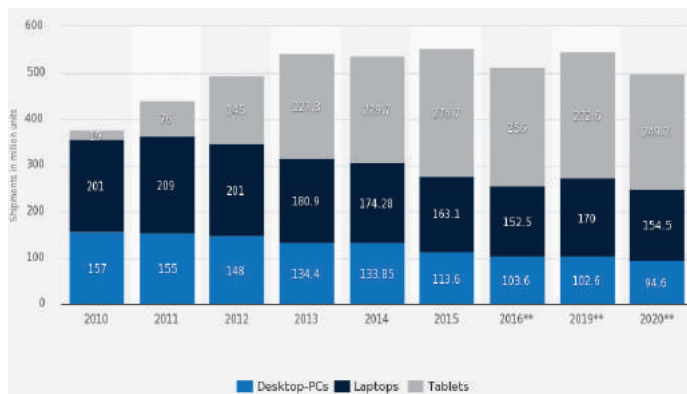


Figure 170: Shipment forecast by device, 2010-2020

Source: Statista

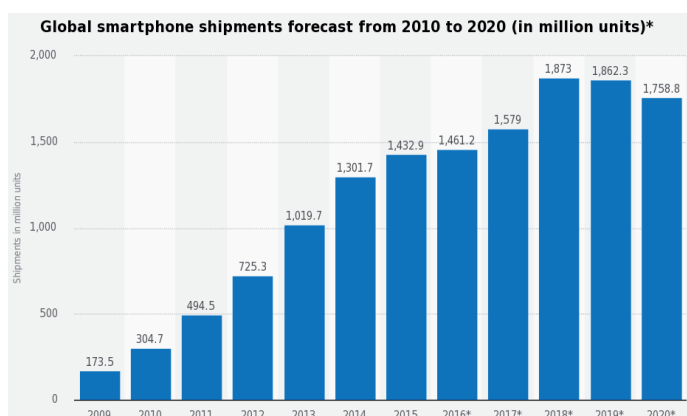


Figure 171: Smartphone shipment forecast, 2010-2020

Source: Statista

2.6.3.4.3. Computing initiatives in China, Russia, Japan

A threat to the European computing industry is the rapid development of the computing industry in China, Russia and Japan. Many countries understand that computing is a key-enabling technology of strategic importance, and invest in their own research, products and companies (see section 2.3.12). If Europe fails to do the same, it might eventually become dependent on technology which is designed, developed, produced and controlled outside Europe. The same holds for the cybersecurity solutions. As shown in the picture below indicating total R&D expenditure (public and business), R&D expenditure in China is growing very fast, which will eventually result in more innovative products and services brought to the market, and hence more players in this competitive market (Figure 173).

Chinese companies are growing fast, without the rest of the world necessarily noticing.

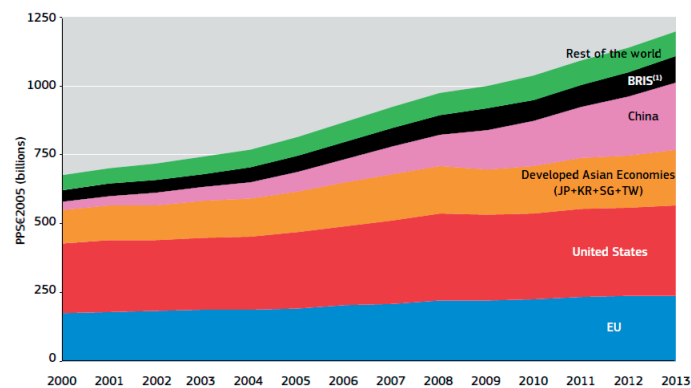


Figure 172: Evolution of world GERD, 2000-2013

Source: European Commission

		Value (billion \$)	Turnover (billion \$)	Profit (billion \$)	Personnel	User/ Units sold (million)
Social network	Facebook	370.0	17.9	3.7	14,495	1130
	Tencent	252.0	15.8	4.8	30,160	697
Web shop	Amazon	354.0	107.0	0.6	230,800	304
	Alibaba	242.0	15.7	11.1	36,450	410
Search engine	Alphabet	539.0	73.6	15.8	66,575	1000
	Baidu	64.0	10.6	5.3	41,467	667
Microblog	Twitter	12.6	2.2	0.5	3,900	313
	Weibo	9.8	0.5	0.3	6,400	282
Taxi	Uber	68.0	0.5	-1.0	6,700	8
	Didi Chuxing	33.8			6,000	250
Electric car	Tesla	29.7	4.1	-0.9	14,000	0.080
	BYD	21.5	11.6	0.4	200,000	0.120
Smartphone	Apple	595.7	231.3	53.4	110,000	1000
	Huawei		60.8	5.6	176,000	290

Figure 173: Vital statistics of leading global technology companies

It is certain that several of them will enter the top 10 of most valuable technology companies in the next decade and that several of the current top 10 companies will be disrupted by companies that are still under the radar at present [218].

2.6.3.4.4. China is building a huge patent portfolio

The number of patent filings at the State Intellectual Property Office (SIPO) of China is increasing exponentially and in 2011 surpassed all other patent offices in the world in number of filings. This strategy could make it in the future more difficult for foreign companies to introduce new products to the Chinese market [345].

2.6.3.4.5. Political instability (Brexit, immigration crisis, ...)

Another threat is the political instability that Europe and the world are currently experiencing. Terrorist attacks, Brexit, unforecasted results for upcoming elections in various countries, financial problems and the refugee crisis influence business and consumer confidence. The current uncertainty on how and when the UK will leave the European Union is having an impact on the international relations of certain UK companies and universities.

From 1883 to 1963, the USPTO was the leading office in world filings. Application numbers at the JPO and the USPTO were stable until the early 1970s, when the JPO began to see rapid growth, a pattern also observed for the USPTO from the 1980s onwards.

Among the top five offices, the JPO surpassed the USPTO in 1968 and maintained the top position until 2005. Since 2006,

the number of applications at the JPO has trended downward. Both the EPO and KIPO have seen increases each year since the early 1980s, as has SIPO since 2001. SIPO surpassed the EPO and KIPO in 2005, the JPO in 2010 and the USPTO in 2011 – and it now receives the largest number of applications worldwide. There has been a gradual upward trend in the combined share of the top five offices in the world total – from 70% in 2000 to 82% in 2014.

Trend in patent applications for the top five offices

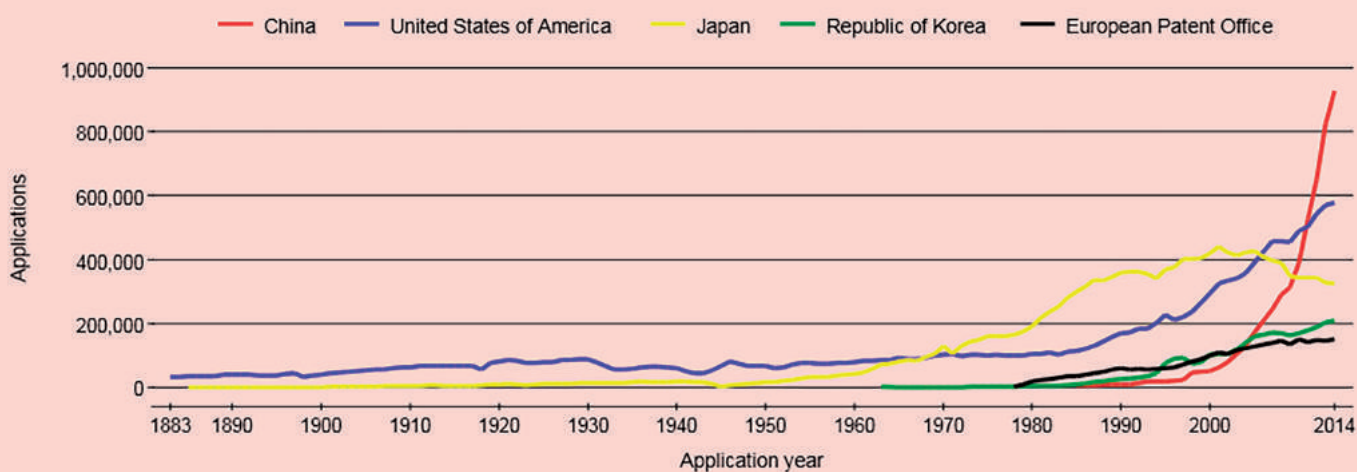


Figure 174: Patent filings since 1883

Source: European Patent Office

3

GLOSSARY

AGI	Artificial General Intelligence
API	Application Programming Interface
ASIC	Application-Specific Integrated Circuits are integrated circuits designed for a particular purpose, as opposed to being applicable for general use in many different situations.
Bayesian computing	Bayesian computing refers to computational methods that are based on Bayesian (probabilistic) statistics.
CAGR	Compound annual growth rate is a specific business and investing term for the smoothed annualised gain of an investment over a given time period.
Cloud computing	Cloud computing is a paradigm whereby computing power is abstracted as a virtual service over a network. Executed tasks are transparently distributed.
CMOS	Complementary Metal–Oxide–Semiconductor is a common technology for constructing integrated circuits. CMOS technology is used in microprocessors, microcontrollers, static RAM, and other digital logic circuits.
CPS	Cyber-Physical Systems combine computing resources and sensors/actuators that directly interact with and influence the real world. Robotics is one of the primary fields that works on such systems.
CPU	Central Processing Unit
Data analytics	Data analytics examines large amounts of data to uncover hidden patterns, correlations and other insights.
Declarative programming	Declarative programming is a programming paradigm that expresses the logic of a computation without describing its control flow. Many languages applying this style attempt to minimize or eliminate side effects by describing what the program should accomplish, rather than describing how to go about accomplishing it (the how is left up to the language's implementation). The opposite concept is imperative programming.
Edge computing	Edge Computing is pushing the frontier of computing applications, data, and services away from centralized nodes to the logical extremes of a network. It enables analytics and knowledge generation to occur at the source of the data.
EUV	Extreme ultraviolet lithography is a next-generation lithography technology using an extreme ultraviolet (EUV) wavelength, currently expected to be 13,5 nm.
FDSOI	Fully Depleted Silicon On Insulator (MOSFETs). For a FDSOI MOSFET the sandwiched p-type film between the gate oxide (GOX) and buried oxide (BOX) is very thin so that the depletion region covers the whole film. In FDSOI the front gate (GOX) supports less depletion charges than the bulk transistors so an increase in inversion charges occurs resulting in higher switching speeds. Other drawbacks in bulk MOSFETs, like threshold voltage roll off, higher sub-threshold slop body effect, etc. are reduced in FDSOI since the source and drain electric fields cannot interfere, due to the BOX (adapted from Wikipedia).

FinFet	The term FinFet was coined by University of California, Berkeley researchers (Profs. Chenming Hu, Tsu-Jae King-Liu and Jeffrey Bokor) to describe a nonplanar, double-gate transistor built on an SOI substrate.... The distinguishing characteristic of the FinFet is that the conducting channel is wrapped by a thin silicon 'fin', which forms the body of the device. In the technical literature, FinFet is used somewhat generically to describe any fin-based, multigate transistor architecture regardless of number of gates (from Wikipedia).
Fog computing	Fog computing is an architecture that uses one or more end-user clients or near-user edge devices to carry out a substantial amount of storage (rather than stored primarily in cloud data centres), communication (rather than routed over the internet backbone), control, configuration, measurement and management.
FPGA	Field-Programmable Gate Array
Generative Design	Generative design is a technology that starts with design goals and then explores all of the possible permutations of a solution to find the best option. The process lets designers generate brand new options, beyond what a human alone could create, to arrive at the most effective design.
GPU	A Graphics Processing Unit refers to the processing units on video cards. In recent years, these have evolved into massively parallel execution engines for floating point vector operations, reaching performance peaks of several gigaflops.
HiPEAC	The European Network of Excellence on High Performance and Embedded Architecture and Compilation coordinates research, facilitates collaboration and networking, and stimulates commercialization in the areas of computer hardware and software research.
Homomorphic encryption	Homomorphic systems send encrypted data to an application (generally executed on a remote server) and let application perform its operations without ever decrypting the data. As a result the application never knows the actual data, nor the results it computes.
ICT	Information & Communication Technology is a generic term used to refer to all areas of technology related to computing and telecommunications.
Imperative programming	Imperative programming is a programming paradigm that describes computation in terms of statements that change a program state. In much the same way that the imperative tense in natural languages expresses commands to take action, imperative programs define sequences of commands for the computer to perform. The opposite concept is declarative programming.
Internet of Things	The Internet of Things (IoT) is a computing concept that describes a future where everyday physical objects will be connected to the Internet and will be able to identify themselves to other devices.
IP block	Intellectual property block, is a reusable unit of logic, cell, or chip layout design that is the intellectual property of one party. IP cores may be licensed to another party or can be owned and used by a single party alone. IP blocks can be used as building blocks within ASIC chip designs or FPGA logic designs.
ISA	An Instruction Set Architecture is the definition of the machine instructions that can be executed by a particular family of processors.
LIDAR	<i>Light Detection And Ranging</i> is a technology that measures distance by illuminating a target with a laser.
MCU	Micro Controller Unit
MPU	Micro Processor Unit
NAS	Network attached storage
Neural networks	Neural networks are computational entities that operate in a way that is inspired by how neurons and synapses in an organic brain are believed to function. They need to be trained for a particular application, during which their internal structure is modified until they provide adequately accurate responses for given inputs.

Neuromorphic	Analog, digital, or mixed-mode analogue/digital VLSI and software systems that implement models of neural systems.
NRE	Non-Recurring Engineering costs refer to one-time costs incurred for the design of a new chip, computer program or other creation, as opposed to marginal costs that are incurred per produced unit.
OLED	An organic light-emitting diode (OLED) is a light-emitting diode (LED) in which the emissive electroluminescent layer is a film of organic compound that emits light in response to an electric current.
Programming model	A programming model is a collection of technologies and semantic rules that enable the expression of algorithms in an efficient way. Often, such programming models are geared towards a particular application domain, such as parallel programming, real-time systems, image processing ...
Pseudo-quantum computing	Pseudo-quantum computing is a term used to refer to machines that allegedly are quantum computers, but that in practice have not been proven to be actually faster than regular computers executing very optimized algorithms.
QoS	Quality of Service
Reservoir computing	Reservoir Computing is similar to neural networks, but rather than modifying the internal structure during the training phase, the way to interpret the output is adjusted until the desired accuracy has been obtained.
RFID	Radio-Frequency Identification is the use of a wireless non-contact system that uses radio-frequency electromagnetic fields to transfer data from a tag attached to an object, for the purposes of automatic identification and tracking.
SAN	Storage area network, a dedicated network that connects a set of storage devices that are able to share low-level data with each other.
SME	Small and Medium-sized Enterprise, a company of up to 250 employees.
SoC	A System on Chip refers to integrating all components required for the operation of an entire system, such as processors, memory, and radio, on a single chip.
Spike computations	A programming model where large collections of devices, modelled after neurons, interact through the transmission of spike signals.
STDP	Spike-Timing-Dependent Plasticity is a biological process that adjusts the strength of connections between neurons in the brain. The process adjusts the connection strengths based on the relative timing of a particular neuron's input and output action potentials (or spikes).
Streaming analytics	Streaming analytics, also called event stream processing, is the analysis of large, in-motion data called event streams. The growing number of connected devices – the Internet of Things – will exponentially increase the volume of events that surround business activity. The more data is generated, the greater the potential benefits from streaming analytics.
TRL	Technology Readiness Level
TSV	Through Silicon Via, a (vertical) electrical interconnect that goes through a silicon die or wafer (“via” = vertical interconnect access).
UML	Unified Modelling Language is a general-purpose, developmental, modelling language in the field of software engineering, that is intended to provide a standard way to visualize the design of a system.
VLSI	Very-large-scale integration is the process of creating integrated circuits by combining thousands of transistors into a single chip.

4

REFERENCES

(Note: all web references were available as of November 2016)

- [1] A. Aliane et al. Large area printed temperature sensors on flexible substrate. 5th IEEE International Workshop on Advances in Sensors and Interfaces (IWASI), June 2013.
- [2] https://www.whitehouse.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf
- [3] A. Nathan, A. Ahnood, M. T. Cole, L. Sungsik, Y. Suzuki, P. Hiralal, et al. Flexible electronics: The next ubiquitous platform. Proceedings of the IEEE, vol. 100, pp. 1486–1517, 2012.
- [4] C.H. Bennett, and G. Brassard. Advances in Cryptology. Proceedings of Crypto’84, 1984.
- [5] B. Belhadj, A. Joubert, Zh. Li, R. Hélot, and O. Temam. Continuous real-world inputs can open up alternative accelerator designs. SIGARCH Comput. Archit. News 41, 3 (June 2013), pp. 1–12, 2013.
- [6] Software tools for next generation computing, Brussels 24 June 2014. <https://ec.europa.eu/digital-single-market/news/software-tools-next-generation-computing-o>.
- [7] C. Gentry. Fully homomorphic encryption using ideal lattices. In Proc. STOC’09, pp. 169–178, 2009.
- [8] Changmoo Kim, Mookyoung Chung, Yeongon Cho, Mario Konijnenburg, Soojung Ryu, and Jeongwook Kim. 2014. ULP-SRP: Ultra low-power Samsung reconfigurable processor for biomedical applications. ACM Trans. Reconfig. Technol. Syst. 7, 3, Article 22 (August 2014), 15 pages. DOI: <http://dx.doi.org/10.1145/2629610>.
- [9] OECD. Commercialising Public Research, New Trends and Strategies. 2013
- [10] H. Esmaeilzadeh, et al. Dark silicon and the end of multicore scaling. 38th Annual International Symposium on Computer Architecture (ISCA), 2011.
- [11] S. Debnath, N. M. Linke, C. Figgatt, K. A. Landsman, K. Wright, and C. Monroe. Demonstration of a small programmable quantum computer with atomic qubits. Nature, pp. 63–66, 2016.
- [12] E. W. Dijkstra. The humble programmer. Communications of the ACM, Vol. 15, Num. 10, pp. 859–866, 1972.
- [13] D. E. Nikonow and I. A. Youg. Benchmarking of Beyond-CMOS Exploratory Devices for Logic Integrated Circuits. IEEE Journal on Exploratory Solid-State Computational Devices and Circuits, July 2015.
- [14] B. Meyer. Eiffel, The Language. Prentice Hall, 1991. ISBN-13: 978-0132479257.
- [15] R. P. Feynman. Simulating physics with computers. International journal of theoretical physics 21.6, pp. 467–488, 1982.
- [16] A. Smith, AI, Robotics, and the Future of Jobs, August 2014. <http://www.pewinternet.org/2014/08/06/future-of-jobs/>
- [17] OpenAI. <https://openai.com/about>
- [18] C. Mooney, 'Range anxiety' is scaring people away from electric cars – but the fear may be overblown, October 2016. https://www.washingtonpost.com/news/energy-environment/wp/2016/08/15/range-anxiety-scares-people-away-from-electric-cars-why-the-fear-could-be-overblown/?utm_term=.50e8da396650
- [19] M. Pickavet et al., “Worldwide energy needs for ICT: The rise of power-aware networking,” 2008 2nd International Symposium on Advanced Networks and Telecommunication Systems, Mumbai, 2008, pp. 1-3. doi:10.1109/ANTS.2008.4937762
- [20] H. D. Lee et al. Integration of 4F2 selector-less crossbar array 2Mb ReRAM based on transition metal oxides for high density memory applications. In 2012 Symposium on VLSI Technology (VLSIT), pp. 151–152, 2012.
- [21] H.-E. Nilsson et al. System Integration of Electronic Functions in Smart Packaging Applications. IEEE Transactions on Components, Packaging and Manufacturing Technology, Vol. 2, No. 10, Oct. 2012.
- [22] BeagleBoard.org Foundation. <http://beagleboard.org>
- [23] Dot-Bit. <http://bit.namecoin.info>
- [24] A. Linn. Microsoft releases CNTK, its open source deep learning toolkit, on GitHub. Microsoft, January 2016. <http://blogs.microsoft.com/next/2016/01/25/microsoft-releases-cntk-its-open-source-deep-learning-toolkit-on-github>
- [25] A. Linn. Project Malmo: Using Minecraft to build more intelligent technology. Microsoft, March 2016. <http://blogs.microsoft.com/next/2016/03/13/project-aix-using-minecraft-build-intelligent-technology>
- [26] J. Neidlinger. The Horrifying Truth About Multitasking And Productivity. CoSchedule Blog, November 2014. <http://coschedule.com/blog/multitasking-and-productivity/>
- [27] D. J. Bernstein. Cost analysis of hash collisions: Will quantum computers make SHARCS obsolete? SHARCS’09 Special-purpose Hardware for Attacking Cryptographic Systems, 2009.

- [28] S. C. Morgan. Cybersecurity Market Report. Cybersecurity Ventures, 2016. <http://cybersecurityventures.com/cybersecurity-market-report/>
- [29] J. Planas, R. M. Badia, E. Ayguadé, and J. Labarta. Hierarchical Task-Based Programming With StarSs. *Int. J. High Perform. Comput. Appl.* 23, 3, pp. 284–299, 2009.
- [30] J. Van Cleemput, B. Coppens, and B. De Sutter. Compiler mitigations for time attacks on modern x86 processors. *ACM Trans. Archit. Code Optim.* 8, 4, Article 23, January 2012.
- [31] Dogecoin. <http://dogecoin.com>
- [32] F. Schroth. Keyshare Exec on Entering the Competitive Drone Market. *Dronelife*, June 2016. <http://dronelife.com/2016/06/29/keyshare-exec-entering-competitive-drone-market>
- [33] M. McNabb. The First Dominoes Pizza Delivered by Drone. *Dronelife*, August 2016. <http://dronelife.com/2016/08/26/first-dominoes-pizza-delivered-drone>
- [34] European Commission. 2020 climate & energy package. Last update November 2016. http://ec.europa.eu/clima/policies/strategies/2020/index_en.htm
- [35] European Commission. Open Source Software Strategy 2014-2017. Last update June 2016. http://ec.europa.eu/dgs/informatics/oss_tech/strategy/strategy_en.htm
- [36] European Commission. European Innovation Scoreboard 2016. July 2016. <http://ec.europa.eu/DocsRoom/documents/17822>
- [37] European Commission. e-Skills for Jobs in Europe. 2014. <http://ec.europa.eu/DocsRoom/documents/4992/attachments/1/translations/en/renditions/native>
- [38] European Commission. Ageing and welfare state policies. Last update July 2015. http://ec.europa.eu/economy_finance/structural_reforms/ageing/index_en.htm
- [39] EMR Health Alliance of BC. Zigbee & Smart Appliances. March 2015. http://emrabc.ca/?page_id=4884
- [40] The State Council of the People's Republic of China. China's Sunway-TaihuLight named world's fastest supercomputer. June 2016. http://english.gov.cn/news/photos/2016/06/20/content_281475376099575.htm
- [41] The State Council of the People's Republic of China. Most Chinese students return to China for jobs after studying abroad. 2016. http://english.gov.cn/news/video/2016/03/28/content_281475316337701.htm
- [42] Wikipedia. Communications of the ACM. https://en.wikipedia.org/w/index.php?title=Communications_of_the_ACM&oldid=728809419
- [43] Wikipedia. Dennard scaling. https://en.wikipedia.org/w/index.php?title=Dennard_scaling&oldid=736493278
- [44] Wikipedia. Edsger W. Dijkstra. https://en.wikipedia.org/w/index.php?title=Edsger_W._Dijkstra&oldid=749656033
- [45] J. Fernandez-Ramil, D. Izquierdo-Cortazar, and T. Mens. What does it take to develop a million lines of open source code? IFIP International Conference on Open Source Systems. Springer Berlin Heidelberg, 2009.
- [46] C. Morris. Ordinary Home Appliances Are About to Get Really Sexy. *Fortune*, January 2016. <http://fortune.com/2016/01/06/home-appliances-ces-2016>
- [47] R. Salomon. Why Uber Couldn't Crack China. *Fortune*, August 2016. <http://fortune.com/2016/08/07/uber-china-didi-chuxing>
- [48] Future of Life Institute. An Open Letter: Research Priorities for Robust and Beneficial Artificial Intelligence. <http://futureoflife.org/ai-open-letter>
- [49] Vinclu Inc. Gatebox. 2016. <http://gatebox.ai/en>
- [50] N. Thibieroz. It's Time to Open up the GPU. *GPUOpen*, January 2016. <http://gpuopen.com/welcometogpuopen>
- [51] B. Linder. Yes, you can run desktop Linux apps in Windows 10 thanks to Ubuntu on Windows. *Liliputing*, April 2016. <http://liliputing.com/2016/04/yes-can-run-desktop-linux-apps-windows-10-thanks-ubuntu-windows.html>
- [52] A. Allan. ESP8266: This \$5 Microcontroller with Wi-Fi is now Arduino-Compatible. *Make.*, April 2015. <http://makezine.com/2015/04/01/esp8266-5-microcontroller-wi-fi-now-arduino-compatible>
- [53] D. Beres. Autonomous cars will make your data plan look tiny. *Mashable*, August 2016. <http://mashable.com/2016/08/17/intel-autonomous-car-data/#XVApWHDc6Oqt>
- [54] D. Genkin et al. Physical key extraction attacks on PCs. *Communications of the ACM* 59.6, pp. 70–79, 2016. <http://m.cacm.acm.org/magazines/2016/6/202646-physical-key-extraction-attacks-on-pcs/fulltext>
- [55] MIAUOW GPU. <http://miaowgpu.org>
- [56] D. Rossignol. POKÉMON GO Brought 2,000 Players Together at the Sydney Opera House. *Nerdist*, July 2016. <http://nerdist.com/pokemon-go-brought-2000-players-together-at-the-sydney-opera-house>
- [57] L. Hardesty. Toward hack-proof RFID chips, *MIT News*, February 2016. <http://news.mit.edu/2016/hack-proof-rfid-chips-0203>
- [58] OpenPOWER Foundation. <http://openpowerfoundation.org/>
- [59] A. Murabayashi. We Don't Understand Privacy. *PetaPixel*, April 2016. <http://petapixel.com/2016/04/13/dont-understand-privacy/>
- [60] M. Schirber. Focus: Graphyne May Be Better than Graphene. *Physics*, February 2012. <http://physics.aps.org/articles/v5/24>
- [61] The Arduino Playground. <http://playground.arduino.cc>
- [62] A. Lewis, M. Kraemer, and M. Travagnin. Quantum Technologies: Implications for European Policy. Issues for debate. Publications Office of the European Union, 2016. <http://publications.jrc.ec.europa.eu/repository/handle/JRC101632>

- [63] O. Staley. Harvey Mudd College took on gender bias and now more than half its computer-science majors are women. Quartz, August 2016. <http://qz.com/730290/harvey-mudd-college-took-on-gender-bias-and-now-more-than-half-its-computer-science-majors-are-women/>
- [64] RISC-V Foundation. <http://riscv.org/>
- [65] ARTEMIS Industry Association. ARTEMIS Strategic Research Agenda 2016. 2016. <https://artemis-ia.eu/publication/download/sra2016.pdf>
- [66] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM review 41.2, pp. 303–332, 1999.
- [67] S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system. 2008.
- [68] S. Niyogi. Microsoft joins the Eclipse Foundation and brings more tools to the community. The Visual Studio Blog, March 2016. <https://blogs.msdn.microsoft.com/visualstudio/2016/03/08/microsoft-joins-the-eclipse-foundation>
- [69] NVIDIA. The New NVIDIA TITAN X: The Ultimate. Period. July 2016. <https://blogs.nvidia.com/blog/2016/07/21/titan-x>
- [70] The PRET Project. PRecision Timerd (PRET) Machines. Center for Hybrid and Embedded Software Systems, 2014. <https://chess.eecs.berkeley.edu/pret>
- [71] N. Jouppi. Google supercharges machine-learning tasks with TPU custom chip. Google Cloud Platform Blog, May 2016. <https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html>
- [72] K. Lee. Facebook to open-source AI hardware design. Facebook, December 2015. <https://code.facebook.com/posts/1687861518126048/facebook-to-open-source-ai-hardware-design>
- [73] W. Abbey. ARM IP and Intel Custom Foundry collaboration: A new era for premium mobile design. ARM Connected Community Blog, August 2016. <https://community.arm.com/groups/processors/blog/2016/08/16/arm-ip-and-intel-custom-foundry-collaboration-a-new-era-for-premium-mobile-design>
- [74] CureCoin. <https://curecoin.net>
- [75] Microsoft. Kinect. Microsoft Developer Resources, 2016. <https://developer.microsoft.com/en-us/windows/Kinect>
- [76] Sony. Sony SmartEyeglass at CES, Jan 6–9th. Sony Developer World, 2016. <https://developer.sony.com/2015/12/22/sony-smarteyeglass-at-ces-jan-6-9th/>
- [77] M. Rouse. Edge Computing. SearchDataCenter, 2016. <http://searchdatacenter.techtarget.com/definition/edge-computing>
- [78] European Commission. Next Generation Computing Roadmap. August 2014. <https://ec.europa.eu/digital-single-market/en/news/next-generation-computing-roadmap>
- [79] European Commission. Policies for Ageing Well with ICT. January 2016. <https://ec.europa.eu/digital-single-market/en/policies-ageing-well-ict>
- [80] European Commission. Towards 5G. April 2016. <https://ec.europa.eu/digital-single-market/en/towards-5g>
- [81] European Commission. The ICT sector represents 4.8% of the European economy. <https://ec.europa.eu/programmes/horizon2020/en/area/ict-research-innovation>
- [82] P. Paganini. OVH hosting hit by 1Tbps DDoS attack, the largest one ever seen. Security Affairs, September 2016. <http://securityaffairs.co/wordpress/51640/cyber-crime/tbps-ddos-attack.html>
- [83] P. Paganini. The hosting provider OVH continues to face massive DDoS attacks launched by a botnet composed at least of 150000 IoT devices. Security Affairs, September 2016. <http://securityaffairs.co/wordpress/51726/cyber-crime/ovh-hit-botnet-iot.html>
- [84] Bitcoin Wiki. Category: Mining contractors. May 2016. https://en.bitcoin.it/w/index.php?title=Category:Mining_contractors&oldid=61047
- [85] Wikipedia. Arduino. <https://en.wikipedia.org/w/index.php?title=Arduino&oldid=750568017>
- [86] Wikipedia. Artificial Intelligence. https://en.wikipedia.org/w/index.php?title=Artificial_intelligence&oldid=751080647
- [87] Wikipedia. Authoritarianism. <https://en.wikipedia.org/w/index.php?title=Authoritarianism&oldid=745604461>
- [88] Wikipedia. Automotive Head-up Display. https://en.wikipedia.org/w/index.php?title=Automotive_head-up_display&oldid=729228400
- [89] Wikipedia. Autonomous Car. https://en.wikipedia.org/w/index.php?title=Autonomous_car&oldid=750932349
- [90] Wikipedia. BBC Micro. https://en.wikipedia.org/w/index.php?title=BBC_Micro&oldid=747912382
- [91] Wikipedia. BeagleBoard. https://en.wikipedia.org/w/index.php?title=BeagleBoard&oldid=749399982#BeagleBone_Black
- [92] Wikipedia. D-Wave Systems. https://en.wikipedia.org/w/index.php?title=D-Wave_Systems&oldid=750902993
- [93] Wikipedia. Field-Programmable Gate Array. https://en.wikipedia.org/w/index.php?title=Field-programmable_gate_array&oldid=750223949
- [94] Wikipedia. Fog Computing. https://en.wikipedia.org/w/index.php?title=Fog_computing&oldid=751138994
- [95] Wikipedia. Google Glass. https://en.wikipedia.org/w/index.php?title=Google_Glass&oldid=751023525

[96] Wikipedia. In-memory Processing. https://en.wikipedia.org/w/index.php?title=In-memory_processing&oldid=748875982

[97] Wikipedia. Loongson. <https://en.wikipedia.org/w/index.php?title=Loongson&oldid=750324425>

[98] Wikipedia. LyteShot. <https://en.wikipedia.org/w/index.php?title=LyteShot&oldid=730308488>

[99] Wikipedia. Machine Translation. https://en.wikipedia.org/w/index.php?title=Machine_translation&oldid=747383439#History

[100] Wikipedia. Massive Open Online Course. https://en.wikipedia.org/w/index.php?title=Massive_open_online_course&oldid=750942132

[101] Wikipedia. Micro_Bit. https://en.wikipedia.org/w/index.php?title=Micro_Bit&oldid=750030448

[102] Wikipedia. Microsoft_HoloLens. https://en.wikipedia.org/w/index.php?title=Microsoft_HoloLens&oldid=750373123

[103] Wikipedia. Multiple Independent Levels of Security. https://en.wikipedia.org/w/index.php?title=Multiple_Independent_Levels_of_Security&oldid=702651787

[104] Wikipedia. OpenSPARC. <https://en.wikipedia.org/w/index.php?title=OpenSPARC&oldid=741234869>

[105] Wikipedia. Pok%C3%A9mon_Go. https://en.wikipedia.org/w/index.php?title=Pok%C3%A9mon_Go&oldid=750882721

[106] Wikipedia. Raspberry Pi. https://en.wikipedia.org/w/index.php?title=Raspberry_Pi&oldid=750769719

[107] Wikipedia. Smartglasses. <https://en.wikipedia.org/w/index.php?title=Smartglasses&oldid=750939001>

[108] Wikipedia. Smartwatch. <https://en.wikipedia.org/w/index.php?title=Smartwatch&oldid=751013127>

[109] Wikipedia. Tensor Processing Unit. https://en.wikipedia.org/w/index.php?title=Tensor_processing_unit&oldid=738196720

[110] Wikipedia. TrueNorth. <https://en.wikipedia.org/w/index.php?title=TrueNorth&oldid=743413538>

[111] Wikipedia. Weak AI. https://en.wikipedia.org/w/index.php?title=Weak_AI&oldid=742631291

[112] European Union. Measuring the EU's economy. Statistics for 2014. https://europa.eu/european-union/about-eu/figures/economy_en#tab-6-12

[113] Follow My Vote. Blockchain Technology in Online Voting. 2016. <https://followmyvote.com/online-voting-technology/blockchain-technology/>

[114] Wikipedia. Cyberpunk. <https://en.wikipedia.org/w/index.php?title=Cyberpunk&oldid=749926594>

[115] Git. <https://git-scm.com>

[116] A. Zaharia. 10 Alarming Cyber Security Facts that Threaten Your Data [Updated]. Heimdal Security, May 2016. <https://heimdalsecurity.com/blog/10-surprising-cyber-security-facts-that-may-affect-your-online-safety/>

[117] A. Frank. The World Depends on Technology No One Understands. SingularityHub, July 2017. <http://singularityhub.com/2016/07/17/the-world-will-soon-depend-on-technology-no-one-understands/>

[118] B. Krebs. The Democratization of Censorship. Krebs on Security, September 2016. <https://krebsonsecurity.com/2016/09/the-democratization-of-censorship>

[119] A. Courbot. [PREVIEW]

[120] Lotecoin. <https://litecoin.com>

[121] S. Schubert. Self Built MC HCK for \$5. mchck.org, August 2013. <https://mchck.org/blog/2013-08-06-self-built-mchck-for-5-dollars>

[122] B. Bozhanov. Bulgaria Got a Law Requiring Open Source. The Policy, July 2016. <https://medium.com/@bozhobg/bulgaria-got-a-law-requiring-open-source-98bf626cf70a#.bdse7wtjh>

[123] M. Hearn. The resolution of the Bitcoin experiment. Medium, January 2016. <https://medium.com/@octskyward/the-resolution-of-the-bitcoin-experiment-dabb30201f7>

[124] METIS II. <https://metis-ii.5g-ppp.eu/>

[125] R. Sanders. Sprinkling of neural dust opens door to electroceuticals. Berkely News, August 2016. <https://news.berkeley.edu/2016/08/03/sprinkling-of-neural-dust-opens-door-to-electroceuticals/>

[126] Intel. Intel Announces Tools for RealSense Technology Development. August 2016. <https://newsroom.intel.com/chip-shots/intel-announces-tools-realsense-technology-development>

[127] Intel. Make Amazing Things Happen in IoT and Entrepreneurship with Intel Joule. August 2016. <https://newsroom.intel.com/chip-shots/make-amazing-things-happen-iot-entrepreneurship-intel-joule>

[128] B. Krzanich. Brian Krzanich: Our Strategy and The Future of Intel. Intel, April 2016. <https://newsroom.intel.com/editorials/brian-krzanich-our-strategy-and-the-future-of-intel>

[129] Peercoin. <https://peercoin.net>

[130] S. Chintala. FAIR open sources deep-learning modules for Torch. Research at Facebook, January 2015. <https://research.facebook.com/blog/fair-open-sources-deep-learning-modules-for-torch>

[131] A. Waterman, Y. Lee, D. Patterson, and K. Asanović. The RISC-V Instruction Set Manual. RISC-V, May 2016. <https://riscv.org/specifications>

- [132] T. Scott. A. E. Rung. Federal Source Code Policy. US Government, 2016. <https://sourcecode.cio.gov/>
- [133] Starting Electronics. Arduino Programming Course. September 2014. <https://startingelectronics.org/software/arduino/learn-to-program-course>
- [134] Apple. About Touch ID security on iPhone and iPad. November 2015. <https://support.apple.com/en-us/HT204587>
- [135] Google. Transale images. 2012. https://support.google.com/translate/answer/6142483#languages_available
- [136] R. Dillet. Coding school 42 plans to educate 10,000 students in Silicon Valley for free. TechCrunch, May 2016. <https://techcrunch.com/2016/05/17/coding-school-42-plans-to-educate-10000-students-in-silicon-valley-for-free/>
- [137] J. Constine. Facebook open sources Surround 360 camera with Ikea-style instructions. July 2016. <https://techcrunch.com/2016/07/26/if-you-source-it-they-will-build>
- [138] S. Biddle. Privacy Scandal Haunts Pokemon Go's CEO. The Intercept, August 2016. <https://theintercept.com/2016/08/09/privacy-scandal-haunts-pokemon-gos-ceo/>
- [139] Unity. Company Facts. <https://unity3d.com/public-relations>
- [140] Oculus VR, LLC/Facebook. Oculus Rift. <https://www3.oculus.com/en-us/rift/>
- [141] Amazon. Amazon Prime Air. <https://www.amazon.com/b?node=8037720011>
- [142] Arduino. <https://www.arduino.cc>
- [143] Arduino. Language Reference. <https://www.arduino.cc/en/Reference/HomePage>
- [144] M. Vanhoef and T. Van Goethem. HEIST: HTTP Encrypted Information can be Stolen Through TCP-Windows. Black Hat USA 2016, August 2016. <https://www.blackhat.com/us-16/briefings/schedule/#heist-http-encrypted-information-can-be-stolen-through-tcp-windows-3379>
- [145] Canalys. Tablets down 20% in Q3 after four quarters of negative PC growth. October 2015. <https://www.canalys.com/newsroom/tablets-down-20-q3-after-four-quarters-negative-pc-growth>
- [146] CB Insights. Artificial Intelligence Explodes: New Deal Activity Record For AI Startups. CB Insights Blog, June 2016. <https://www.cbinsights.com/blog/artificial-intelligence-funding-trends>
- [147] CB Insights. The Race For AI: Google, Twitter, Intel, Apple In A Rush To Grab Artificial Intelligence Startups. CB Insights Blog, October 2016. <https://www.cbinsights.com/blog/top-acquirers-ai-startups-ma-timeline>
- [148] D. Genkin, L. Pachmanov, I. Pipman, and E. Tromer. ECDH Key-Extraction via Low-Bandwidth Electromagnetic Attacks on PCs. In RSA Conference Cryptographers' Track (CT-RSA) 2016, March 2016. <https://www.cs.tau.ac.il/~tromer/ecdh>
- [149] D. Genkin, L. Pachmanov, I. Pipman, E. Tromer, and Y. Yarom. ECDSA Key Extraction from Mobile Devices via Nonintrusive Physical Side Channels. Cryptology ePrint Archive, Report 2016/230, 2016. <https://www.cs.tau.ac.il/~tromer/mobilesc>
- [150] European Aviation Safety Agency. 'Prototype' Commission Regulation on Unmanned Aircraft Operations. August 2016. <https://www.easa.europa.eu/system/files/dfu/UAS%20Prototype%20Regulation%20final.pdf>
- [151] Eclipse Foundation. <https://www.eclipse.org/>
- [152] S. Buckley. Intel announces Edison: a 22nm dual-core PC the size of an SD card. Engadget, January 2014. <https://www.engadget.com/2014/01/06/intel-edison>
- [153] D. Hardawar. Intel shows off Project Alloy, an all-in-one VR headset. Engadget, August 2016. <https://www.engadget.com/2016/08/16/intel-announces-project-alloy-an-all-in-one-vr-headset>
- [154] N. Lee. Windows Holographic coming to all Windows 10 PCs next year. Engadget, August 2016. <https://www.engadget.com/2016/08/16/windows-holographic-coming-to-all-windows-10-pcs-next-year>
- [155] M. Moon. Intel intros a ready-to-fly drone for software developers. Engadget, August 2016. <https://www.engadget.com/2016/08/17/intel-aero-drone-for-developers>
- [156] W. Xu, F. Lin, C. Song and Zh. Ba. Smartphone hacks 3-D printer by measuring 'leaked' energy and acoustic waves. University at Buffalo, September 2016. https://www.eurekalert.org/pub_releases/2016-09/uab-sh3090716.php
- [157] EuroLab-4-HPC. <https://www.eurolab4hpc.eu>
- [158] ExpressVPN. An Open Infographic to Police: Stop Worrying and Learn to Love Encryption! 2016. <https://www.expressvpn.com/Internet-privacy/smartphone-data-encryption-police-infographic/>
- [159] Fairphone. <https://www.fairphone.com>
- [160] M. Huisken. Guest blog. iFixit on Fairphone 2: The first truly smart smartphone. Fairphone, November 2015. <https://www.fairphone.com/2015/11/18/guest-blog-ifixit-on-fairphone-2-the-first-truly-smart-smartphone>
- [161] T. Wheeler. Leading towards Next Generation "5G" Mobile Services. Federal Communications Commission, August 2015. <https://www.fcc.gov/news-events/blog/2015/08/03/leading-towards-next-generation-5g-mobile-services>
- [162] Global Market Insights. Embedded Systems Market. March 2016. <https://www.gminsights.com/industry-analysis/embedded-system-market>
- [163] Google. Google Self-Driving Car Project. <https://www.google.com/selfdrivingcar>
- [164] J. Russell. Intel's Fryman: "It's not that we love CMOS; it's the only real choice." HPCwire, September 2016. <https://www.hpcwire.com/2016/09/01/intels-fryman-not-love-cmos-real-choice/?eid=328378889&bid=1516705>

- [165] J. Vaughan-Brown. Open source adoption: tread carefully but don't hold back. AppDynamics, June 2016. <https://www.linkedin.com/pulse/open-source-adoption-tread-carefully-dont-hold-back-vaughan-brown>
- [166] ARM. ARMmbed. <https://www.mbed.com/en>
- [167] National Science Foundation. Directorate for Computer and Information Science and Engineering (CISE) Funding. February 2015. https://www.nsf.gov/about/budget/fy2016/pdf/18_fy2016.pdf
- [168] The New York State Senate. Senate Bill S161. <https://www.nysenate.gov/legislation/bills/2015/s161>
- [169] K. Heiner. Optimizing Performance For Intel OpenGL On Linux. Phoronix forums, August 2013. <https://www.phoronix.com/forums/forum/linux-graphics-x-org-drivers/intel-linux/37332-optimizing-performance-for-intel-opengl-on-linux?p=462410#post462410>
- [170] Raspberry Pi Foundation. <https://www.raspberrypi.org>
- [171] Raspberry Pi Foundation. Raspberry Pi Zero. <https://www.raspberrypi.org/products/pi-zero>
- [172] Raspberry Pi Foundation. Raspberry Pi 3 Model B. <https://www.raspberrypi.org/products/raspberry-pi-3-model-b>
- [173] Technopedia. Maker Movement. <https://www.techopedia.com/definition/28408/maker-movement>
- [174] Tensorflow. <https://www.tensorflow.org>
- [175] Tesla Motors. Tesla Model S. <https://www.tesla.com/models>
- [176] N. Kobie. How we talk about privacy matters. The Guardian, January 2016. <https://www.theguardian.com/media-network/2016/jan/14/how-we-talk-about-privacy-matters>
- [177] J. Borra. Digital obesity: our high-tech lives may be bad for our health. The Guardian, April 2013. <https://www.theguardian.com/sustainable-business/digital-obesity-high-tech-health>
- [178] C. Arthur. How to keep on selling smartphones when we've nearly all got one already. The Guardian, April 2014. <https://www.theguardian.com/technology/2014/apr/27/smartphone-market-saturation-apple-samsung>
- [179] D. Matthews. Chinese student market: can the West weather a perfect storm? The Times Higher Education, May 2016. <https://www.timeshighereducation.com/features/chinese-student-market-can-the-west-weather-a-perfect-storm>
- [180] The Times Higher Education World University Rangings. 2016. <https://www.timeshighereducation.com/world-university-rankings>
- [181] TOP500.org. TOP 10 Sites for June 2016. June 2016. <https://www.top500.org/lists/2016/06>
- [182] M. Feldman. Chinese Chipmaker Unveils Speedy 64-Core ARM Processor. TOP500.org, August 2016. <https://www.top500.org/news/chinese-chipmaker-unveils-speedy-64-core-arm-processor/>
- [183] Trade Union Congress. UK workers experienced sharpest wage fall of any leading economy, TUC analysis finds. July 2016. <https://www.tuc.org.uk/economic-issues/labour-market/uk-workers-experienced-sharpest-wage-fall-any-leading-economy-tuc>
- [184] Y. Cao, Zh. Qian, Zh. Wang, T. Dao, S. V. Krishnamurthy, and L. M. Marvel. Off-Path TCP Exploits: Global Rate Limit Considered Dangerous. 25th USENEX Security Symposium, August 2016. <https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/cao>
- [185] HTC. Vive headset. <https://www.vive.com/us/product/>
- [186] M. Rhodes. The Bizarre, Bony-Looking Future of Algorithmic Design. Wired, September 2013. <https://www.wired.com/2015/09/bizarre-bony-looking-future-algorithmic-design>
- [187] A. Greenberg. Apple's 'Differential Privacy' Is About Collecting Your Data—But Not Your Data. Wired, June 2016. <https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/>
- [188] A. Davies. Uber's Self-Driving Truck Makes Its First Delivery: 50,000 Beers. October 2016. <https://www.wired.com/2016/10/ubers-self-driving-truck-makes-first-delivery-50000-beers/>
- [189] http://techon.nikkeibp.co.jp/english/NEWS_EN/20141001/379923/?ST=msbe&P=2
- [190] G. Patterson et al. 5G Manifesto for timely deployment of 5G in Europe. July 2016. <http://telecoms.com/wp-content/blogs.dir/1/files/2016/07/5GManifestofortimelydeploymentof5GinEurope.pdf>
- [191] Thin Film Electronics ASA. Thinfilm Demonstrates First Integrated Printed Electronic System with Rewritable Memory. December 2012. <http://thinfilm.no/2012/12/20/thinfilm-demonstrates-first-integrated-printed-electronic-system-with-rewritable-memory/>
- [192] Thin Film Electronics ASA. Thinfilm Demonstrates First Printed NFC-Enabled Smart Label. May 2014. <http://thinfilm.no/2014/05/28/thinfilm-demonstrates-first-printed-nfc-enabled-smart-label/>
- [193] MEMS Industry Group. TSensors Summit for a Trillion Sensor Roadmap. October 2013. <http://tsensorssummit.org/Resources/Why%20TSensors%20Roadmap.pdf>
- [194] N. Malone. 11 everyday things your smartphone has made obsolete. Business Insider UK, November 2014. <http://uk.businessinsider.com/11-things-the-iphone-has-made-obsolete-2015-11?r=US&IR=T>
- [195] B. Bryan. Europe is bigger than the US. Business Insider UK, June 2015. <http://uk.businessinsider.com/charts-eu-economy-is-bigger-than-the-us-2015-6?r=US&IR=T>
- [196] J. Greenbough. 10 million self-driving cars will be on the road by 2020. Business Insider UK, June 2016. <http://uk.businessinsider.com/report-10-million-self-driving-cars-will-be-on-the-road-by-2020-2015-5-6?r=US&IR=T>

- [197] B. Carson. Uber's self-driving truck went on a 120-mile beer run to make history. Business Insider UK, October 2016. <http://uk.businessinsider.com/uber-otto-self-driving-truck-completes-first-run-2016-10?r=US&IR=T>
- [198] J. Novet. Amazon open-sources its own deep learning software, DSSTNE. VentureBeat, May 2016. <http://venturebeat.com/2016/05/11/amazon-open-sources-its-own-deep-learning-software-dsstne>
- [199] VisualGDB. Developing a Raspberry PI app with Visual Studio. SysProgs, March 2014. <http://visualgdb.com/tutorials/raspberry>
- [200] K. Moammer. Intel Expects to Launch 10nm Chips in 2017. Wccf Tech, Feb 2015. <http://wccftech.com/intel-expects-launch-10nm-2017>
- [201] H. Mujtaba. Intel Confirms Launch of 10nm Cannonlake Processors in 2H 2017. Wccf Tech, Feb 2016. <http://wccftech.com/intel-kaby-lake-q3-2016-cannonlake-q3-2017>
- [202] K. Moammer. US Government Bans Intel, NVIDIA and AMD From Selling High End Chips To The Chinese Government. Wccf Tech, April 2015. <http://wccftech.com/us-government-bans-intel-nvidia-amd-chips-china/#ixzz4HfzB874b>
- [203] What When How. Ambient Intelligence Environments (Artificial Intelligence). <http://what-when-how.com/artificial-intelligence/ambient-intelligence-environments-artificial-intelligence>
- [204] M. Fink. Discover Day Two: The Future Is Now—The Machine from HP. HP, June 2014. https://web.archive.org/web/20151002082437/http://www8.hp.com/hpnext/posts/discover-day-two-future-now-machine-hp#.WD172Gco_mE
- [205] R. Smith and I. Cutress. Intel's Changing Future: Smartphone SoCs Broxton & SoFIA Officially Cancelled. AnandTech, April 2016. <http://www.anandtech.com/show/10288/intel-broxton-sofia-smartphone-socs-cancelled>
- [206] A. Shilov. SiFive Unveils Freedom Platforms for RISC-V-Based Semi-Custom Chips. AnandTech, July 2016. <http://www.anandtech.com/show/10488/sifive-unveils-freedom-platforms-for-riscvbased-semicustom-chips>
- [207] R. Smith. AMD's 2016 Linux Driver Plans & GPUOpen Family of Dev Tools: Investing In Open Source. AnandTech, December 2015. <http://www.anandtech.com/show/9853/amd-gpuopen-linux-open-source>
- [208] S. Arbesman. Overcomplicated. Current, July 2016. <http://www.arbesman.net/overcomplicated>
- [209] The Barrelfish Operating System. <http://www.barrelfish.org>
- [210] BBC News. Oxfam says wealth of richest 1% equal to other 99%. January 2016. <http://www.bbc.com/news/business-35339475>
- [211] S. Coughlan. Digital dependence 'eroding human memory'. BBC News, October 2015. <http://www.bbc.com/news/education-34454264>
- [212] BBC News. Uber to deploy self-driving cars in Pittsburgh. August 2016. <http://www.bbc.com/news/technology-37117831>
- [213] BBC News. Five reasons we reckon Google Glass was pulled. January 2015. <http://www.bbc.co.uk/newsbeat/article/30830265/five-reasons-we-reckon-google-glass-was-pulled>
- [214] Zh. Xinying. More students returning after overseas studies. China Daily, March 2016. http://www.chinadaily.com.cn/china/2016-03/18/content_23931407.htm
- [215] C. Dancy. Mindful Cyborg. <http://www.chrisdancy.com/about>
- [216] Cisco Public. Cisco Global Cloud Index: Forecast and Methodology, 2015–2020. Cisco, 2016. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/Cloud_Index_White_Paper.html
- [217] NEI. How Intel's acquisition of Altera could transform IoT and the data centre. CloudTech, April 2016. <http://www.cloudcomputing-news.net/news/2016/apr/07/intel-acquires-altera-what-their-fpga-capabilities-could-mean-for-data-centers-and-iot>
- [218] A. Kharpal. Apple, Google, Facebook, Amazon, Microsoft wont' be biggest in 20 years: Top VC. CNBC, June 2016. <http://www.cnbc.com/2016/06/30/apple-google-facebook-amazon-microsoft-wont-be-biggest-in-20-years-top-vc.html>
- [219] C. Cooper. Dead and buried: Microsoft's holy war on open-source software. CNET, June 2014. <http://www.cnet.com/news/dead-and-buried-microsofts-holy-war-on-open-source-software>
- [220] K. Collins. Man seeking robot: One inventor's quest to cure loneliness. CNET, June 2016. <http://www.cnet.com/news/man-seeking-robot-one-inventors-quest-to-cure-loneliness>
- [221] C. Saran. Cisco predicts machine-to-machine traffic will increase 22-fold. ComputerWeekly.com, February 2012. <http://www.computerweekly.com/news/2240117424/Cisco-predicts-machine-to-machine-traffic-will-increase-22-fold>
- [222] K. Mearian. Flash memory's density surpasses hard drives for first time. Computerworld, February 2016. <http://www.computerworld.com/article/3030642/data-storage/flash-memorys-density-surpasses-hard-drives-for-first-time.html>
- [223] S. Engell et al. Proposal of a European Research and Innovation Agenda on Cyber-Physical Systems of Systems - 2016-2025. CPSoS, April 2016. <http://www.cpsos.eu/roadmap>
- [224] J. Lee. Coarse Grained Reconfigurable Array Architecture. School of Computer Science and Engineering, University of New South Wales, April 2016. <http://www.cse.unsw.edu.au/~cs4601/16s1/slides/lee-cgras.pdf>
- [225] D. Genkin, I. Pipman, and E. Tromer. Get Your Hands Off My Laptop: Physical Side-Channel Key-Extraction Attacks On PCs. CHES 2014, September 2014. <http://www.cs.tau.ac.il/~tromer/handsoff>

- [226] J. Ianelle. Video Shows Zombielike Mob of FIU Students Swarming Campus for Pokemon. Miami New Times, July 2014. <http://www.miaminewtimes.com/news/video-shows-zombielike-mob-of-fiu-students-swarming-campus-for-pokemon-8600277>
- [227] Outreach@darpa.mil. President Obama Highlights New DARPA Program Aimed at Developing Novel Therapies Customized to Individual Patients. DARPA, August 2014. <http://www.darpa.mil/news-events/2014-o8-26>
- [228] Outreach@darpa.mil. Another Big Shrink: Tiling Chiplets into Next-Generation Microsystems. DARPA, July 2016. <http://www.darpa.mil/news-events/2016-07-19>
- [229] DeepSpec. The science of deep specification. <http://www.deepspec.org>
- [230] K. Parrish. Speedy new Intel Atom processors might not be used in consumer devices. Digital Trends, August 2016. <http://www.digitaltrends.com/computing/intel-joule-atom-broxtom-m-goldmont-processors>
- [231] The Disappearing Computer. 2003. <https://web.archive.org/web/20160306160949/http://www.disappearing-computer.eu/>
- [232] T. Romero. The Real Reason Uber is Failing in Japan. Disrupting Japan, August 2016. <http://www.disruptingjapan.com/real-reason-uber-failing-japan>
- [233] D-Wave Systems. The D-Wave 2X System. <http://www.dwavesys.com/d-wave-two-system>
- [234] Accrington. Printed electronics - On a roll. The Economist, July 2016. <http://www.economist.com/news/science-and-technology/21702741-printing-conventional-rotary-presses-will-create-cheaper-electronics>
- [235] L. Chow. 5 Ways Vertical Farms Are Changing the Way We Grow Food. EcoWatch, March 2015. <http://www.ecowatch.com/5-ways-vertical-farms-are-changing-the-way-we-grow-food-1882019986.html>
- [236] European Cyber Security Organisation. <http://www.ecs-org.eu/>
- [237] European Cyber Security Organisation. European Cybersecurity Industry Proposal for a contractual Public-Private Partnership. June 2016. <http://www.ecs-org.eu/documents/ecs-cppp-industry-proposal.pdf>
- [238] European Environment Agency. Ecological footprint of European countries. September 2015. <http://www.eea.europa.eu/data-and-maps/indicators/ecological-footprint-of-european-countries/ecological-footprint-of-european-countries-2>
- [239] P. Ricoux. EESl2 Final Report - 2015 Update Vision & Recommendations. EESl2, September 2015. http://www.eesi-project.eu/wp-content/uploads/2015/05/EESl2_D73_Final-report-on-EESl2-exascale-vision-roadmap-and-recommendations.pdf
- [240] R. Merrit. ndia Preps RISC-V Processors. EETimes, January 2016. http://www.eetimes.com/document.asp?doc_id=1328790
- [241] R. Merrit. Ethernet Flexes Network Muscles. EETimes, October 2016. http://www.eetimes.com/document.asp?doc_id=1330553
- [242] T. P. Morgan. How Microsoft Is Using FPGAs To Speed Up Bing Search . EnterpriseTech, September 2014. <http://www.enterprisetech.com/2014/09/03/microsoft-using-fpgas-speed-bing-search>
- [243] European Space Agency. Leading up to LEON: ESA's First Microprocessors. January 2013. http://www.esa.int/Our_Activities/Space_Engineering_Technology/Leading_up_to_LEON_ESA_s_first_microprocessors
- [244] M. Gustlin and F. Dada. What is FlexEthernet? Ethernet Technology Summit, April 2015. http://www.ethernetsummit.com/English/Collaterals/Proceedings/2015/20150415_1C_Gustlin.pdf
- [245] ETP4HPC. The Strategic Research Agenda. <http://www.etp4hpc.eu/en/sra.html>
- [246] ETP4HPC. Strategic Agenda 2015 Update. 2015. <http://www.etp4hpc.eu/pujades/files/ETP4HPC%20SRA%202%20Single%20Page.pdf>
- [247] D. K. Taft. Microsoft Eyes Visual Studio-Like Toolset for Machine Learning. eWeek, August 2016. <http://www.eweek.com/developer/microsoft-eyes-visual-studio-like-toolset-for-machine-learning.html>
- [248] J. Hruska. TSMC will begin 10nm production this year, claims 5nm by 2020. ExtremeTech, January 2016. <http://www.extremetech.com/computing/221532-tsmc-will-begin-10nm-production-this-year-claims-5nm-by-2020>
- [249] <http://www.extremetech.com/extreme/230458-meet-the-new-worlds-fastest-supercomputer-chinas-taihulight>
- [250] D. Munro. New Survey Highlights Startling Erosion Of Online Trust. Forbes, May 2016. <http://www.forbes.com/sites/danmunro/2016/05/15/new-survey-highlights-startling-erosion-of-online-trust/#685cc9f85e67>
- [251] E. Mack. Elon Musk, Silicon Valley Elite Launch 'Open' Artificial Intelligence With \$1 Billion. Forbes, December 2015. <http://www.forbes.com/sites/ericmack/2015/12/11/elon-musk-sam-altman-peter-thiel-others-launch-open-a-i-with-1-billion-donation/#544863a363ef>
- [252] Trefis Team. Seagate's Enterprise Storage: Helium-Filled Drives, New Technology To Drive Future Growth. Forbes, September 2015. <http://www.forbes.com/sites/greatspeculations/2015/09/10/seagates-enterprise-storage-helium-filled-drives-new-technology-to-drive-future-growth/2/#6025c43f6dfd>
- [253] J. Gorzelany. Volvo Will Accept Liability For Its Self-Driving Cars. Forbes, October 2015. <http://www.forbes.com/sites/jimgorzelany/2015/10/09/volvo-will-accept-liability-for-its-self-driving-cars/#31900fo33d80>
- [254] J. Kang. How China's New \$2.8 Billion Chip Maker Will Affect The Global Semiconductor Industry. Forbes, August 2016. <http://www.forbes.com/sites/johnkang/2016/08/06/how-chinas-new-2-8-billion-chip-maker-will-affect-the-global-semiconductor-industry/#294c71ac3ecd>

- [255] K. Hill. How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did. Forbes, February 2012. <http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did>
- [256] K. Freund. Google's TPU Chip Creates More Questions Than Answers. Forbes, May 2015. <http://www.forbes.com/sites/moorinsights/2016/05/26/googles-tpu-chip-creates-more-questions-than-answers/#950771ed96af>
- [257] S. Morgan. Cyber Crime Costs Projected To Reach \$2 Trillion by 2019. Forbes, January 2016. <http://www.forbes.com/sites/stevemorgan/2016/01/17/cyber-crime-costs-projected-to-reach-2-trillion-by-2019>
- [258] T. Coughlin. 100 TB HDDs and A New Spin on Storage. Forbes, November 2014. <http://www.forbes.com/sites/tomcoughlin/2014/11/22/100-tb-hdds-and-a-new-spin-on-storage/#18e43598279f>
- [259] Freightliner. Freightliner Inspiration Truck. <http://www.freightlinerinspiration.com/>
- [260] H. Warrell. Overseas students contribute £2.3bn a year to UK, says study. Financial Times, May 2015. <https://www.ft.com/content/f058ce3c-fd4f-11e4-9e96-00144feabdco>
- [261] Gadget Review. 16 of the Coolest Smartphone Connected Appliances (list). <http://www.gadgetreview.com/16-of-the-coolest-smartphone-connected-appliances>
- [262] Cobham Gaisler AB. <http://www.gaisler.com>
- [263] Gartner, Inc. Gartner Says Worldwide Smartphone Sales Recorded Slowest Growth Rate Since 2013. August 2015. <http://www.gartner.com/newsroom/id/3115517>
- [264] Global Foundries. GLOBALFOUNDRIES Extends FDX Roadmap with 12nm FD-SOI Technology. September 2016. <http://www.globalfoundries.com/newsroom/press-releases/2016/09/07/globalfoundries-extends-fdx-roadmap-with-12nm-fd-soi-technology>
- [265] T. Lin. Why overseas Chinese students return. Global Times, March 2016. <http://www.globaltimes.cn/content/973022.shtml>
- [266] Manchester University. The Home of Graphene. <http://www.graphene.manchester.ac.uk/>
- [267] TOP500.org. The Green 500. June 2016. <https://www.top500.org/green500/lists/2016/06/>
- [268] S. Kim, Y.-H. Park, J. Kim, M. Kim, W. Lee, and S. Kee. Flexible Video Processing Platform for 8K UHD TV. Hot Chips, August 2015. http://www.hotchips.org/wp-content/uploads/hc_archives/hc27/HC27.26-Posters/HC27.26.p50-FlexibleVideoPlatformFor8KUHDTV-Kin-Samsung.pdf
- [269] IBM. Go beyond artificial intelligence with Watson. <http://www.ibm.com/watson>
- [270] K. Bora. Worldwide Tablet Shipment Down 20% In Q3 As Big-Screen Smartphones Cannibalize Sales. International Business Times, October 2015. <http://www.ibtimes.com/worldwide-tablet-shipment-down-20-q3-big-screen-smartphones-cannibalize-sales-2163784>
- [271] ICT Energy. ICT Energy Strategic Research Agenda. March 2016. <http://www.ict-energy.eu/sites/ict-energy.eu/files/ICTEnergySRA5.3.pdf>
- [272] Ellie Zolfagharifard, May 2016, Dailymail.com, "Do YOU have 'low battery anxiety'? 90% of us panic about losing power on our phones", <http://www.dailymail.co.uk/sciencetech/article-3607598/Do-low-battery-anxiety-90-panic-losing-power-phones.html>
- [273] R. Das and P. Harrop. Printed, Organic & Flexible Electronics Forecasts, Players & Opportunities 2016-2026. IDTechEx, September 2016. <http://www.idtechex.com/research/reports/printed-organic-and-flexible-electronics-forecasts-players-and-opportunities-2016-2026-000457.asp?viewopt=desc>, <http://www.idtechex.com/research/reports/printed-organic-and-flexible-electronics-forecasts-players-and-opportunities-2016-2026-000457.asp?viewopt=desc>,
- [274] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papmanthou. Hawk: The Blockchain Model of Cryptography and Privacy-Preserving Smart Contracts. In 2016 IEEE Symposium on Security and Privacy, May 2016. <http://www.ieee-security.org/TC/SP2016/papers/o824a839.pdf>
- [275] International Federation of Robotics. Executive Summary World Robotics 2016 Service Robots. October 2016. <http://www.ifr.org/service-robots/statistics>
- [276] Institute of International Education. Report on International Educational Exchange. November 2015. <http://www.iie.org/~media/Files/Corporate/Open-Doors/Open-Doors-Presentation-2015.pdf?la=en>
- [277] Institute of International Education. Special Reports: Economic Impact of International Students. 2015. <http://www.iie.org/Research-and-Publications/Open-Doors/Data/Economic-Impact-of-International-Students#.V7djMY9OJeM>
- [278] Intel. Ultra-fast, energy-sipping devices powered by Intel. <http://www.intel.com/content/www/us/en/silicon-innovations/intel-14nm-technology.html>
- [279] K. Akdemir et al. Breakthrough AES performance with intel AES new instructions. Intel. White paper, June 2010.
- [280] ISORG. Large Area Image Sensors. http://www.isorg.fr/default.asp?cat_id=78
- [281] V. Thomson. Battery Sales Surge Due To 'Pokemon Go' Players' Purchases. iTechPost, August 2016. <http://www.itechpost.com/articles/23634/20160806/battery-sales-surge-due-pokemon-go-players-purchases.htm>
- [282] Juniper Research. Robots to Reside in more than 1 in 10 American Households by 2020, Finds Juniper Research. December 2015. <http://www.juniperresearch.com/press/press-releases/robots-to-reside-in-more-than-1-in-10-houses>
- [283] Kalray. Kalray Products. <http://www.kalrayinc.com/kalray/products>

- [284] Lancaster University. World should consider limits to future Internet expansion to control energy consumption. August 2016. <http://www.lancaster.ac.uk/news/articles/2016/world-should-consider-limits-to-future-Internet-expansion/>
- [285] C. Dembach. The Legend of Steve Jobs – His Life and Career. Mac History. <http://www.mac-history.net/steve-jobs/2012-10-30/the-legend-of-steve-jobs-his-life-and-career/5>
- [286] Markets and Markets. Printed Electronics Markets. October 2016. <http://www.marketsandmarkets.com/Market-Reports/printed-electronics-market-197.html>
- [287] R. Dobbs, A. Madgavkar, J. Manyika, J. Woetzel, J. Bughin, E. Labaye, and P. Kashyap. Poorer than their parents? A new perspective on income inequality. McKinsey Global Institute, July 2016. <http://www.mckinsey.com/global-themes/employment-and-growth/poorer-than-their-parents-a-new-perspective-on-income-inequality>
- [288] Microsoft. Microsoft HoloLens. <http://www.microsoft.com/microsoft-hololens/en-us>
- [289] Movidius. <http://www.movidius.com>
- [290] Movidius. Movidius Announces Deep Learning Accelerator and Fathom Software Framework. <http://www.movidius.com/news/movidius-announces-deep-learning-accelerator-and-fathom-software-framework>
- [291] National Nanotech Initiative. A Federal Vision for Future Computing: A Nanotechnology-Inspired Grand Challenge. US Government, July 2016. http://www.nano.gov/sites/default/files/pub_resource/federal-vision-for-nanotech-inspired-future-computing-grand-challenge.pdf
- [292] X. Han, G. Stocking, M. A. Gebbie, and R. P. Appelbaum. Will They Stay or Will They Go? International Graduate Students and Their Decisions to Stay or Leave the U.S. upon Graduation. PLoS One, 10(3): e0118183, 2015. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4356591/>
- [293] J. Brodtkin. IBM's Jeopardy-playing machine can now beat human contestants. Network World, February 2010. <http://www.computerworld.com/article/2520980/app-development/ibm-s-jeopardy-playing-machine-can-now-beat-human-contestants.html>
- [294] Gautham. A Brief History of Bitcoin and Bitcoin Mining. NewsBTC, May 2016. <http://www.newsbtc.com/2016/05/26/brief-history-bitcoin-bitcoin-mining>
- [295] News.com/au. Researchers identify Busy Lifestyle Syndrome as new condition that makes you forgetful. December 2012. <http://www.news.com.au/lifestyle/health/why-are-we-so-forgetful-again/story-fneuz9ev-1226544547362>
- [296] N. Hemsoth. Can Open Source Hardware Crack Semiconductor Industry Economics? The Next Platform, May 2016. <http://www.nextplatform.com/2016/05/16/can-open-source-hardware-crack-semiconductor-industry-economics>
- [297] N. Hemsoth. A Look Inside China's Chart-Topping New Supercomputer. The Next Platform, June 2016. <http://www.nextplatform.com/2016/06/20/look-inside-chinas-chart-topping-new-supercomputer>
- [298] T. P. Morgan. Inside Japan's Future Exascale ARM Supercomputer. The Next Platform, June 2016. <http://www.nextplatform.com/2016/06/23/inside-japans-future-exaflops-arm-supercomputer>
- [299] A. Selyukh. FAA Expects 600,000 Commercial Drones In The Air Within A Year. NPR, August 2016. <http://www.npr.org/sections/thetwo-way/2016/08/29/491818988/faa-expects-600-000-commercial-drones-in-the-air-within-a-year>
- [300] N. Popper. How China Took Center Stage in Bitcoin's Civil War. The New York Times, June 2016. http://www.nytimes.com/2016/07/03/business/dealbook/bitcoin-china.html?_r=0
- [301] OE-A. Organic and Printed Electronics Association. <http://www.oe-a.org/home>
- [302] OECD. Inequality and Income. <http://www.oecd.org/inequality.htm#income>
- [303] Open Compute Project. <http://www.opencompute.org>
- [304] P. J. Winzer. Scaling Optical Fiber Networks: Challenges and Solutions. Optics & Photonics News, March 2015. http://www.osa-opn.org/home/articles/volume_26/march_2015/features/scaling_optical_fiber_networks_challenges_and_solu
- [305] OSRAM Opto Semiconductors. Osram Opto Semiconductors' lasers and photodiodes power Phantom Intelligence LIDAR. March 2015. http://www.osram-os.com/osram_os/en/press/press-releases/ir-devices-and-laser-diodes/2015/low-cost-lidar--a-key-technology-to-enable-autonomous-driving-in-urban-environments/index.jsp?src=rss_pressreleases_134138
- [306] M. Riofrio. Ford will mass-produce self-driving cars by 2021—but for sharing, not buying. PCWorld, August 2016. <http://www.pcworld.com/article/3108471/consumer-electronics/ford-will-mass-produce-self-driving-cars-by-2021-but-for-sharing-not-buying.html>
- [307] B. H. Frank. Microsoft PowerShell goes open source, lands on Linux and Mac. PCWorld, August 2016. <http://www.pcworld.com/article/3109176/microsoft-powershell-goes-open-source-and-lands-on-linux-and-mac.html>
- [308] Pipeline Workspaces. Unhealthy Smartphone? You're undateable! <http://www.pipelineworkspaces.com/unhealthy-smartphone-youre-undateable>
- [309] M. MacWilliams. The One Weird Trait That Predicts Whether You're a Trump Supporter. Politico, January 2016. <http://www.politico.com/magazine/story/2016/01/donald-trump-2016-authoritarian-213533>
- [310] M. Guest et al. PRACE Scientific Case for HPC in Europe 2012 – 2020. PRACE, October 2012. <http://www.prace-ri.eu/prace-the-scientific-case-for-hpc>
- [311] Pragmatic Printing. Electronics for a flexible world . <http://www.pragmaticprinting.com/technology/>
- [312] A. Hadhazy. New microchip demonstrates efficiency and scalable design. Princeton University, August 2016. <http://www.princeton.edu/main/news/archive/S47/19/67G69/?section=topstories>

- [313] Printed Electronics World. Beer bottle lights up when touched. December 2015. <http://www.printedelectronicsworld.com/articles/8806/beer-bottle-lights-up-when-touched>
- [314] R. S. Olson. Percentage of Bachelor's degrees conferred to women, by major (1970-2012). June 2015. <http://www.randalolson.com/2014/06/14/percentage-of-bachelors-degrees-conferred-to-women-by-major-1970-2012/>
- [315] J. Jackson. The future fallout from the death of GSM. RCRWirelessNews, June 2015. <http://www.rcrwireless.com/20150616/network-infrastructure/the-future-fallout-from-the-death-of-gsm-tag20>
- [316] D. Modha. Introducing a Brain-inspired Computer. IBM, 2016. <http://www.research.ibm.com/articles/brain-chip.shtml>
- [317] IBM. The DeepQA Research Team. <http://www.research.ibm.com/deepqa/faq.shtml>
- [318] G. Alioto, C. Avare, P. Carpenter, M. Leich, O. Unsal. Roadmap v23. RETHINK big, December 2015. http://www.rethinkbig-project.eu/sites/default/files/u273/D5.3RoadmapV23_o.pdf
- [319] M. Yamazaki. Softbank, Honda: Sit back, relax, let the car do the talking. Reuters, July 2016. <http://www.reuters.com/article/us-softbank-honda-ai-idUSKCN1010PR>
- [320] S. Crowe. Leka Social Robot Helps Special Needs Children. Robotics Trends, January 2016. http://www.roboticstrends.com/article/leka_social_robot_helps_special_needs_children
- [321] S. Crowe. Pepper's Robots US Debut Planned for 2016. Robotics Trends, May 2016. http://www.roboticstrends.com/article/peppers_robots_us_debut_planned_for_2016
- [322] J. Van Grove. Voice assistants are taking over. San Diego Union-Tribune, May 2016. <http://www.sandiegouniontribune.com/business/technology/sdut-voice-personal-assistant-amazon-google-apple-2016may27-story.html>
- [323] Semiconductor Industry Association and Semiconductor Research Corporation. Rebooting the IT Revolution: A Call to Action. September 2015. <http://www.semiconductors.org/clientuploads/Resources/RITR%20WEB%20version%20FINAL.pdf>
- [324] Russia Insider. Baikal Technologies Announces the Baikal-T1 Microprocessor with MIPS P5600 CPU. May 2015. <http://russia-insider.com/en/technology/baikal-technologies-announces-baikal-t1-microprocessor-mips-p5600-cpu/ri7451>
- [325] ECSEL PMB. 2016 Multi Annual Strategic Research and Innovation Agenda for ECSEL Joint Undertaking. December 2015. <http://www.smart-systems-integration.org/public/documents/publications/ECSEL%20MASRIA%202016.pdf>
- [326] K. Arlington. Smartphones are ruling our lives and killing our imaginations. The Sidney Morning Herald, March 2016. <http://www.smh.com.au/technology/mobiles/smartphones-are-ruling-our-lives-and-killing-our-imaginations-20160316-gnqliq.html>
- [327] MIT Media Lab, President Obama discusses artificial intelligence with Media Lab Director Joi Ito, <http://news.mit.edu/2016/president-obama-discusses-artificial-intelligence-media-lab-joi-ito-1014>
- [328] Global Industry Analysts, Inc. The Global Embedded Systems Market - Trends, Drivers & Projections. March 2015. http://www.strategyr.com/MarketResearch/Embedded_Systems_Market_Trends.asp
- [329] R. Adhikari. DARPA Challenges Researchers to Link Human Brains With Computers. Tech News World, January 2016. <http://www.technewsworld.com/story/83029.html>
- [330] J. Osborne. Google's Tensor Processing Unit explained: this is what the future of computing looks like. TechRadar, August 2016. <http://www.techradar.com/news/computing-components/processors/google-s-tensor-processing-unit-explained-this-is-what-the-future-of-computing-looks-like-1326915>
- [331] A. Wright and P. De Filippi. Decentralized Blockchain Technology and the Rise of Lex Cryptographica. Available at SSRN 2580664 (2015). <http://www.the-blockchain.com/docs/Decentralized%20Blockchain%20Technology%20And%20The%20Rise%20Of%20Lex%20Cryptographica.pdf>
- [332] T. C. Greene. Ballmer: "Linux is a cancer". The Register, June 2001. http://www.theregister.co.uk/2001/06/02/ballmer_linux_is_a_cancer
- [333] D. Pauli. Work begins on Russian rival to Android. The Register, May 2016. http://www.theregister.co.uk/2016/05/16/work_begins_on_russian_android_rival_os
- [334] R. Chirgwin. Software-defined networking is dangerously sniffable. The Register, August 2016. http://www.theregister.co.uk/2016/08/23/sdns_normal_behaviour_is_sniffable_say_researchers
- [335] J. Leyden. Sniffing your storage could lead to sensitive leaks, warn infosec bods. The Register, September 2016. http://www.theregister.co.uk/2016/09/12/storage_device_data_sniffing
- [336] S. Hollister. The age of the iPod is over. The Verge, January 2014. <http://www.theverge.com/2014/1/27/5351918/apples-ipod-rides-into-the-sunset>
- [337] J. Koebler. Elon Musk Says There's a 'One in Billions' Chance Reality Is Not a Simulation. Motherboard, June 2016. <http://motherboard.vice.com/read/elon-musk-simulated-universe-hypothesis>
- [338] J. Vincent. Baidu follows US tech giants and open sources its deep learning tools. The Verge, September 2016. <http://www.theverge.com/2016/9/1/12725804/baidu-machine-learning-open-source-paddle>
- [339] D. Bohm. Intel is buying the computer vision company that powers Tango and DJI's drones. The Verge, September 2016. <http://www.theverge.com/2016/9/6/12803652/intel-buying-movidius-tango-dji-phantom-computer-vision-chip>
- [340] P. Alcorn. IBM Challenges 3D XPoint With TLC Phase-Change Memory (PCM). Tom's Hardware, May 2016. <http://www.tomshardware.com/news/ibm-pcm-tlc-3d-xpoint,31811.html>
- [341] United Nations. Temperature in Kuwait hits 54 Celsius, sets possible record amid Middle East heatwave. July 2016. <http://www.un.org/apps/news/story.asp?NewsID=54559#.V7luWY9OJCo>

- [342] United Nations. World Population Ageing. 2015. http://www.un.org/en/development/desa/population/publications/pdf/ageing/WPA2015_Report.pdf
- [343] E. C. Baig. Personal digital assistants are on the rise (and they want to talk). USA Today, May 2016. <http://www.usatoday.com/story/tech/columnist/baig/2016/05/08/personal-digital-assistants-rise-and-they-want-talk/83715794>
- [344] Whirlpool. Smart Appliances. <http://www.whirlpool.com/smart-appliances>
- [345] World Intellectual Property Organisation. World Intellectual Property Indicators. 2015. http://www.wipo.int/edocs/pubdocs/en/wipo_pub_941_2015.pdf
- [346] K. Finley. Facebook Open-Sources a Trove of AI Tools. Wired, January 2015. <http://www.wired.com/2015/01/facebook-open-sources-trove-ai-tools>
- [347] D. Clark. Intel Completes Acquisition of Altera. The Wall Street Journal, December 2015. <http://www.wsj.com/articles/intel-completes-acquisition-of-altera-1451338307>
- [348] S. E. Ante. IBM Set to Expand Watson's Reach. The Wall Street Journal, January 2014. <http://www.wsj.com/articles/SB10001424052702303754404579308981809586194>
- [349] T. Mochizuki and Y. Koshino. Sony Re-Commits to Roots With Return to Robotics. The Wall Street Journal, June 2016. <http://www.wsj.com/articles/sony-re-commits-to-roots-with-return-to-robotics-1467303437>
- [350] R. Goodwins. The future of storage: 2015 and beyond. WDNNet, January 2015. <http://www.zdnet.com/article/the-future-of-storage-2015-and-beyond/>
- [351] William M. Holt, Moore's Law: A Path Forward, Keynote at International Solid-State Circuits Conference, February 2017. http://isscc.org/videos/2016_plenary.html
- [352] Semiconductor Industry Association. 2015 International Technology Roadmap for Semiconductors (ITRS). http://www.semiconductors.org/main/2015_international_technology_roadmap_for_semiconductors_itrs/
- [353] J. Gosling, B. Joy, and G. L. Steele Jr. The Java Language Specification. Addison Wesley Publishing Company, 1996, ISBN 0-201-63451-1.
- [354] P. Hockenos. The climate wars are coming – and more refugees with them. Aljazeera America, September 2015. <http://america.aljazeera.com/opinions/2015/9/the-climate-wars-are-coming--and-more-refugees-with-them.html>
- [355] K. Myny, et al. A thin-film microprocessor with inkjet print-programmable memory. Scientific Reports, 4, Article No. 7398, 2014.
- [356] Katherine Tweed, July 2014, GTM, "Network-Connected Electronics Wasted the Energy Equivalent of 133 Coal Plants Last Year", <https://www.greentechmedia.com/articles/read/connected-electronics-waste-80b-annually-and-growing>
- [357] Will Knight, October 2016, MIT Technology Review, "Obama: My Successor Will Govern a Country Being Transformed by AI", <https://www.technologyreview.com/s/602612/obama-my-successor-will-govern-a-country-being-transformed-by-ai/>
- [358] L. Xie et al. Integration of f-MWCNTs Sensor and Printed Circuits on Paper Substrate. IEEE Sensors Journal, vol.13, no.10, Oct. 2013.
- [359] <http://www.esterel-technologies.com/products/scade-suite/>
- [360] M. Mantysalo et al. System Integration of Smart Packages using Printed Electronics. 62nd Electronic Components and Technology Conference (ECTC), May-June 2012.
- [361] M. Zhang et al. Fabrication of organic electrochemical transistor arrays for biosensing. Biochimica et Biophysica Acta (BBA), 1830, 4402, 2013.
- [362] N. Carr. The Shallows: What the Internet Is Doing to Our Brains. 2010.
- [363] M. Nisen, Robot Economy Could Cause Up To 75 Percent Unemployment, January 2013. <http://www.businessinsider.com/50-percent-unemployment-robot-economy-2013-1>
- [364] Quote from Paolo Faraboschi, HP Labs during his keynote "The Perfect Storm" at HiPEAC2014 Conference, January 2014.
- [365] C. Joachim, J.K. Gimzewski, and A. Aviram, Nature (London) 408, 541 (2000), Diamond
- [366] Adriano Barenco, Charles H Bennett, Richard Cleve, David P DiVincenzo, Norman Margolus, Peter Shor, Tycho Sleator, John A Smolin, Harald Weinfurter, November 1995, Elementary gates for quantum computation, Physical review A, Vol. 52, N. 5, page 3457
- [367] <https://www.iad.gov/iad/programs/iad-initiatives/cnsa-suite.cfm>
- [368] Robbin A. Miranda, William D. Casebeer, Amy M. Hein, Jack W. Judy, Eric P. Krotkov, Tracy L. Laabs, Justin E. Manzo, Kent G. Pankratz, Gill A. Pratt, Justin C. Sanchez, Douglas J. Weber, Tracey L. Wheeler, Geoffrey S.F. Ling, DARPA-funded efforts in the development of novel brain-computer interface technologies, Journal of Neuroscience Methods, Volume 244, 15 April 2015, Pages 52-67, ISSN 0165-0270, <http://dx.doi.org/10.1016/j.jneumeth.2014.07.019>
- [369] S. Khan et al., "Screen printed flexible pressure sensors skin", 25th Annual Advanced Semiconductor Manufacturing Conference (ASMC), May 2014
- [370] S. Khan, L. Lorenzelli and R. Dahiya, "Technologies for Printing Sensors and Electronics over Large Flexible Substrates: A Review", IEEE Sensors Journal, vol. 15, pp. 3164-3181, 2015.
- [371] The new political divide, <http://www.economist.com/news/leaders/21702750-farewell-left-versus-right-contest-matters-now-open-against-closed-new>
- [372] Thomas Friedman, The World Is Flat 3.0: A Brief History of the Twenty-first Century, 2007.
- [373] University of Tokyo, "Flexible, 3D Printed, Solar Powered Thermal Alarm for Patient Monitoring", <https://3dprint.com/47116/3d-printed-fever-alarm/>, Feb. 2015

- [374] B. Valison, Quantum Computation: from a Programmer's Perspective, *New Generation Computing*, 31(1):1-26, January 2013.
- [375] M. Veldhorst et al - A two-qubit logic gate in silicon, *Nature* 526, 410–414, October 2015. doi:10.1038/nature15263.
- [376] P. Caserta, O. Zendra. Visualization of the Static aspects of Software: a survey. *IEEE Transactions on Visualization and Computer Graphics*, Institute of Electrical and Electronics Engineers, 2011, 17 (7), pp.913-933
- [377] L. Pasini et al. "High performance CMOS FDSOI devices activated at low temperature." *VLSI Technology*, 2016 IEEE Symposium on. IEEE, 2016.
- [378] S. Wiesner, *Conjugate Coding*, *SIGACT News* 15:1, pp. 78–88, 1983.
- [379] Wind et al. Vertical Scaling of Carbon Nanotube Field-Effect Transistors Using Top Gate Electrodes, IBM's T.J. Watson Research Center, *Journal of Applied Physics Letters*, 2002
- [380] Y.V. Pershin and M. Di Ventra. Memory effects in complex materials and nanoscale systems. *Advances in Physics*, vol. 60, no. 2, p. 145, 2011
- [381] Gen-Z Consortium. Data Access - A New Approach. October 2016. http://genzconsortium.org/wp-content/uploads/2016/05/Gen-Z-Consortium-Briefing-Deck_Final.pdf
- [382] P. Diamandis. Why the Cost of Living Is Poised to Plummet in the Next 20 Years. *SingularityHub*, July 2016. <http://singularityhub.com/2016/07/18/why-the-cost-of-living-is-poised-to-plummet-in-the-next-20-years/>
- [383] M. Curtis. Catalyzing the Collapse: The Computer and the Fall of the Soviet Union. *DigitalCommons@Olin*, April 2006. http://digitalcommons.olin.edu/cgi/viewcontent.cgi?article=1004&context=ahs_capstone_2006
- [384] European Commission. Growing the Silver Economy. February 2015. <http://ec.europa.eu/research/innovation-union/pdf/active-healthy-ageing/silvereco.pdf#view=fit&pagemode=none>
- [385] GLOBALFOUNDRIES. GLOBALFOUNDRIES Launches Industry's First 22nm FD-SOI Technology Platform. July 2015. <http://www.globalfoundries.com/newsroom/press-releases/2015/07/13/globalfoundries-launches-industry-s-first-22nm-fd-soi-technology-platform>
- [386] ASML. <https://www.asml.com/asml/en/s427>
- [387] NXP. S32VLS2-RDB: BlueBox: Autonomous Driving Platform (S32VLS2-RDB). <http://www.nxp.com/products/microcontrollers-and-processors/arm-processors/s32-arm-processors-microcontrollers/bluebox-autonomous-driving-platform-s32vls2-rdb:S32VLS2-RDB>
- [388] Apple. The most personal technology must also be the most private. <http://www.apple.com/privacy/approach-to-privacy/>
- [389] Wikipedia. FBI–Apple encryption dispute. https://en.wikipedia.org/w/index.php?title=FBI%E2%80%93Apple_encryption_dispute&oldid=752294694
- [390] IBM Institute for Business Value. Device democracy- Saving the future of the Internet of Things. IBM, July 2015. http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=XB&infotype=PM&appname=GBSE_GB_TI_USEN&htmlfid=GBE03620USEN&attachment=GBE03620USEN.PDF#loaded
- [391] Wikipedia. Optical head-mounted display. https://en.wikipedia.org/w/index.php?title=Optical_head-mounted_display&oldid=749748905
- [392] W. S. Harrison, N. Hanebutte, P. W. Oman, and J. Alves-Foss. The MILS Architecture for a Secure Global Information Grid. *Crosstalk: The Journal of Defense Software Engineering* 18.10. 2005.
- [393] If This Then That. <https://ifttt.com/>
- [394] C. H-L. Kao, A. Roseway, C. Holz, P. Johns, A. Calvo, and C. Schmandt. DuoSkin. MIT Media Lab, September 2016. <http://duoskin.media.mit.edu/>
- [395] D. J. Hill. 7 Key Factors Driving the Artificial Intelligence Revolution. *SingularityHub*, August 2016. <http://singularityhub.com/2016/08/29/7-factors-driving-the-artificial-intelligence-revolution/>
- [396] I. Barajas. NVIDIA's New 'Super Chip' Will Make Autonomous Vehicles Smarter. *Newegg*, January 2015. <http://blog.newegg.com/nvidias-new-super-chip-will-make-autonomous-vehicles-smarter/>
- [397] Wikipedia. Comparison of deep learning software. https://en.wikipedia.org/w/index.php?title=Comparison_of_deep_learning_software&action=history
- [398] IBM. Watson - A System Designed for Answers. February 2011. <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=SA&subtype=WH&htmlfid=POW03061USEN>
- [399] IBM. Watson in healthcare. http://www-05.ibm.com/innovation/uk/watson/watson_in_healthcare.shtml
- [400] IBM. What is Watson? <http://www-07.ibm.com/innovation/in/watson/what-is-watson/index.html>
- [401] Wikipedia. Watson (computer). [https://en.wikipedia.org/w/index.php?title=Watson_\(computer\)&oldid=749104844](https://en.wikipedia.org/w/index.php?title=Watson_(computer)&oldid=749104844)
- [402] M. Duranton, J. Hoogerbrugge, G. Al-kadi, S. Guntur, A. Terechko. Rapid Technology-Aware Design Space Exploration for Embedded Heterogeneous Multiprocessors. In *Processor and System-on-Chip Simulation*. Springer, August 2010. http://link.springer.com/chapter/10.1007%2F978-1-4419-6175-4_16
- [403] C. Joseph. Best iPhone voice control | Best Mac voice control. *Macworld UK*, August 2016. <http://www.macworld.co.uk/feature/iosapps/best-iphone-voice-control-siri-alternatives-siri-vs-cortana-vs-google-now-3511811/>
- [404] Z. Or-Bach. 28nm Was Last Node of Moore's Law. *EETimes*, August 2016. http://www.eetimes.com/author.asp?section_id=36&doc_id=1330366

- [405] Blue Frog Robotics. Buddy the first companion robot. <http://www.bluefrogrobotics.com/en/buddy/>
- [406] A. Geffer. The Man Who Tried to Redeem the World with Logic. February 2015. <http://nautil.us/issue/21/information/the-man-who-tried-to-redeem-the-world-with-logic>
- [407] G. Fiori, F. Bonaccorso, G. Iannaccone, T. Palacios, D. Neumaier, A. Seabaugh, S. Banerjee, and L. Colombo. Electronics based on two-dimensional materials. *Nature Nanotechnology* 9, pp. 768-779, October 2014. http://www.nature.com/nnano/journal/v9/n10/fig_tab/nnano.2014.207_T2.html
- [408] A. de Touzalin, C. Marcus, F. Heijman, I. Cirac, R. Murray, and T. Calarco. Quantum Manifesto for Quantum Technologies. European Commission, 2016. <https://ec.europa.eu/futurium/en/content/quantum-manifesto-quantum-technologies>
- [409] E. Szuman, M. Drosback, W. T. Polk, and B. Obama. The National Strategic Computing Initiative Turns One. July 2016. [https://www.whitehouse.gov/blog/2016/07/29/national-strategic-computing-initiative-turns-one\](https://www.whitehouse.gov/blog/2016/07/29/national-strategic-computing-initiative-turns-one/)
- [410] N. C. Moore. New \$28M Center Will Develop Computers of 2025. University of Michigan, January 2013. http://www.eecs.umich.edu/eecs/about/articles/2013/Austin_directs_C-FAR.html
- [411] Center for Future Architectures Research. <https://www.futurearchs.org>
- [412] IEEE. Rebooting Computing. <http://rebootingcomputing.ieee.org/>
- [413] IEEE. Rebooting Computing Summits. <http://rebootingcomputing.ieee.org/rc-summits>
- [414] European Commission. Science, Research and Innovation Performance of the EU. February 2016. http://bookshop.europa.eu/en/science-research-and-innovation-performance-of-the-eu-pbKl0415512/downloads/Kl-04-15-512-EN-N/Kl0415512ENN_002.pdf;pgid=Iq1Eknio.1lSRoOOK4MycO9BooooQ69Pgo_o;sid=gcmpAHWmu-WpLSIfysE3pxeDlwrVUHJ1hx0=?FileName=Kl0415512ENN_002.pdf&SKU=Kl0415512ENN_PDF&CatalogueNumber=Kl-04-15-512-EN-N
- [415] Deloitte. Technology and people: The great job-creating machine. August 2015. <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/finance/deloitte-uk-technology-and-people.pdf>
- [416] NVIDIA. NVIDIA DGX-1 AI Supercomputer. <http://www.nvidia.com/object/deep-learning-system.html>
- [417] M. Duranton et al. HiPEAC Vision, 2015 edition, chapter 3.3 “Silicon based technology: more and more roadblocks”, pp. 29-31.
- [418] Kirk M. Bresniker, Sharad Singhal, R. Stanley Williams, “Adapting to Thrive in a New Economy of Memory Abundance”, *Computer*, vol. 48, no., pp. 44-53, Dec. 2015.
- [419] Siemens to buy Mentor Graphics in \$4.5 billion deal, November 2016. <http://www.cnbc.com/2016/11/14/siemens-to-buy-mentor-graphics-in-45-billion-deal.html>
- [420] R. Koolen, and J. Schmaltz. Modeling Information Routing with Noninterference. January 2016. From <http://dx.doi.org/10.5281/zenodo.47980>
- [421] Standby Power Summary Table, <http://standby.lbl.gov/summary-table.html>
- [422] EURO-MILS Project. <http://euromils.eu>
- [423] S. Tverdyshev, H. Blasum, B. Langenstein, J. Maebe, B. De Sutter, B. Leconte, B. Triquet, K. Müller, M. Paulitsch, A. Söding-Freiherr von Blomberg, and A. Tillequin. MILS Architecture. 2013. From <http://dx.doi.org/10.5281/zenodo.45164>
- [424] J. Zhou, and J.-A. Foss. Security policy refinement and enforcement for the design of multi-level secure systems. *Journal of Computer Security*, 16, pp. 107-131. 2008.
- [425] Common Criteria Sponsoring Organizations. Common Criteria for Information Technology Security Evaluation. Version 3.1, revision 4. September 2012.
- [426] I. Furgel, V. Saftig, T. Wagner, K. Müller, and A. Söding-Freiherr von Blomberg. Non-Interfering Composed Evaluation. January 2016. From <http://dx.doi.org/10.5281/zenodo.47979>
- [427] Information Assurance Directorate. U.S. Government Protection Profile for Separation Kernels in Environments Requiring High Robustness. Version 1.03. June 2007.
- [428] I. Furgel, and V. Saftig. Common Criteria Protection Profile, “Multiple Independent Levels of Security: Operating System” (MILS PP: Operating System). 2016. From <http://dx.doi.org/10.5281/zenodo.51582>
- [429] Common Criteria Portal. Certified Products List - Statistics. From <http://www.commoncriteriaportal.org/products/stats/>
- [430] R. DeLong. The MILS Architecture - a Foundation for Dependable Systems. 2012.
- [431] P. H. Feiler, D. P. Gluch, and J. J. Hudak. The Architecture Analysis & Design Language (AADL): An Introduction. 2006. From http://resources.sei.cmu.edu/asset_files/TechnicalNote/2006_004_001_14678.pdf
- [432] D-MILS. D2.1 Specification of MILS-AADL. 2014. From <http://www.d-mils.org/page/results>
- [433] A. Arbesman, It’s complicated. January 2014. <https://aeon.co/essays/is-technology-making-the-world-indecipherable>
- [434] <http://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>
- [435] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat and R. Sirdey, “Recent advances in homomorphic encryption”, *IEEE Signal Processing Magazine* 30:108-117, 2013
- [436] <https://www.greenbiz.com/article/greenbiz-101-get-smarter-artificial-intelligence>
- [437] Carlo Reita Prospects for nanoelectronics beyond 7nm node – Nano Korea 2016, July 13th 2016

5 PROCESS

The HiPEAC Vision is a bi-annual document that presents the trends that have an impact on the community of High Performance and Embedded Architecture and Compilation. The document is based on information collected through different channels:

- Meetings with teachers and industrial partners during the ACACES 2015 and ACACES 2016 Summer Schools;
- A survey circulated to all HiPEAC members, and which received more than 30 responses;
- A dedicated workshop during the HiPEAC Computing Systems Week in Porto on 21 April 2016;
- A workshop with HiPEAC members and external invitees in Brussels on 27 June 2016;
- Three workshops organised in Brussels in cooperation with the DGConnect - Technology & Systems for Digitising Industry:
 - Workshop on “Computing for cyber-physical systems in 2025” on 27 April 2016;
 - Workshop on “Smart Anything Everywhere 2016: Enhancing digital transformation in European SMEs” on 13 June 2016;
 - Workshop on “Advanced Computing and Cyber-Physical Systems 2016” on 14 June 2016;
- A dedicated feedback workshop during the HiPEAC Computing Systems Week in Dublin on 8 November 2016.
- Presentations at DG Connect University on 10 October 2016, meeting with Electronics Industry on 12 October 2016, Road4CPS meeting on 15 November 2016.

The document is called a ‘Vision’ because it is the result of the interpretation of the trends and directions as seen by the HiPEAC community. As HiPEAC has no direct power to enforce the recommendations, the timeline associated with the potential implementation of the recommendations is uncertain; this is why the document is not a roadmap *per se*.



Roadmap Feedback Session

6

ACKNOWLEDGEMENTS

This document is based on the valuable inputs from the HiPEAC members. The editorial board, composed of Marc Duranton (CEA), Koen de Bosschere (Ghent University), Christian Gamrat (CEA), Jonas Maebe (Ghent University), Harm Munk (Astron/Radboud University) and Olivier Zendra (Inria), would like to thank particularly: Catherine Roderick (BSC), Vicky Wandels (Ghent

University), Eneko Illarramendi (Ghent University), Albert Cohen (Inria), Carlo Reita (CEA-Leti), Holger Blasum (Sysgo), Benoît De Dinechin (Kalray) and all the numerous people that provided support and information during the summer schools, Computing Systems Weeks and other events.

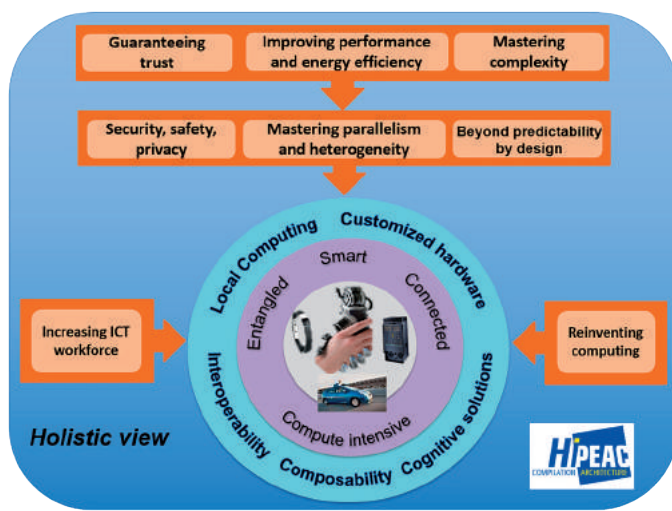


HiPEAC Roadmaps, since 2008.

HIGHLIGHTS OF THE HIPEAC VISION 2017

Information technology is one of the cornerstones of modern society and it is evolving rapidly: while the main challenges identified in the HiPEAC Vision 2015 remain valid and have even increased in importance, new challenges are ahead of us.

Computers are disappearing from view. They are taking on new forms, such as cars, smart meters, thermostats, and so on. They communicate with their users using voice, sound, pictures and video, closely resembling human interaction. We are entering the Artificial Intelligence era. This will not only change how we interact with machines, but it will also redefine how we instruct a machine what to do: less programming and more learning.



The function of the computer is shifting from carrying out computational tasks to provide answers to numerical problems, to working together with humans (what we call the **beginning of the Centaur Era***), augmenting reality to assist us, or even creating virtual worlds for us to explore: the **cyber-physical entanglement** between the physical and virtual world.

Computers will increasingly interact with the physical world, leading to a **expansion from security into safety**. Humans need

to **trust** both the machines and the information that they keep about us, and therefore **enforcement of security and privacy** is of paramount importance.

For compute-intensive tasks, we will continue to use the cloud and supercomputers (HPC); this means that connectivity is crucial, yet **local processing** is becoming increasingly important. The increasing computational requirements are making computer system architects look for **accelerators for specialized tasks**, diverting in many cases from the traditional Von Neumann architecture.

Energy efficiency of computing systems remains a major challenge for the coming years.

As the cost per transistor is no longer decreasing, we might see **diversified paths for using silicon technology**: many designs will not use the latest technology node, but the more mature (and cheaper) one. It is also the right time to **revisit the basic assumptions** in order to open new tracks and approaches and to eventually **reinvent computing**.

With the flood of new systems and new system architectures, increasing attention must be paid to **composability and interoperability** between systems. The complexity of the new systems will be so high that human designers will only be able to master it with the help of computers using AI-based techniques. Innovative approaches will be required to **ensure that the systems will do what they are supposed to do**, both at the functional and at the non-functional level (e.g. timing requirement or reliability). We need to develop design techniques that go **beyond predictability by design** and allow the **building of reliable systems from unreliable parts**.

Finally, **holistic approaches**, implying multi-disciplinary techniques, will be needed in order to meet all the requirements of trustability, efficiency and cost.

* In Advanced Chess, a 'Centaur' is a man/machine team. Advanced Chess (sometimes called cyborg chess or centaur chess) was first introduced by grandmaster Garry Kasparov, with the objective of a human player and a computer chess program playing as a team against other such pairs (from Wikipedia).

ISBN 978-90-9030182-2



9 789090 301822 >