Technology session

Marc Duranton (CEA)
Jo De Boeck (imec)
15:00 Introduction, goals of the workshop

15:15 The Research Agenda for process technology, equipment, material and manufacturing (Jo de Boeck)

15:45 The Research Agenda for computing and storage (Marc Duranton)

16:15 Open discussion on
  • new trends,
  • gaps,
  • connecting application needs with tech challenges and key drivers (HPC, Robotics)
  • what should be done in Europe?

16:55 Close of the session
Electronics value chain
ECS Strategic Research Agenda

KEY APPLICATION AREAS

- Transport & Smart Mobility
- Health & Well-Being
- Energy
- Digital Industry
- Digital Life

ESSENTIAL CAPABILITIES

- Systems and Components: Architecture, Design & Integration
- Connectivity & Interoperability
- Safety, Security & Reliability
- Computing & Storage
- Process Technology, Equipment, Materials & Manufacturing for ECS

OUR DIGITAL FUTURE
The Research Agenda for process technology, equipment, material and manufacturing

Jo De Boeck

Content highlights

• Major challenges
  1. Develop advanced logic and memory technology for nanoscale integration and application-driven performance
  2. Develop technology for heterogeneous SoC integration
  3. Develop technology for advanced packaging and heterogeneous SiP integration
  4. Extend world leadership in semiconductor equipment, materials and packaging

• Strategy
  1. Explore new avenues of applications for ECS
  2. Implement pilot lines and test beds for ECS
  3. Demonstrate manufacturing-line capabilities for flexible, high-quality, competitive and ‘green’ semiconductors
  4. Invest in workforce education

Arco.Krijgsman@ASML.com
Gerhard.Doell@zeiss.com
Jo.DeBoeck@imec.be
Changes in the SRA

• Executive Summary strengthen the link to advancing applications, most importantly quantum information processing.

• The impact section:
  • On FET flagship (Graphene and Quantum technologies computing),
  • On advanced education and training and
  • On technology sovereignty in light of the growing strategic importance of a local base for KET.

• The following changes were made in terms of the major challenges:
  • Challenge 1 - added detail on nanowires, functional oxides, 2D and high-performance switches.
  • Challenge 2 – added new building blocks, materials and applications.
  • Challenge 3 – added new applications, such as quantum computing and testing.
  • Challenge 4 – used more technology-based terminology for the sub-challenges, and the addition of new technology challenges.
Co-optimisation

• Managing critical, autonomous, cooperating, evolvable systems
• Managing complexity
• Managing diversity
• Managing multiple constraints
• Integrating features of various technologies and materials into miniaturised smart components
• Effectively integrating modules for highly demanding environments
• Increasing compactness and capabilities by functional and physical systems integration
Co-optimization of system needs and technology development
Long term vision
The Research Agenda for computing and storage

Marc Duranton
Computing and Storage

Content highlights

• 9.3. Major Challenges
  • 9.3.1. Increasing performance at acceptable cost
  • 9.3.2. Making computing systems more integrated with the real world
  • 9.3.3. Making "intelligent" machines
  • 9.3.4. Developing new disruptive technologies

• 9.4. Make it happen
  • 9.4.1. Educational Challenge
  • 9.4.2. Standardization
  • 9.4.3. Advices for policy makers
    • Boosting innovation and education in computing and storage in Europe
    • Organize the computing community - "synergies -"
    • Foreign export restriction & sovereignty
    • Favorite emergence of AI-enabling technology to maintain and make industry
Exponential increase of performances in 33 years

Production car of 1985
Lamborghini Countach 5000QV
Max speed 300 Km/h

27 times the speed of light
Warp 3?
Star Trek Enterprise
(Year: about 2290)
WHAT WILL BE THE NEXT TECHNOLOGY?

And after CMOS?

- Quantum computing?
- Neuromorphic?
- Optical computing?
- Spin-based computing?
The Hype cycle - 2018

- Deep Learning
- Virtual assistants
- DNN Asics
- Smart Robots
- Autonomous Mobile Robots
- Autonomous Driving

- More “CPS” (interaction with the real world)
- More “Intelligence”
Computing and Storage

Content highlights

• 9.3. Major Challenges
  • 9.3.1. Increasing performance at acceptable cost
  • 9.3.2. Making computing systems more integrated with the real world
  • 9.3.3. Making "intelligent" machines
  • 9.3.4. Developing new disruptive technologies

• 9.4. Make it happen
  • 9.4.1. Educational Challenge
  • 9.4.2. Standardization
  • 9.4.3. Advices for policy makers
    • Boosting innovation and education in computing and storage in Europe
    • Organize the computing community – “synergies –
    • Foreign export restriction & sovereignty
    • Favorite emergence of AI-enabling technology to maintain a high-stake industry
Computing and Storage

Changes: several additions

- Computing now forms a continuum between extreme edge devices, edge devices, IoT, fog, cloud and HPC.
- Energy consumption of cloud/computing is a growing concern.
- AI requires much computing power (and therefore energy) for the learning and applications require highly efficient accelerators for inference.
- GAFA and BATX are now developing their own hardware, and Europe has to act quickly to remain competitive.
- Open source hardware, such as RISC V, is becoming increasingly attractive.
- CPS implies multi-domain/multi-paradigm design and global system view, which is becoming also an educational challenge.
- Software complexity is a real bottleneck.
- AI techniques can be used to help design better hardware and software.
Computing now forms a continuum enabling edge intelligence.

CPS + AI: Processing, Understanding as soon as possible.

Transforming data into information as early as possible.

Enabling intelligent data processing at the edge:
- Fog computing
- Edge computing
- Stream analytics
- Fast data...

True collaboration between edge devices and the HPC/cloud -> creating a continuum.

aka Intelligent Embedded Systems
aka Autonomous CPS (ACPS)
aka Cognitive CPS (C²PS)
The problem:
IT projected to challenge future electricity supply

Exponential power consumption

Expected case scenario

From “Total Consumer Power Consumption Forecast”, Anders S.G. Andrae, October 2017
EXPONENTIAL INCREASE OF COMPUTING POWER
FOR AI TRAINING

“Since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.5 month-doubling time
(by comparison, Moore’s Law had an 18-month doubling period)**

Peta= $10^{15}$

* [https://blog.openai.com/ai-and-compute/](https://blog.openai.com/ai-and-compute/)
## Common carbon footprint benchmarks

<table>
<thead>
<tr>
<th>Activity</th>
<th>Emission (lbs of CO2 equivalent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Roundtrip flight b/w NY and SF (1 passenger)</td>
<td>1,984</td>
</tr>
<tr>
<td>Human life (avg. 1 year)</td>
<td>11,023</td>
</tr>
<tr>
<td>American life (avg. 1 year)</td>
<td>36,156</td>
</tr>
<tr>
<td>US car including fuel (avg. 1 lifetime)</td>
<td>126,000</td>
</tr>
<tr>
<td>Transformer (213M parameters) w/ neural architecture search</td>
<td>626,155</td>
</tr>
</tbody>
</table>

Chart: MIT Technology Review • Source: Strubell et al. • Created with Datawrapper

“Artificial Intelligence at the Edge” hardware is well recognized… in USA

Intel® Neural Compute Stick 2 at $99
Powered by the Intel® Movidius™ Myriad™ X Vision Processing Unit

Google Edge TPU
The $75 USB Accelerator provides the Edge TPU’s high-performance ML inferencing

NVIDIA Announces Jetson Nano:
$99, 472 GFLOPS, 5 W, 69.6 mm x 45 mm
NVIDIA CUDA-X AI Computer
(Performances x236 compared to Cray 2)

For example, the European car industry which is very good up to know, is threatened by control of the mandatory-in-near-future of AI inside dominated by non EU countries...
Evolution of society: data (information) is the new fuel

Most of them are also developing (AI) hardware!
Hardware developments for AI accelerators will grow
Are we able to estimate the impact of Open Source?

Pivotal acquisition foreshadows strategic shifts at VMware; VMworld announcements confirm them

NVDA

The NVIDIA Deep Learning Accelerator (NVDA) is a free and open architecture that promotes a standard way to design deep learning inference accelerators. With its modular architecture, NVDA is scalable, highly configurable, and designed to simplify integration and portability. The hardware supports a wide range of IoT devices. Delivered as an open source project under the NVIDIA Opensource NVDA License, all of the software, hardware, and documentation will be available on GitHub. Contributions are welcome.

- Open Source: Developed on GitHub in an open directed community where contributions are encouraged.
- Complete Solution: Comes complete with a Verilog and C-model compiler, Linux drivers, test benches and test suites, kernel- and user-mode software, and software development tools. Easily portable to other operating systems.
- Scalable: Well-suited to scale across a wide range of IoT devices.
- Proven Hardware Architecture: Based on Xavier — the world’s first autonomous processor that NVIDIA designed for automotive products and more — and backed by a full verification suite.
- Deep Learning Savvy: Smart, efficient, and ready to work with the wide range of NVIDIA supported solutions.

MOST MICROPROCESSORS—the chips that do the grunt work in computers—are built around designs, known as instruction-set architectures (ISAs), which are owned either by Intel, an American giant, or by Arm, a Japanese one. Intel’s ISA powers desktop computers, servers and...
At the hardware level, the good old Von Neumann/CMOS partnership can act as a computing substrate, or *orchestrator* of various accelerators/technologies

- Acting as coordination/communication node
- Allowing Hardware/Software integration
PROCESSOR ARCHITECTURE EVOLUTION?

End of Dennard’s scaling

Mono-core architecture for single thread performance

Many-core architecture for parallelism

Memory invades logic

Moore’s law slow-down

Disruptive Architecture for data management

In Memory Computing, Neuromorphic Computing.

Data Centric Interconnect

~2006

~2016

~2026...

Original slide from Denis Dutoit, CEA
MANAGING COMPLEXITY

Cognitive solutions for complex computing systems:

- Using AI and optimization techniques for computing systems
  - Creating new hardware
  - Generating code
  - Optimizing systems
- Similar to *Generative design* for mechanical engineering

“And that’s why we need a computer.”
Open discussion

introduction
LOW POWER & increased functionality

- **LOW POWER**
  - Ultra-low power
  - EDGE

- **HIGH PERFORMANCE**
  - Increased performance at constant power density
  - MOBILE

- **IOT**
  - Increased performance at constant leakage

**SYSTEM PERFORMANCE (Giga-Operations/sec)**

**POWER (Watt)**

- 0,0001
- 0,01
- 1
- 10
- 100
- 500

**GRAPHIC**: Various devices illustrate the concepts of low power and increased functionality across different performance levels.
ARTIFICIAL INTELLIGENCE IS A KEY STRATEGIC DRIVER

Augmenting human intelligence by technology

Pre-processing of data & understanding data patterns

Making useful predictions & decisions

Triggering action and self-regulation
ARTIFICIAL INTELLIGENCE IS A KEY STRATEGIC DRIVER

- More energy-efficient computation
- Higher bandwidth networking
- More intelligent sensors & actuators
- More efficient powering

Compute Store Connect Sense Actuate Power
AI CHIPSET sales to double over 5 years

Total Annual Revenue from AI Chipset Sales
World Markets, Forecast: 2017 to 2024

Source: ABI Research
Next Generation Computing MISSION

- **Drive** high-impact technology research from system perspective
- **Evaluate** technology opportunities for system companies

NGC

- Prototype Demonstration
- Hardware exploration
- System-Technology Evaluation
- Advanced Technology Modules
Open Source hardware leading to Open Innovation?

- Remove upfront barriers
- Share cost among users
- OS Business models encourage innovations
- Owned by a community

Original slide from Arnaud Samama, Thalès
Open discussion

- New trends
- Gaps
- Application <> Tech challenges <> Key drivers (HPC, Robotics)
- Europe?

Strengths / Weakness / Threat / Opportunity
New trends....
Gaps....
Application <> Tech challenges
<> Key drivers (HPC, Robotics)
Position of Europe, recommendations...