WEBINAR

SRIA* - Chapter 10
Opportunities for Devices and Components

André Bourdoux

IMEC

* SRIA = Strategic Research and Innovation Agenda
"Smart Networks in Context of Next Generation Internet"

Chapter 10
“Opportunities for Devices and Components ”

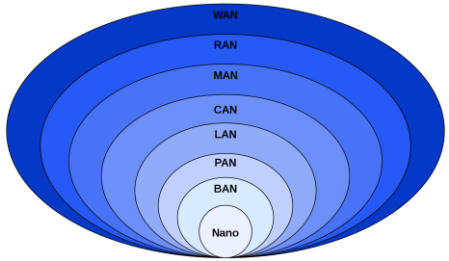
Table of content

Chapter 10
"Opportunities for Devices and Components "

- 10.5 High-speed Transceivers, Wireline and Optical
  - 10.5.1 Radio-over-fibre communication, sub-systems and components for B5G and 6G networks
  - 10.5.2 Terabaud capable opto-electronic transceivers
  - 10.5.3 Ultra low-cost and low-power coherent "lite" transceivers
  - 10.5.4 Optically assisted wireless subsystems
- 10.6 Baseband Modems
- 10.7 Processors for Cloud-AI, Edge-AI and on-device-AI
- 10.8 Memories
  - 10.8.1 Memory technologies towards 2030
  - 10.8.2 Compute-in-Memory
- 10.9 Hardware for Security
- 10.10 Opportunities for IoT Components and Devices
  - 10.10.1 Approach for components
  - 10.10.2 Approach for devices
  - 10.10.3 Requirements for IoT devices

10.1 - Sub-10GHz

https://commons.wikimedia.org/w/index.php?curid=95509836



- Standards trend: more and more standards!
  - Dominance of Cellular (xG), Wi-Fi, Bluetooth, GPS
  - Many other systems: NFC, IoT, WAN, ...
- More efficient use of spectrum
  - MIMO, multiple bands
- Higher integration
  - SoC (RF, analog, ADC/ADC, DSP, CPU)
- Front-end module and antenna nightmare
  - RF filters, switches, multi-band antennas, multiple channels
- Power, autonomy
- Technology choice
  - Best RF (linearity, filtering), fastest digital (digital RF), lowest power
- New opportunities, new challenges
  - e-health: wearables, implantables, ingestibles, ...
  - Human-machine interface, brain-machine interface
  - Mobile display technologies: AR/VR/XR, glasses, ...
  - Sensing: passive, active (! self-interference)

This may look like "business as usual" but the requirements and constraints keep growing and so does the number of different platforms

10.2 - Millimeter-wave and TeraHertz

- The 90 to 300GHz range has great potential
  - Access, P2P and fronthaul/backhaul
- At higher frequencies:
  - Higher free-space loss must be compensated by higher antenna gain
  - Front-end becomes more challenging
  - Ultra-wide bandwidth → multi-GHz baseband and 10+Gsps ADC/DACs
  - Improvements in circuit design needed
    - Phase noise, noise figure, IQ mismatch, ...
    - Frequency dependent effects: group delay distortion in all components
    - Beamforming: phase shifters vs true-time delay.
    - Efficient beamforming remains a challenge, especially for high gain, large bandwidth
  - Chip, chip interconnect and antenna module must be co-designed
    - Minimize interconnect lengths, losses
    - 2D, 2.5D and 3D electromagnetic simulations
- Air interface design exploiting ultra-wide bandwidth, very directional beamforming and “front-end friendly”
- Move digital processing to analog (equalization, synchronization, ...)

$$P_r = \frac{P_T G_T G_R \lambda^2}{(4\pi R)^2} = \frac{P_T A_T A_R}{(\lambda R)^2}$$

10.2 - Millimeter-wave and TeraHertz

- Solid-state technologies
  - CMOS no longer the panacea: must be replaced or *complemented* with III-V
  - Huge trade-off:
    - Chip partitioning
    - Improve RF circuits vs calibration vs digital compensation
    - Technology choice
  - Many options:
    - Silicon-based: RF-SOI, FD-SOI, SiGe BiCMOS
    - III-V on silicon substrates: GaAs/Si, GaN/Si
    - III-V on native substrates: InP
    - III-V on CMOS
  - With further scaling, CMOS will transition from FinFET to gate-all-around structure
    - Impact on 10+Gsps ADC/ADC
- Not only wireless comm: convergence of communications and sensing
  - Passive THz imaging
    - Above-IC bolometer: better performance but expensive
    - Monolithic CMOS-based imagers: much lower performance but cost-effective
  - Active mm-wave and THz imaging
    - Higher frequencies enable smaller devices/better angular resolution and larger bandwidth/range resolution
    - Antenna options include on-chip and on-package
  - Imaging at >100GHz expected to boom and help driving circuit and technology research towards higher performances, smaller form factors and lower cost

10.3 - Ultra-low Power Wireless

- IoT to grow to 100 billion by 2030
  - ULP sensors
  - e-health: wearables, implantables, ingestibles, brain-machine interface, active eye lenses, ...
- Huge challenge towards zero or near-zero power
  - Profound impact on the complete transceiver architecture and design and the protocol
- Battery-free operation
  - Energy scavenging
  - Wake-up receivers
  - Back-scattering devices
- Degradable devices
  - Huge # devices → huge e-waste at end of life
  - Bio-degradable substrates
  - Renewable materials

Moving to >100GHz brings new challenges

- Packaging for consumer equipment is a challenge
- Lossy interconnects → avoid interconnects → on-chip or in-package antenna
- Lens can help to increase the gain but reduces the field-of-view
- Dual polarization (because of mobility)
- Wide bandwidth
  - Frequency-dependent behaviour
- Coupling between antenna elements
- MIMO arrays, hybrid beamforming architectures
- Array calibration
  - Can be expensive
  - Fully-automatic, both start-time and run-time

Metamaterials and metasurfaces

- Allow to manipulate electromagnetic waves
- → huge potential for antennas and surfaces (“Intelligent Reflective Surfaces”)
- Coverage increase, smart radio environment, better cell edge coverage, less interference and electromagnetic pollution
- Can be combined with antenna technologies, massive MIMO, mm-wave and THz communication, D2D
- Applications include:
  - Antenna design, absorbers, reflectors, superlenses, cloaking devices, RCS manipulation (radar),
- Very active field of research, many innovations and disruptions

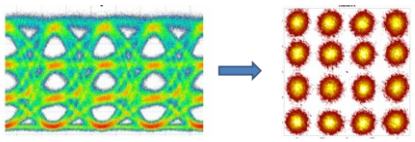
Radio-over-fibre communication, sub-systems and components for B5G and 6G networks

- Fronthauling needs explode with coordinated BF, CoMP, massive MIMO and cell-free MIMO
- CPRI and OBSAI are expected to saturate
- Example:
  - 2GHz BW, 4 carriers, 3 sectors each with 32 antennas, 8bits I&Q, 8B/10B encoding, 10% overhead → sustained throughput of 25Tb/s
- Innovative fronthauling solutions are needed
  - Analog RoF (high linearity needed)
  - RF Sigma-delta modulation
  - ...
- Split processing trade-offs between RRH and BBU

10.5 - High-speed Transceivers, Wireline and Optical

Towards Terabaud capable opto-electronic transceivers

- Traffic to Data Centres is exploding
  - Higher telecom needs + cloud-based ML/AI + ...
- Need for new generations of optical transceivers with ever higher capacity
- Deployment of optical links at ever shorter distances
- More pervasive use of coherent transceiver technologies
  - From long-haul to metro to data centers to access



- Need for electro-photonic Systems-in-Package and co-packaged optics
  - Optical transceiver chiplets + CMOS data processing in one package
- Increase
  - Symbol rate: 100G → 200G → 400G → 800G → 1.6T → 3.2T ... transceivers
  - Number of parallel lanes: (multiple wavelengths and/or fibres)
  - Higher spectral efficiencies: 4-PAM → complex modulation
  - Integration: denser integration e.g. 3D modules
- Enabling technologies:
  - CMOS → SiGe, InP
  - Novel materials for ultra-broadband optical modulators and detectors: e.g Organic hybrid material, Ferro-electric materials, Lithium Niobate ($LiNbO_3$)
  - Monolithically integrated optics and electronics
  - Optically assisted analog-to-digital and digital-to-analog conversion

10.5 - High-speed Transceivers, Wireline and Optical

Ultra low-cost and low-power coherent “lite” transceivers

- Need for coherent detection for shorter ranges and at very low-cost
- Potential enabling technologies:
  - Integrated narrow linewidth laser sources
  - Integrated optical phase locked loops
    - For carrier recovery
  - Novel equalization approaches relying on co-developed opto-electronics
    - Move compute-intensive digital functions to optical e.g. passive optical filter

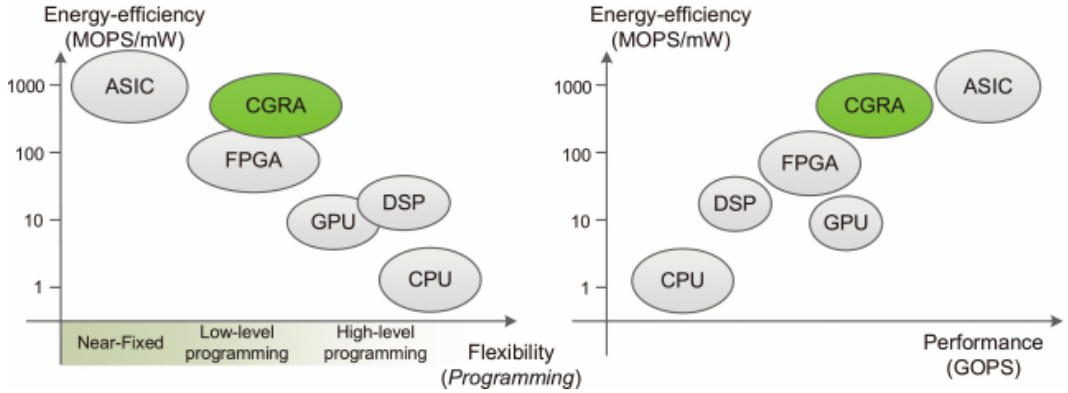
Optically assisted wireless subsystems

- Microwave photonic techniques to replace conventional beamforming
- ...

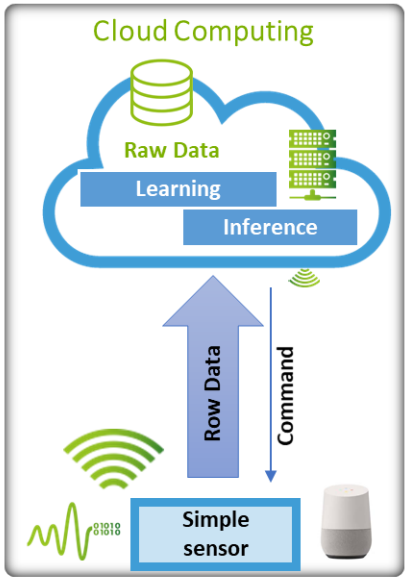
10.6 - Baseband Modems

Picture from : ["A Survey of Coarse-Grained Reconfigurable Architecture and Design: Taxonomy, Challenges, and Applications", Liu et al., 2019]

- Architectural trade-off
  - Flexibility/Programmability
  - Performance
- According to application
  - IoT, UE, Infrastructure
  - SISD, SIMD, MIMD
- Trends
  - Coarse-grained reconfigurable architecture
    - Near-ASIC performance with SW-like programmability
  - Array of COTS CPUs
    - Better suited for infrastructure
  - Deep-learning architectures for PHY processing
- Challenge for very high throughput: from nJ/bit to pJ/bit
- Challenge for very low throughput: from sub-mW to sub-µW
- Memory architecture/hierarchy!
  - A lot of power goes in data exchange



10.7 - Processors for Cloud-AI, Edge-AI and on-device-AI



2012

AlexNet

Cloud computing is still the workhorse today

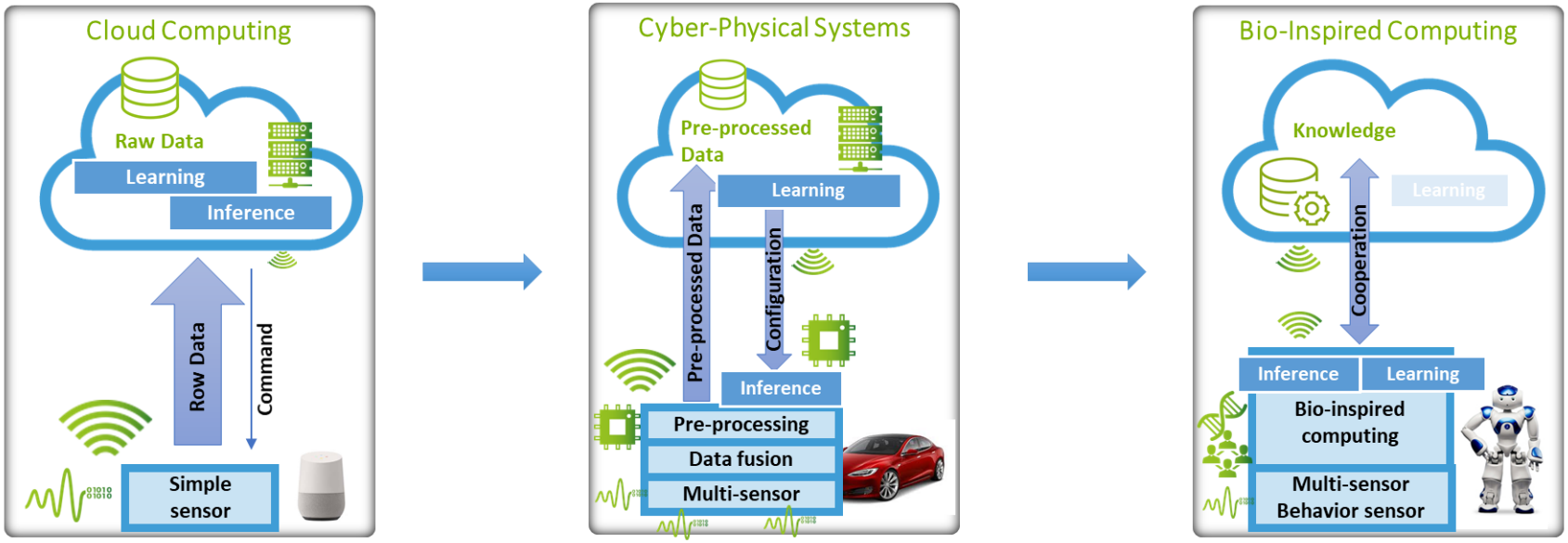
Need for very **high compute power for training** large networks (ex. GPT-3 model with 175B parameters)

→ Large GPUs, TPUs, scalability over multiple nodes

Need for **low latency inference** (on a batch of 1 piece of data)

→ FPGA, CPU

10.7 - Processors for Cloud-AI, Edge-AI and on-device-AI



2012 — AlexNet

2019 — Edge TPU, Movidius …

2025 — ?

But there are ever larger needs for edge computing

Because of safety of operation, latency, privacy, power dissipation …

→ Need for dedicated ASICs, with sensor integration

10.7 - Processors for Cloud-AI, Edge-AI and on-device-AI

2020 NETWORLD

Increased computing efficiency

Weight quantization

Reduced bit accuracy
- Smaller memory footprint
- Lighter operations

Variable bit precision

Handling higher bit accuracy when needed
- For higher inference precision

Sparsity

Skip MAC operations
- When weight or intermediate result is 0

Increased storage efficiency

Near memory computing

Avoid external memory accesses

Weights
- Embedded Non-Volatile Memory

Intermediate results
- SRAM or Embedded DRAM

In-Memory computing

SRAM or Embedded NVM

Digital or analog

Several techniques can be employed at architecture and circuit levels

Keys are to avoid moving data, to quantize weights and activations, increase sparsity …

A completely different paradigm, related to CIM:

- Spiking neural networks (SNN), analog and digital flavour

10.8 – Memories
Memories for processing and storage

We are entering the zettabyte and soon the yottabyte eras: yearly growth rate: 1.2 ... 1.4x

- Yottabyte predicted in 2030
- Data and traffic generated through
  - Apps such as Amazon, YouTube, Facebook, Netflix ...
  - IoT such as autonomous cars, smart buildings, smart city, e-health, ...
- Huge environmental problem: heat and power consumption
- 3-axis performance improvements
  - Density
  - Time: latency and speed/bandwidth
  - Energy
  - ... and Cost ...

10.8 – Memories
Memories for processing and storage

- Growth rate for SRAM and DRAM is saturating and not sufficient towards yottabyte era
- Novel approaches and technologies are needed to sustain the growth rate
  - 3D stacking, already largely exploited
  - Emerging storage class memories to fill the gap betweem DRAM and NAND
    - PC-RAM, VMCO, CB-RAM, OxRAM, ...
  - MRAMs: many variants
    - STT-MRAM (for L3 cache), SOT-MRAM (for L1-L2 cache), VCMA-based MRAM, ...
- DNA storage
  - Highest density *potential* by orders of magnitude
  - Challenges: speed, reliability ... and cost
  - Very long term

	ENERGY REDUCTION FOR A GIVEN THROUGHPUT	DENSITY IMPROVEMENT	SPEED (AT DEVICE LEVEL)
CACHE (SRAM)	1.12x	1.15x	1.1x
MEMORY (DRAM)	1.1X Cs reduction,Vdd reduction	1.2X → 1.1X Slow down with C scaling	1x
STORAGE (FLASH)	1X	1.4X → 1.2X 3D log trend cannot last	1x performance increase from 2D to 3D
ARCHIVAL (TAPES)	1.1x	1.4X Doubling at each LTO node	1.1x

https://www.imec-int.com/en/imec-magazine/imec-magazine-september-2018/emerging-memories-for-the-zettabyte-era

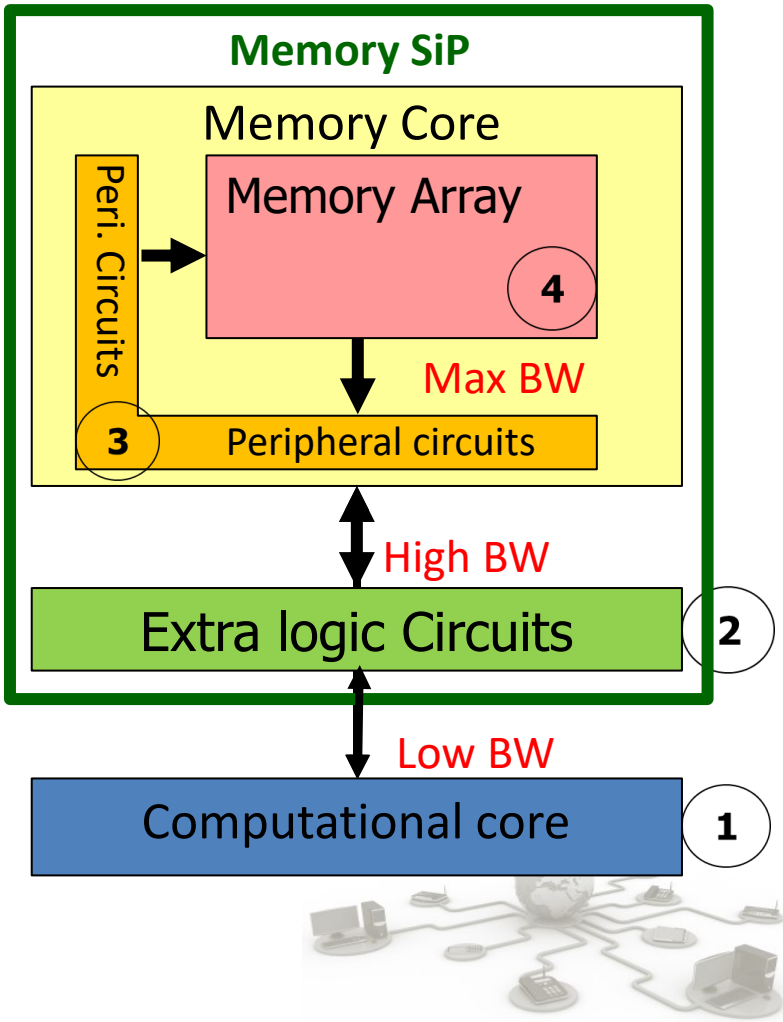
10.8 – Memories
Computation-in-Memory (CIM)

Computer architectures: Classification

- Computation-outside-Memory (COM)
    1. Far
    2. Near
- Computation-in-Memory (CIM)
    3. Periphery
    4. Array

CIM relies on memristive devices

Not a mature technology



10.8 – Memories
Computation-in-Memory (CIM)

Challenges

Technology

- Multi-state behaviour
- Energy switching
- Threshold behaviours
- Endurance
- Fault tolerance
- Variability, R_{OFF}/R_{ON} ratio
- Integration, yield

Circuit design

- High precision programming of NVM
- Fast and energy efficient signal conversion circuits (DAC, ADC)
- Precise measurement of current
- Vector x matrix: output as current
- Control complexity

Architecture

- Micro vs macro architectures
- Intra- and inter-communication
- System accuracy
- Design exploration
- Simulation tools

Tools & applications

- Mapping applications on architecture
- Compilers
- EDA tool chains
- Bridging device characteristics to circuit and to algorithm design
- Simulators

Research on Sustainable Security and Privacy - Motivation

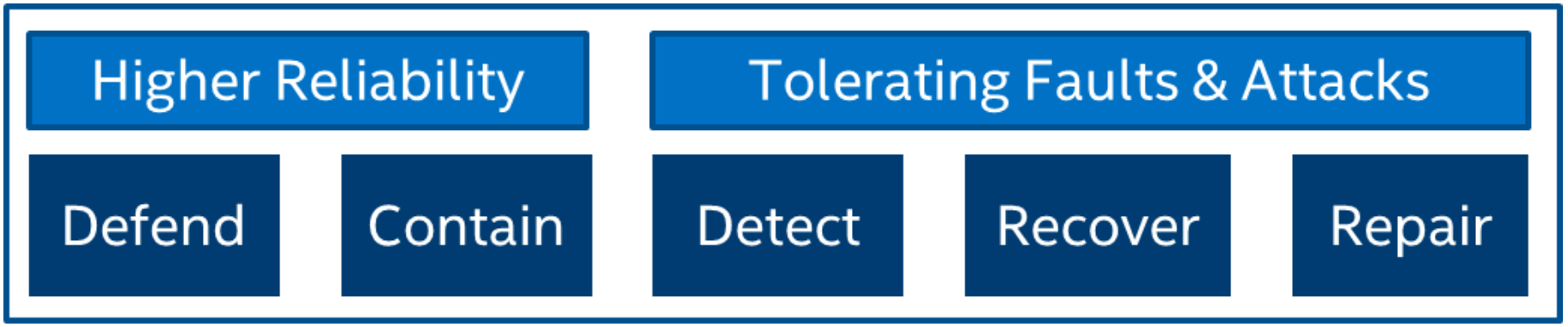
Today's Limitations:

- Devices with limited security life-time
- Devices do not survive attacks – require manual recovery
- Manual mitigation of risks and frequent patching
- Crypto is degrading and suffers against quantum computing attacks

Desired Future:

- Devices that survive for years in the field
- With minimal maintenance and automated recovery
- Guaranteed long-term survival of crypto mechanisms

Research on Sustainable Security and Privacy - Research Vectors



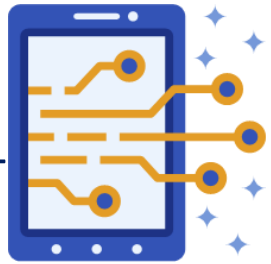
Research: Maintaining Security and Surviving Attacks

- Graceful degradation into fail-secure states, maintaining critical services
- Systems survive attacks with automated recovery

Research: Post-Quantum Cryptography with Hardware Support

- Range of crypto that are robust against quantum computing attacks
- Toolboxes for wide range of usages

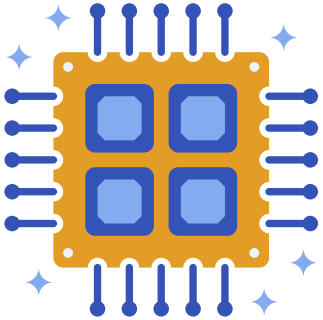
10.10 - Opportunities for IoT Components and Devices



IoT - Components and Devices – Research areas

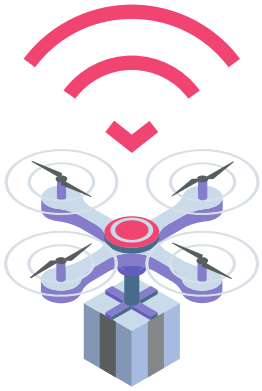
- **Pervasive wireless connectivity** as a major component behind the IoT technology and one of the key layers in IoT and IIoT architecture.
- Research challenges in the development of IoT components and devices for IT/OT integration using **multi-frequency/multi-protocol heterogeneous wireless communication and networking** for IoT/IIoT and edge computing with built-in end-to-end distributed security.
- **Ultra-low power IoT,** extended to Tactile IoT components and on-IoT device AI techniques and methods.
- **Wide frequency range** from sub-1GHz to THz
  - Use of CMOS and III-V semiconductors-based GaAs, GaN, InGaAs, SiC semiconductor technologies. Integrate microwave and analogue front-end technology and millimetre wave monolithic integrated circuits (MMIC).
  - Requires alignment between SNS and KDT.

10.10 - Opportunities for IoT Components and Devices



Approach for IoT devices

- Specialized IoT devices and sensors enabled and validated especially for vertical sectors
  - Leveraging system on chip activities.
  - Specifying the way to communicate in the network/systems .
  - Integrating them in their operational systems in vertical (and as well cross- vertical) application domains.
- Sustainable growth for energy efficient IoT devices development, battery efficiency and battery-free operation.
- Degradable devices and energy autonomous devices that uses ultra-low power radios and harvest the needed energy.



SRIA Chapter 10 Contributors

- André Bourdoux (editor)
- Jan Craninckx
- Piet Wambacq
- Aarno Pärssinen
- Didier Belot
- Yao-Hong Liu
- Ullrich Pfeiffer
- Antonio Manzalini
- Peter Ossieur
- Maziar Nekovee
- Alexandre Valentian
- Said Hamdioui
- Francky Catthoor
- Gouri Sankar Kar
- Valerio Frascolla
- Ovidiu Vermesan
- Luis Perez-Freire
- Georgios Karagiannis