The role of Artificial Intelligence in Telecoms

Ioannis Tomkos
Research Director, Next Generation Networks, AIT

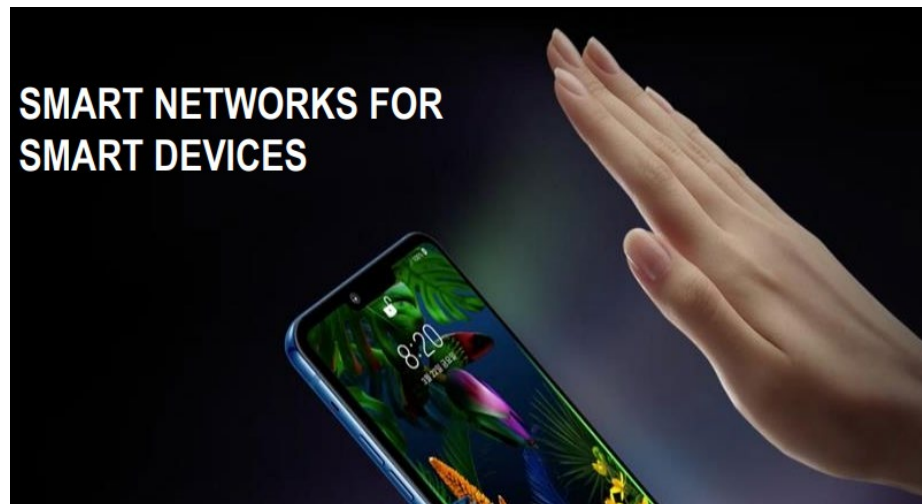Infocom Conference, Cyprus
November 5th, 2019

Overview

- Introduction
  - Which are the key technology areas that introduce significant changes in our economy/society?
    - Next Generation 5G and IoT Connectivity
    - Artificial Intelligence
    - Edge Computing

- Evolution of advance 5G mobile networks towards 6G Smart Networks enabled by the Convergence of Key Digital Technologies

- AI for Telecoms and Telecoms for AI
  - Two sides of the same coin…
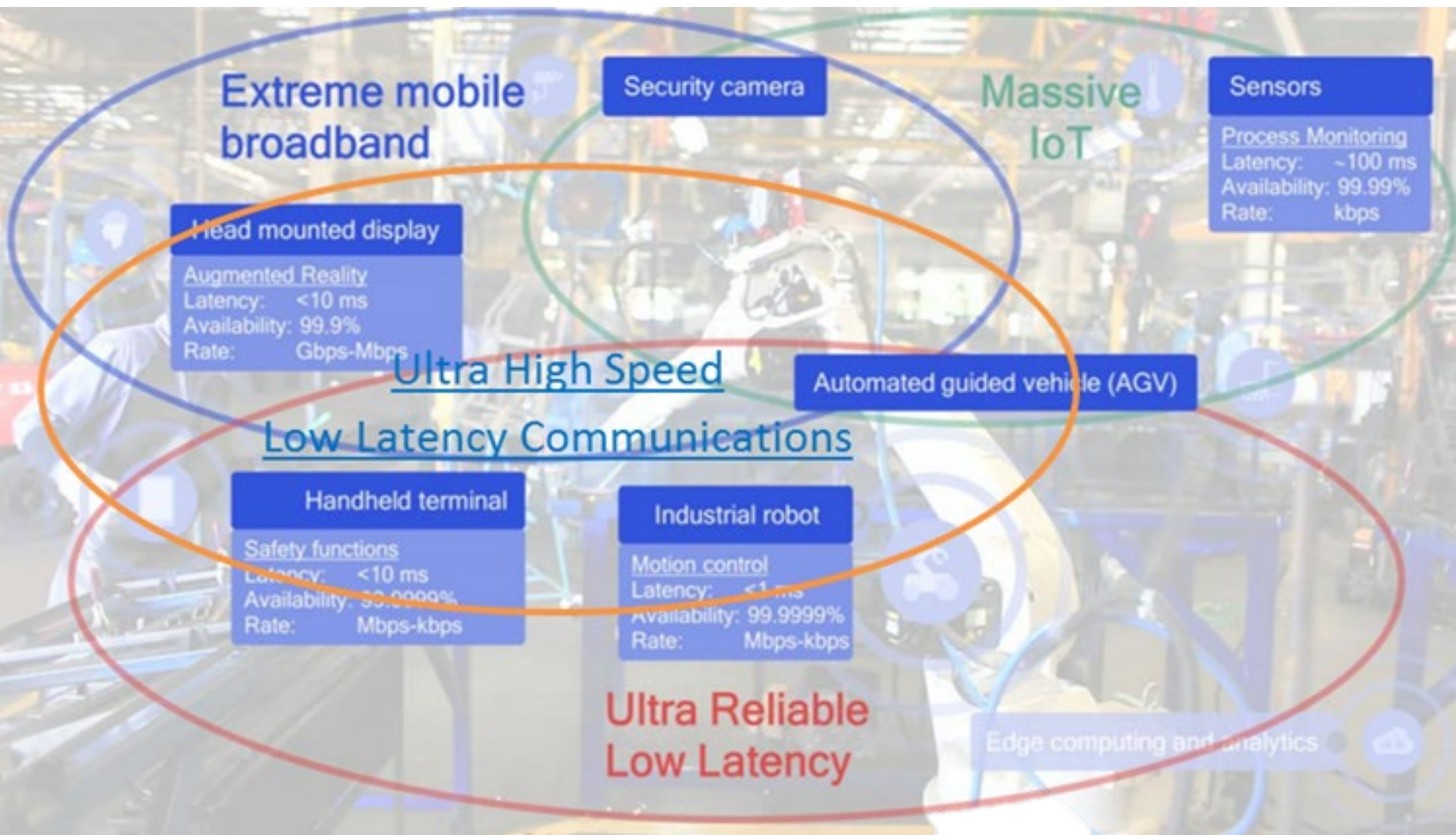
- Summary & Conclusions

We live in a connected world

- The proliferation of smart devices has created a world that is connected more than ever!
    - How much data are we producing on a daily basis?
        - About **2.5 quintillion** (i.e. 2.5 followed by 18 0's) bytes of data are created daily
    - What is the average growth rate for the amount of digital data generated worldwide?
        - On average the estimated rate across all network segments is about of **50% per year**
   - **What is needed?**

SMART NETWORKS FOR SMART DEVICES

Smart Networks require AI!

Next Generation Connectivity

Extreme mobile broadband

Security camera

Massive IoT

Sensors

Process Monitoring
Latency: ~100 ms
Availability: 99.99%
Rate: kbps

Head mounted display

Augmented Reality
Latency: <10 ms
Availability: 99.9%
Rate: Gbps-Mbps

Ultra High Speed
Low Latency Communications

Automated guided vehicle (AGV)

Handheld terminal

Safety functions
Latency: <10 ms
Availability: 99.9999%
Rate: Mbps-kbps

Industrial robot

Motion control
Latency: <1 ms
Availability: 99.9999%
Rate: Mbps-kbps

Ultra Reliable Low Latency
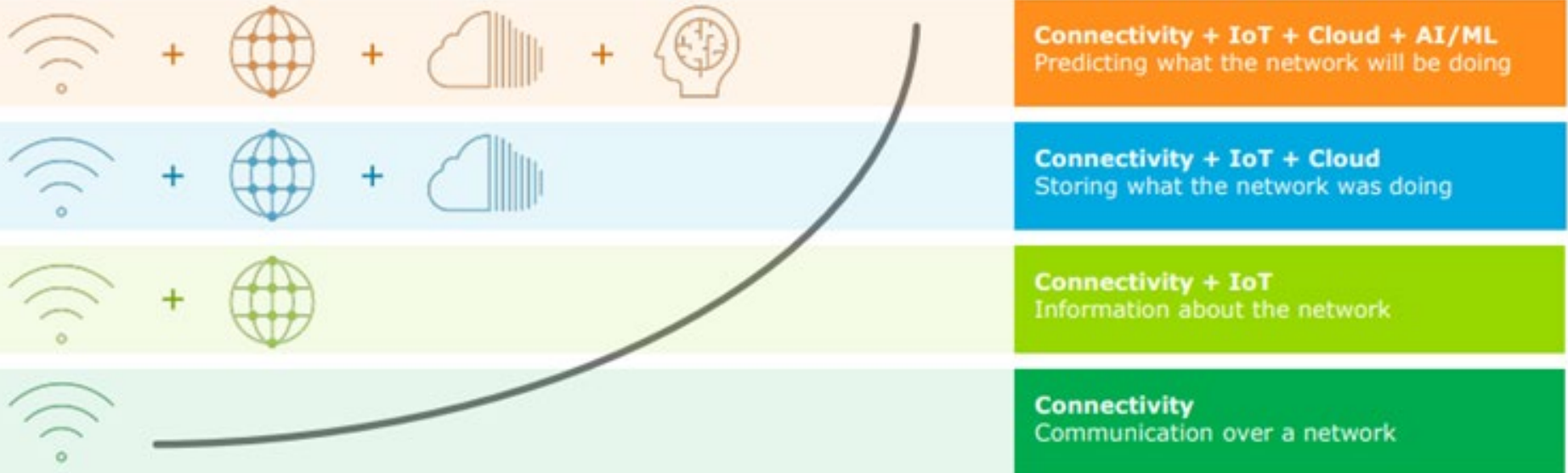
Edge computing and analytics

Besides 5G/IoT connectivity which are the technologies that introduce significant changes in our society?
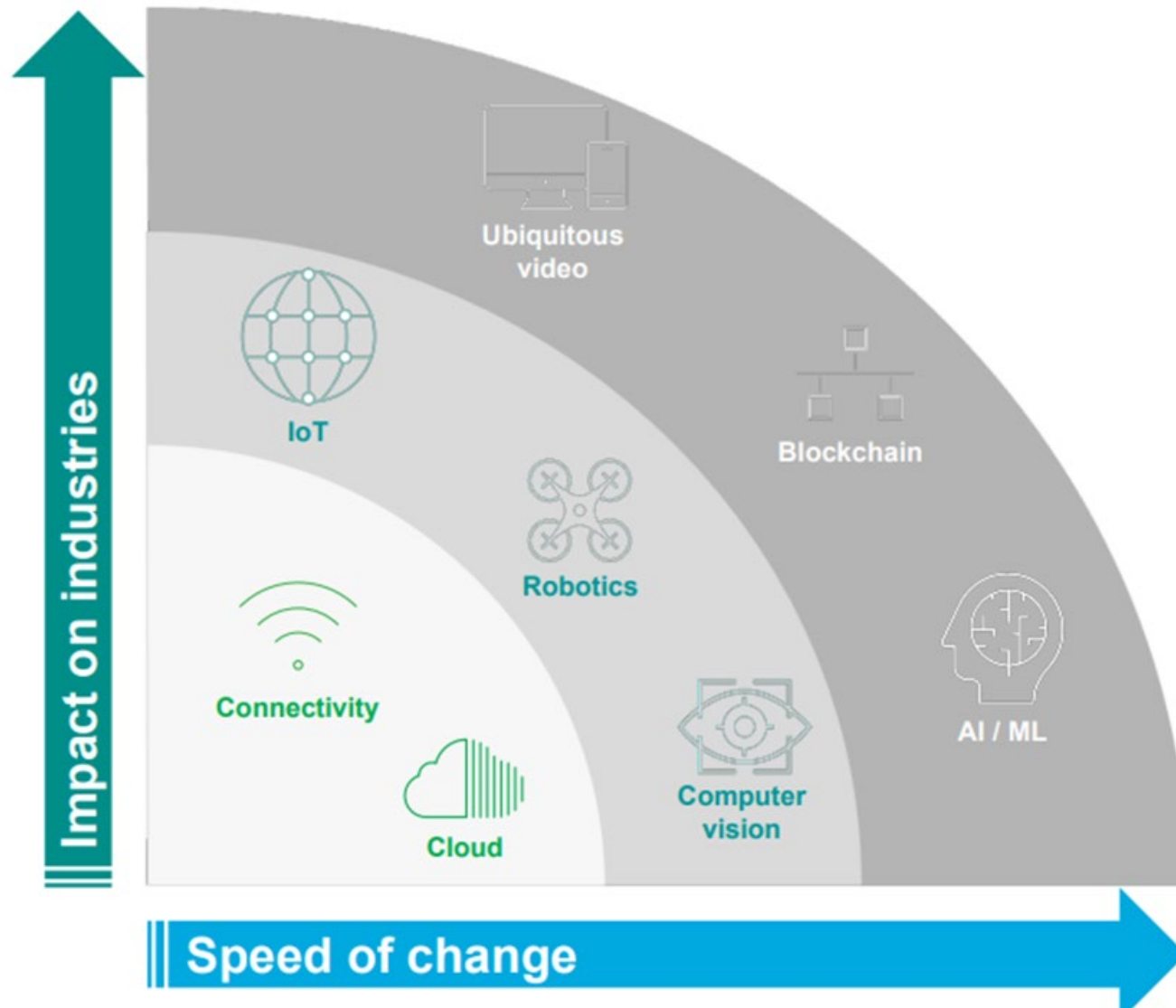
The convergence of key technologies will drive exponential impact for our society
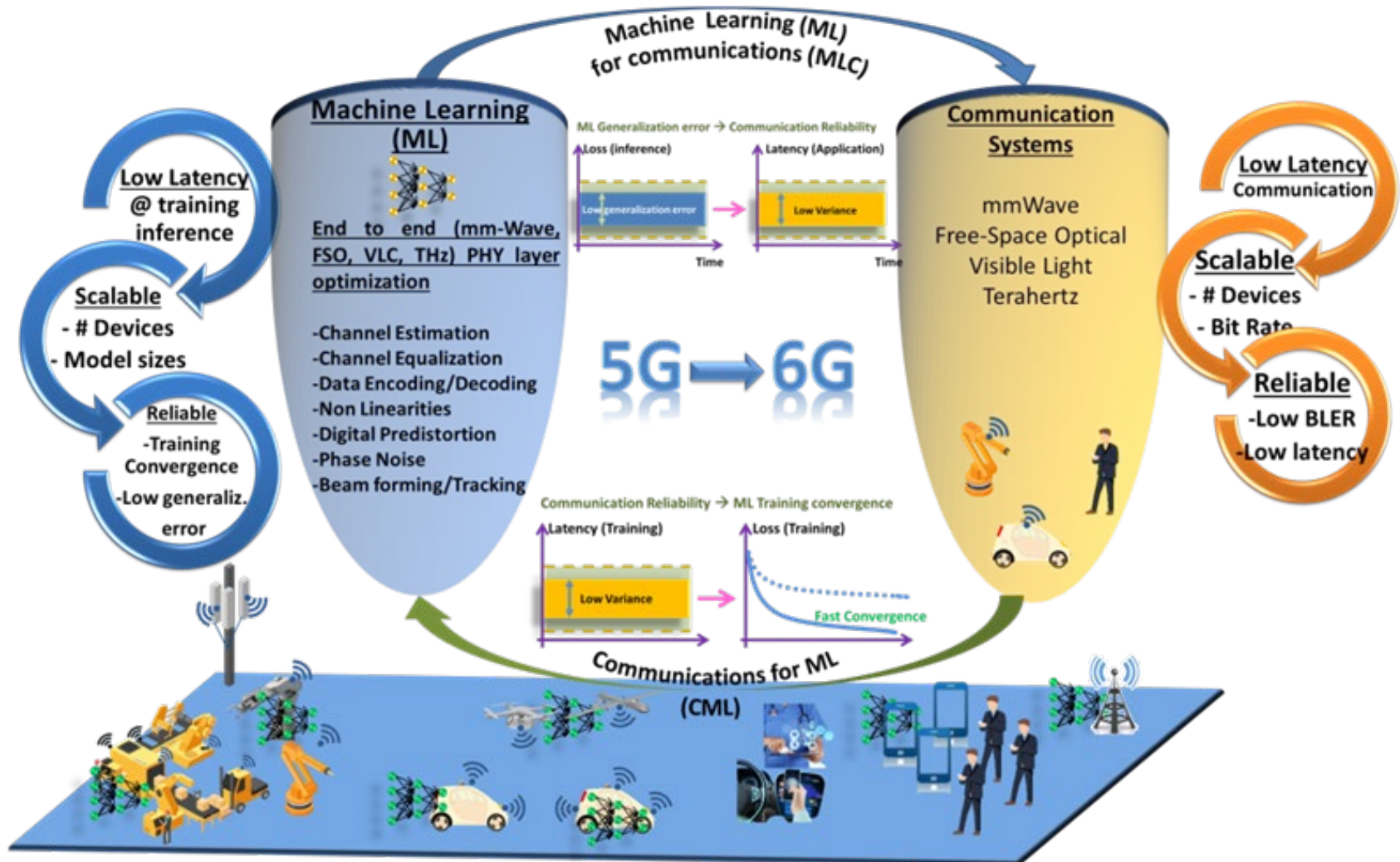
As different transformative technologies come together, impact goes from linear to exponential

Connectivity + IoT + Cloud + AI/ML
Predicting what the network will be doing

Connectivity + IoT + Cloud
Storing what the network was doing

Connectivity + IoT
Information about the network

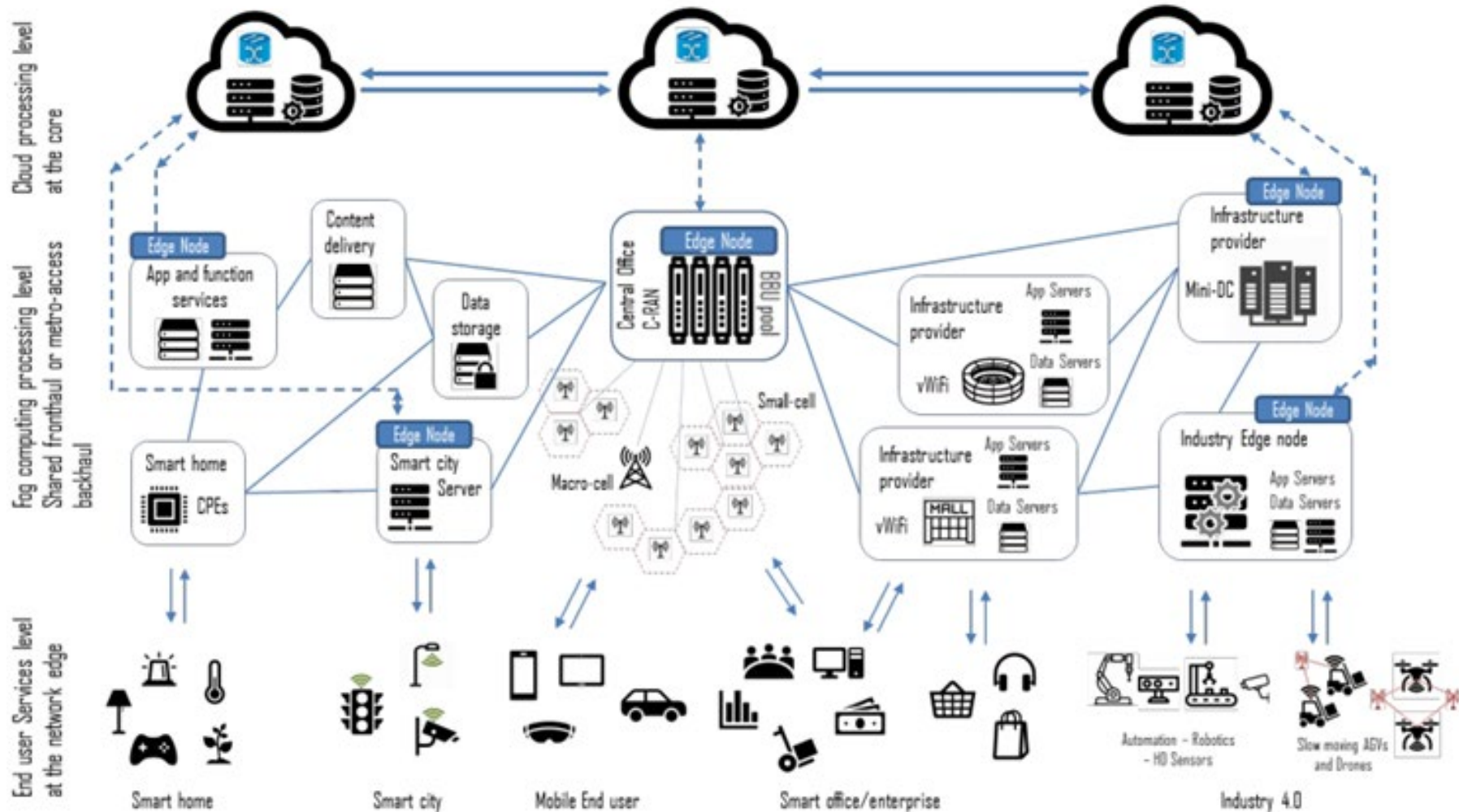**Connectivity**
Communication over a network

Vision for 5G evolution and convergence with NG-IoT and AI towards the 6G of communication systems

- As we move beyond current fifth generation (5G) networks, operating at low (sub-6GHz) radio frequency bands, towards mm-wave ones operating at higher and higher carrier frequencies, (initially at the high radio frequency window, above 20GHz, and then even higher towards optical frequencies in the visible and invisible electromagnetic waves spectrum), the true capabilities of 5G networks will be gradually revealed.

- In the subsequent evolution stage, the technologies of both 5G networks and Next-Generation Internet of Things (NG-IoT) will move towards some form of convergence where the use of mm-wave frequencies will become commonplace.

- Besides the evolution of wireless connectivity at the edge of the network, we anticipate major advances also on the control and management of the edge network that will be supported by the introduction of new technologies like Artificial Intelligence (AI) and Blockchain (BC), which in-turn will make the edge networks much more smart, trustworthy and secure
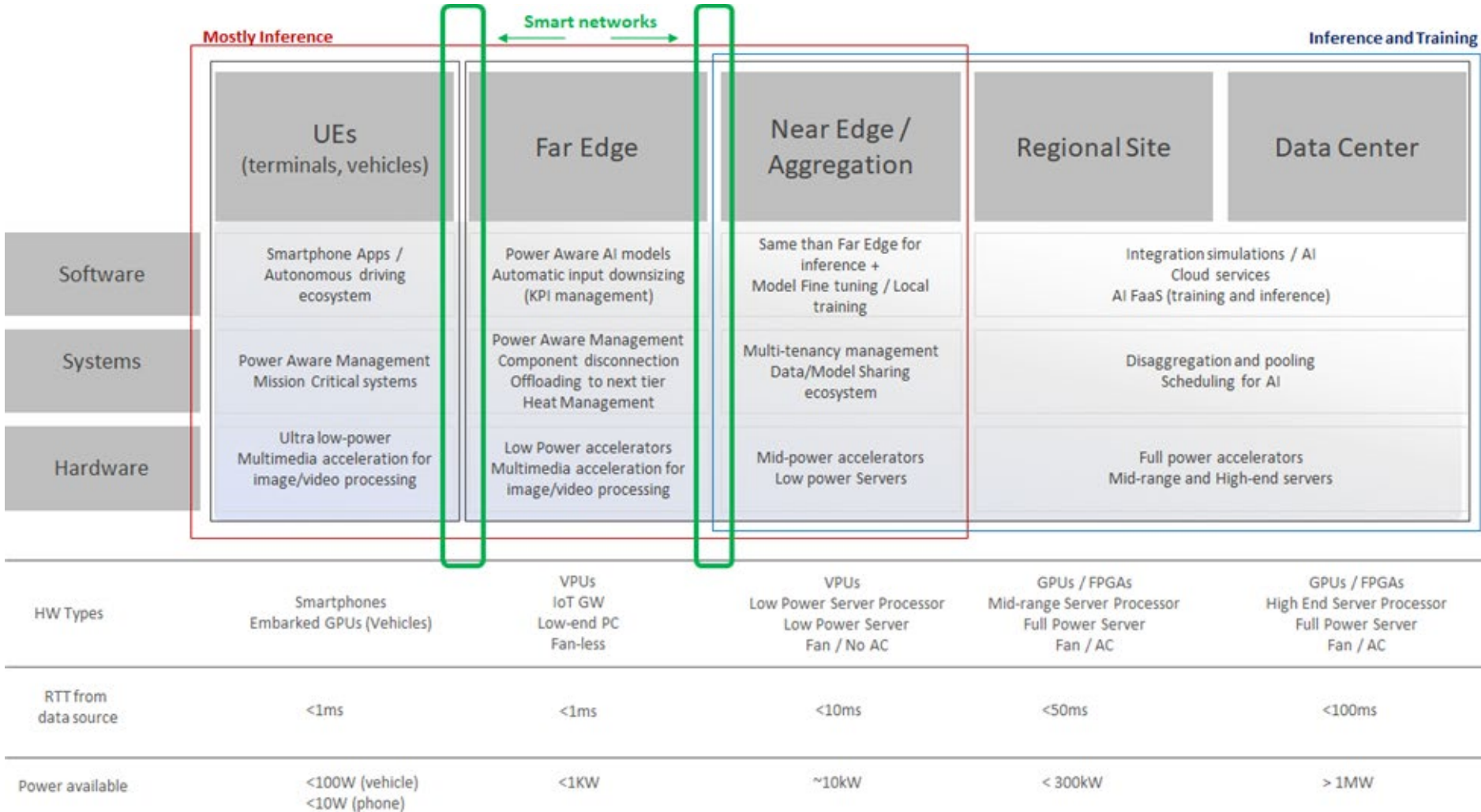
AI/ML for Comms vs Comms for AI/ML

Requirements for evolving Computing paradigms: From Cloud Computing to Edge Computing

	UEs (terminals, vehicles)	Far Edge	Near Edge / Aggregation	Regional Site	Data Center
Software	Smartphone Apps / Autonomous driving ecosystem	Power Aware AI models Automatic input downsizing (KPI management)	Same than Far Edge for inference + Model Fine tuning / Local training	Integration simulations / AI Cloud services AI FaaS (training and inference)	
Systems	Power Aware Management Mission Critical systems	Power Aware Management Component disconnection Offloading to next tier Heat Management	Multi-tenancy management Data/Model Sharing ecosystem	Disaggregation and pooling Scheduling for AI	
Hardware	Ultra low-power Multimedia acceleration for image/video processing	Low Power accelerators Multimedia acceleration for image/video processing	Mid-power accelerators Low power Servers	Full power accelerators Mid-range and High-end servers	
HW Types	Smartphones Embarked GPUs (Vehicles)	VPUs IoT GW Low-end PC Fan-less	VPUs Low Power Server Processor Low Power Server Fan / No AC	GPUs / FPGAs Mid-range Server Processor Full Power Server Fan / AC	GPUs / FPGAs High End Server Processor Full Power Server Fan / AC
RTT from data source	<1ms	<1ms	<10ms	<50ms	<100ms
Power available	<100W (vehicle) <10W (phone)	<1KW	~10kW	< 300kW	> 1MW

Mostly Inference — Smart networks — Inference and Training

Towards AI at the Edge

- All the applications/use-cases of 5G networks, can be enhanced with the capability of cognition offered by computing systems that use the vast amounts of available data and Machine Learning algorithms

- The anticipated future requirements for introducing intelligence at the edge devices will lead to the evolution towards an AI-enabled platform that will be capable to offer intelligent services delivered over the access network to the edge devices.

- Such devices can include in the future computationally efficient dedicated hardware capable of running locally Machine Learning/AI Algorithms at the edge devices, deviating from the classical concept of ML/AI which focuses mainly on offline and centralized AI/ML, implemented by a cloud computing model.

- AI at the Edge is a nascent research field whose system design is entangled with communication and on-device resource constraints (i.e., energy, memory and computing power). Enabling ML at the network edge introduces novel fundamental research problems in terms of jointly optimizing training, communication, and control under end-to-end (E2E) latency, reliability, privacy, as well as devices' hardware requirements.

Smart Networks for AI at the edge

- Much of today's computation remains tied to hardware built for conventional centralized cloud-computing applications.
  - However, AI workloads are different from the calculations most of our current computers are built to perform.
  - Therefore, when used for AI targeting fog and edge computing applications, current platforms are power-hungry and inefficient.

- We need to pushing the physics and computer science of AI platforms in order to deliver radical improvements over the next decade, with innovations ranging from algorithms to systems to devices:
  a) Algorithms and software platforms for AI-based edge resource allocation and services orchestration,
  b) Dedicated computing hardware implementations for AI-based edge computing,
  c) Programmable edge hardware and IoT infrastructures for AI-based edge networking

Federated learning (FL) approach for practical AI at the Edge implementation

- Enabling AI/ML at the network edge introduces novel fundamental research problems in terms of jointly optimizing training, communication, and control under end-to-end (E2E) latency, reliability, privacy, as well as devices' hardware requirements.

- For practical Edge AI/ML, online decentralized training can simplify the implementation and preserve privacy by exchanging not the entire dataset but key model parameters. Such an interesting ML architecture implemented with a decentralized approach and enabled by the edge-computing architecture is the *Federated learning (FL) approach*, in which mobile devices periodically exchange their weights and gradients during local training.

$$\sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

Δw^1 $\qquad$ Δw^2 $\qquad$ Δw^3 $\qquad$ Δw^K
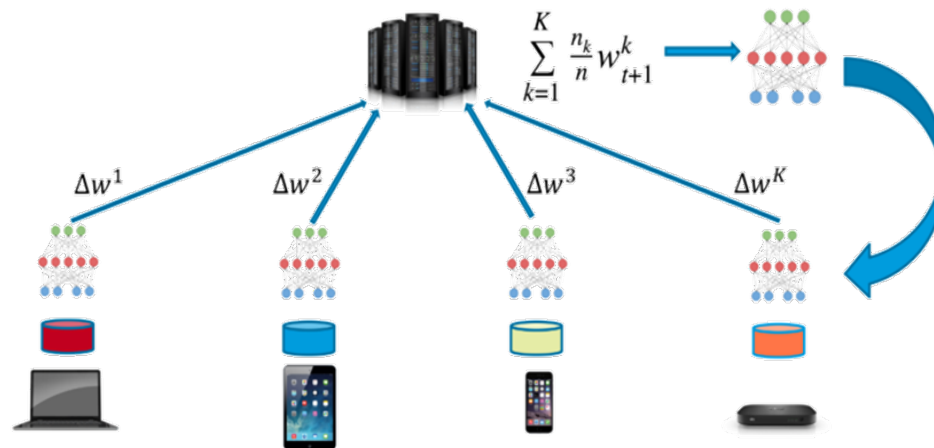
Towards Distributed Edge Computing
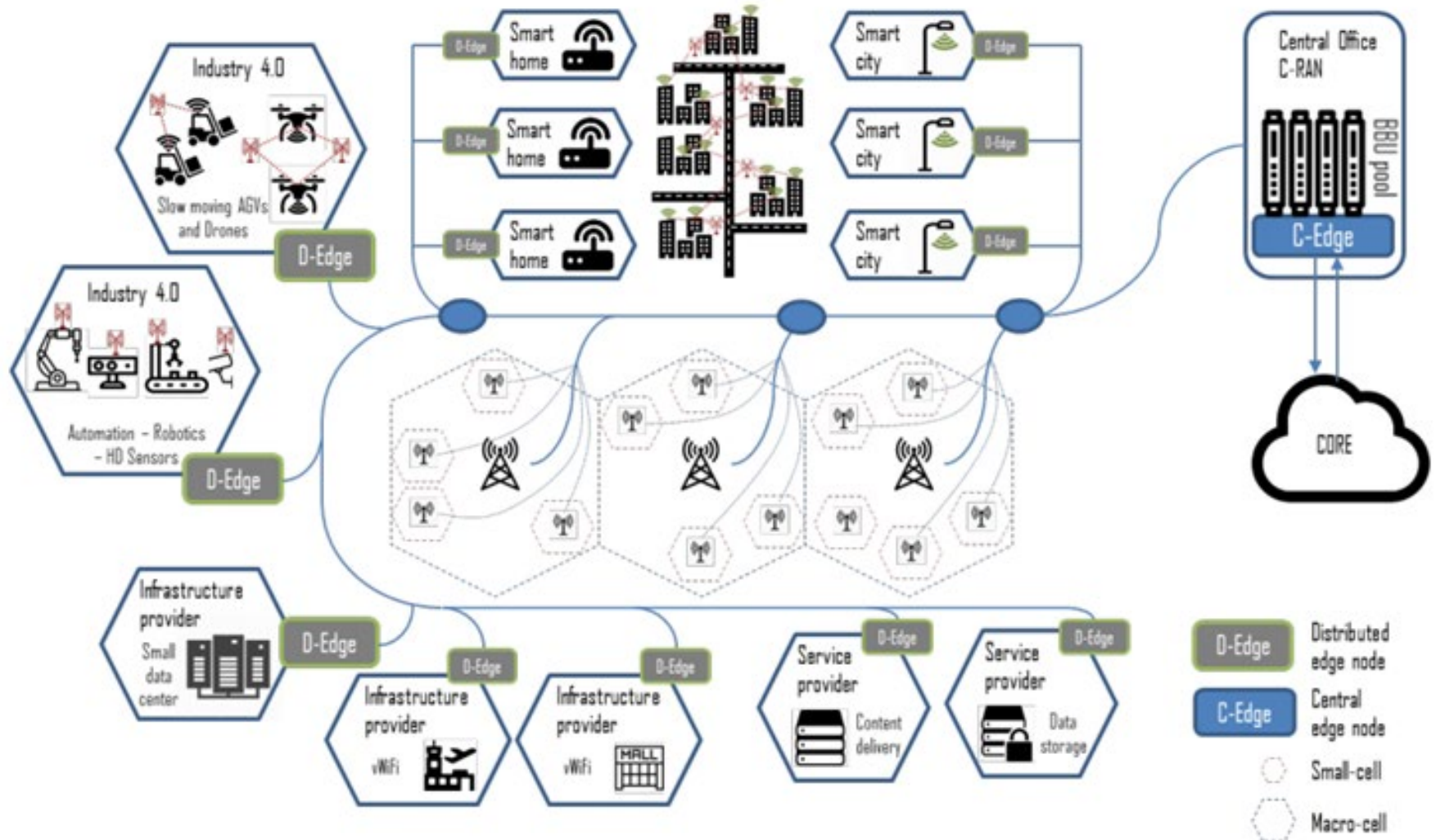
Conclusions

- The smart 6G networks of the future will integrate intelligent 5G and NG-IoT connectivity infrastructures, with advanced edge-computing hardware that will support the computational-resources heavy execution of the AI algorithms to bring new capabilities at the edge networks and services that will be offered over them to humans and machines that will be connected to this converged infrastructure.

- In such smart 6G network, the communications infrastructure connecting the foreseen billions of smart edge devices (collecting and distributing data) should be seen as its "nervous system", while the edge computing hardware executing AI, BC and other algorithms over an SDN/NFV software control and management platform, should be seen as its "brain".

- Such evolution, combined with the proliferation of smart (and even wearable) wireless devices at the edge will make the Internet of Everything (IoE) a reality, where billions of smart devices will be connected and managed by the smart converged 6G infrastructure that is expected to emerge after 2025 and will support the emerging 4th industrial revolution (Industry 4.0) in human history.

Thank you!

Dr. Ioannis Tomkos
itom@ait.gr
+30 6973 328 329