

Evolution and Trends in Edge AI Systems and Architectures for the Internet of Things Era



Prof. David Atienza Alonso
Embedded Systems Laboratory (ESL), EPFL

david.atienza@epfl.ch

<https://www.epfl.ch/labs/esl/>

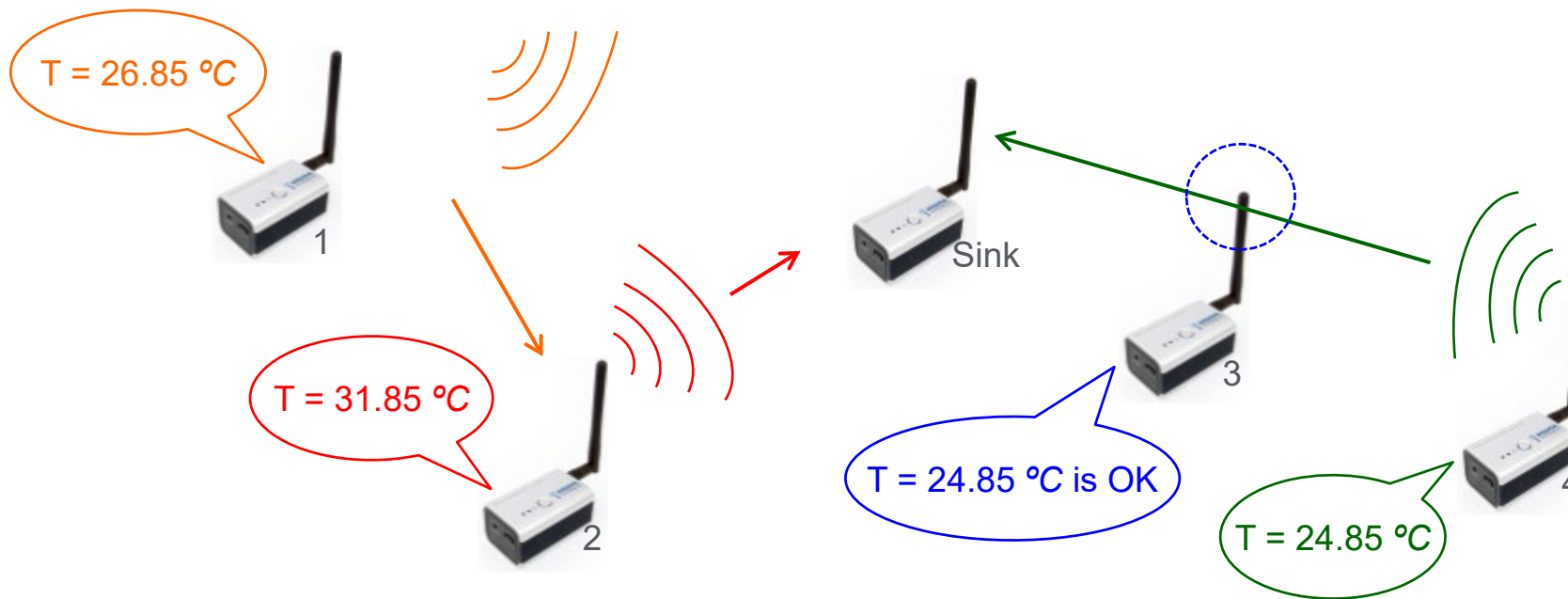
ESL Team and Collaborations

- Team covers IoT and edge AI (both hardware and software co-design)
 - Headed by Prof. David Atienza
- All started in 2008, 45 members today:
 - 8 post-docs
 - 22 PhD students (18 grad)
 - 11 M.Sc. Theses (26 grad)
 - 2 senior engineers
 - 1 administrative assistant, 1 system admin.
- Support from academia and industry for R&D projects
 - 31 companies/R&D centers provide equipment, grants and donations



Origin of IoT: Wireless Sensor Networks

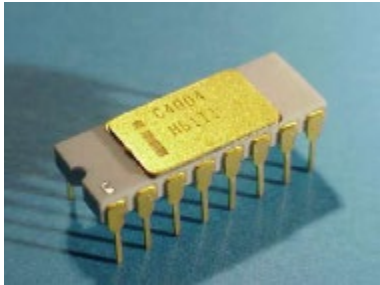
- **Late 80s: Wireless Sensor Networks (WSNs)** are data acquisition and distribution networks consisting of a set of (large) sensor nodes
 - **Simple** architectures (“get” data)
 - Cooperative **applications** and **transmission**



In 2012 - The Internet of Things Concept : “Nano Sensors with Good Connectivity”

- Few more years beyond 1999, but thanks to Moore’s Law, after 45 years...

1971



Intel 4004 [1971]
(2250 p-MOS trans., 108 KHz)

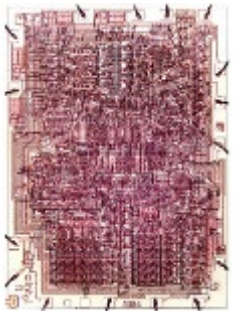
Complexity increase



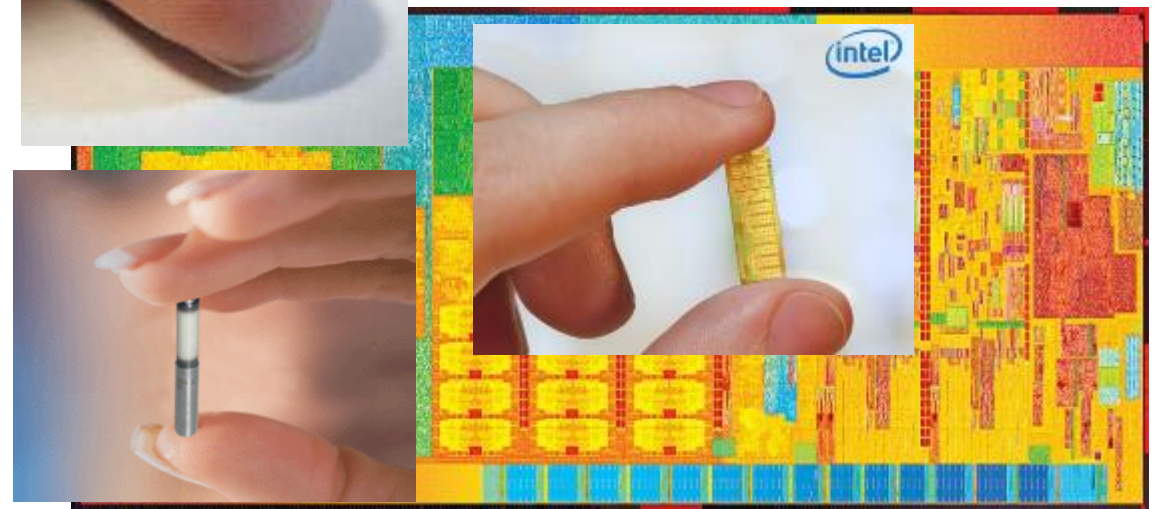
2012



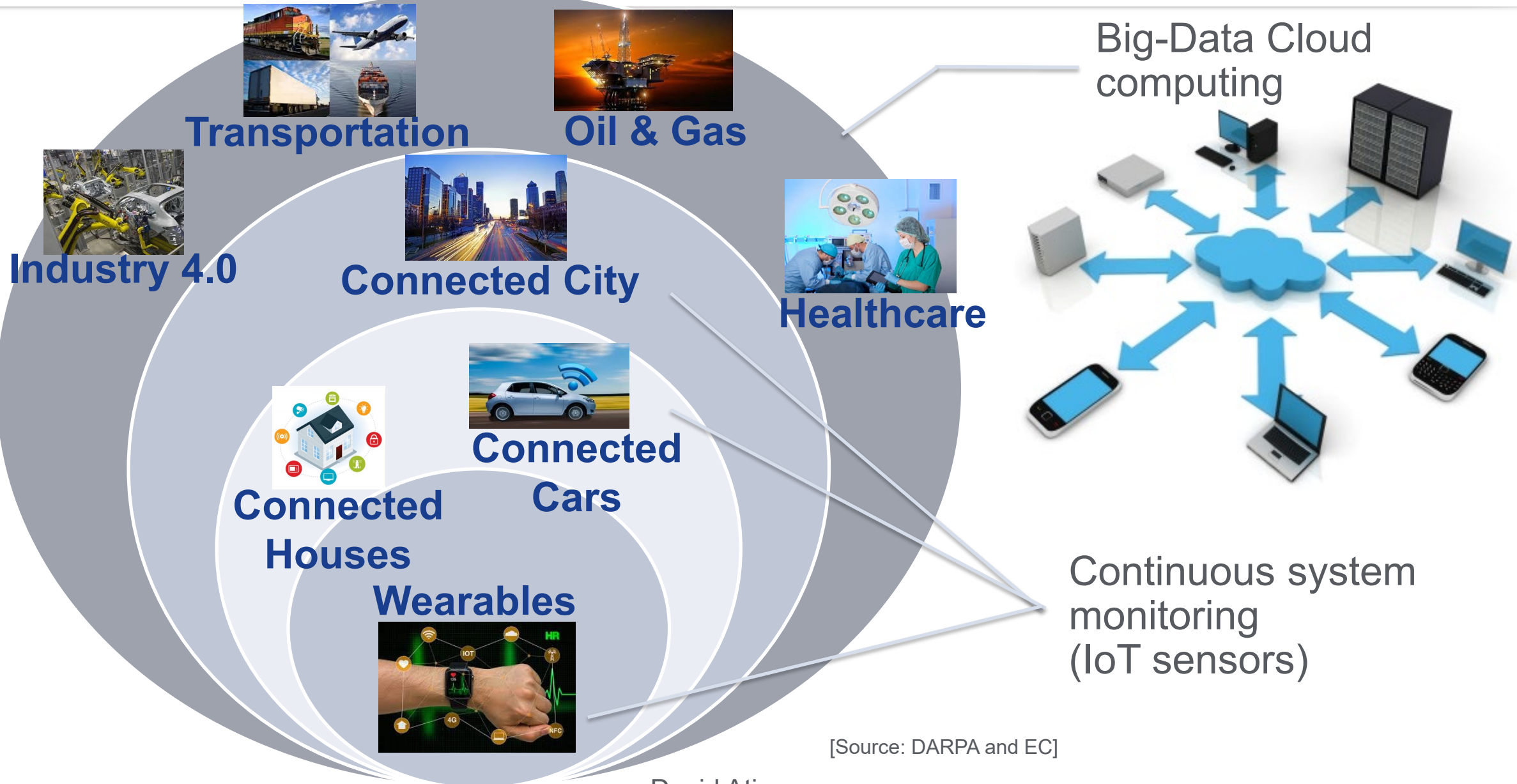
Core i7 vPro [2012]
(**1.3B trans.**, 3.3GHz)



Busicon Calculator 141-PF



IoT System-Level Architecture



IoT Vision: Guardian Angels



See more details at: <https://www.ga-project.eu/>

IoT Enabling Digital Era: Technology Convergence

- IoT concept is the convergence of three technologies within “Digital Era”

1 IoT sensors (Internet of Everything)



+ Digitized Enterprise data
**Able to collect
 unprecedented
 amounts of data**

2 Big data storage and Computing IoT platforms



**Capacity to store
 and handle large
 amounts of data**

3 Analytics and Deep Learning



**Capacity to make
 sense of this data
 in a confident way**

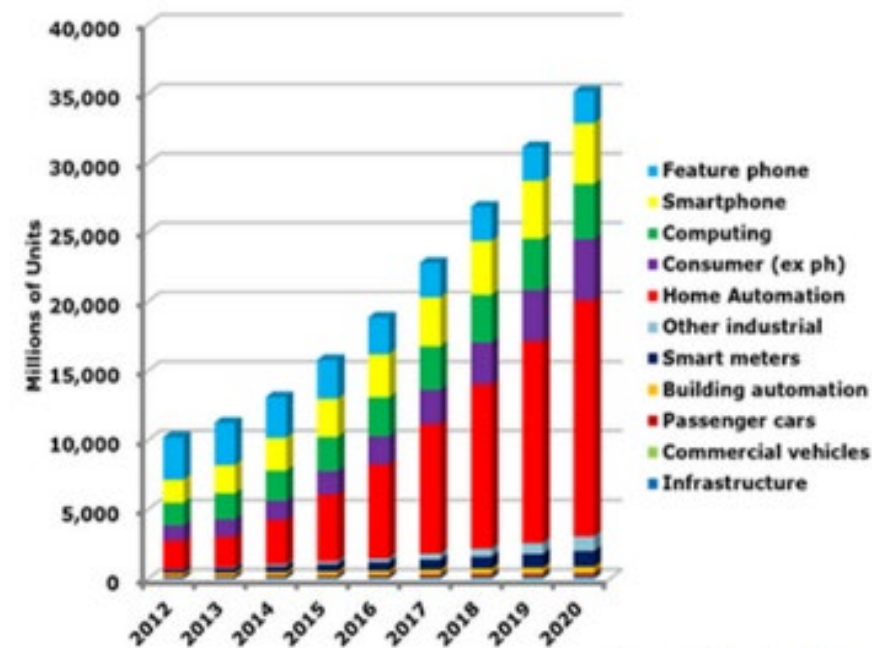
IoT (Potential) Benefits

- IoT long-term economic benefits [McKinsey]
 - Remote healthcare: **\$11.1Trill/year saved (1B people)**
 - Efficient energy: **45TWh/year saved in EU (4M homes)**
 - Automotive industry supply chain: **50% costs reduction**
 - ...



Sources: Merit Metrics, IDC

Total Internet of Things (or IoT) Connected Devices

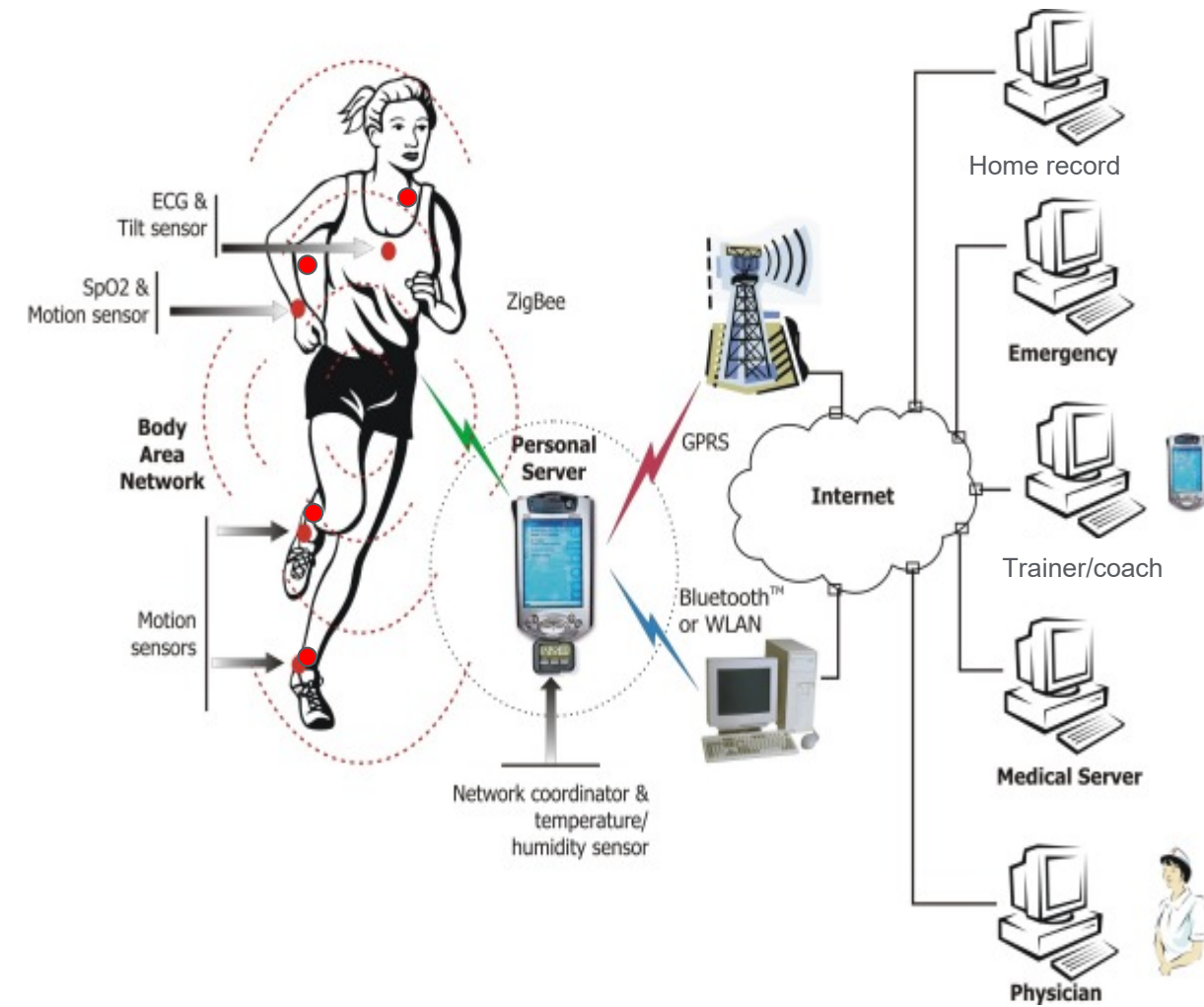


Market Realist

Source: Semco Research

Business-to-business services: 70% added value! But...
How can IoT be designed?

IoT to the Rescue of Our Healthcare System



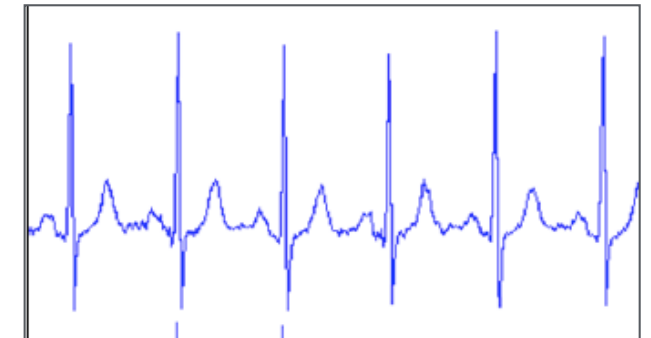
- Two-fold paradigm shift in health delivery

Symptom-based → Preventive healthcare
Hospital-centered → Person-centered

- Cardiovascular monitoring is key today...



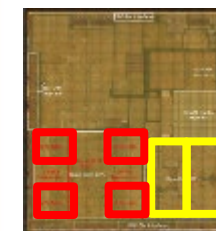
ECG Holter data logger
(clinical practice)



Resting Electrocardiogram
(ECG)

Apple Watch System: Example of (Expensive) IoT Sensor

- April 2015: Combination of Watch + iPhone



iPhone 6: A8 MPCore

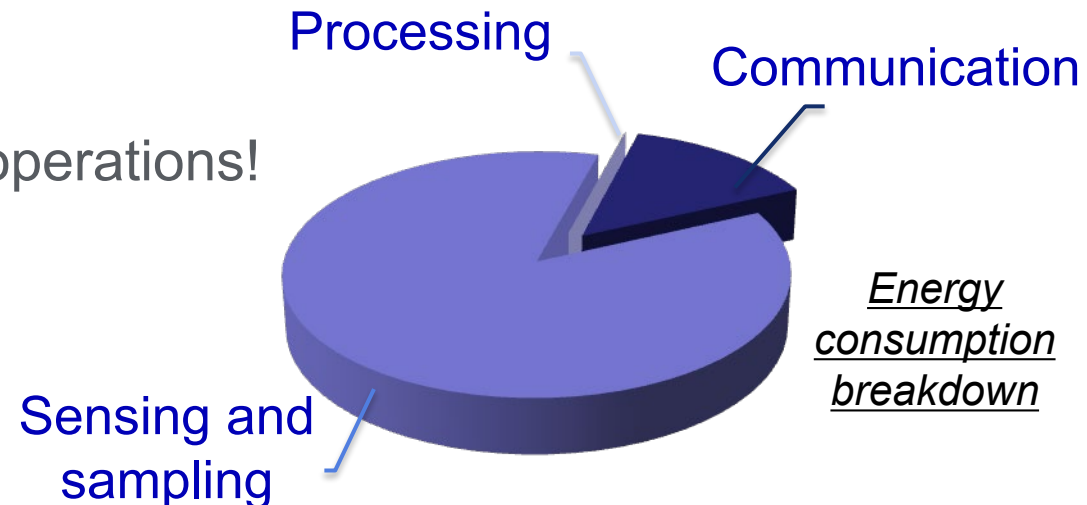
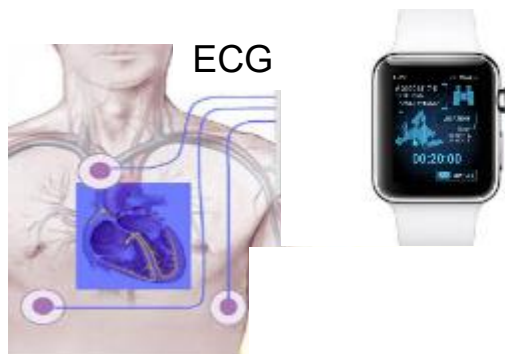
Limited lifetime (< 5 hours for bio-signal analysis) and no real-time...

COMMUNICATION OVERHEAD!

- IoT system design
 - Watch: “Raw” bio-signals acquisition
 - iPhone: User interface and long-range transmission
 - Cloud: Bio-signal (post-)processing

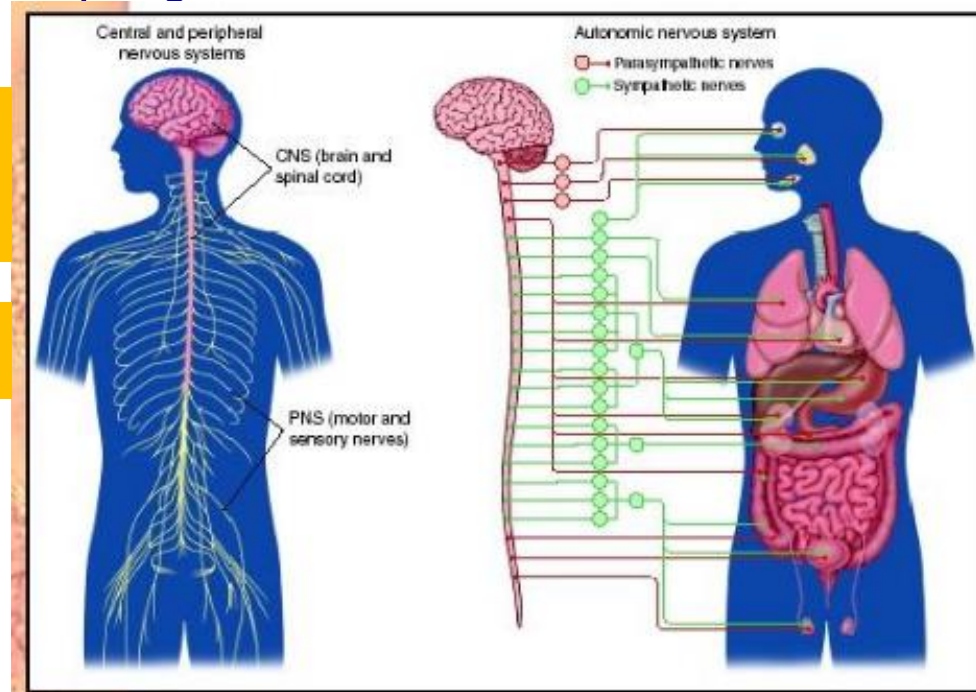
Long-Lived IoT Designs Feasible Today?

- This ECG streaming monitor lasts 5h, why?
 - Send 16 bits: more energy than doing 64'000 operations!



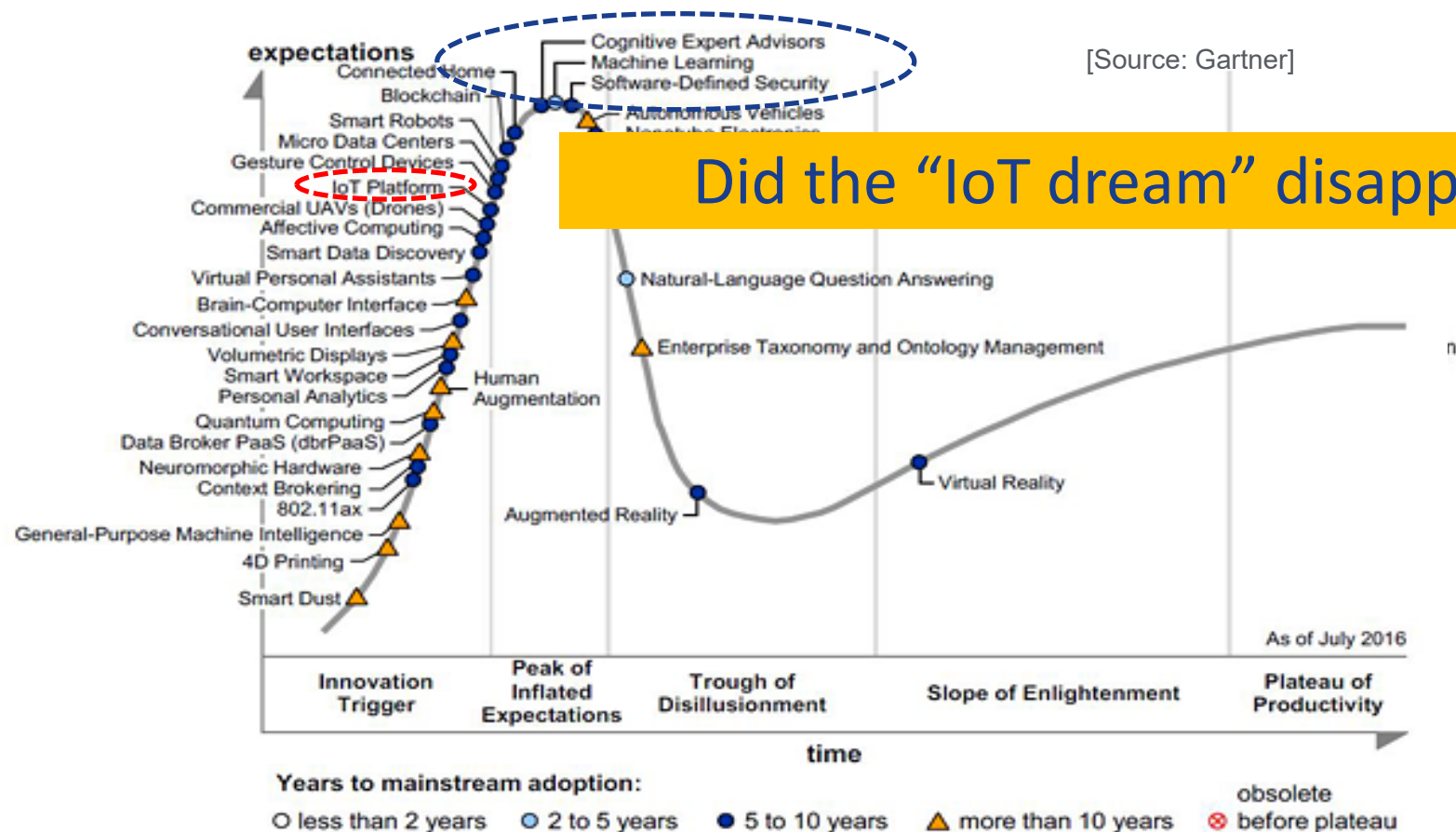
How to reduce IoT data sensed and streamed?
Only get and keep “useful” data from IoT sensors!

IoT age requires complete system revolution!



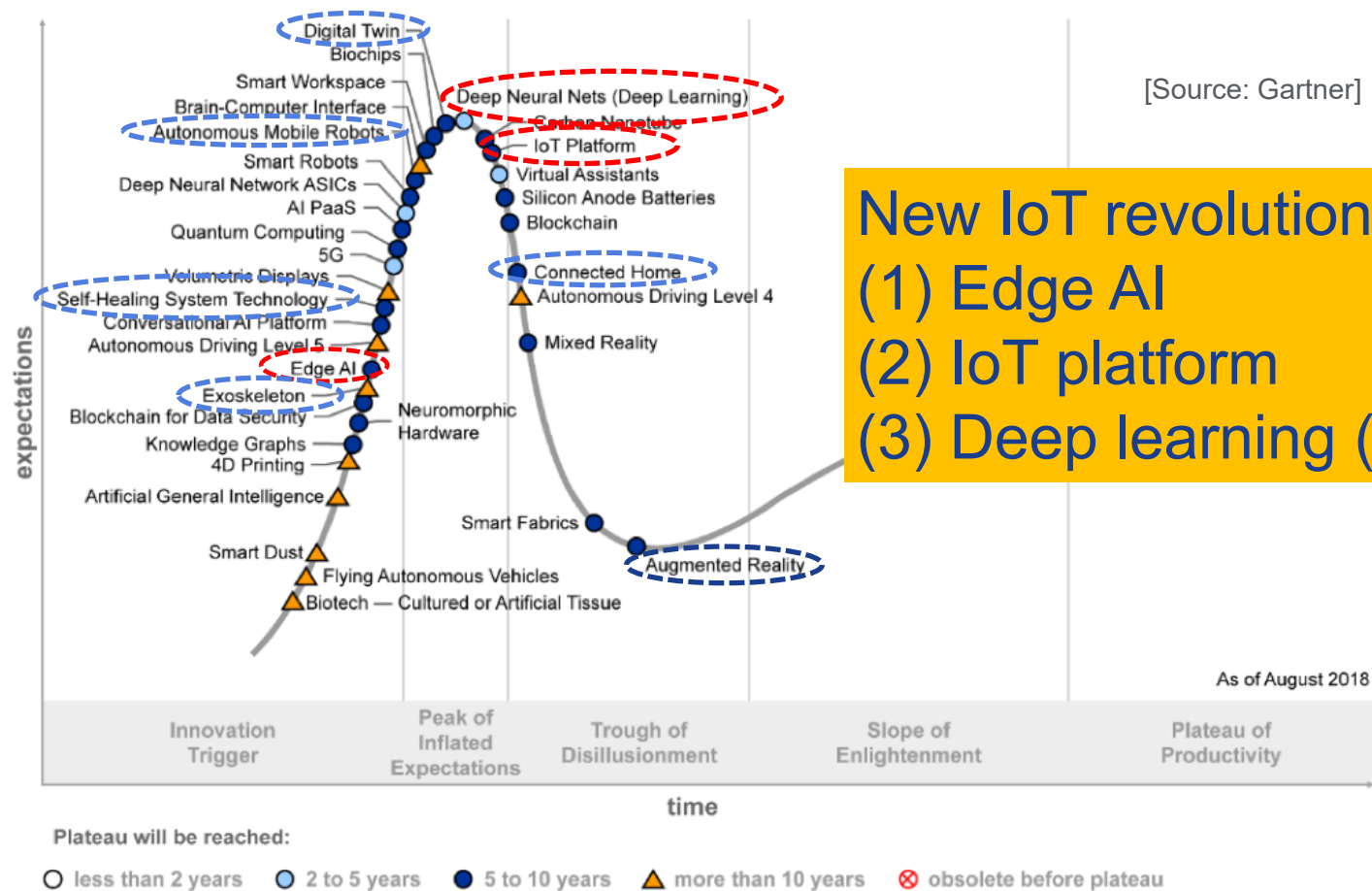
So, Is IoT Still on “Top of the Wave”?

- Very high expectations since 2014...



Source: Gartner (July 2016)

IoT just got “reborn”...

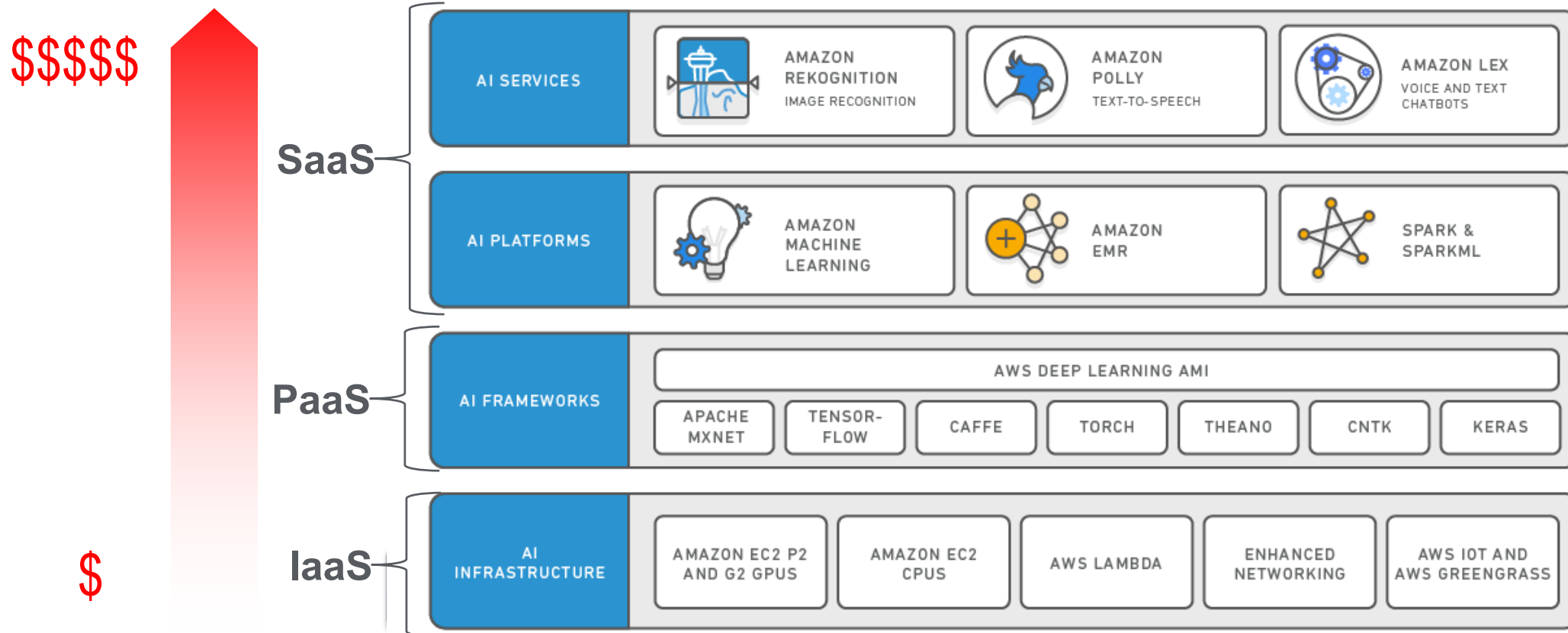


New IoT revolution, 3 technologies:

- (1) Edge AI
- (2) IoT platform
- (3) Deep learning (learn from Big Data!)

New IoT Platform (“Smart Cloud”): 3 Delivery Modes

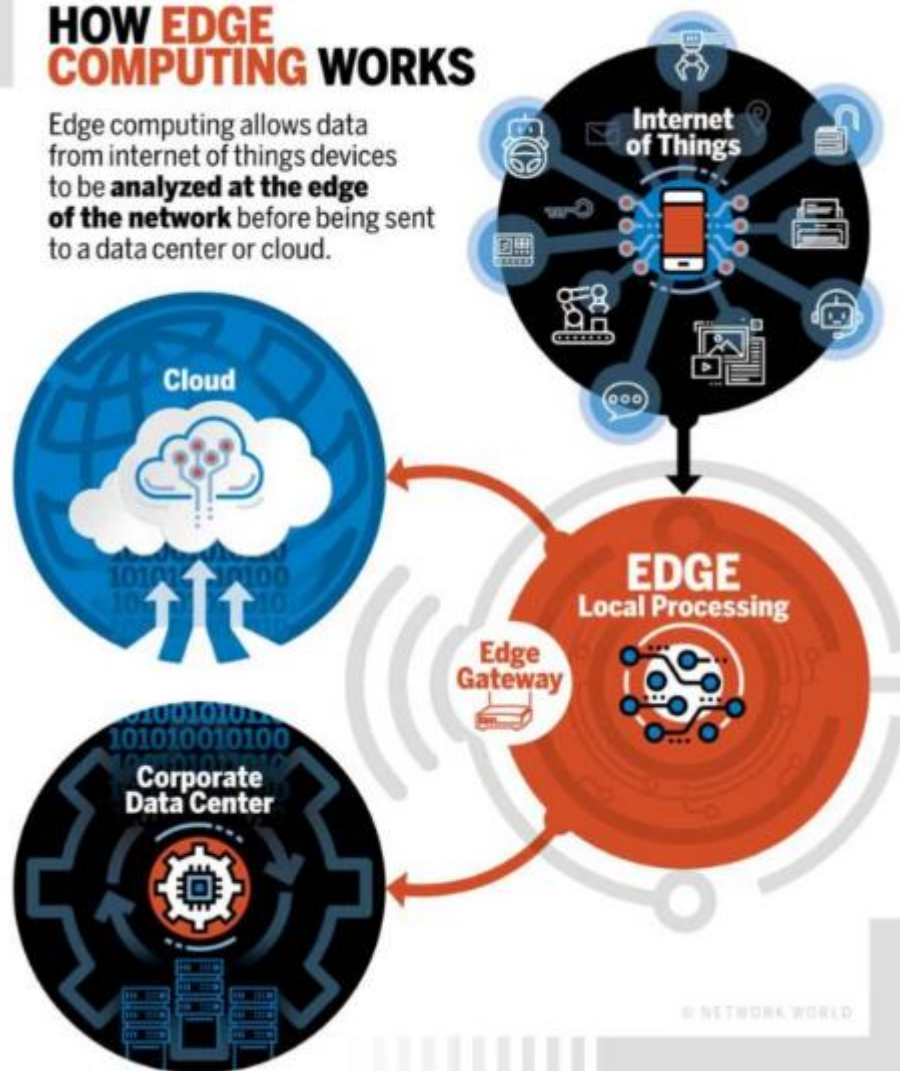
- Infrastructure as a Service (IaaS): IT equipment
- Platform as a Service (PaaS): Big Data access and DL libraries
- **Software as a Service (SaaS): Full AI application services**



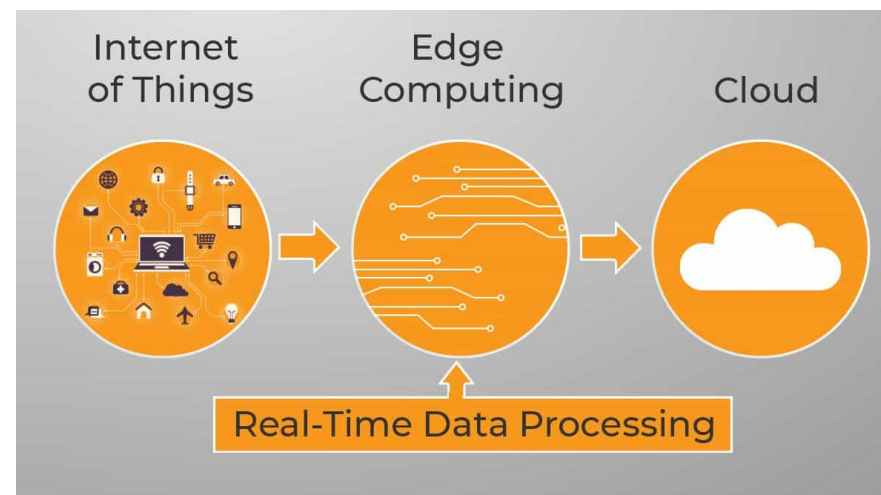
What is “Edge AI / Computing”?

HOW EDGE COMPUTING WORKS

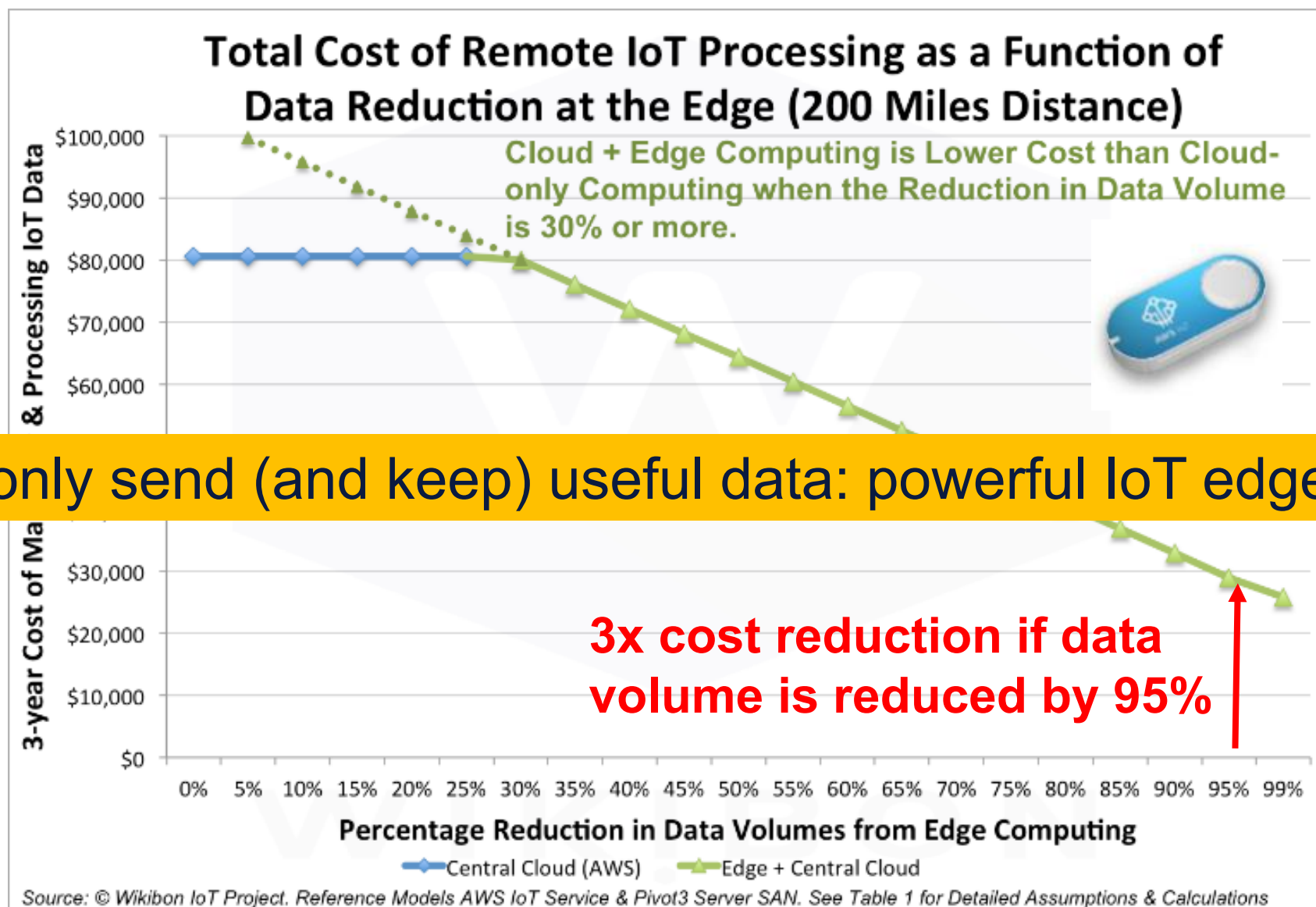
Edge computing allows data from internet of things devices to be **analyzed at the edge of the network** before being sent to a data center or cloud.



- Intelligence is moving towards edge devices
 - Real-time pre-filtering and data processing
- IoT devices send only a portion of the data to the central data repository, or corporate datacenter
 - Reliable results of Big Data AI on IoT Cloud platforms

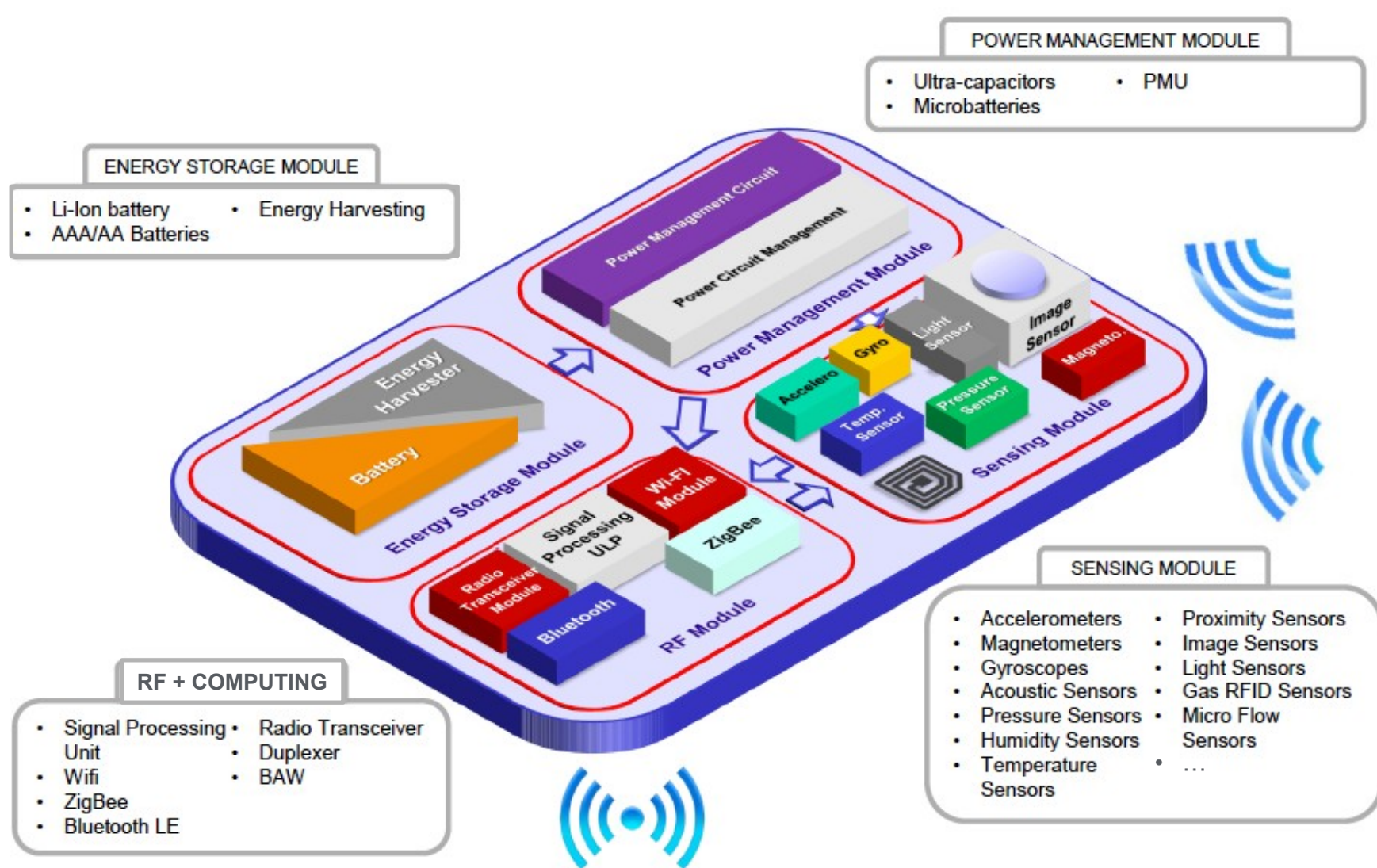


Does Data Size Matters for a Global IoT Business Perspective?



Yes, only send (and keep) useful data: powerful IoT edge AI platforms

IoT Edge AI Concept: Complex Hardware Template



New Edge AI + Patches Can Understand Human Movements

- New collaborative work with multiple patches and one IoT edge AI node
 - Multiple patches to track in real-time: but AI calibration needed



Can you Use the edge AI Concept for new Businesses: Can you be a Ski Coach? Just Add An App and Earbuds!

- Personalized coach with IoT sensors for ski
 - Personalized App
- Create metrics based upon 4 fundamental skiing skills:
 - Edging,
 - Balance,
 - Rotation,
 - Pressure.



Based on few types of sensors and calibrated AI on the App, do you get something you can sell?

Can you Use the edge AI Concept for new Businesses: Can you be a Ski Coach? Just An App and Earbuds!



See more details at: <https://getcarv.com/>

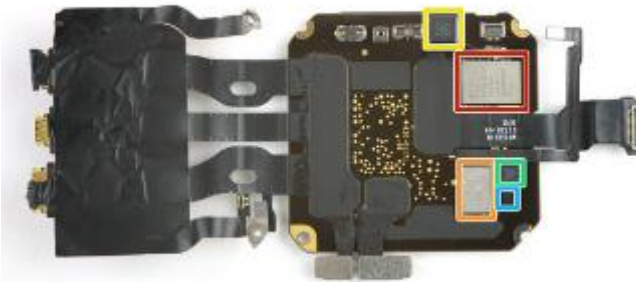
IoT Edge AI Sensing: Few Sensors for Many Application Domains

- Sensors specific per application: but large convergence in similar products



Example (new) IoT Edge AI for Healthcare: Apple Watch Series 6

- Highly-integrated Apple S6 SiP Chip
 - **ARM DualCore SC300 @ 780 MHz**
 - **Integrated graphics and PowerVR (ML accel.)**
 - **32 GB of memory**
- Multiple sensors
 - GPS + GLONASS geolocat.
 - Heart Monitor sensor: PPG-based HR
 - Electrical heart sensor
 - 3-axis accelerometer + gyroscope
 - Ambient light sensor, Built-in compass
- Multiple radios included (+ eSIM)
 - **New IoT Edge AI wearables (with limited set of sensors)**
 - **provide minimal latency and more lifetime**
- Price: \$399

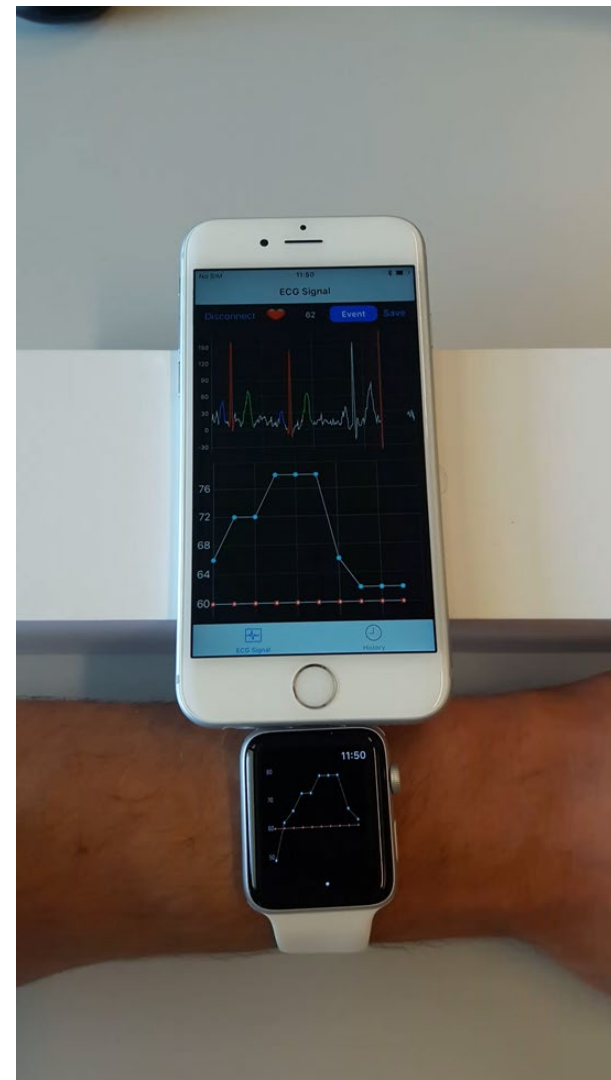


What can new Edge AI Sensors for Healthcare do?

- Same sensors, but algorithms and performance improved with edge AI accelerator chips



**Accurate detection of cardiovascular pathologies
without hospitals and specialized personnel!**

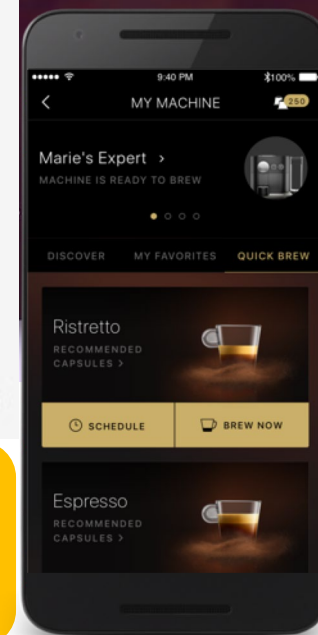
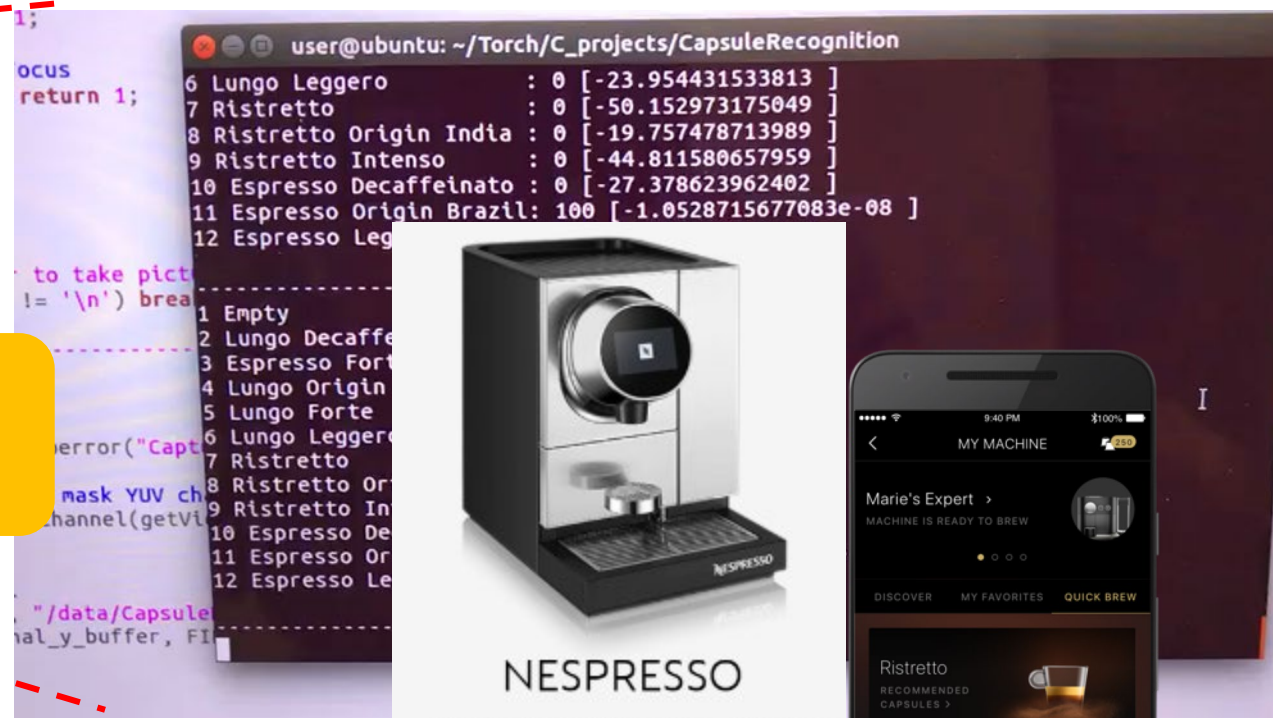


Example IoT Edge AI for Smart Home Devices: “Personalized” Coffee

- New IoT products already on the market... Control Edge AI, Big Data and DL



**Custom IoT Edge
AI design: \$3**



**IoT Edge AI saves energy and gives personalization:
Is then connectivity still relevant?**

New Open-Source Custom Hardware of ULP SoCs for IoT Edge AI Devices

- Case Study: GAP8 Multi-Core SoC Edge Platform
 - RISC-V architecture (Based on ETHZ PULP)

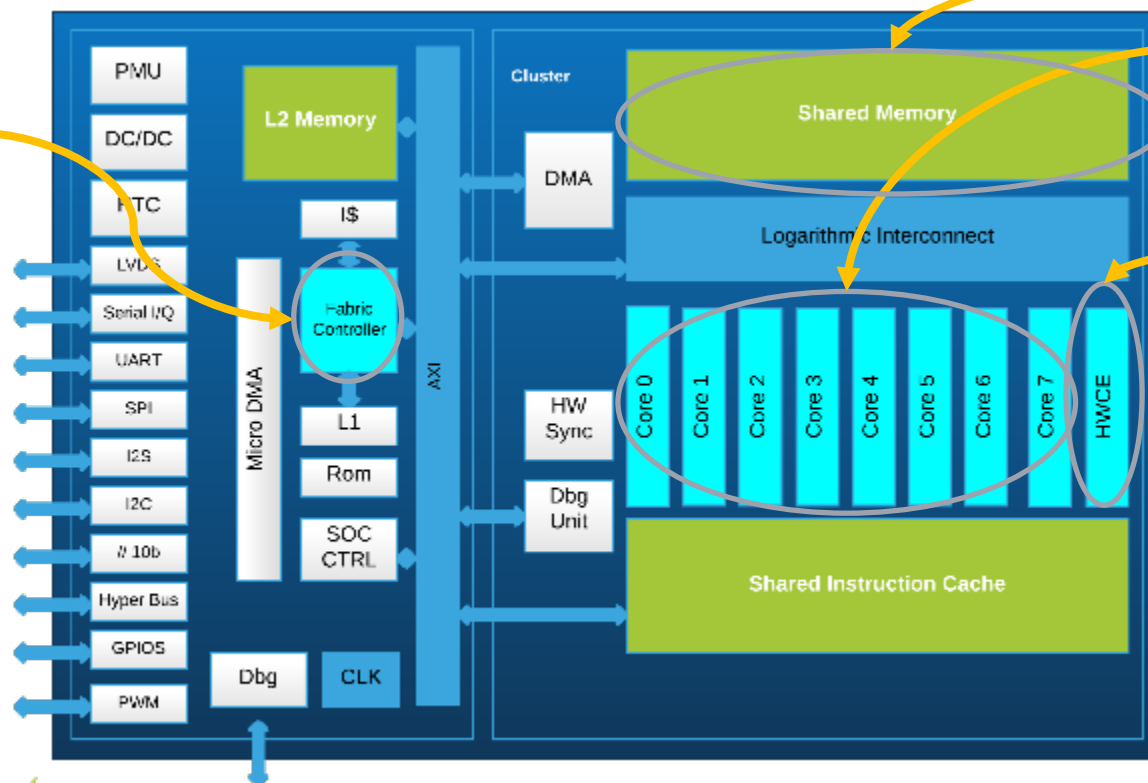
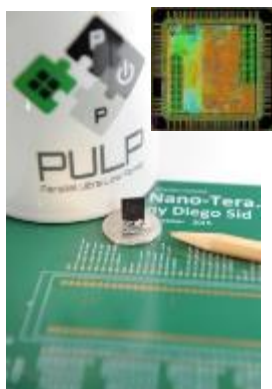
1 general purpose core

A lot of memory

8-core cluster for complex signal processing

New Deep Learning accelerators

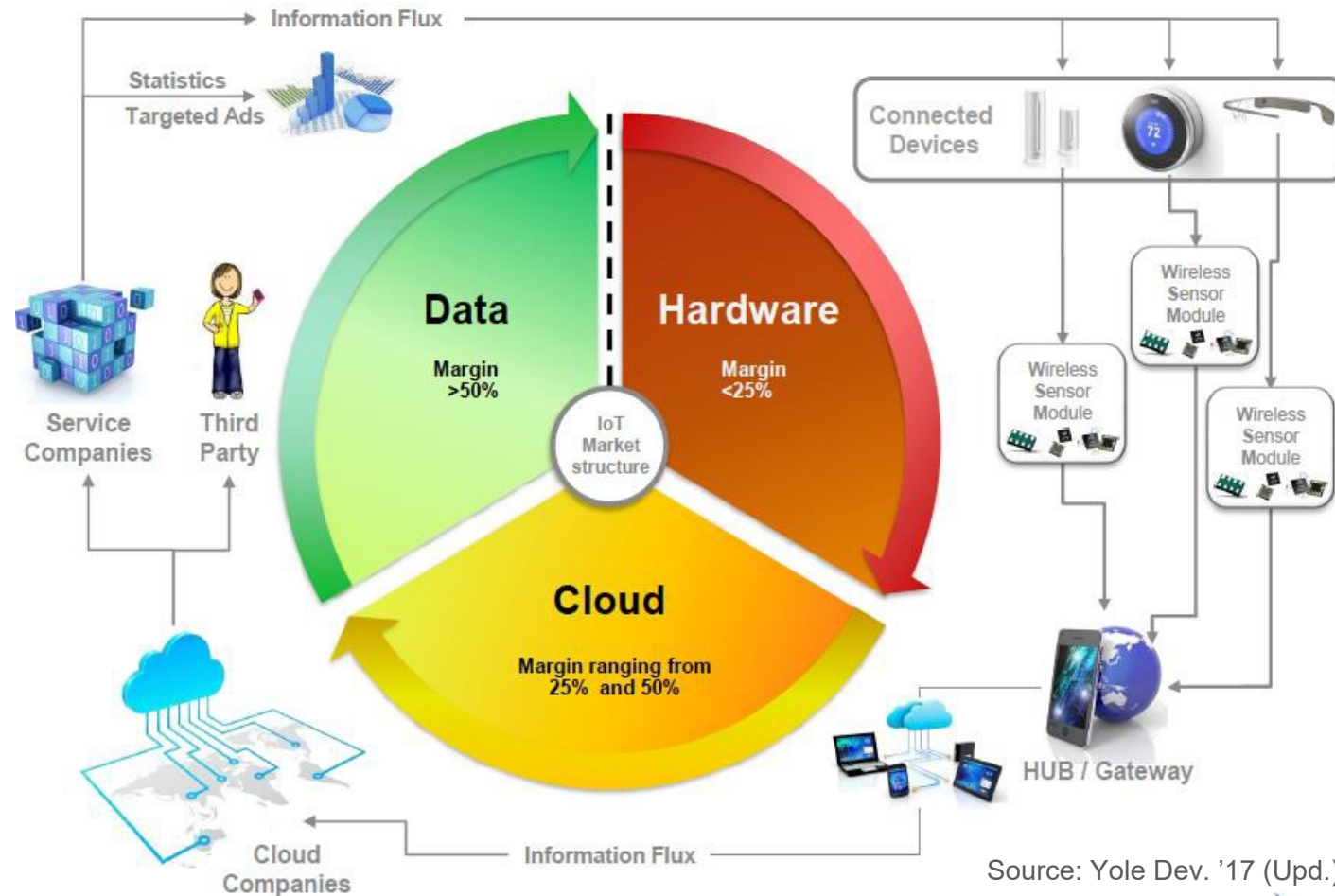
28nm FD-SOI



Custom IoT Edge AI design: \$3

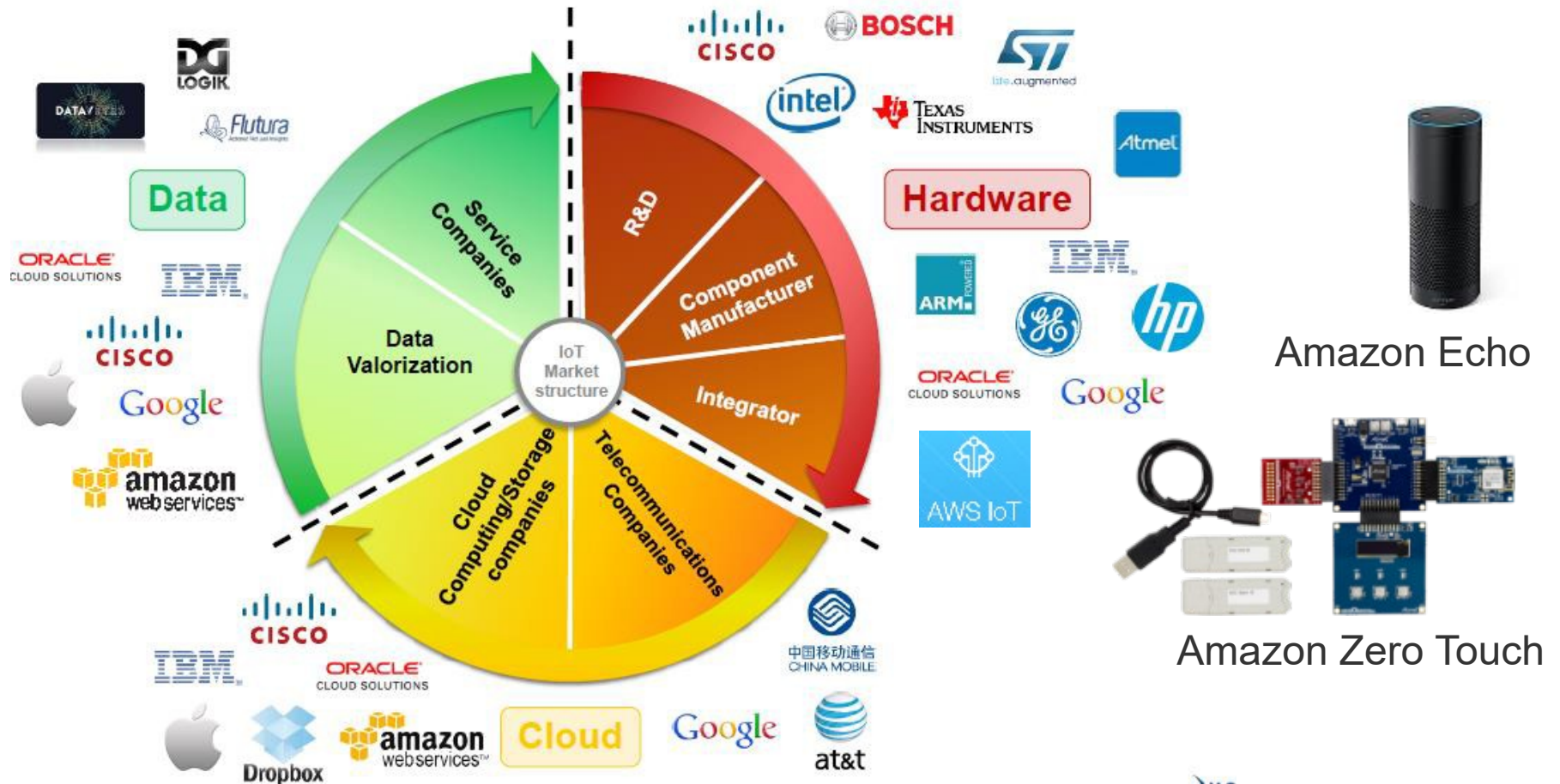
Vertical Companies Created to Control New IoT Space

- IoT business Landscape: Data-related services more profitable



Vertical Companies Created to Control New IoT Space

- IoT business Landscape: Data-related services more profitable
 - But largest companies trying to avoid risks: full chain control



New Extended Support in Cloud for Secure IoT Devices: Example of Amazon Web Services (AWS IoT) and Google Teachable Machines

- Open-source IoT communication standards supported: Message Queuing Telemetry Transport (MQTT)
- Enable new IoT persistence concept: Device Shadow concept (Virtual Twin) when disconnected



New IoT Revolution... Enabling Industry 4.0!

... Collect data from everywhere



Monitoring (all) and maintenance

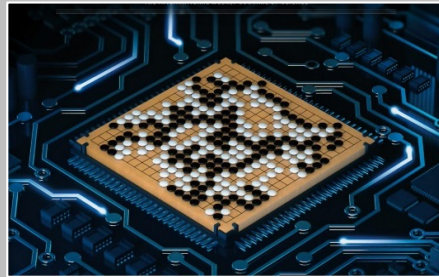


Edge IoT



Deep learning

... Exploit all sources to make decisions



Virtual twins and model predictions



Edge IoT



Big Data analytics



Deep learning

... New robotics help deliver efficiency



Digital autonomy in physical world



Advanced robotics



Edge IoT



Deep learning

... New ways to interact with machines



Human-Machine interfaces



Wearables

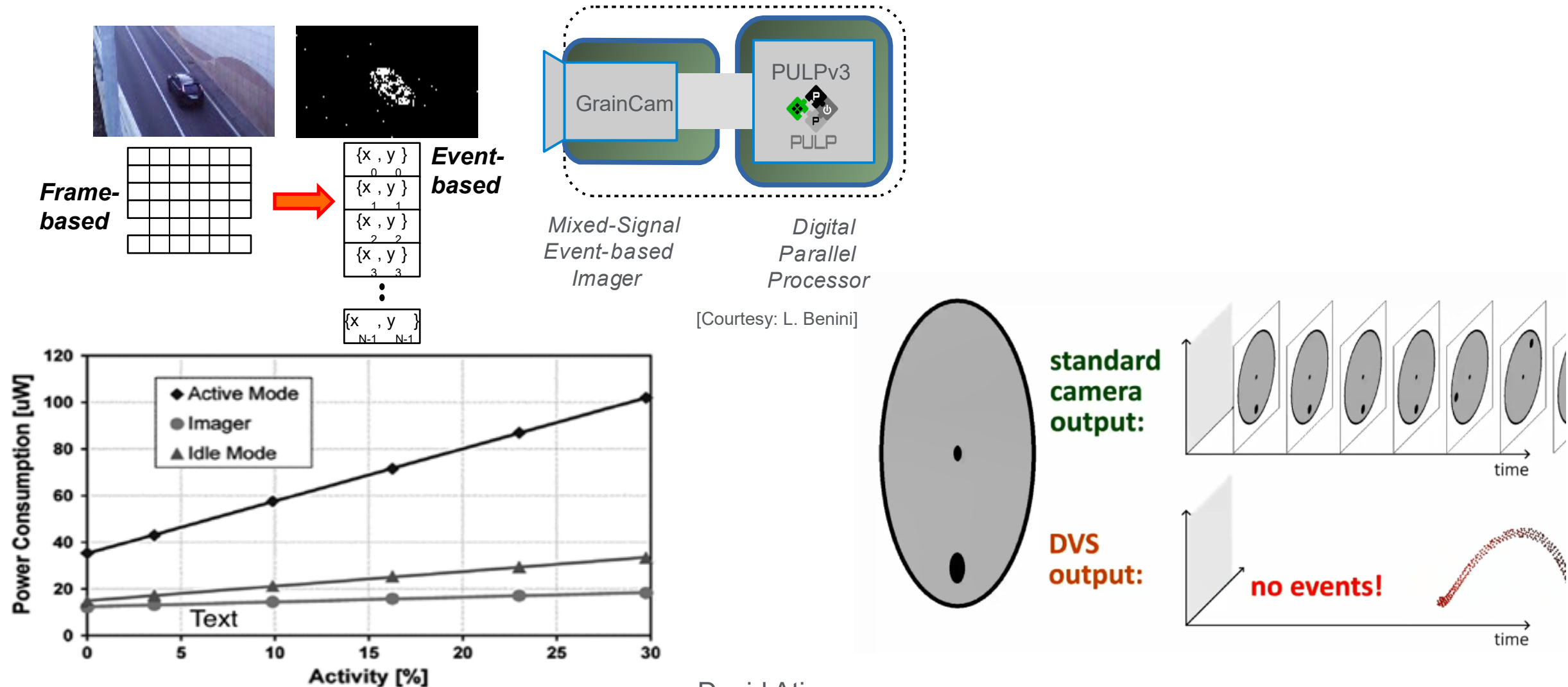


Edge IoT



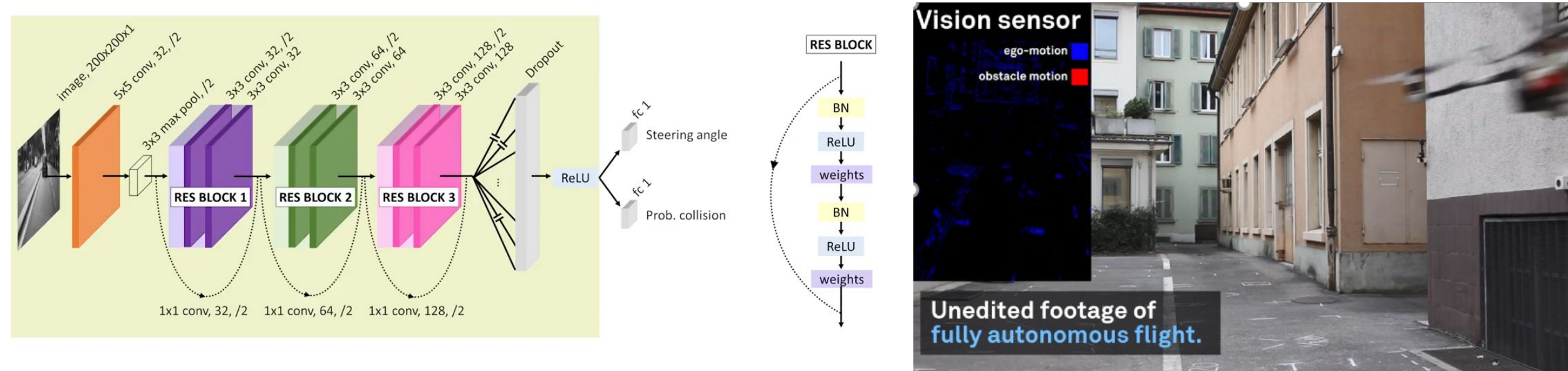
Augmented reality

- From frame- to event-based Dynamic Vision System (DVS): analog frontend off most of the time



Delivery with Drones: React in Real-Time with IoT Edge + Deep Learning

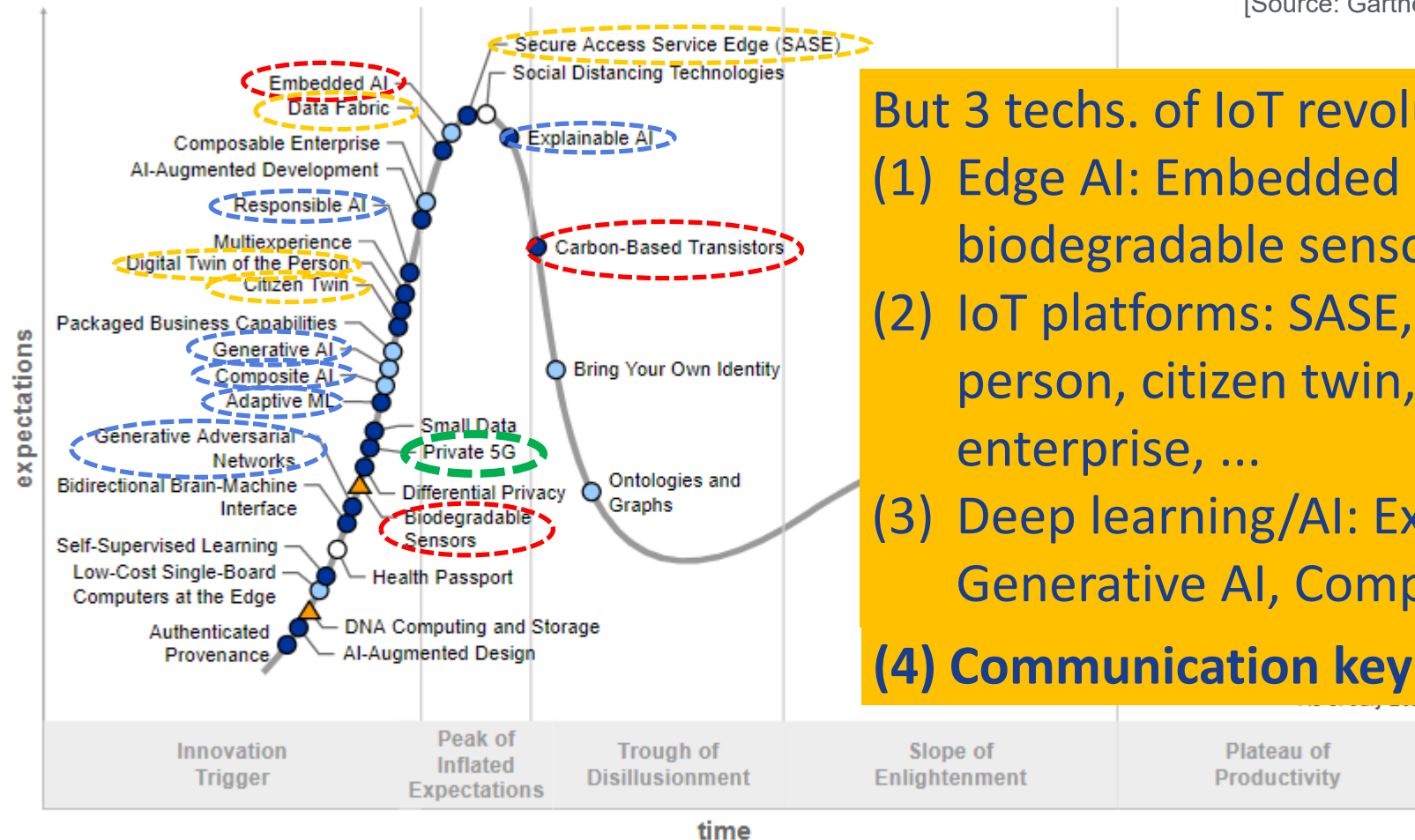
- **Insightness:** Obstacle avoidance with 7-layer Deep neural network
 - Learn with public traffic databases



[Courtesy: Davide Scaramuzza, RPL-Univ. Zurich]

IoT just got “reborn”... But keeps evolving in 2021!

[Source: Gartner]



- But 3 techs. of IoT revolution remain:
- (1) Edge AI: Embedded AI, carbon-based transistors, biodegradable sensors, ...
 - (2) IoT platforms: SASE, data fabric, digital twin of the person, citizen twin, composable infrastructure or enterprise, ...
 - (3) Deep learning/AI: Explainable AI, Responsible AI, Generative AI, Composite AI, Adaptive ML, GANs,...
 - (4) Communication key again: 5G?

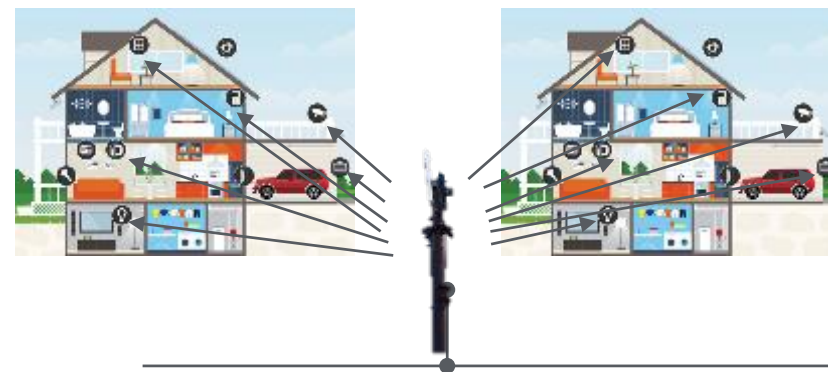
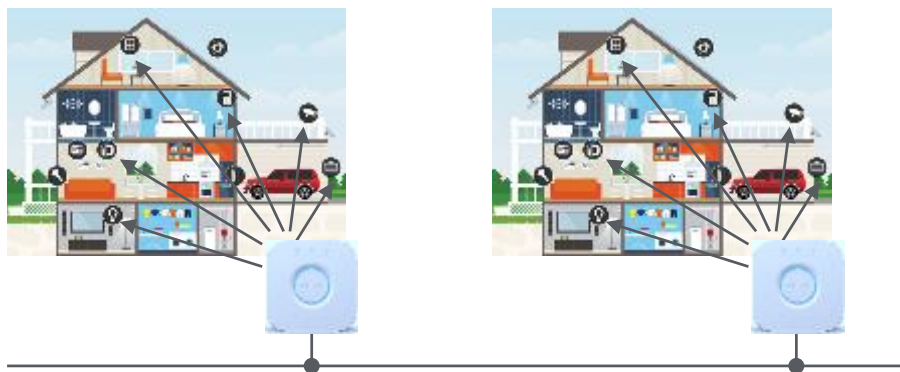
Plateau will be reached:

○ less than 2 years ● 2 to 5 years ● 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau

Today: Local Networks (Legacy) Global Connectivity (5G)

- Simple local networks, **short range links** (scalable so far, but with latency)
- **Requires local infrastructure** devices, setup, and maintenance
- **Transceiver simple/cheap**
- **No** extra service **subscription (low rates)**
- **Reuse of existing** comm. resources

- Complex (cellular) network with **long range links**
- **No local infrastructure** with some QoS by service provider
- **Complex/costly transceiver**
- Requires **service subscription**
- **Limited reuse**, no shared comm. resources



[Courtesy: Andreas Burg (TCL@EPFL)]

Evolution promised/envisioned by 5G

Conclusion

- New IoT concept possible due to techs. convergence: real-time response, safer and better AI

1 Edge AI sensors (and actuators)



**Unprecedented amounts of
(useful) data at the edge/fog**

2 Big data storage and Computing IoT platforms



**Capacity to store and
handle large data volumes**

3 Analytics and Deep Learning



**Capacity to make sense
of it (automatically)**

- Upcoming technology trends:
 - Even more distributed/hierarchical intelligence in future IoT Era
 - “Copy” more from biology (More specialization on processors = energy efficiency, etc.),
 - Finding schemes to secure and activate selectively IoT devices (Self-awareness)

Thank You!



QUESTIONS?

david.atienza@epfl.ch