

The background is a vibrant, abstract composition of overlapping, wavy shapes in various colors including red, orange, yellow, green, blue, and purple. In the center, there are two identical geometric motifs, each consisting of a red outer arc, a green inner arc, and a small green circle below it. The text 'Wireless @ RWTH Aachen' is centered in a white, sans-serif font, with a small green horizontal line underlining the 'e' in 'Wireless' and the 'e' in 'Aachen'.

Wireless @ RWTH Aachen



Next-generation Communications and Computing Systems

AI for PHY: Powerful or insatiable

Haris Gačanin
Chair for Distributed Signal Processing
RWTH Aachen University

About the chair

Computing- and latency-constraint algorithms for wireless communications

Haris Gačanin

Full Professor, IEEE Fellow
RWTH Aachen University

- 14 scientific staff members
- 4 non-scientific staff members
- Algorithms, SW/HW integration, Prototyping



6GEM Leadership in
6G German initiative



DSP@ Scientific events

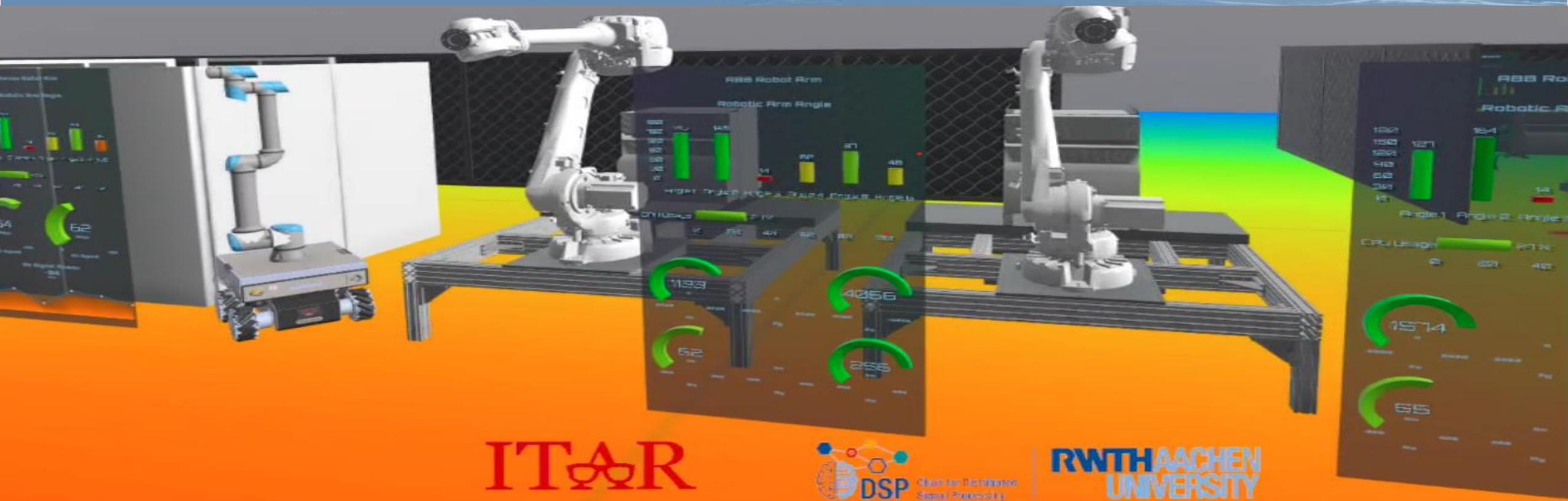


DSP@MWC2023 in Barcelona

Research vision on „Communications and Computing“

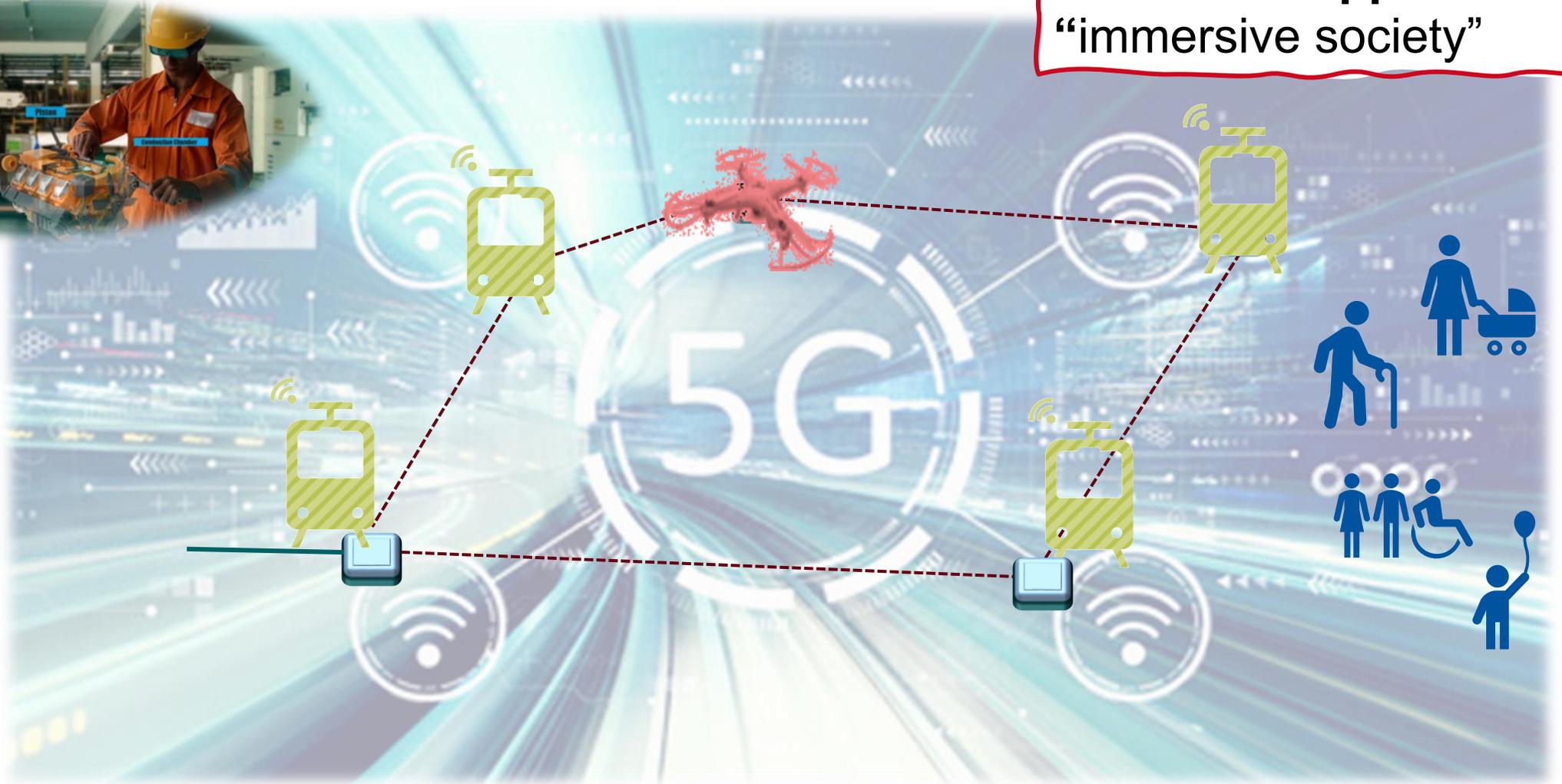
6G and Beyond

Convergence of Communications and Computing:
Wireless for interactive industrial applications



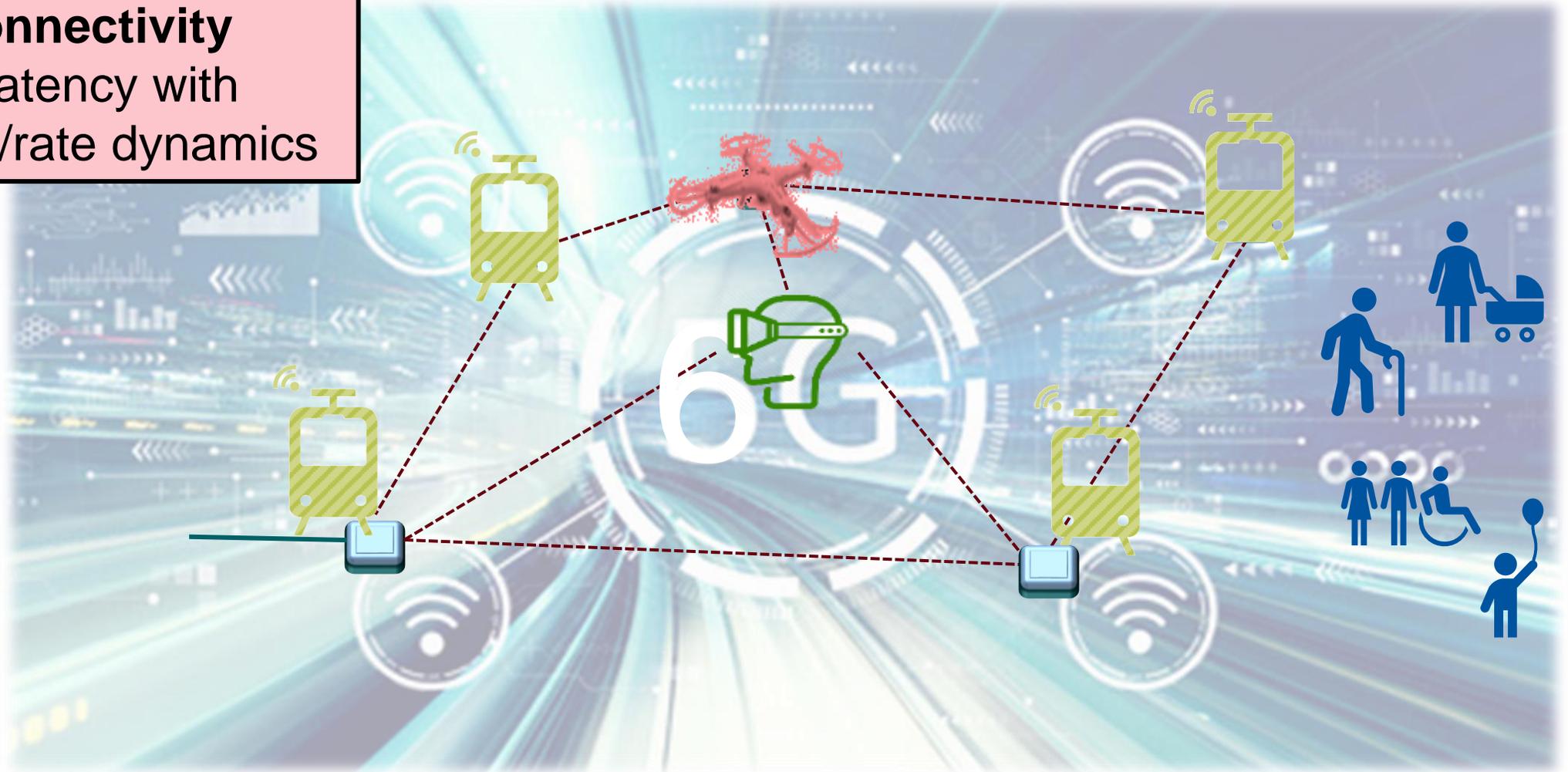
Concept of future interactive industry

**Interactive applications
“immersive society”**



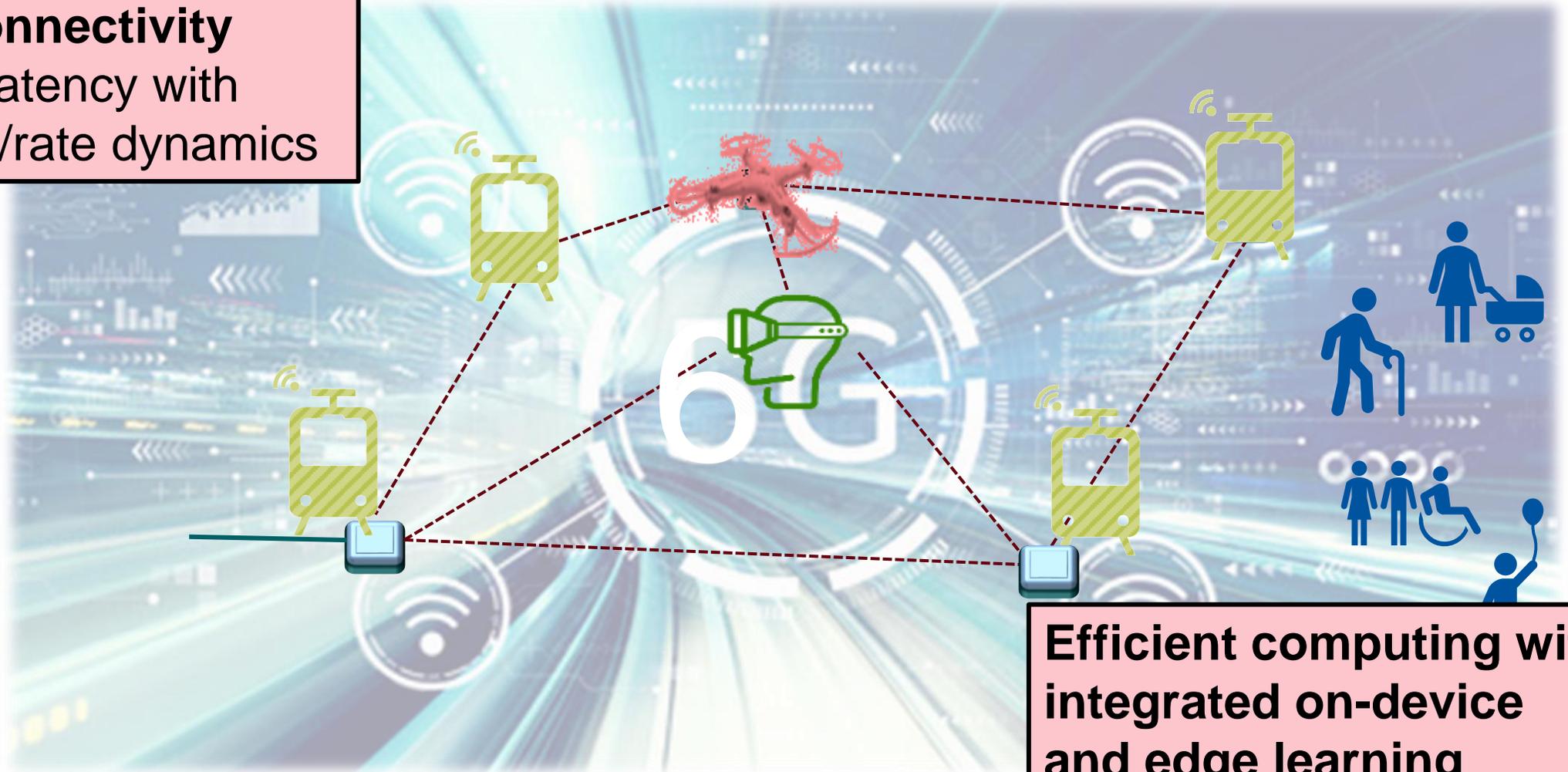
Technology enablers...

Hyperconnectivity
sub-ms latency with
reliability/rate dynamics



Technology enablers...

Hyperconnectivity
sub-ms latency with
reliability/rate dynamics



**Efficient computing with
integrated on-device
and edge learning**

Technology enablers...

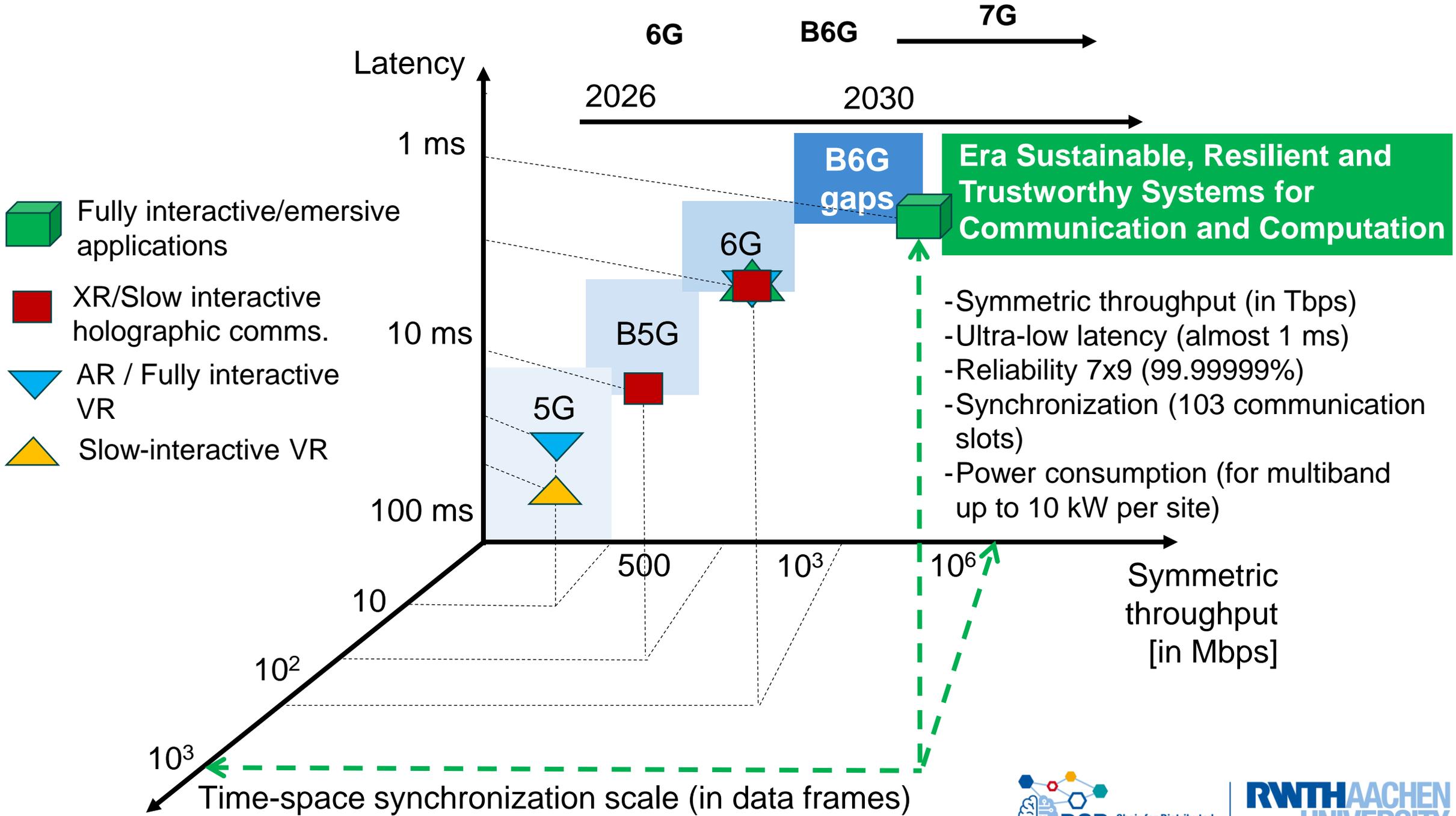
Hyperconnectivity
sub-ms latency with
reliability/rate dynamics



Integrated spatio-temporal localization and ambient intelligence

Efficient computing with integrated on-device and edge learning

Communications and computing challenges



Communications and computing requirements

Online training operation in real-time with ultra-low energy consumption

Real-time data sensing at the speed of the target radio application

Sub-ms decision-making and efficient on-device processing

Use case example: Cooperative carrying with mobile robots

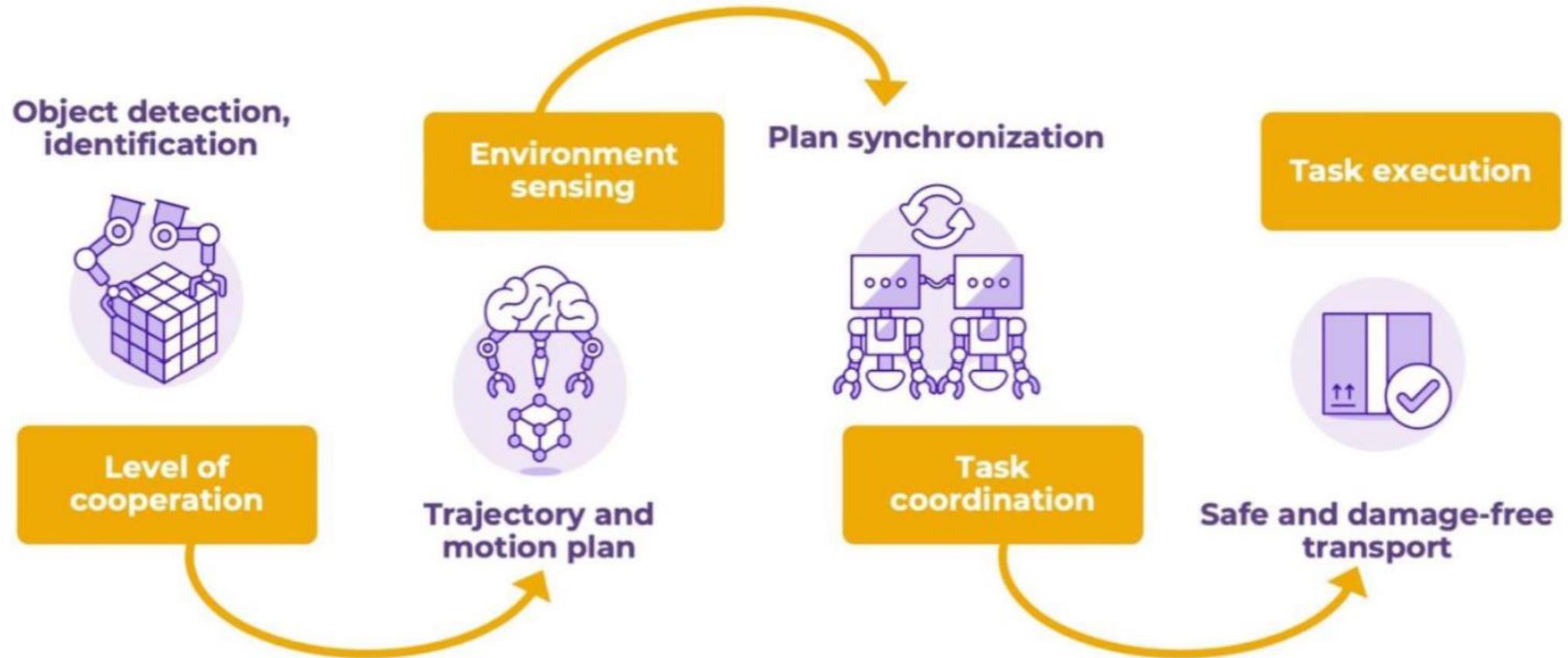
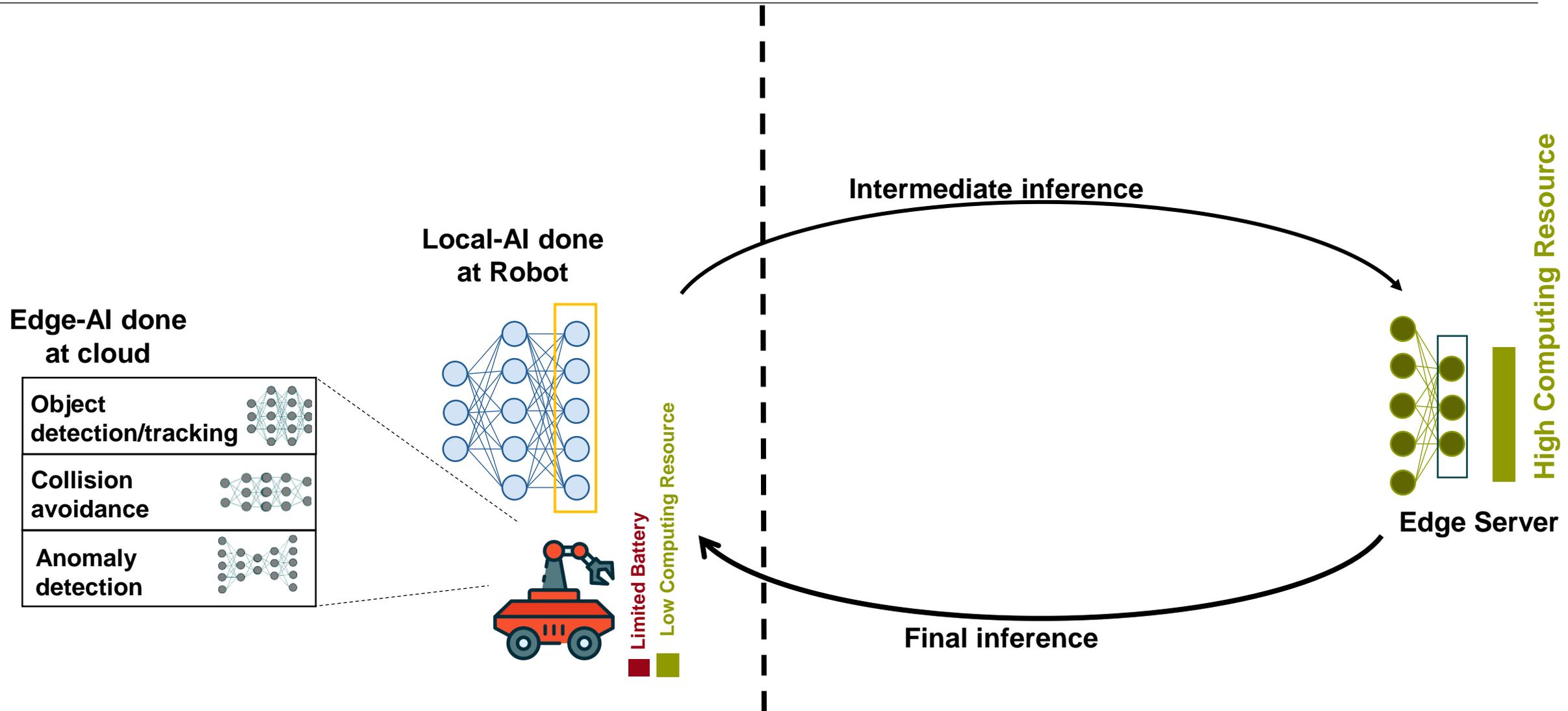
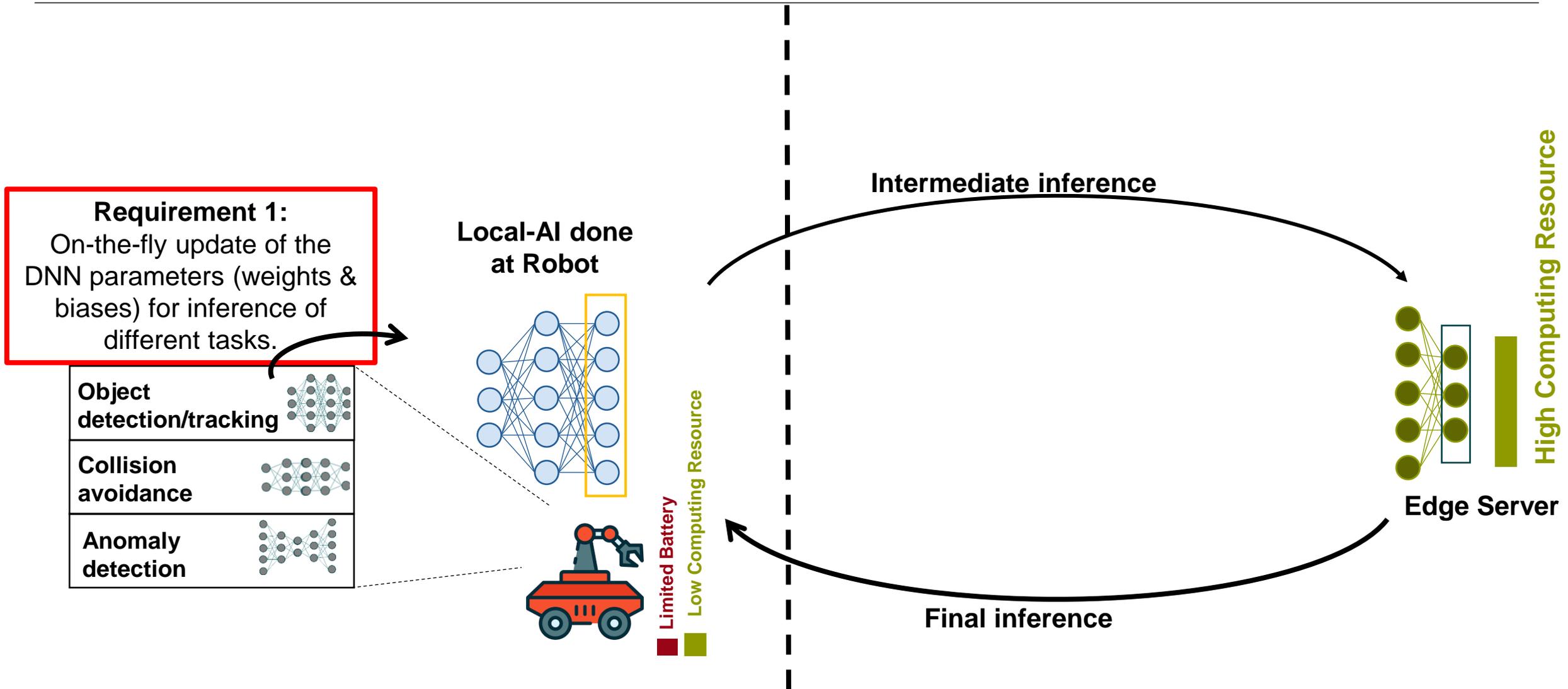


Figure 1: UC1 - Cooperative carrying with robots execution illustration

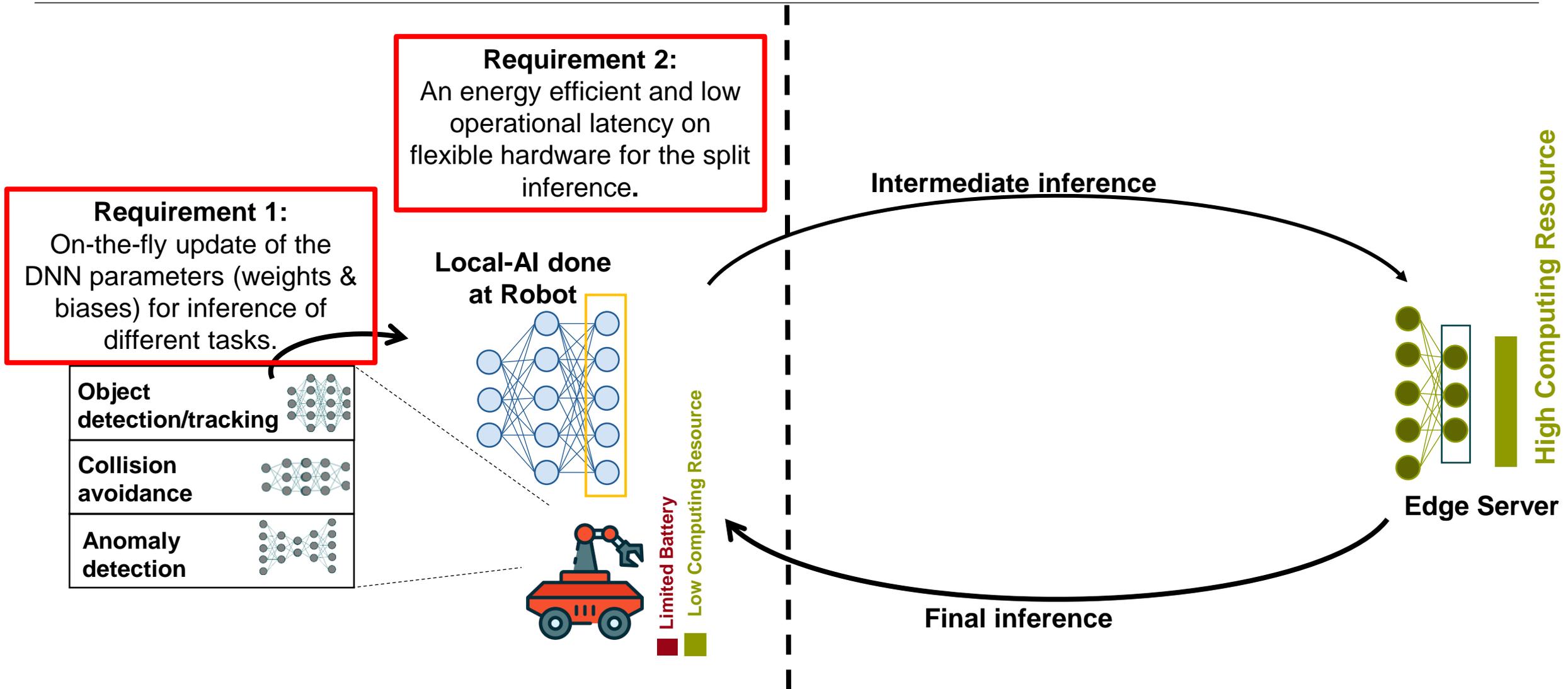
Communications and computing for industrial robotics



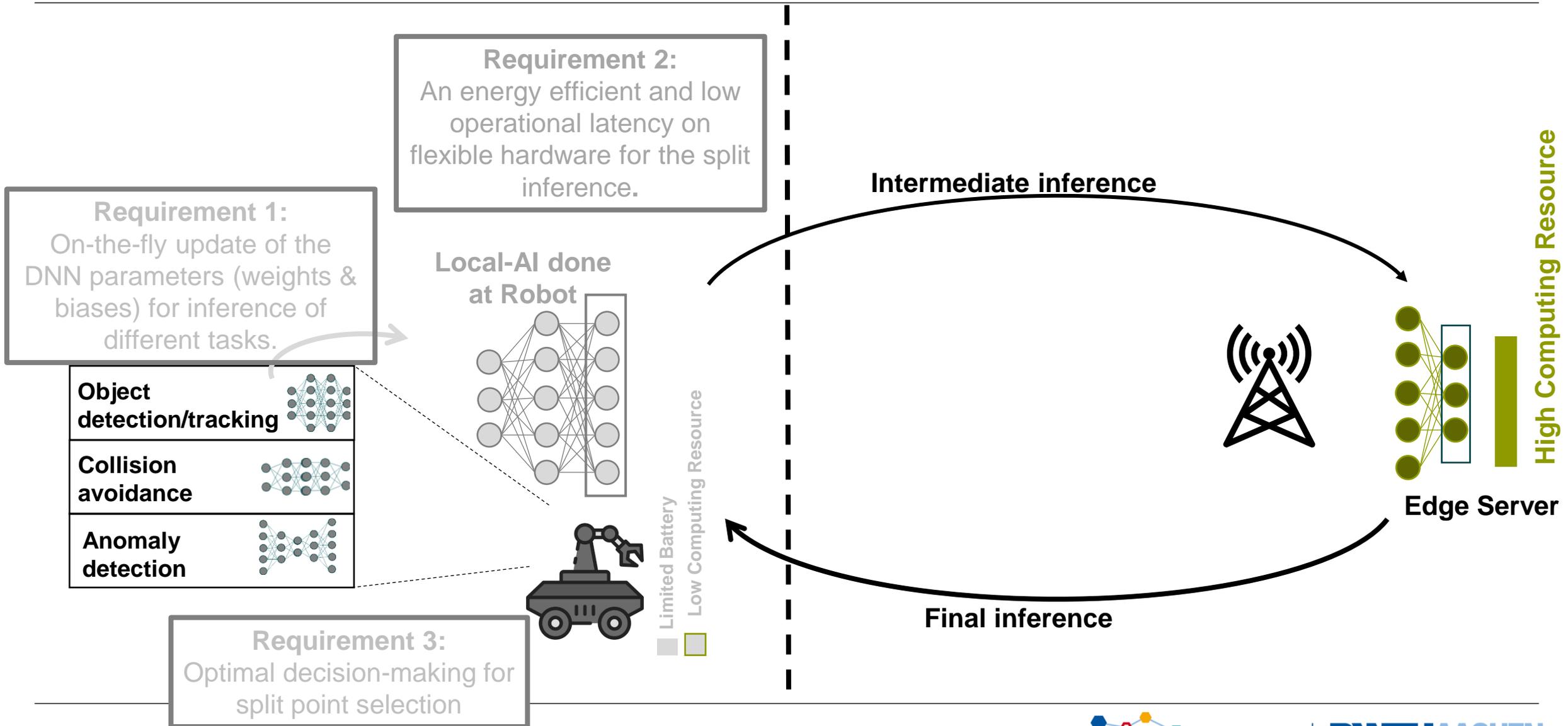
Communications and computing for industrial robotics



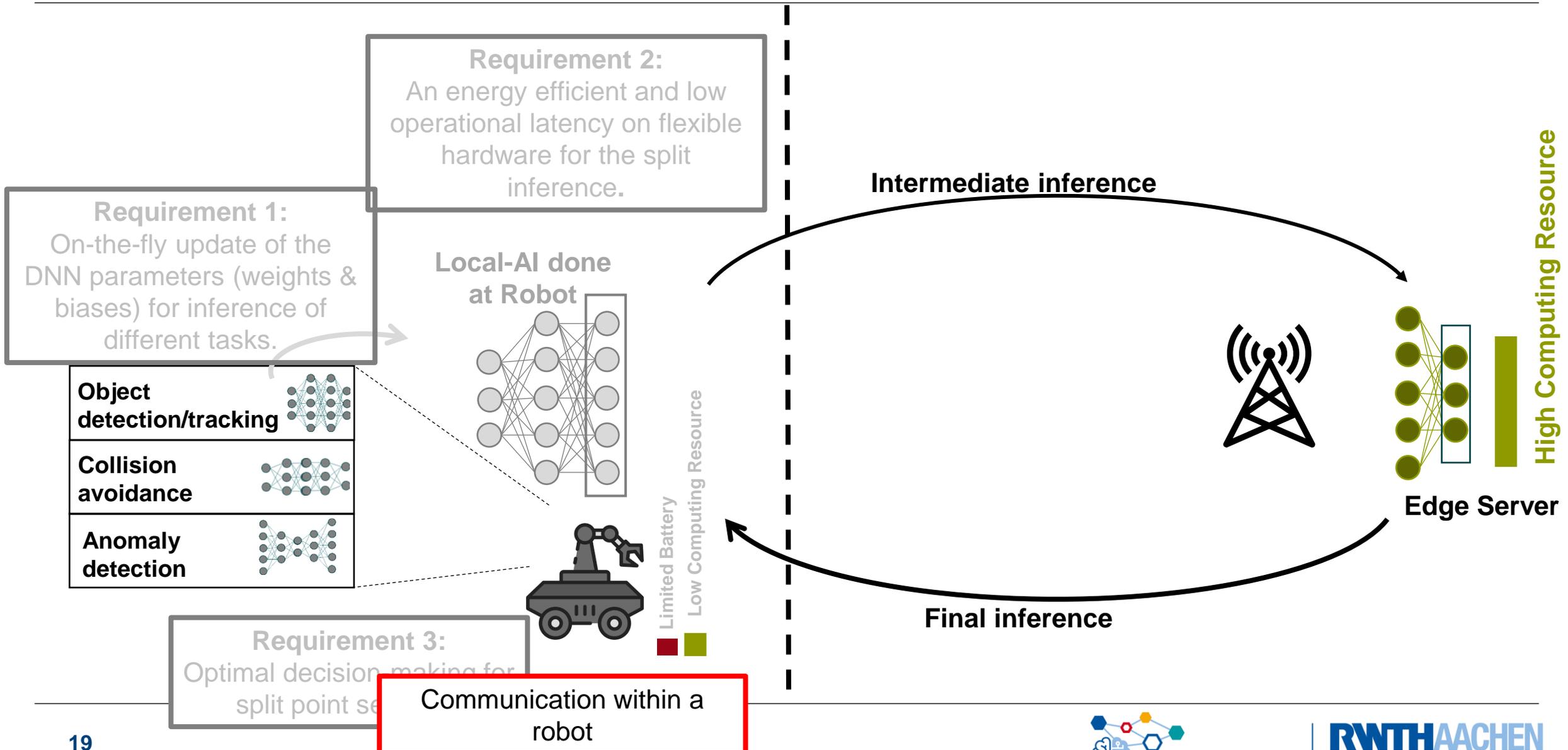
Communications and computing for industrial robotics



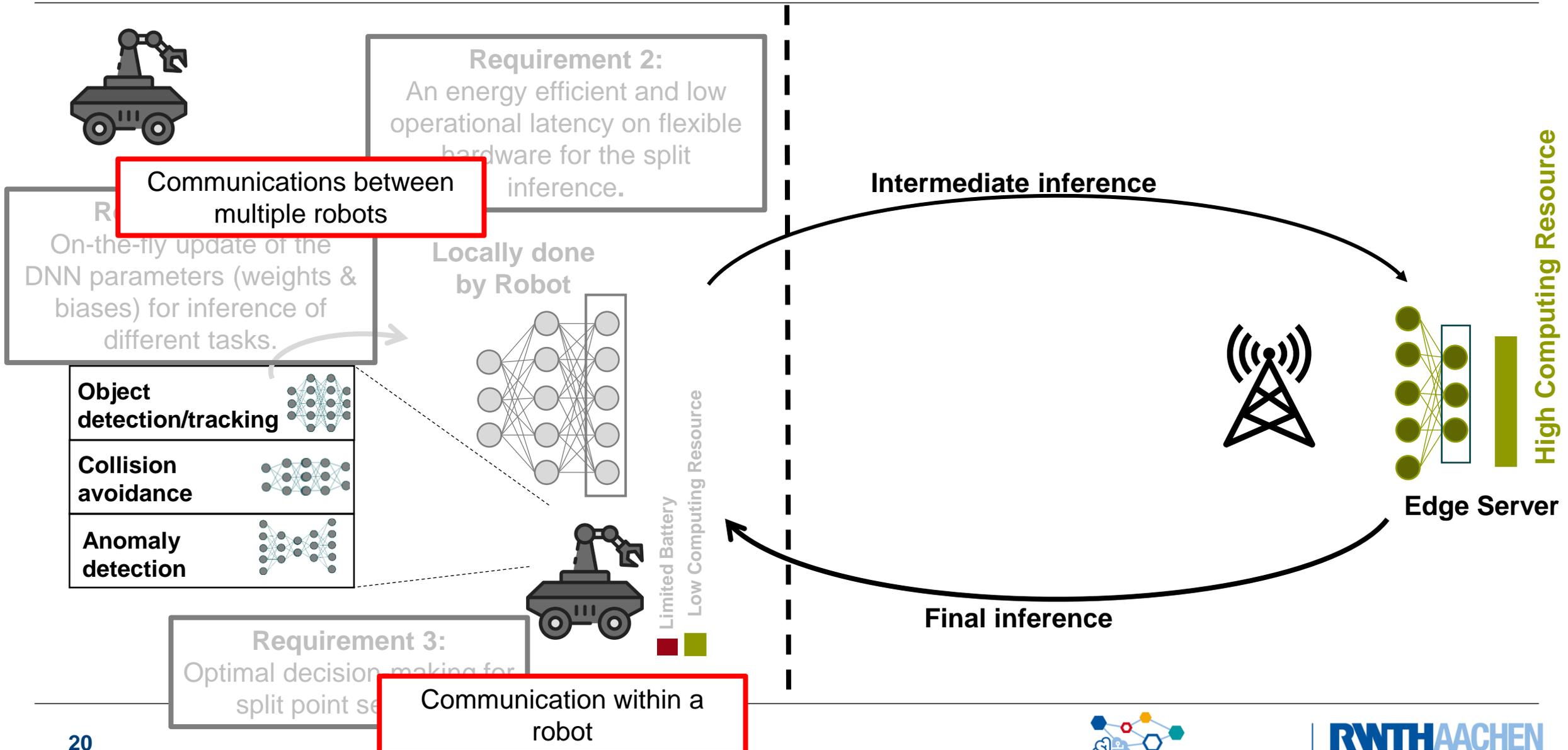
Communications and computing for industrial robotics



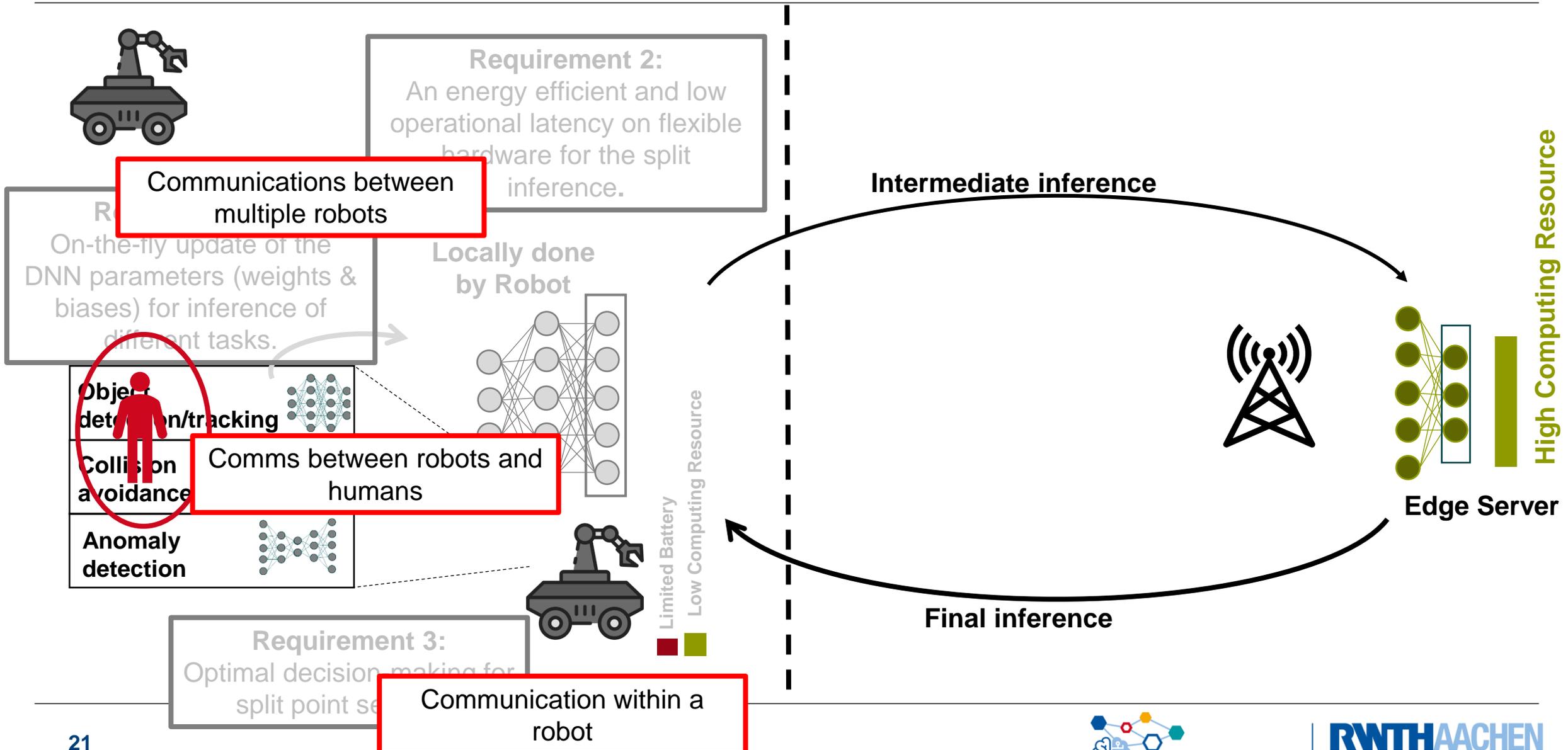
Communications and computing for industrial robotics



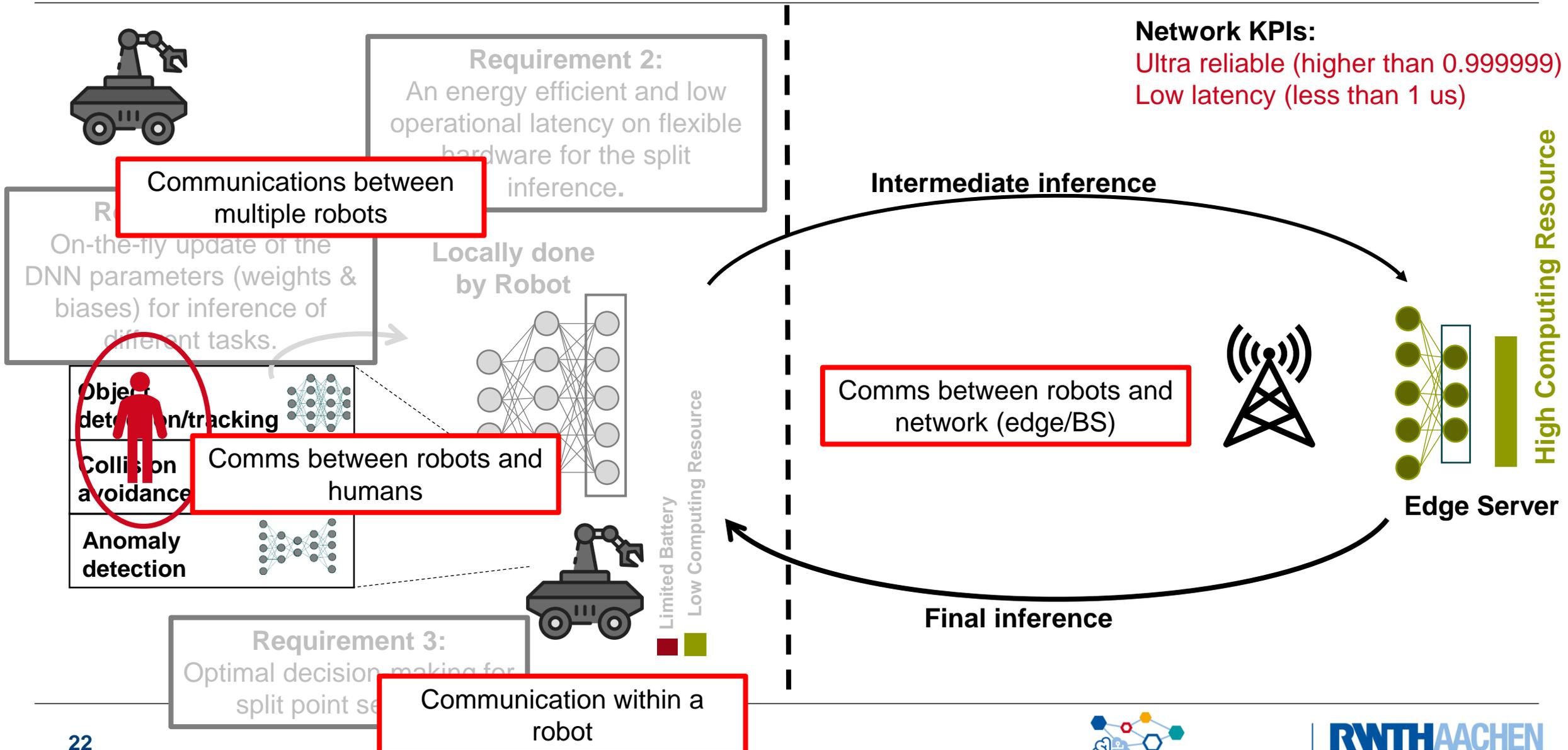
Communications and computing for industrial robotics



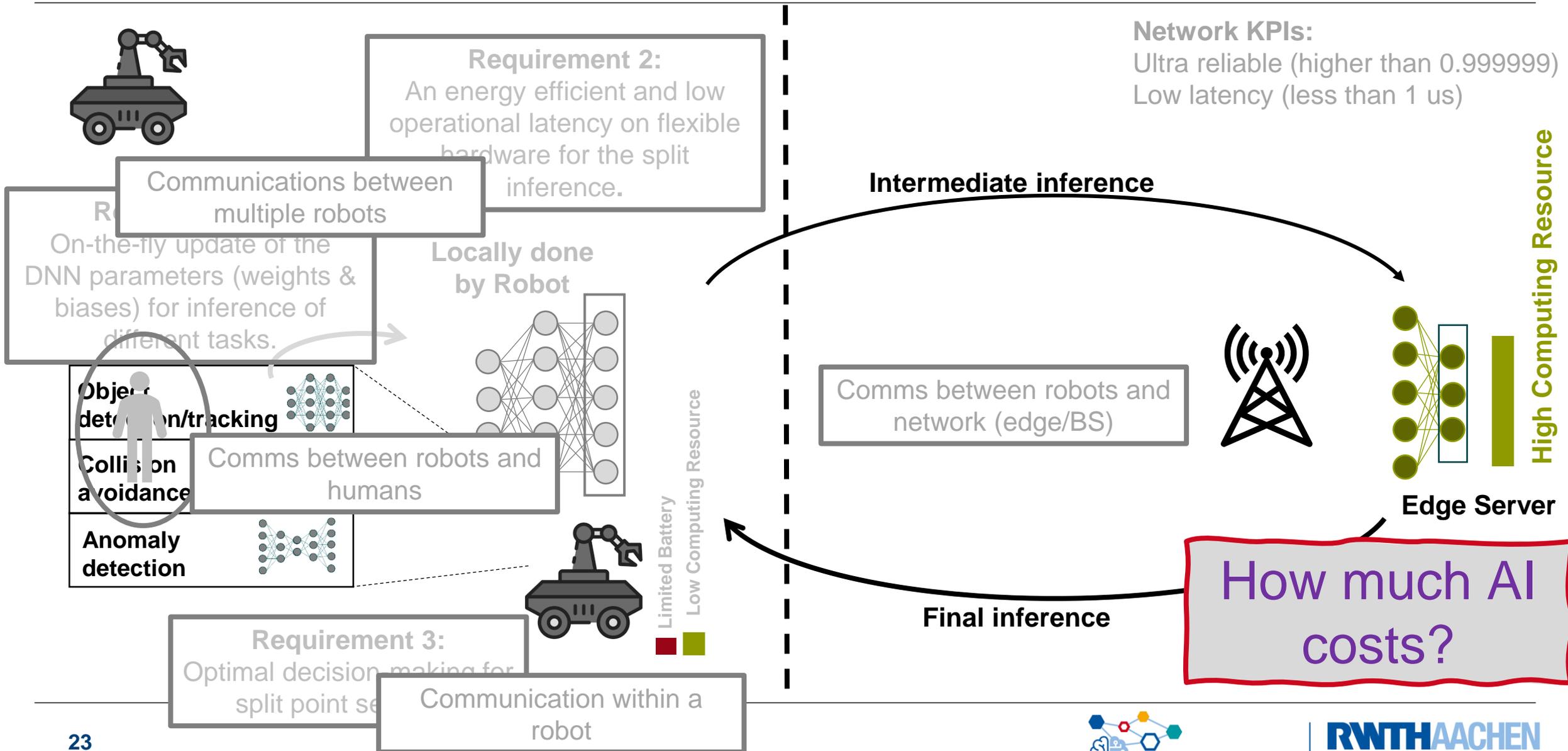
Communications and computing for industrial robotics



Communications and computing for industrial robotics



Communications and computing for industrial robotics



Energy consumption of GPU-based computing

- At 60% annual utilization, an H100 GPU would consume approximately 3,68 megawatt-hours (MWh) of electricity annually
- This is equivalent to the power consumption of the average German household occupant (based on 2-3 people/household)



Energy consumption of GPU-based computing

- At 60% annual utilization, an H100 GPU would consume approximately 3,68 megawatt-hours (MWh) of electricity annually
- This is equivalent to the power consumption of the average German household occupant (based on 2-3 people/household)
- 12,880,000 MWh of electricity per year (ca. 3,456 BEUR)
 - About city of Tokyo per year
 - About the joint consumption of Berlin, Hamburg, Munich, Cologne, Frankfurt
 - Equivalent to 3.5 million of German households



LLM and computing challenges

- Consider an example:
 - 100 million weekly users of chatGPT
 - 10 queries per user per week
- 88,400 MWh per year for operating an LLM like GPT-3.5
 - A European city-like energy consumption per year

LLM and computing challenges

- Consider an example:
 - 100 million weekly users of chatGPT
 - 10 queries per user per week
- 88,400 MWh per year for operating an LLM like GPT-3.5
 - A European city-like energy consumption per year

No preaching „Green Religion“, but please consider to innovate environmental-friendly and energy sustainable technology

Our projects and research directions



6GEM

6G research hub open,
efficient and secure
mobile communications
systems

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



6GEMcubator

6G technology
incubator center

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



Holistic approaches
for 6th generation
mobile networks

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



ITAR project
integrating wireless
and vision

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

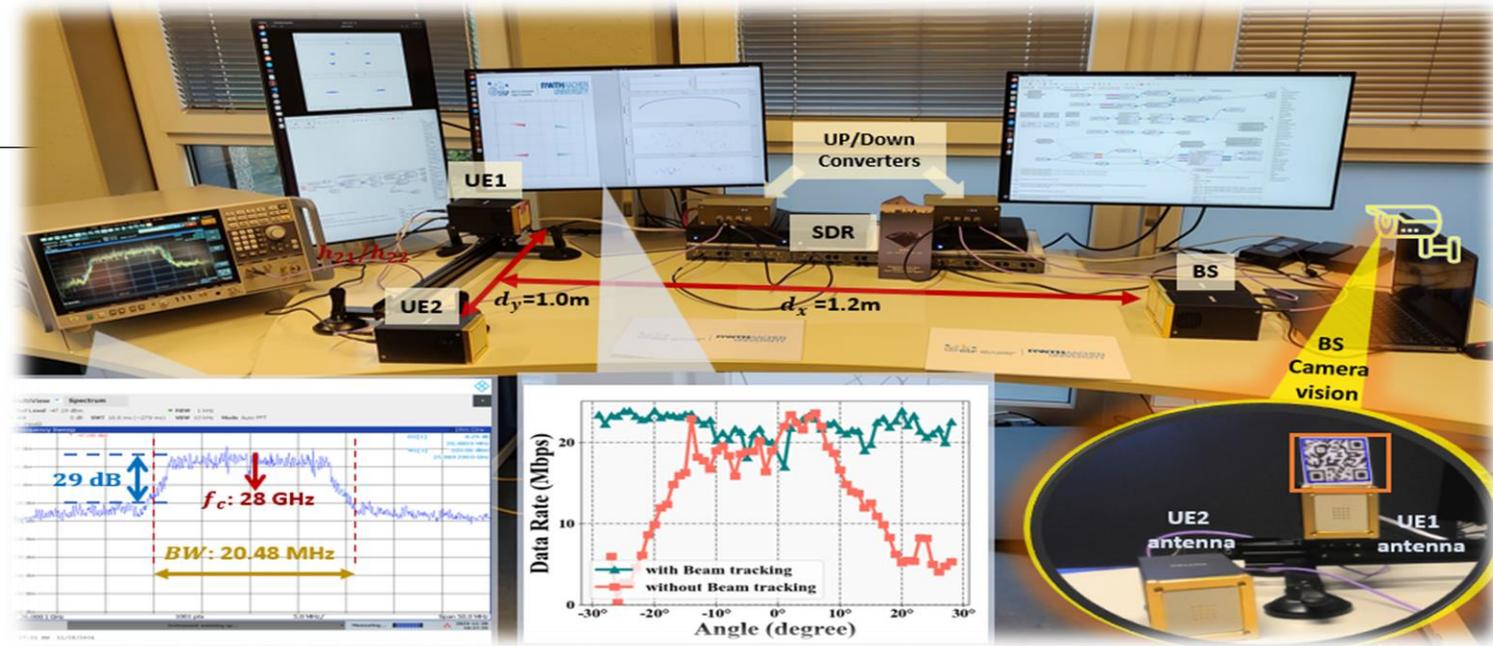
Examples of research topics

- MIMO
- Learning
- Remote control systems

Open research platforms

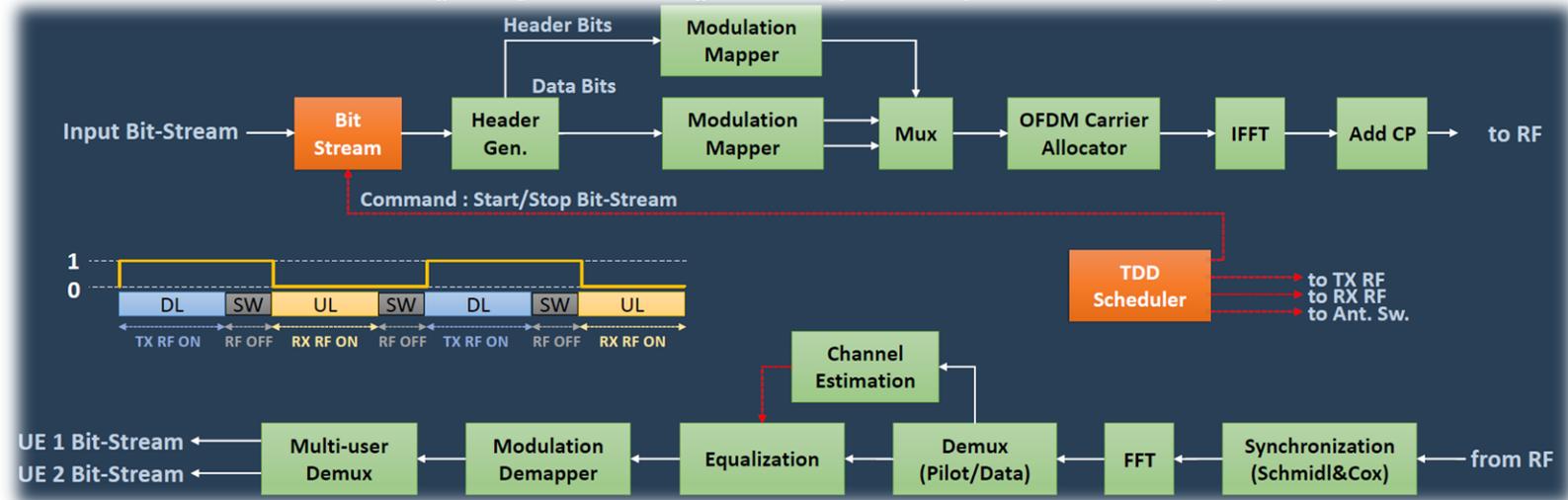
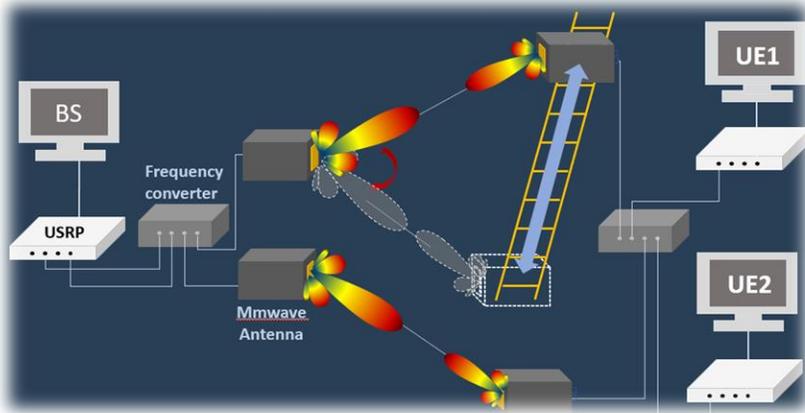
Open and programmable mmWave testbed for industrial automation

- mmWave TDD and beam sweeping
- FPGA channel emulator
- MU-MIMO spatial multiplexing



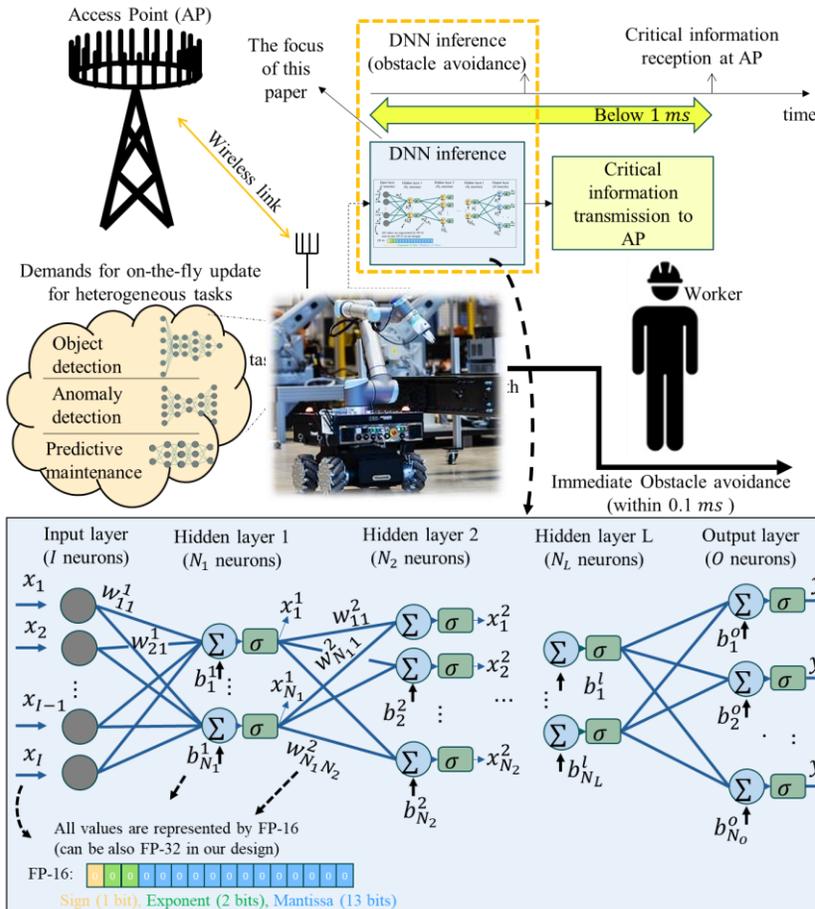
TDD Full chain signal processing development (mmWave PHY)

Hybrid MIMO Beamforming



Energy efficient ML (on-device inference and training)

FPGA Accelerator design for DNN & CNN models



Test results on MPSoC ZCU106

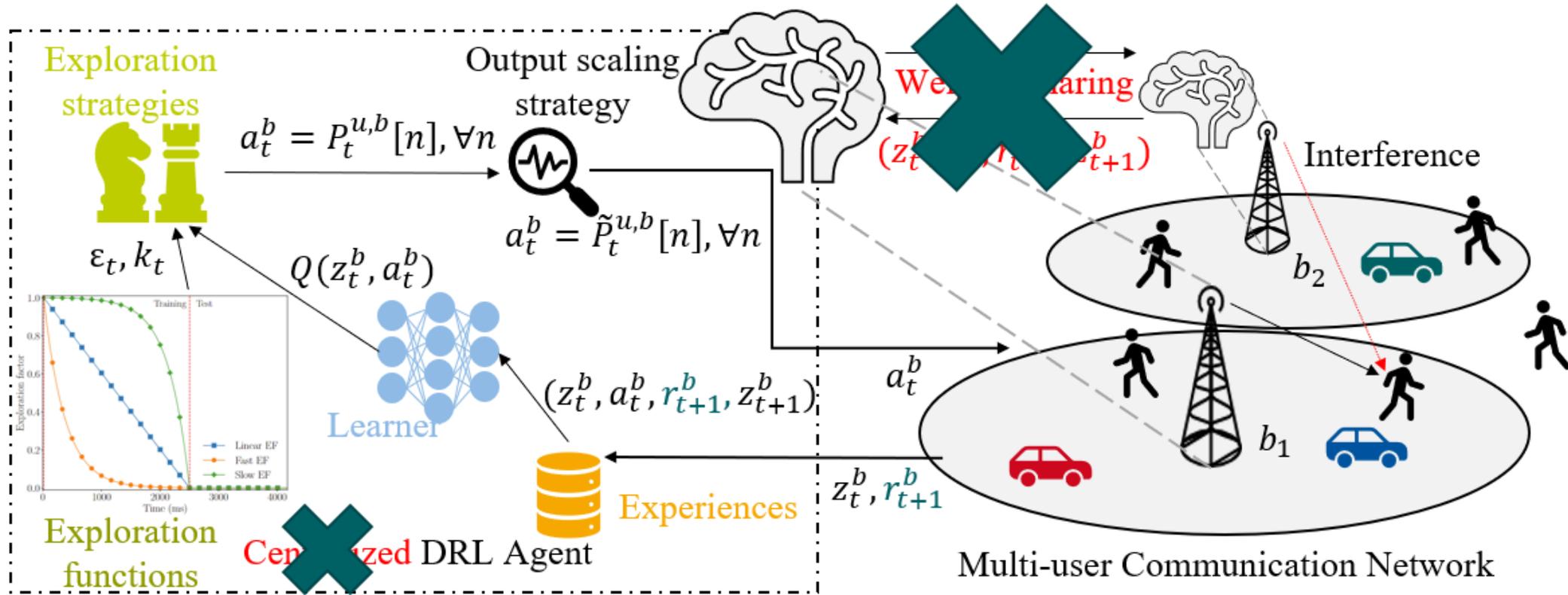
- Power consumption < 3W
- 57% lower latency compared to GPU-RTX 4090
- 80% less power consumption compared to Raspberry Pi
- 29 times faster and 30% more power efficient than AMD Deep Processing Unit (DPU)

Target use case:

- PHY signal processing (e.g., channel estimation & beamforming)
- IoT/edge computing (e.g., object detection & tracking)

Dynamic system adaptation – DRL/LLM

DRL and LLM for wireless power allocation in OFDM systems and for edge computing in IIoT use cases



LLM for prediction problem

■ LLM Architecture (Transformer Model)

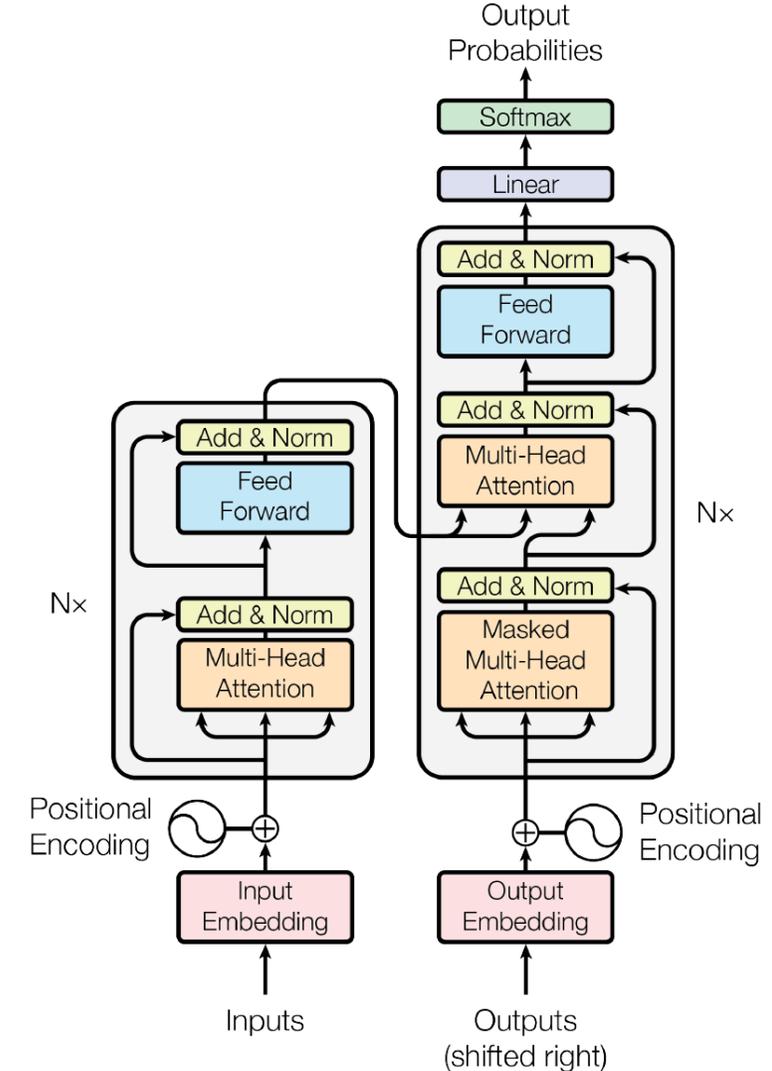
- Input Embedding
- Positional Encoding
- Encoder/Decoder layers
 - Multi-head attention layer
 - Feed forward layer
 - Residual connection
 - Addition and normalization layer

■ LLM Exploitation Strategies

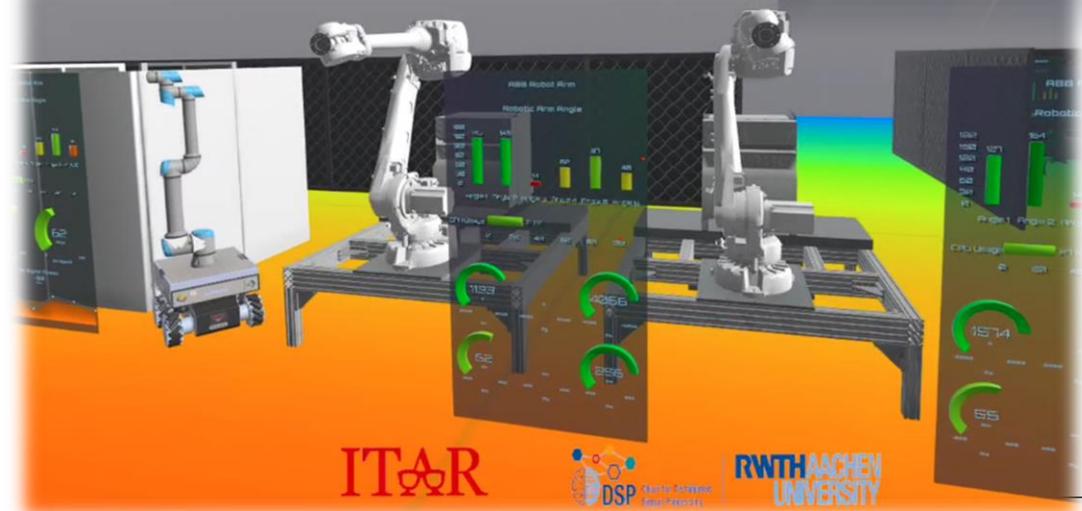
- Prompt Engineering
- RAG
- Fine Tuning

■ Exploitation of LLM in Wireless Networks

- Dynamic Power Allocation
- Channel Estimation



XR remote control systems for interactive industry



Our impact and partnerships

- 7 testfield demos
- 3 patents (1 startup grant)
- 4 conference demos
- 6 invited talks
- 16 conference and journals
- GE-JP workshop organization (2 demo)
- 20+ posters



SmartCom 2024

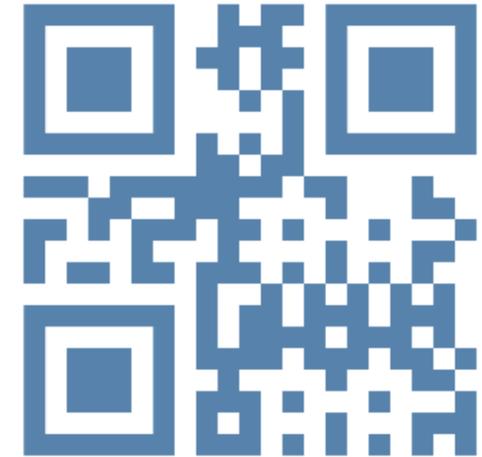


Thank you!

If interested in collaboration please contact us:



RWTHAACHEN
UNIVERSITY



AI's true nature mirrors the intentions and ambitions of those who create and guide it

Personal dilemma

- Powerful or insatiable
 - AI's true nature mirrors the intentions and ambitions of those who create and guide it