Qualcomm

February 6, 2020

@qualcomm_tech

5G+AI:

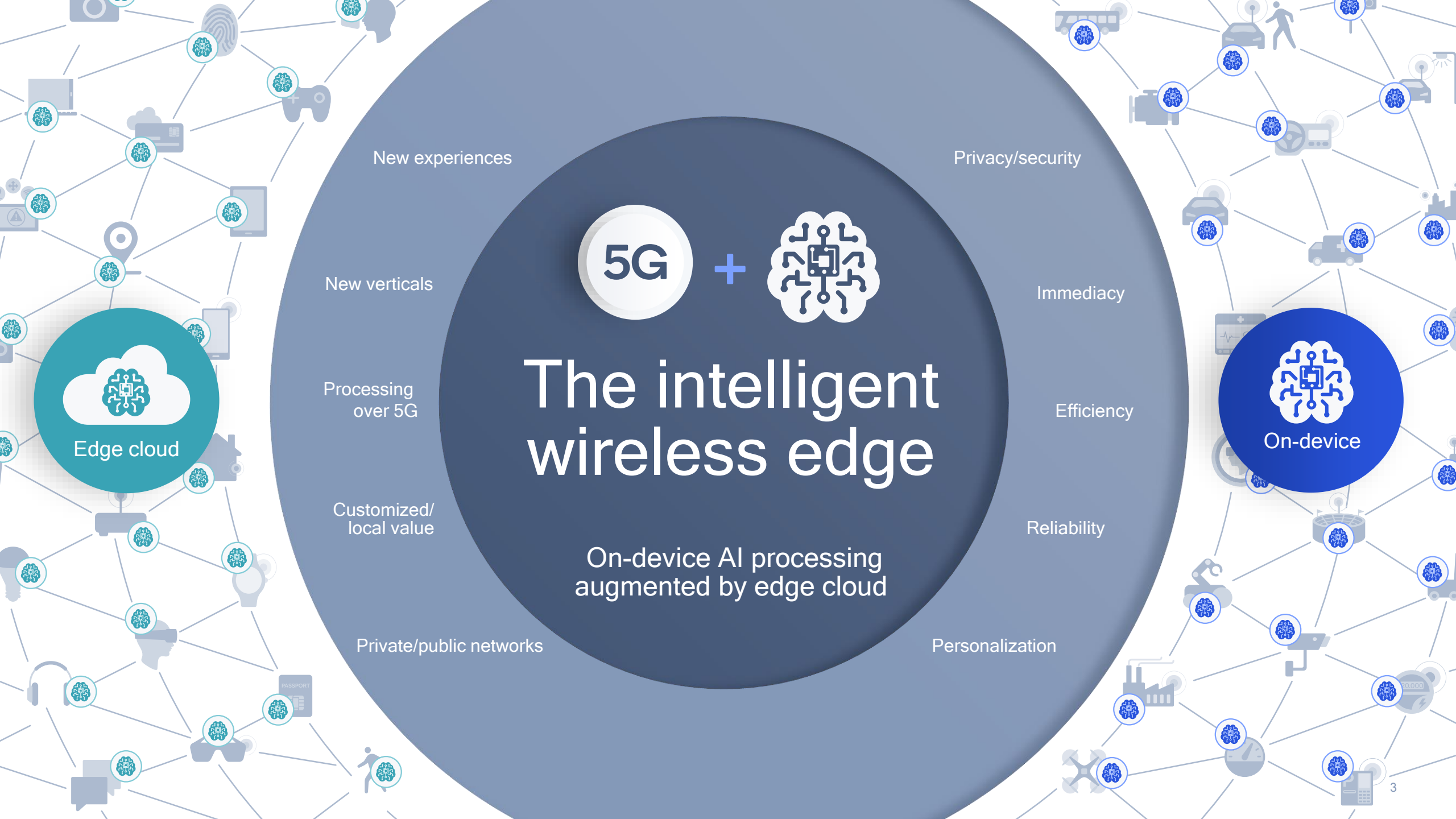
The ingredients fueling tomorrow's technology innovations

5G

Delivering on the 5G vision

$13.2 trillion in goods & services by 2035*

*The 5G Economy, an independent study from IHS Markit, commissioned by Qualcomm


New experiences
Privacy/security
New verticals
Immediacy
Processing over 5G
Efficiency
Customized/ local value
Reliability
Private/public networks
Personalization
5G
+
The intelligent wireless edge
On-device AI processing augmented by edge cloud
Edge cloud
On-device

Process data at the source to scale AI and make sense of a digitized world

Past

Cloud-centric AI

AI training and AI inference
in the central cloud

Today

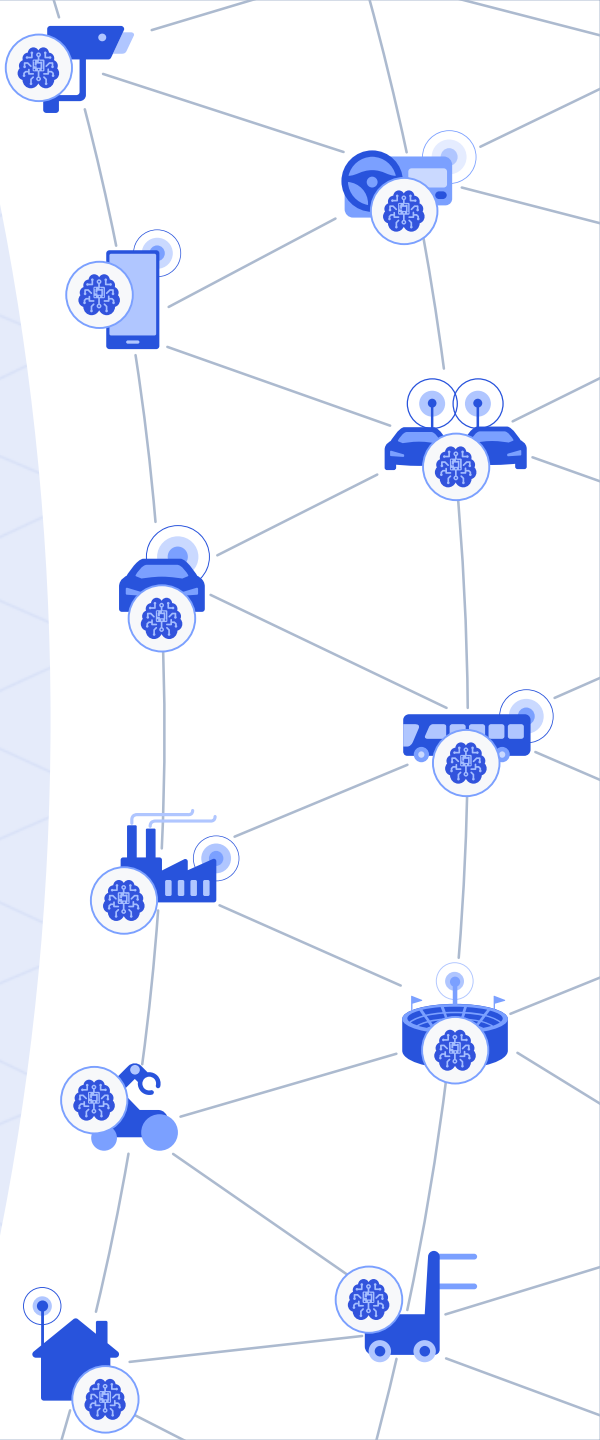
Partially-distributed AI

Power-efficient
on-device AI inference

Future

Fully-distributed AI

With lifelong on-device learning



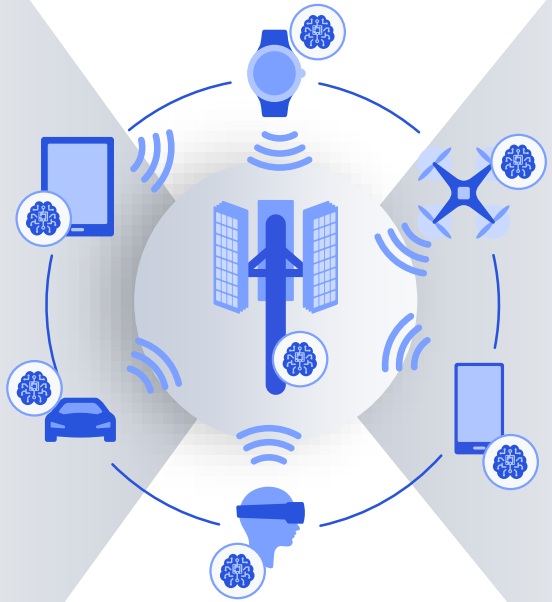
Applying AI to overcome wireless challenges

Applying AI to solve difficult wireless challenges

Deep wireless domain knowledge is required to optimally use AI capabilities

Wireless challenges

- Hard-to-model problems
- Computational infeasibility of optimal solution
- Efficient modem parameter optimization
- Dealing with non-linearity



AI-enhanced wireless communications

AI strengths

- Determining appropriate representations for hard-to-model problems
- Finding near-ideal and computationally realizable solutions
- Modeling non-linear functions

AI enables intelligent 5G network management

Enhanced service quality

Better mobility management, user localization, and user behavior and demand prediction

Simplified deployment

More capable Self Organizing Networks (SON) for e.g., mmWave network densification

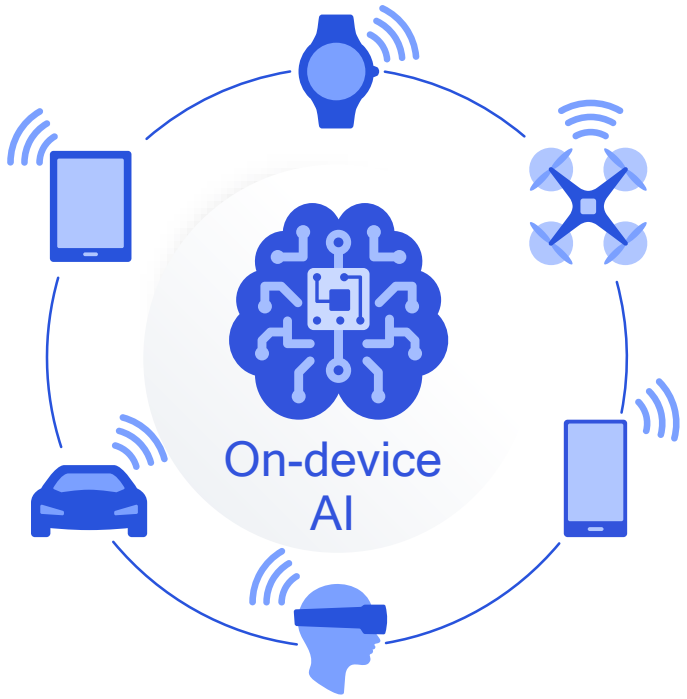
Higher network efficiency

More efficient scheduling, radio resource utilization, congestion control and routing

Improved network security

More effective detection and defense against malicious attacks by analyzing a massive quantity of data

On-device AI improves the 5G end-to-end system

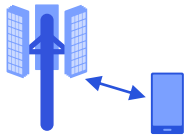


Radio awareness

Environmental and contextual sensing that reduces access overhead and latency

Enhanced device experience

More intelligent beamforming & power management improve throughput, robustness, and battery life



Improved system performance

On-device inference reduces network data traffic for more efficient mobility and spectrum utilization

Better radio security

Detecting and defending against malicious base station spoofing and jamming with fingerprinting

Radio awareness

Achieved by advanced on-device AI algorithms

Spectrum sensing and access

Predict activities of other devices for more efficient access and better scheduling to improve 5G system performance

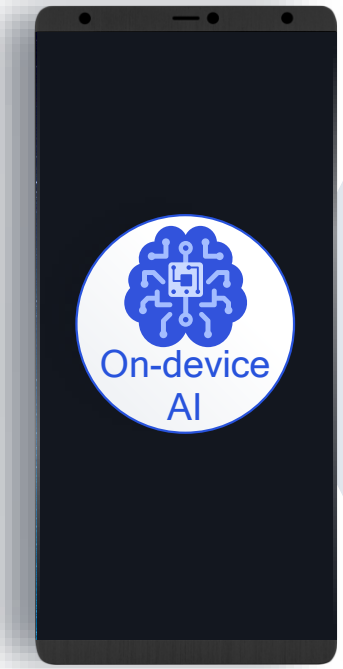
Contextual awareness

Use device context (e.g., position, velocity, or in-car) derived from RF, sensors, traffic activities to improve device experience

Environment (RF) sensing

Detect gestures, movements, and objects by monitoring signal reflection patterns to enable new use cases

On-device AI enhances 5G device experience



Better beam management

Incorporate location, velocity, other aspects of environmental and application awareness to improve robustness and throughput

More power saving

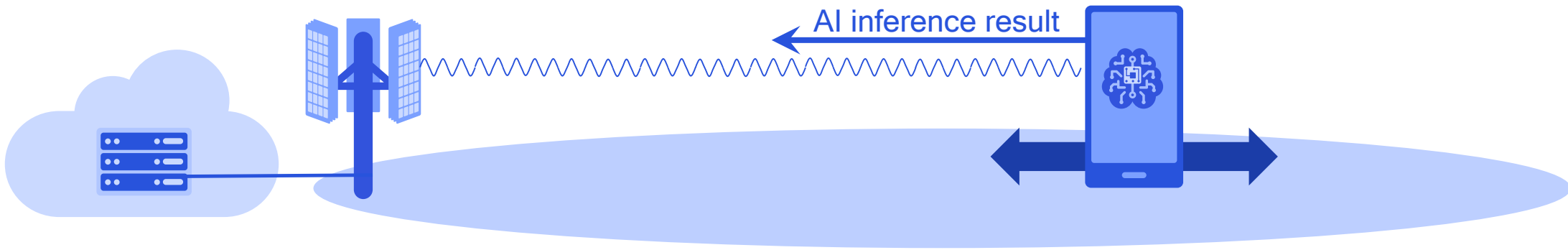
Optimize performance/power consumption tradeoffs by taking advantage of better contextual awareness on device

Deep reinforcement learning example

On-device AI improves 5G system performance

Better link adaptation

Position-aware interference prediction can improve overall system throughput and spectral efficiency



Reduced network loading

On-device AI inference reduces the amount of raw data needed be sent across the network

More seamless mobility

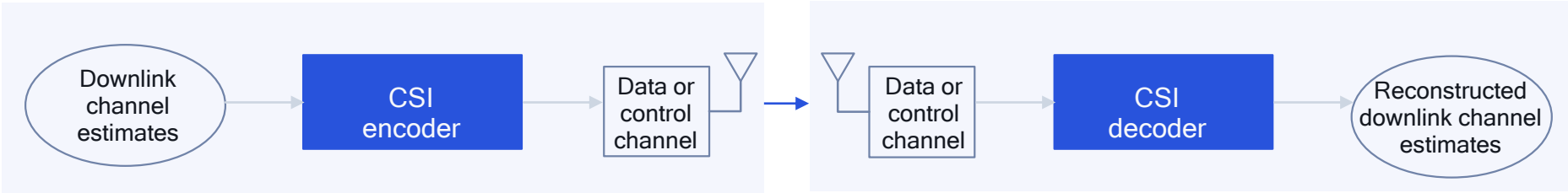
Device-centric mobility utilizes on-device AI and sensors to predict handovers

Applying AI for enhanced 5G air interface efficiency

Example: for uplink transmissions

Improving system spectral efficiency

Implementing a neural network framework for CSI[1] on non-linear temporal encoding & decoding



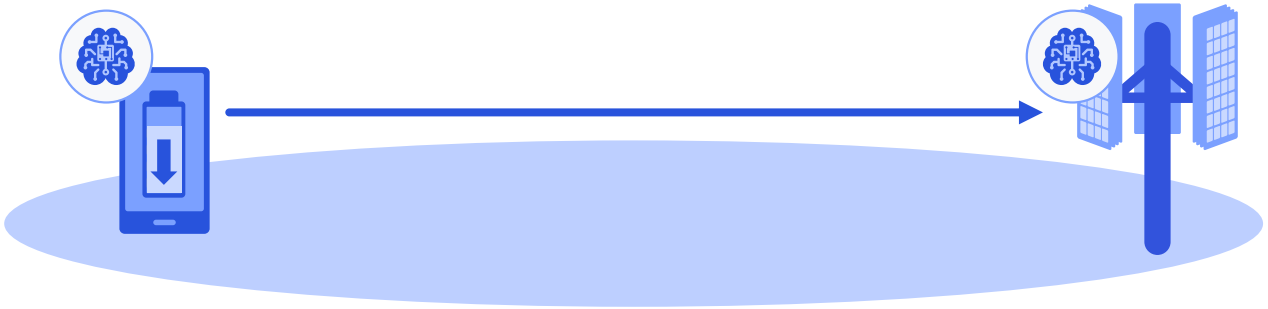
Encoder
at the device

Decoder
at the base station

Improving device power efficiency

Optimizing transmit waveform to reduce peak-to-average power ratio (PAPR)

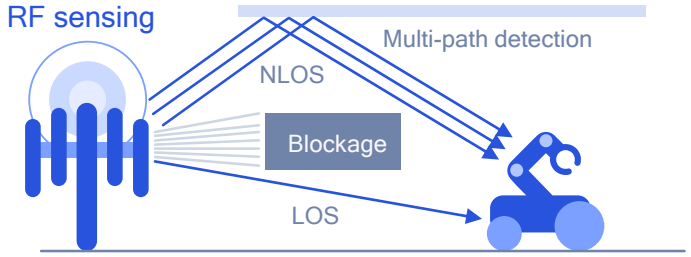
Allowing receiver to recover signal from a device operating in the non-linear PA region



1 Channel State Feedback

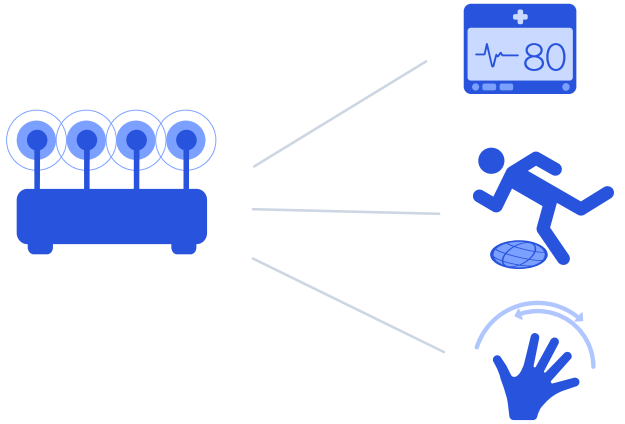
Applying AI for contextual awareness & environmental sensing

More accurate device positioning



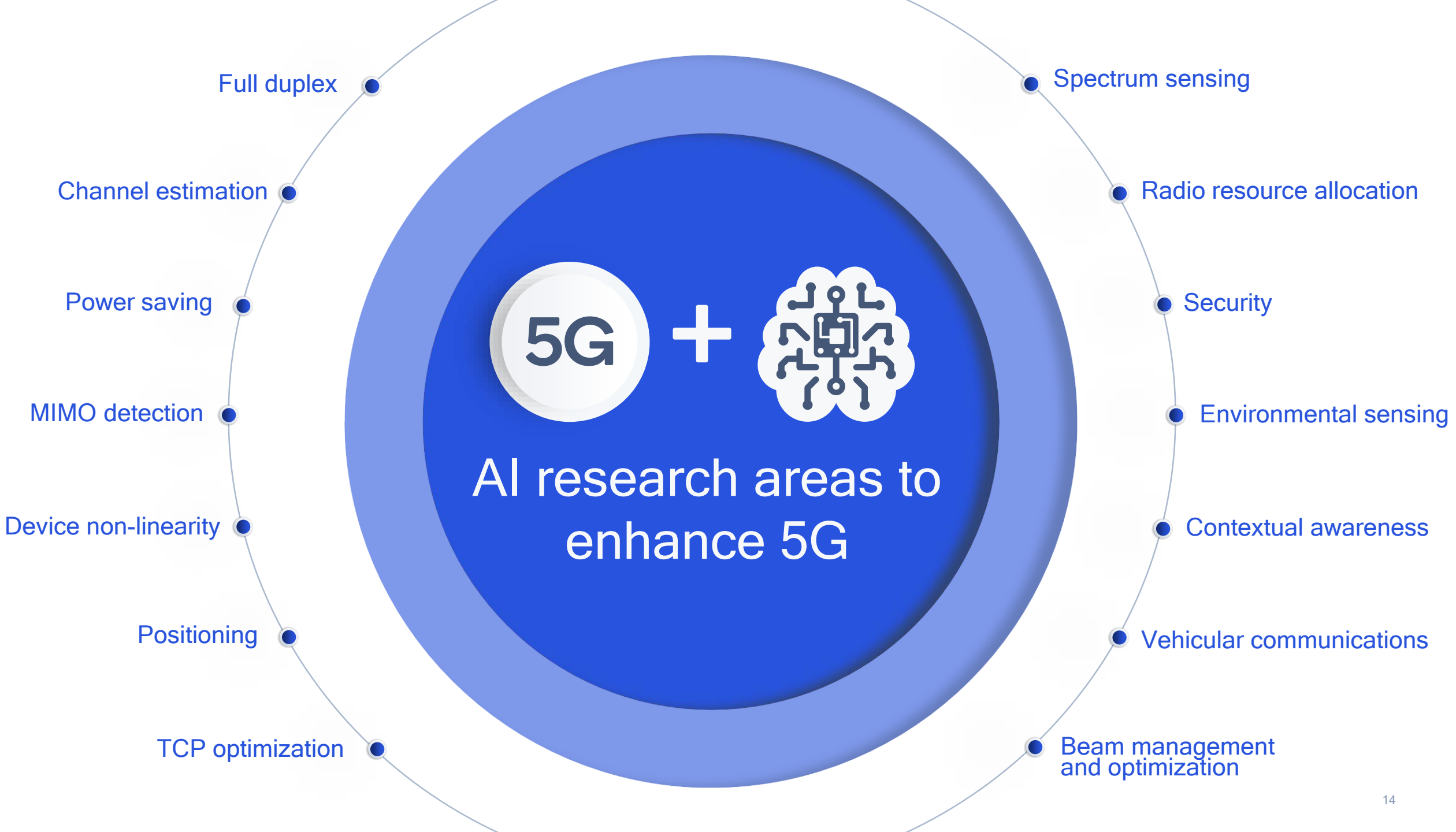
Learning device position over time without prior knowledge with RF sensing – complementing existing positioning methodologies[1]

Motion and gesture detection



Sensing changes in environment to infer location and type of motion for a wide range of use cases (e.g., vital sign tracking, fall detection)

1. For example, Observed Time Difference of Arrival (OTDOA), Multiple Round Trip Time (Multi-RTT), Angle of Arrival (AoA)


5G
AI research areas to enhance 5G
Full duplex
Channel estimation
Power saving
MIMO detection
Device non-linearity
Positioning
TCP optimization
Spectrum sensing
Radio resource allocation
Security
Environmental sensing
Contextual awareness
Vehicular communications
Beam management and optimization

Applying 5G to enhance AI-powered experiences

5G + AI deliver enhanced services and experiences

Distribute AI processing between the device and cloud over wireless

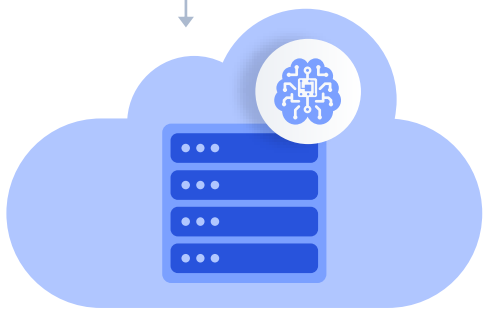
Distributed functionality

5G

Central cloud

Longer latency
Big data/aggregated value
Content/storage/AI/processing

Edge cloud



5G low latency
Customized/local value
Content/storage/AI/processing

Device

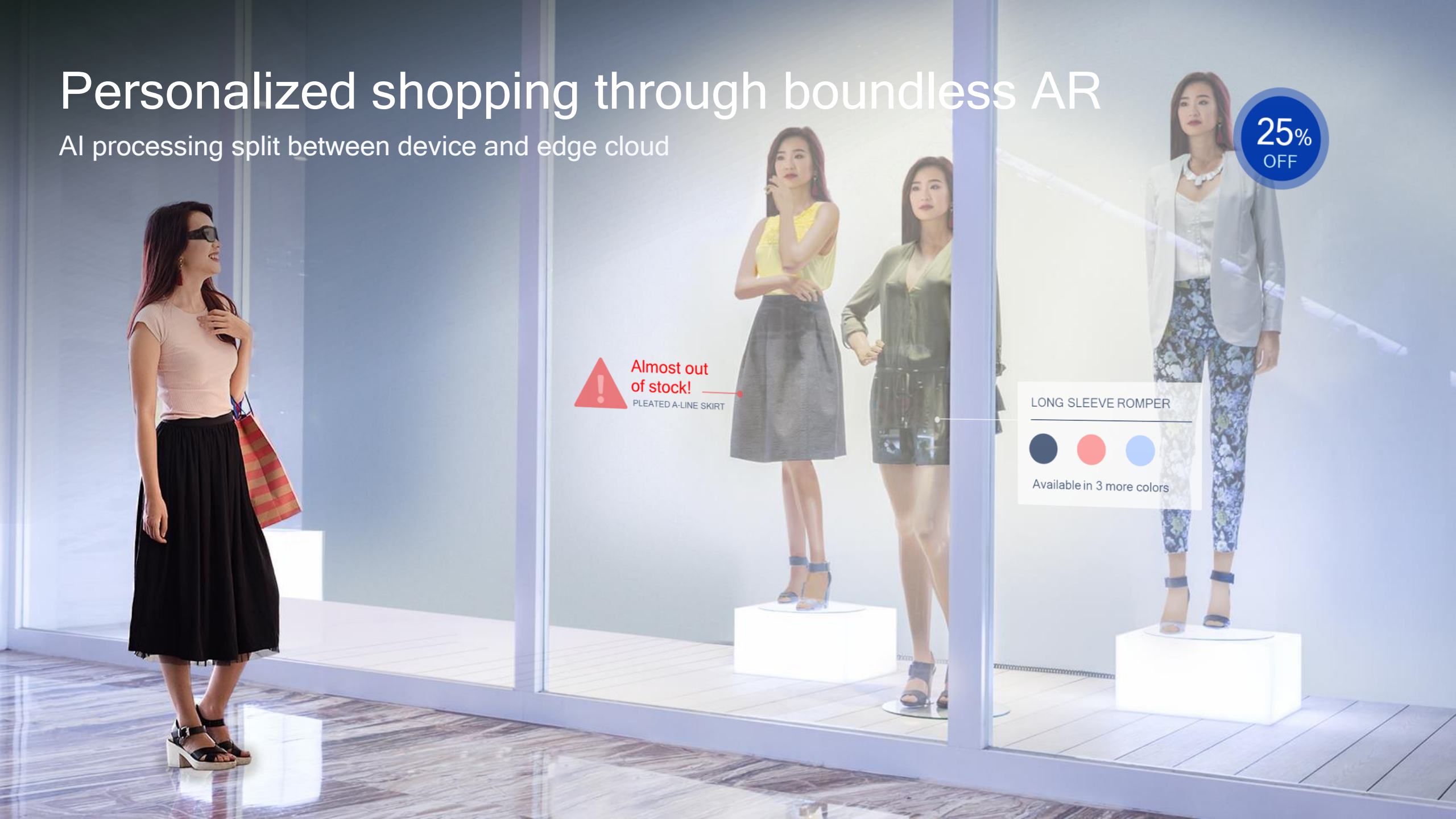
Compute, vision, sensing
AI powered use cases
Internal AI optimizations

Performance and economic tradeoffs per use case may determine how to distribute functions
e.g. to reach required latency or compute

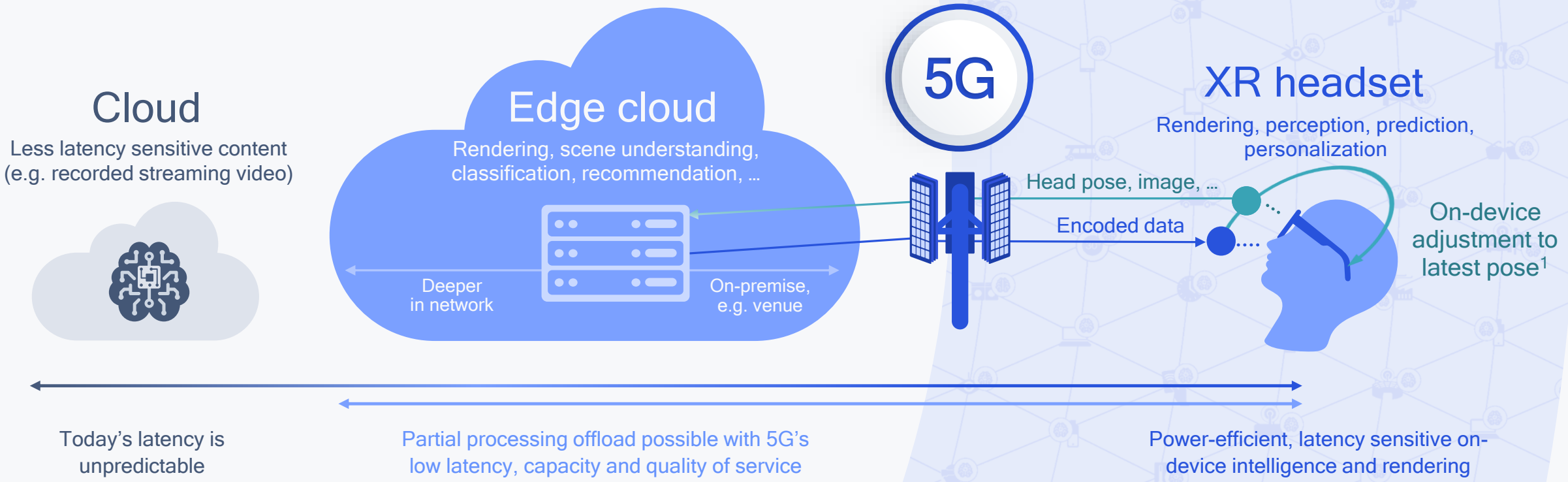
- Distributed processing, like boundless XR
- Real time assisted services like voice UI
- On-premise control for ultra-low latency
- New services
- Cloud computing, storage, instant access
- Low-latency gaming
- On-device intelligence assisted by cloud

Personalized shopping through boundless AR

AI processing split between device and edge cloud



The edge cloud augments on-device processing for boundless XR



1. Asynchronous time warp reduces Motion to Photon (MTP) latency by using on-device processing based on the latest available pose. MTP below 20 ms generally avoids discomfort – has to be processed on the device

Intuitive virtual assistants through vastly improved voice UI

Designed to be:

Always-on

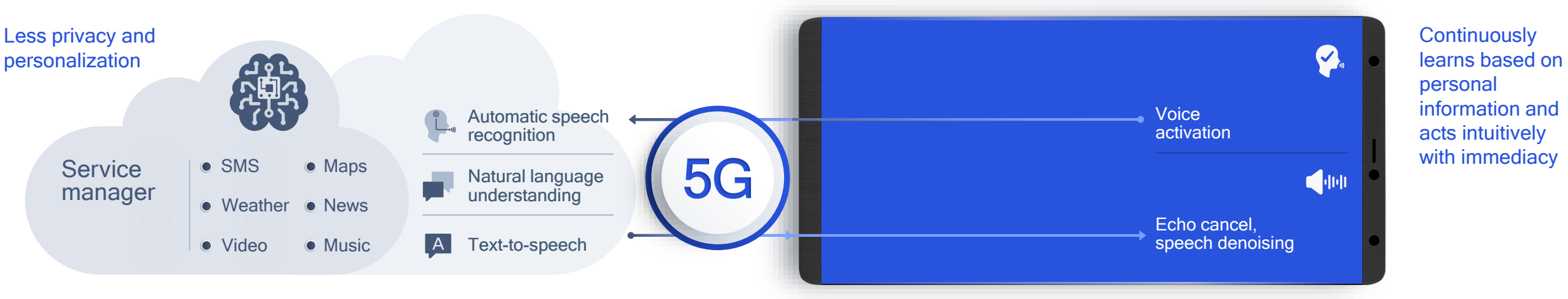
Conversational

Personal

Private

Distributed computing enables a responsive voice UI

5G low latency allows AI tasks to be split between the device and cloud



Both ends are needed – 5G allows various implementation for appropriate tradeoffs

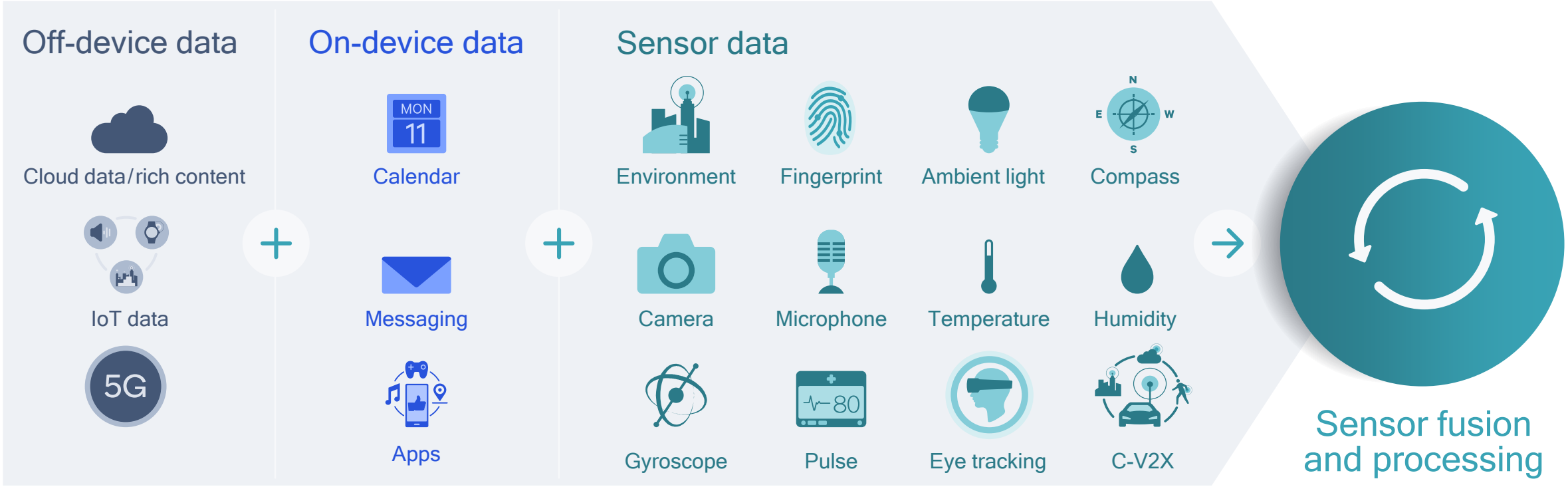
The reconfigurable factory of the future will thrive in the 5G+AI era


Enhanced mobile broadband
Computer vision
Head mounted display
Handheld terminal
Massive IoT
Sensors
Ultra-reliable low-latency
Automated guided vehicle (AGV)
Wireless edge analytics
Industrial robot

Distributed learning over wireless

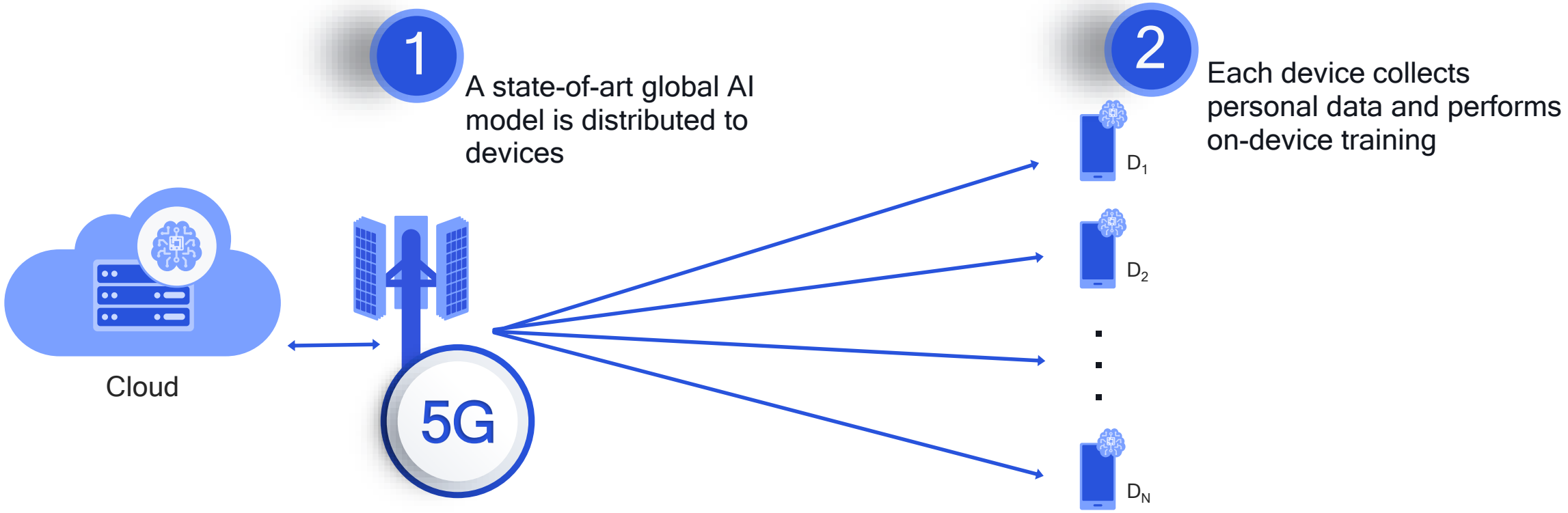
Distributed learning for an intelligent world

Devices generate and possess massive amounts of data



On-device AI processing of sensors and personal information conserves bandwidth while providing contextual intelligence, personalization, and privacy

Distributed learning over 5G is the way to scale intelligence



Scale

Processing is spread over many devices

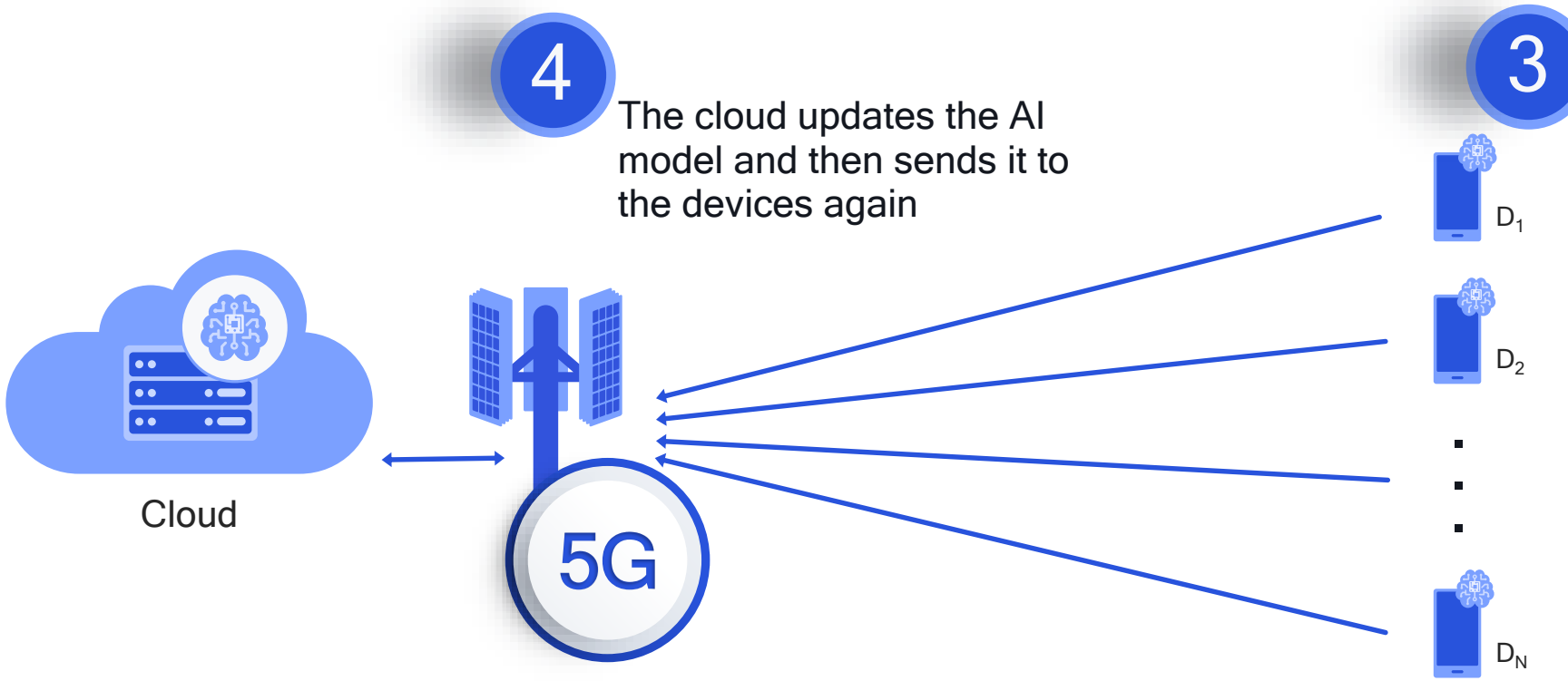
Personalization

Model customized based on your personal data

Privacy

Raw data stays on the device

Distributed learning over 5G is the way to scale intelligence



Each device:

1. Adds perturbation to the parameters of the AI model
2. Compresses the parameters
3. Encrypts them
4. Sends the update to the cloud

Scale

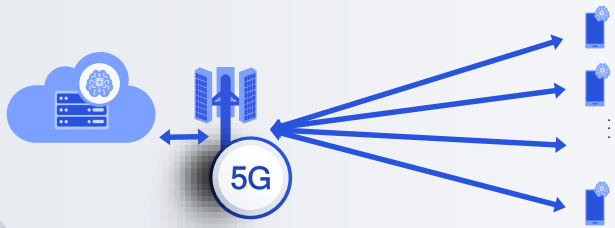
Network bandwidth is conserved

Personalization

Privacy

Only noisy and encrypted weights sent to the cloud

Research directions for distributed learning over wireless



Optimizing the communication network

Compressing information sent on the uplink and downlink

Tackling statistical heterogeneity in data

Smartly combining model updates from a broad distribution

Privacy

Model parameter perturbations with privacy guarantees

Personalization

Meta learning with optimized global model starting point

On-device training efficiency

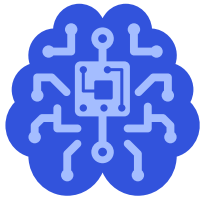
Light-weight models and training

Advanced topologies for distributed learning

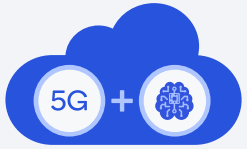
Peer-to-peer, multi-cloud, and hierarchical privacy

Qualcomm

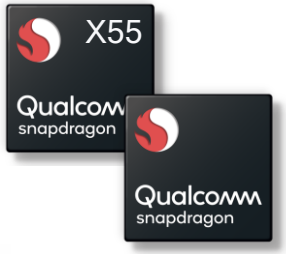
Foundational R&D



Qualcomm AI Research

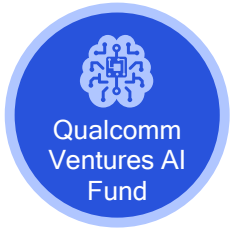


5G + AI technology leadership

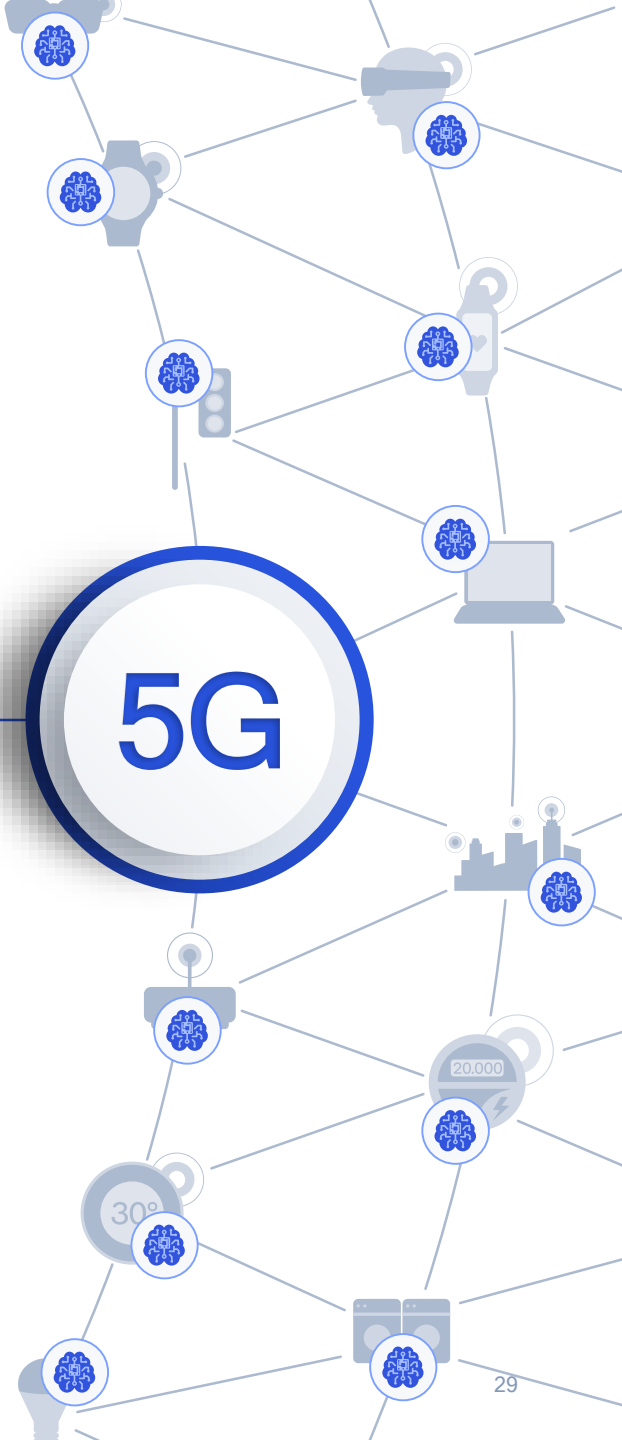


with AI Engine

Advanced silicon



Ecosystem investment



Uniquely positioned to power the intelligently connected future

Qualcomm

AI is making 5G better –
in the network and on the device

On-device AI processing is essential, and even more so in the 5G era

5G is empowering a new computing paradigm:
distributed learning over wireless

Questions?

Connect with Us

www.qualcomm.com/ai

www.qualcomm.com/5g

BLOG www.qualcomm.com/news/onq

@qualcomm_tech

YouTube

SlideShare

Qualcomm

Thank you

Follow us on:

For more information, visit us at:

www.qualcomm.com & www.qualcomm.com/blog