Qualcomm @QCOMResearch San Diego, CA July 13th, 2023

Towards an AI-native communications system design

A closer look at how AI can substantively improve wireless performance starting with 5G Advanced

Wireless + AI

Today’s agenda

ONE

The rise of AI brings a unique opportunity to revolutionize the future of wireless technology

TWO

AI plays an essential role in every part of the cellular system, transforming how it evolves going forward

THREE

5G Advanced starts the era of wireless AI, focusing on use cases that can bring immediate benefits

FOUR

Qualcomm is driving foundational wireless AI innovations leading towards an AI-native 6G

FIVE

Questions?

OUR PRESENTERS

Tingfang Ji

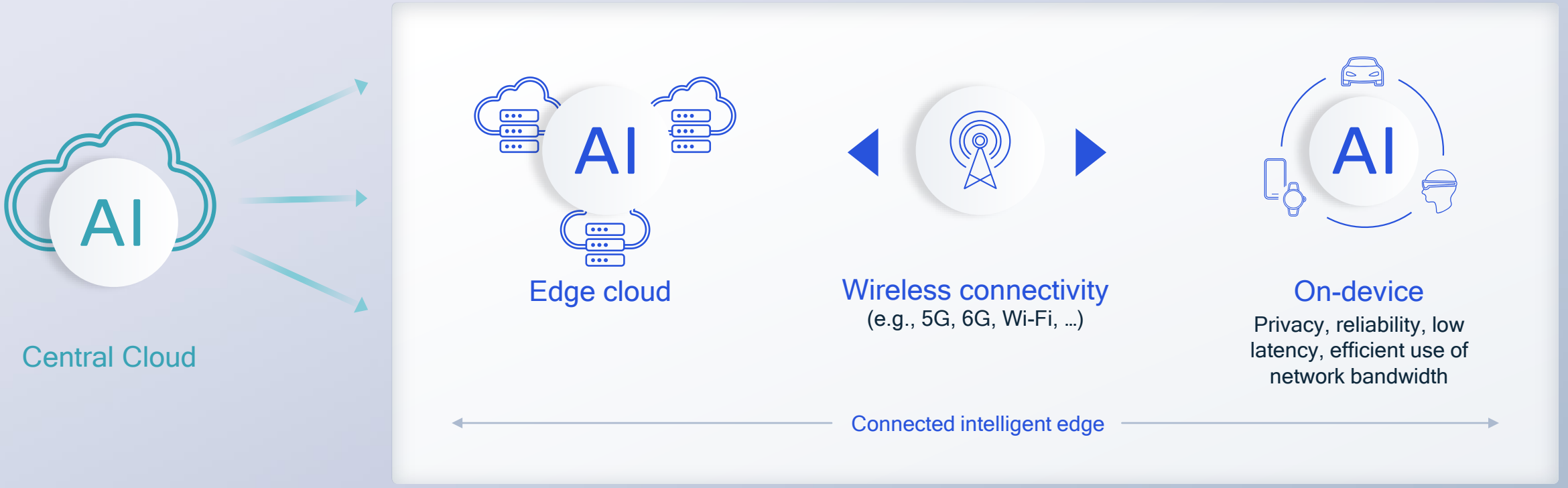
Vice President
Wireless R&D
Qualcomm Technologies, Inc.

Taesang Yoo

Senior Director
Wireless R&D
Qualcomm Technologies, Inc.


The rise of AI brings a unique opportunity to revolutionize the future of wireless technology

To scale efficiently, AI processing is expanding towards the edge



Qualcomm is leading the realization of the connected intelligent edge

CONVERGENCE OF:

- Wireless connectivity
- Efficient computing
- Distributed AI
- Unleashing massive amount of data to fuel our digital future

Wireless

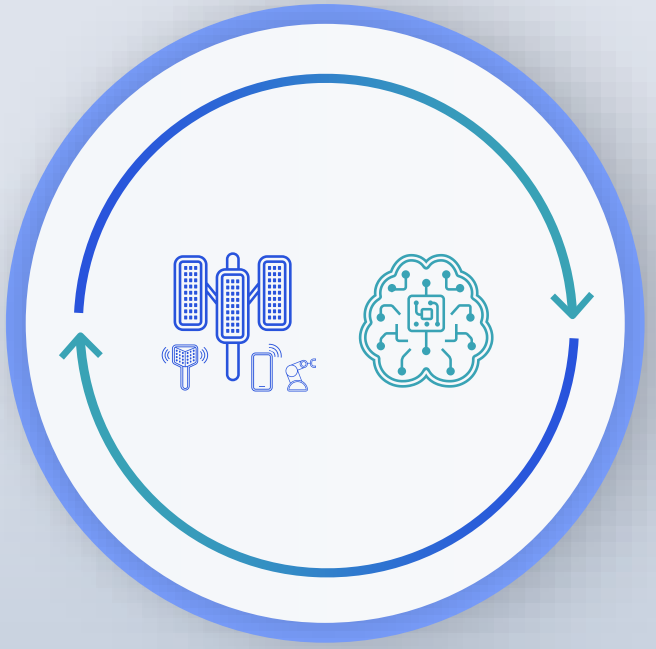
strengths

Design driven by tractable mathematical models

Interpretable solutions

Good generalization under different deployment conditions

Simple model adaptation



AI

strengths

Design with real world priors, fast and flexible models

Accurate prediction in complex tasks

Accurate modeling of generative process

Sensing and perception

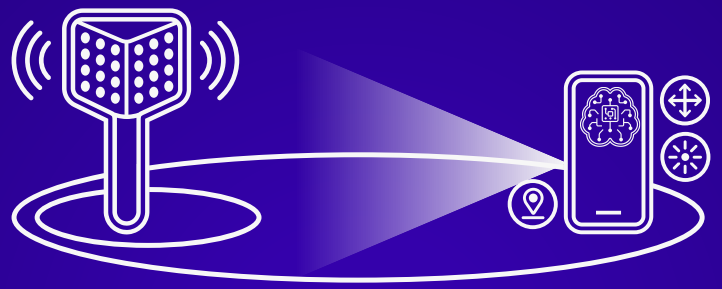
Wireless and AI have complementary strengths

Snapdragon
X75 5G modem-RF

AI hardware acceleration for superior 5G performance

Qualcomm® 5G AI Processor Gen 2 with dedicated tensor accelerator

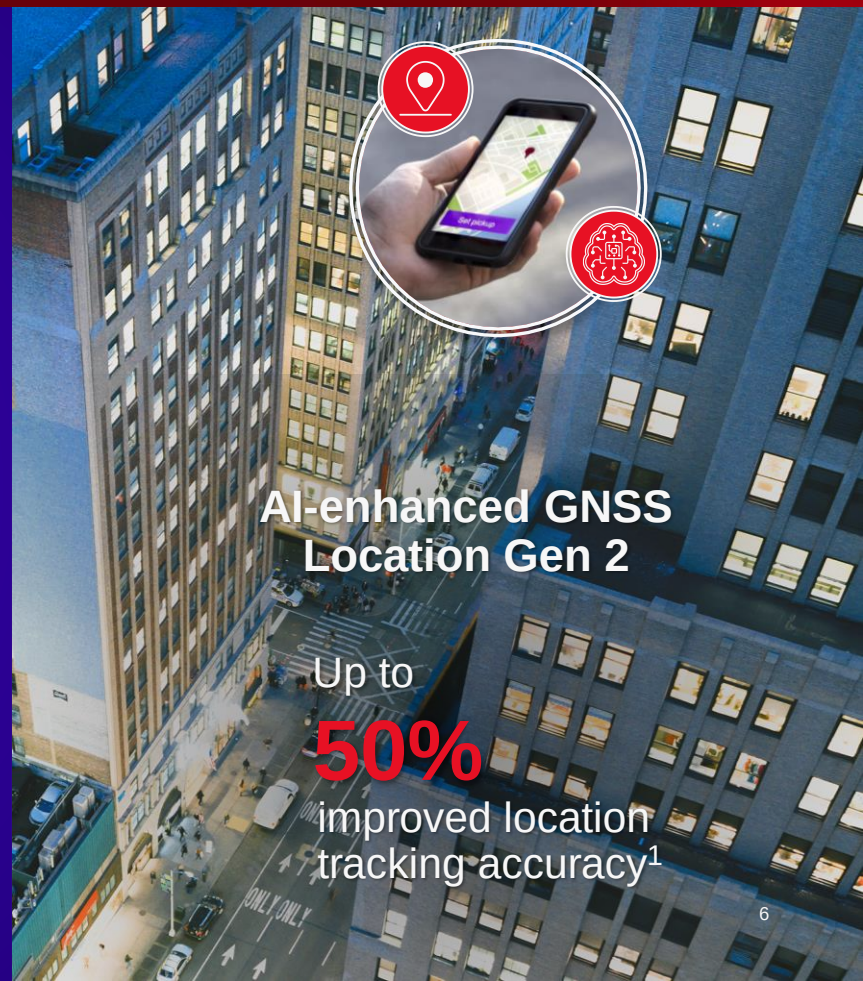
2.5X improved AI performance*



AI-based mmWave beam management

1st sensor-modem-RF fusion solution for mmWave beam processing

Up to 25% higher received power* for increased mmWave robustness



AI-enhanced GNSS Location Gen 2

Up to **50%** improved location tracking accuracy[1]

* Compared to Snapdragon X70 Modem-RF System
[1] Compared to non-AI-based location tracking; Under typical GNSS-challenged dense urban canyon environment

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

Applying AI to solve difficult wireless challenges

Deep wireless domain knowledge is required to optimally use AI capabilities

Our AI research areas to advance wireless communication

5G + AI

Signal intelligence, baseband and medium access

ML-based channel feedback
Channel estimation & pilot optimization
MIMO detection
Link prediction & adaptation
Beam management and optimization
Spectrum sensing and sharing
Radio resource scheduling

Network intelligence and system optimization

Coverage and capacity optimization
Traffic and mobility prediction
Energy saving
Cooperative edge caching
Content-aware X-layer optimization
Enhanced personalized security
TCP optimization

Device intelligence and optimization

Digital front-end optimization
Antenna and RF optimization
Full duplex
Battery saving
Reflective intelligent surface

Vertical intelligence and other capabilities

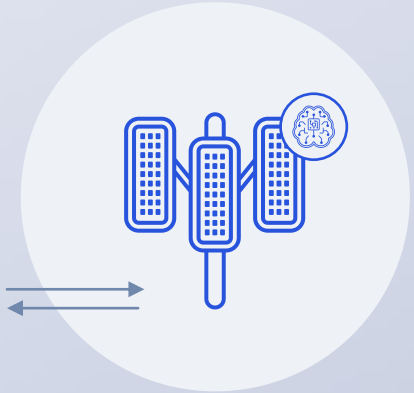
High-precision positioning
Environmental sensing
Contextual awareness
Sensor fusion
Vehicular communication

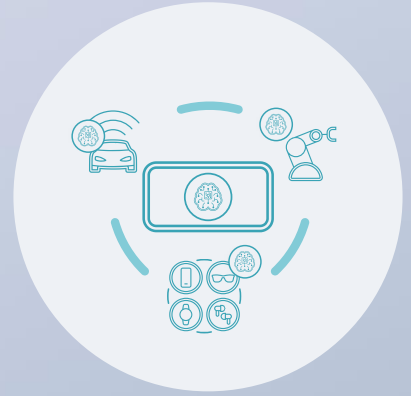
AI plays an essential role in every part of the cellular system

Transforming how it evolves going forward

End-to-end AI working together to realize system gains

AI = the common ingredient of wireless innovations





Optimized distributed clouds

Fully autonomous networks

Predictive and preventive optimization

Reduced network loading

Intelligent disaggregated network

RAN[1] intelligent controller

Interference coordination

Massive MIMO scheduler / CoMP[2]

AI-enabled air interface design

For coding, waveform, multiple access

Dynamic air interface adaptation (signaling, measurements, and feedback)

Joint training, model sharing, and distributed inference across networks and devices

Smart edge devices

Beam management (e.g., prediction)

CSF[3] computation

ML-based positioning and sensing

Efficient network planning | **Continuous operational optimization** | **Enhanced service experience**

[1] Radio Access Network; [2] Coordinated Multi-Point; [3] Channel State Feedback

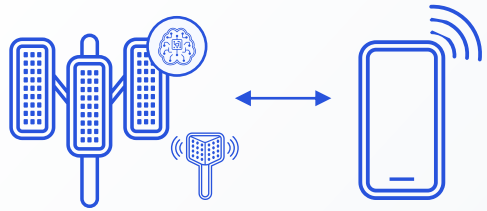
Machine learning can offer continuous wireless enhancements

AI-native air interface design can enable continual system improvements in between major 3GPP releases through self-learning

CURRENT 3GPP RELEASE PROCESS

Release X — Approximately 1.5-year cycle — Release X+1

No standardized improvement during nominal
Work/study item phase towards subsequent release



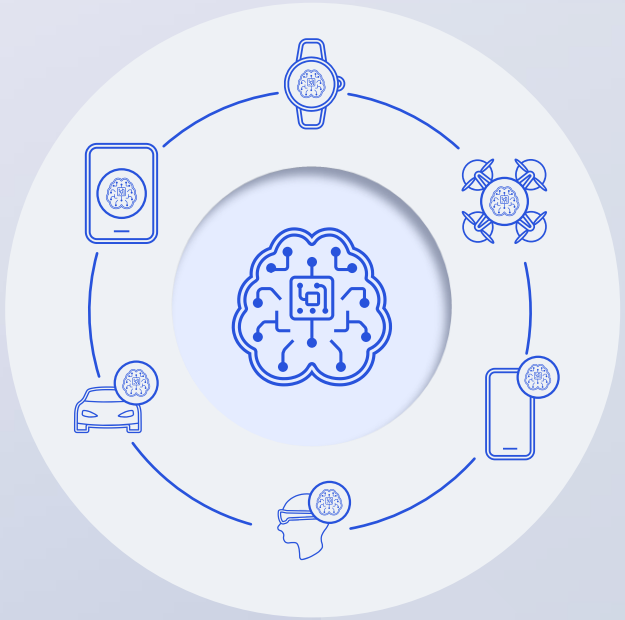
Data-driven communication and network design

Data-driven system configuration provides end-to-end optimizations

Dynamic parameter adaptation based on fast machine learning algorithms

Neural network system design can customize to given wireless environment

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 and 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

AI enables intelligent 5G network management

Enhanced service quality

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

Higher network efficiency

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

Simplified deployment

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

Improved network security

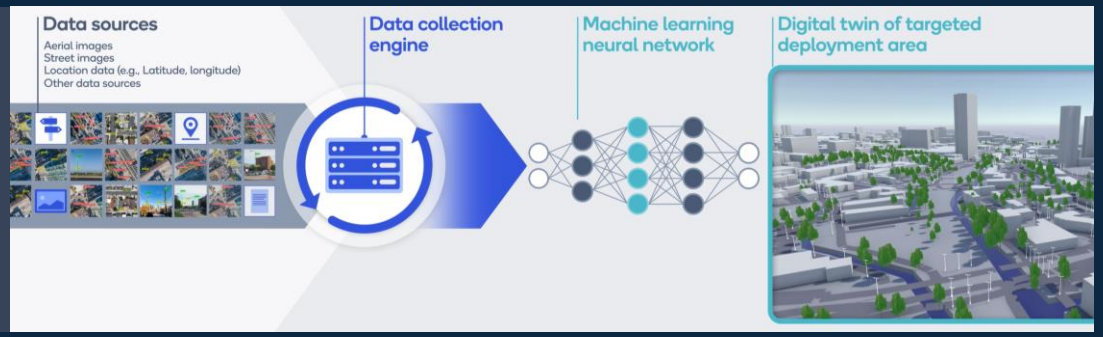
More effective detection and defense against malicious attacks by analyzing a massive quantity of data

Intelligent 5G mmWave network planning

Creating the digital twin

Fusing a variety of GIS and image sources via ML to arrive at such digital twin, in our example, of Manchester, England

Making use of readily available image sources and robust object detection/localization methods to create a representation with consistent locations and dimensions (e.g., poles, buildings, foliage)



Planning for network deployment

Running propagation modeling with pertinent RF parameters and an efficient optimization engine that identifies optimal placement of base stations to achieve a target throughput or coverage objective

Delivering enhanced user experience

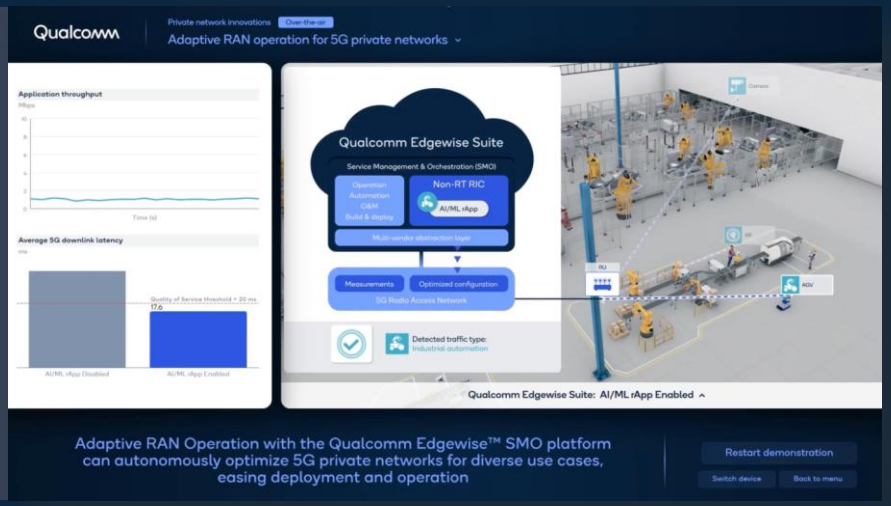
Showcasing exceptional performance (e.g., downlink user throughput) on several walk paths (e.g., from the train station to the Manchester United football stadium

Intelligent network optimization for 5G private networks

Adaptive RAN operation for 5G private networks

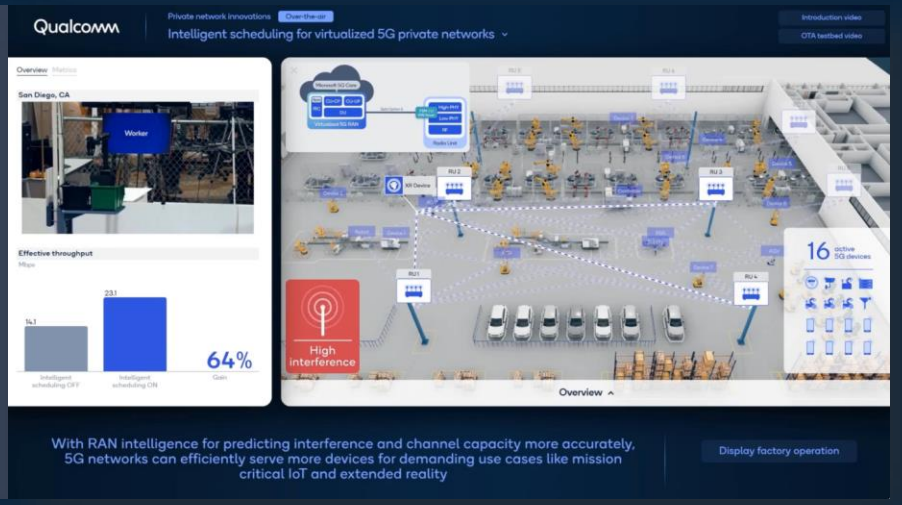
Using our indoor IIoT OTA testbed with Qualcomm Edgewise™ Suite, featuring a SMO platform with a RIC that runs adaptive operations

Highlighting how RAN operation can be optimized for different traffic profiles (e.g., AGV, XR) to meet stringent QoS requirements (e.g., data throughput, latency)



Intelligent scheduling for virtualized 5G private networks

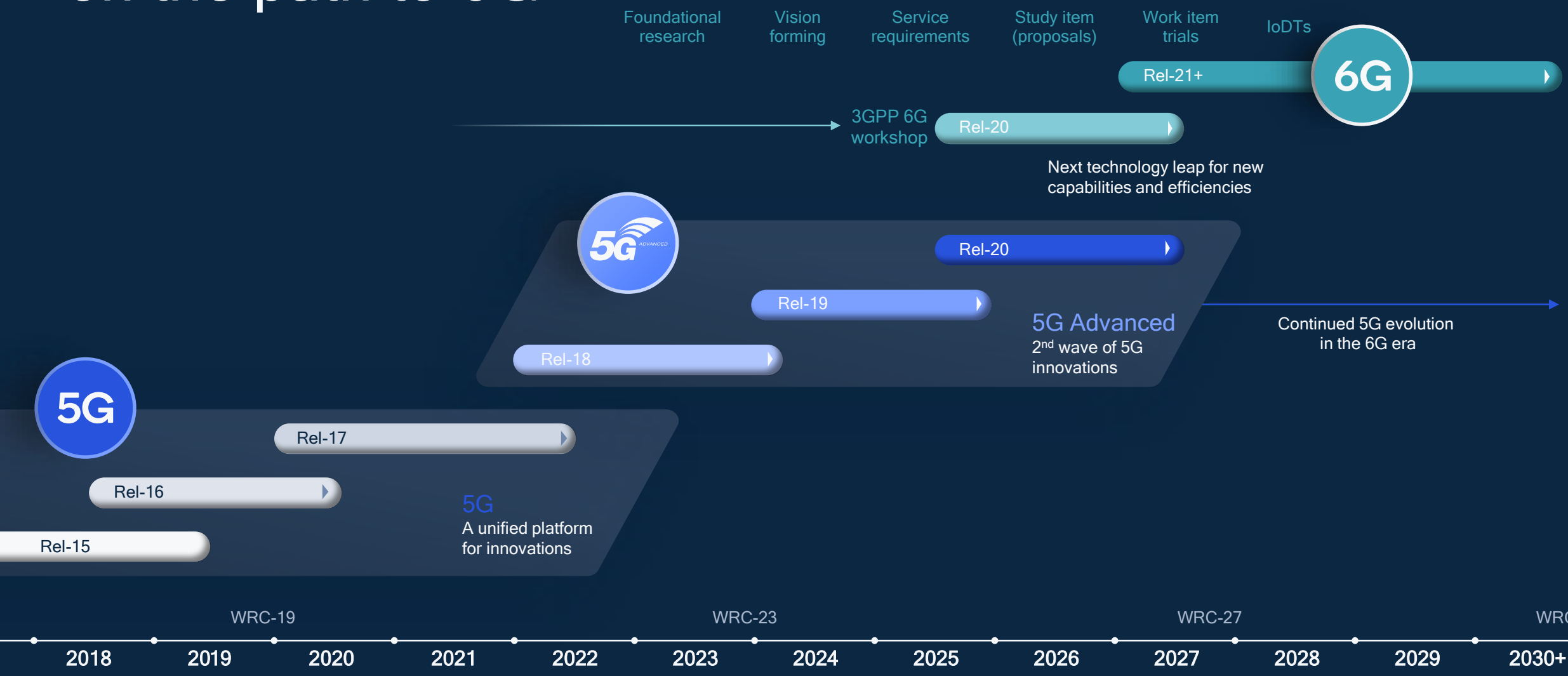
Utilizing our indoor IIoT OTA testbed to showcase a virtualized 5G private network with 3GPP & O-RAN disaggregation, benefits of ML-based interference prediction, and deployment scalability



5G Advanced starts the era of wireless AI

Focusing on key use cases that can bring immediate benefits

5G Advanced on the path to 6G



Completed the groundwork for wireless AI – ready for 5G Advanced

3GPP Release 16

Data collection for network performance enhancements

Enhanced Network Automation (eNA)

Expanding NWDAF[1] for data collection and exposure from/to 5G core NF, AF, OAM[2], data repositories

Minimization of Drive Testing (MDT)

Specifying features for identified use cases, including, QoS verification, coverage optimization, location reporting, sensor data collection

Self Organizing Network (SON)

Specifying device reporting needed to enhance network configurations and inter-node information exchange

3GPP Release 17

Expanding 5G system support for wireless AI

Enhancements for 5G network interfaces

Facilitating machine learning procedures, e.g., model training and inference, as well as actions to enforce model inference output

Augmented network & device data collection

Supporting targeted applications, e.g., energy saving, load balancing, mobility management, operations enhancements, expanded use case[3]

Support for over-the-top AI services

Introducing new QoS definitions that are tailored for machine learning model delivery over 5G

5G ADVANCED

[1] Network Data Analytics Function; [2] Network Function, Application Function, Operations Administration and Maintenance; [3] Such as multicast, broadcast, V2X, sidelink, multi-SIM, RAN slicing, and more

3GPP Release 18 sets off the 5G Advanced Evolution

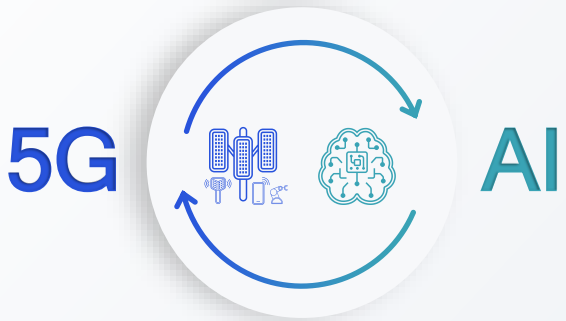
Strengthen the end-to-end 5G system foundation

- Advanced DL/UL MIMO
- Enhanced mobility
- Mobile IAB, smart repeater
- Evolved duplexing
- AI/ML air interface
- Green networks

Proliferate 5G to virtually all devices and use cases

- Boundless extended reality
- NR-Light (RedCap) evolution
- Expanded sidelink
- Expanded positioning
- Drones & expanded satellites comm.
- Multicast & other enhancements

5G ADVANCED



Working together across the connected intelligent edge

5G NR Release 18 Scope

AI/ML-enabled air interface design

Use cases

Enhanced CSI[2] feedback, beam management, and positioning accuracy

AI/ML models

Collaboration models, life cycle management, and algorithms

Evaluation methodology

Existing 3GPP framework and field data to assess performance and identify KPIs

Impact assessment

Spec changes needed to support identified use cases, covering multiple aspects

AI/ML framework for next-gen RAN

Network optimization

Data collection and signaling support for energy saving, load balancing, mobility optimization

Future study

New use cases (e.g., AI/ML for slicing, QoE[1]), network functionality and interface procedures

5G Advanced evolution will expand wireless ML to the end-to-end system across RAN, device, and air interface

Network architecture enhancements

ML to run over different HW/SW and future RAN function split to improve flexibility and efficiency

AI/ML procedure enhancements

Model management, training (e.g., federated and reinforced learning), and inference

Data management enhancements

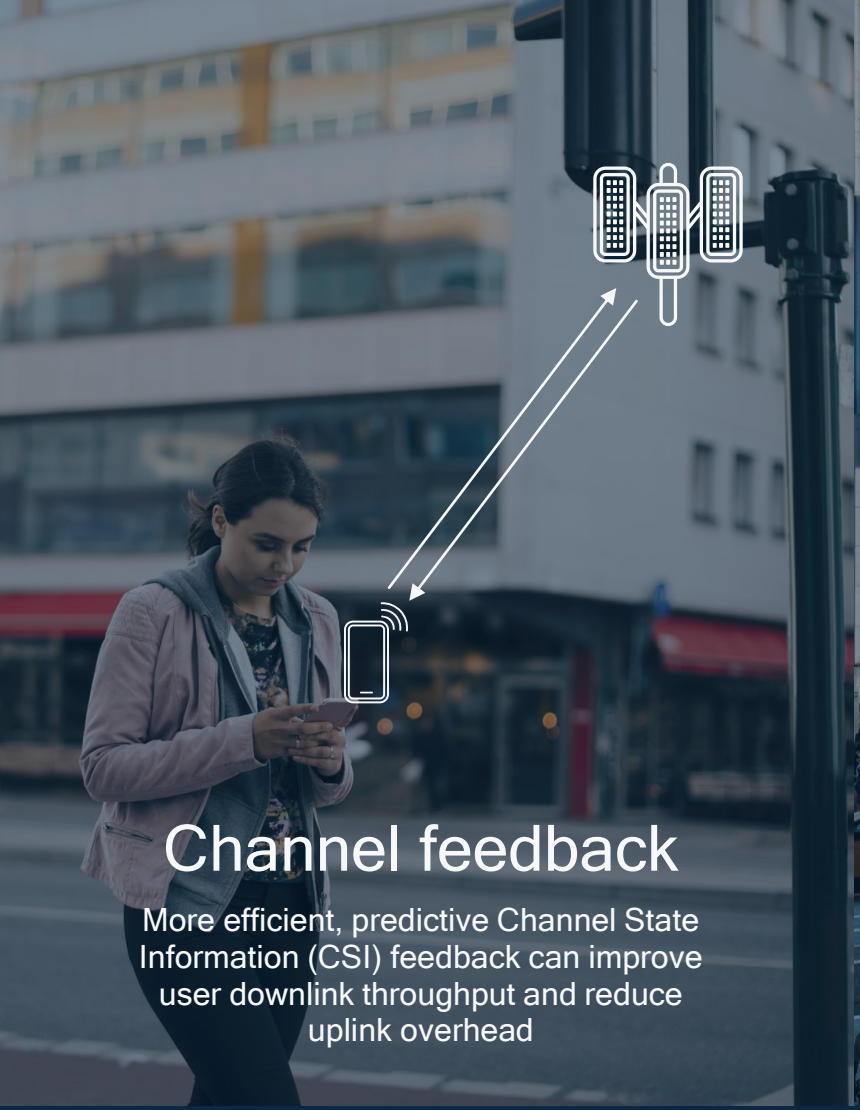
ML data storage/access, data registration/discovery, and data request/subscription

New and expanded use cases

Traffic/mobility prediction, optimized coverage/capacity, massive MIMO, SON, CSI, beam management, …

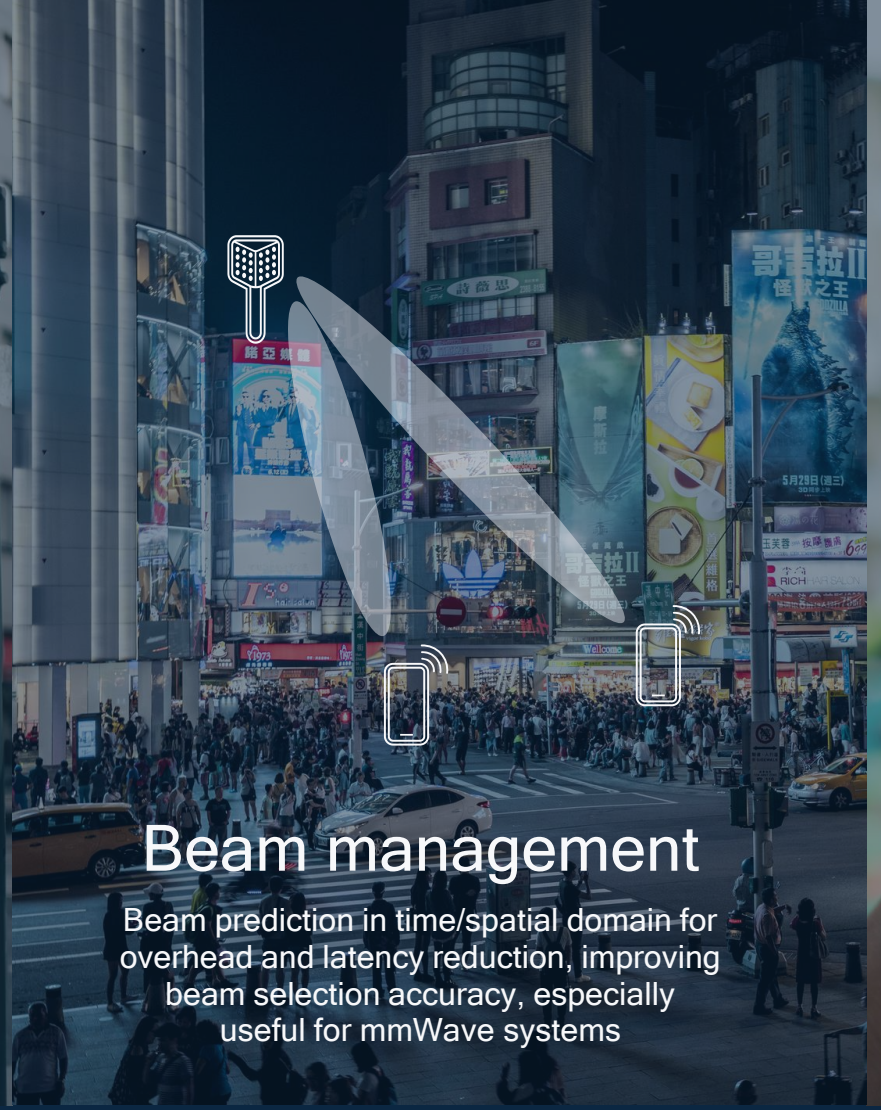
Source: RP-213599 (AI/ML for NR Air Interface), RP-213602 (AI/ML for NG-RAN)

[1] Quality of Experience; [2] Channel State Information



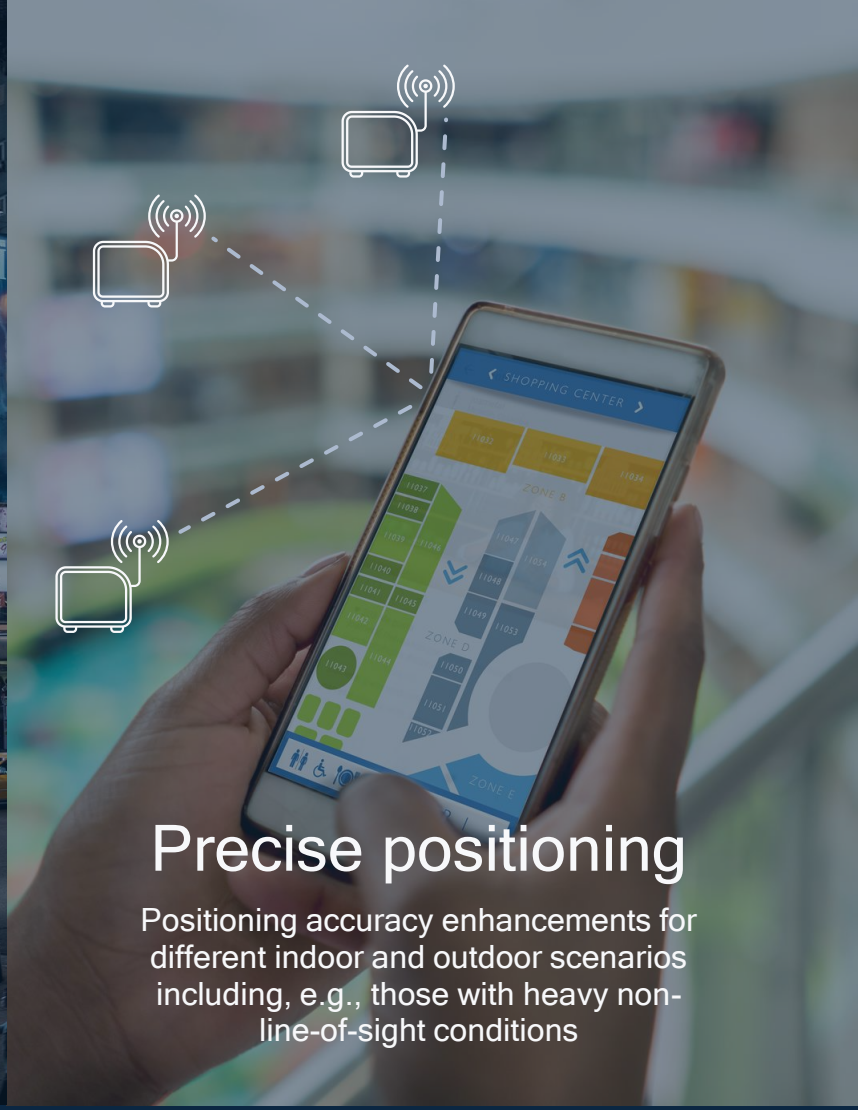
Channel feedback

More efficient, predictive Channel State Information (CSI) feedback can improve user downlink throughput and reduce uplink overhead



Beam management

Beam prediction in time/spatial domain for overhead and latency reduction, improving beam selection accuracy, especially useful for mmWave systems



Precise positioning

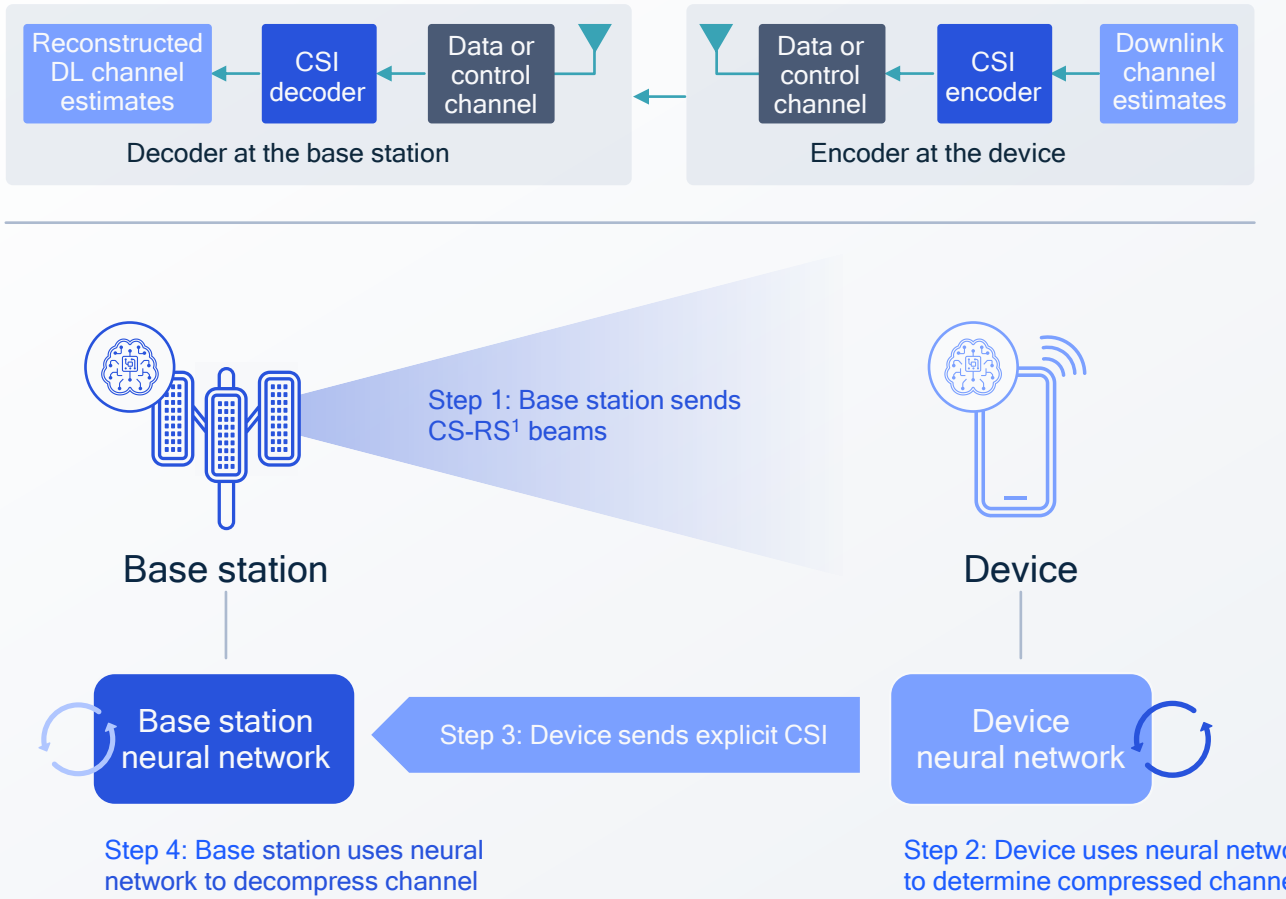
Positioning accuracy enhancements for different indoor and outdoor scenarios including, e.g., those with heavy non-line-of-sight conditions

Release18 focuses on three key wireless AI use cases

But many more potential use cases for the future

Cross-node machine learning based channel state information

Explicit channel feedback framework for CSI compression and prediction utilizing domain knowledge and neural networks



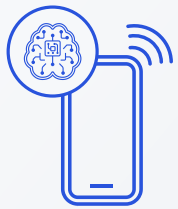
Improve system efficiency with neural network framework for CSI on non-linear encoding and decoding

More effective multi-user multiplexing minimizing interference

Customized, lower overhead feedback based on individual device

[1] Channel State Information Reference Signal

Using AI to improve channel state feedback



PAST DEMONSTRATIONS

MWC'21: System simulation

Simulated ML-based CSF in a wide-area 3.5 GHz massive MIMO environment, demonstrated improved user downlink throughput and reducing uplink overhead

MWC'22: OTA testing

Utilizing our indoor IIoT OTA testbed to showcase a virtualized 5G private network with 3GPP & O-RAN disaggregation, benefits of ML-based interference prediction, and deployment scalability

Qualcomm

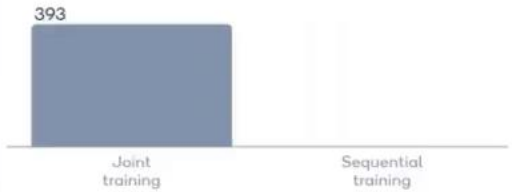
AI-native end-to-end communications | Real-time OTA

Multi-vendor cross-node ML-based channel state feedback

Overview | Throughput trace

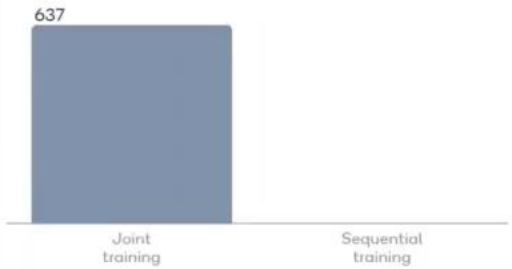
Device 1 - Median downlink throughput

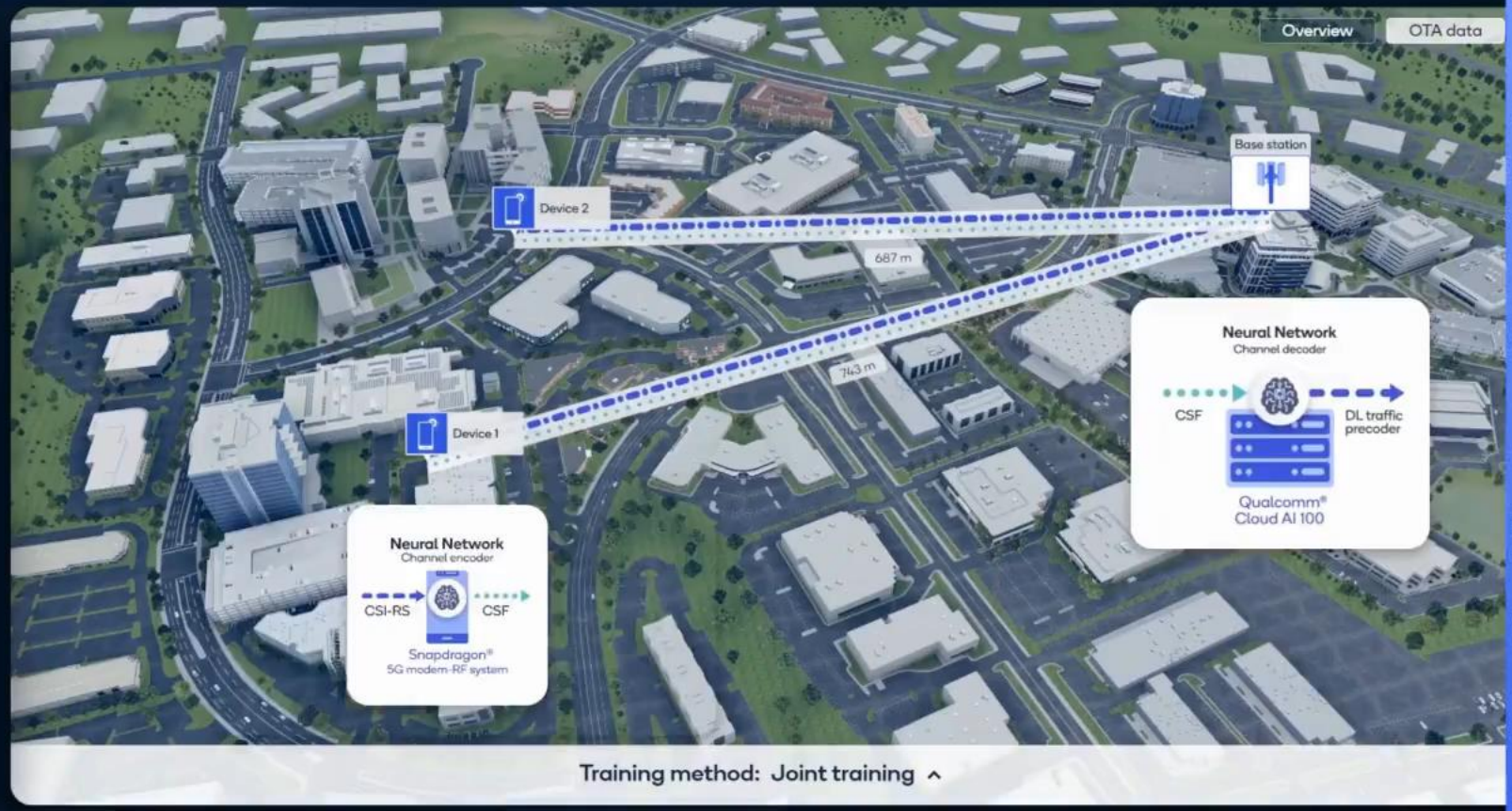
Mbps



Device 2 - Median downlink throughput

Mbps





Training method: Joint training

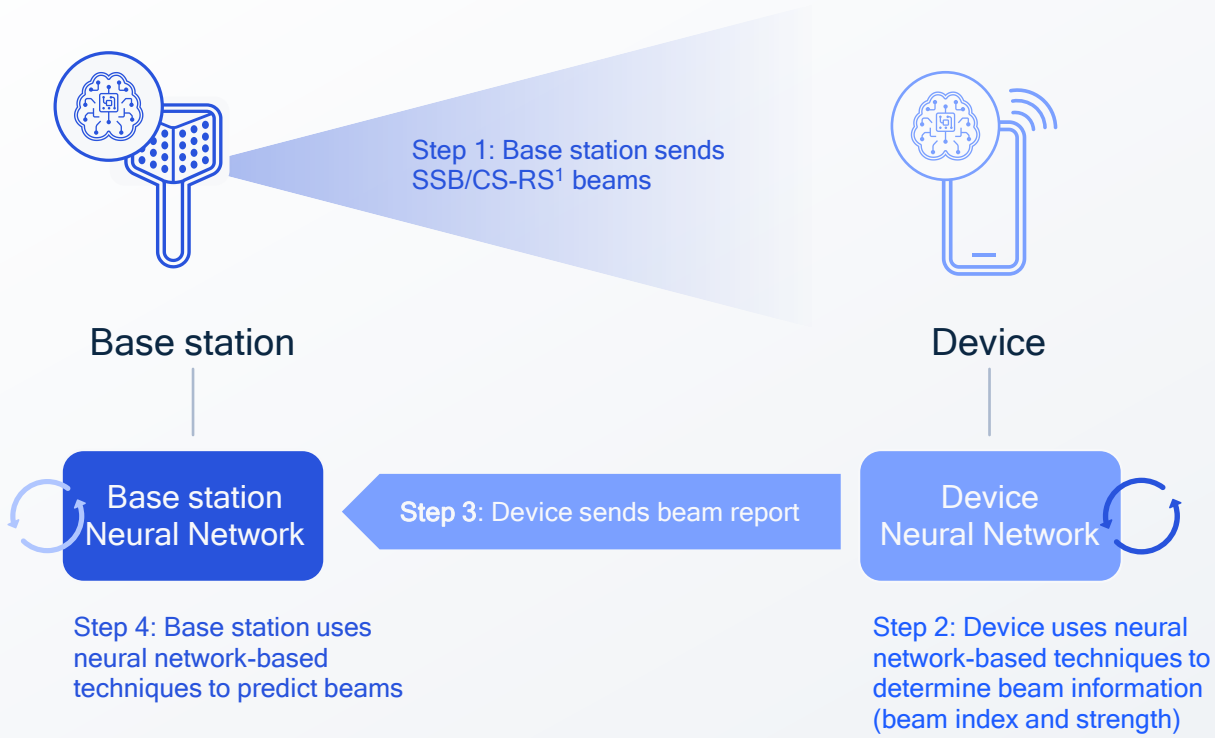
Joint training for neural network encoder and decoder requires knowledge of both neural network structures - not desirable for multi-vendor deployment

Next: Sequential training

Change training method

Machine learning based mmWave beam management

Beam prediction utilizing domain knowledge, location, velocity, other aspects of environmental and application awareness to improve robustness and throughput



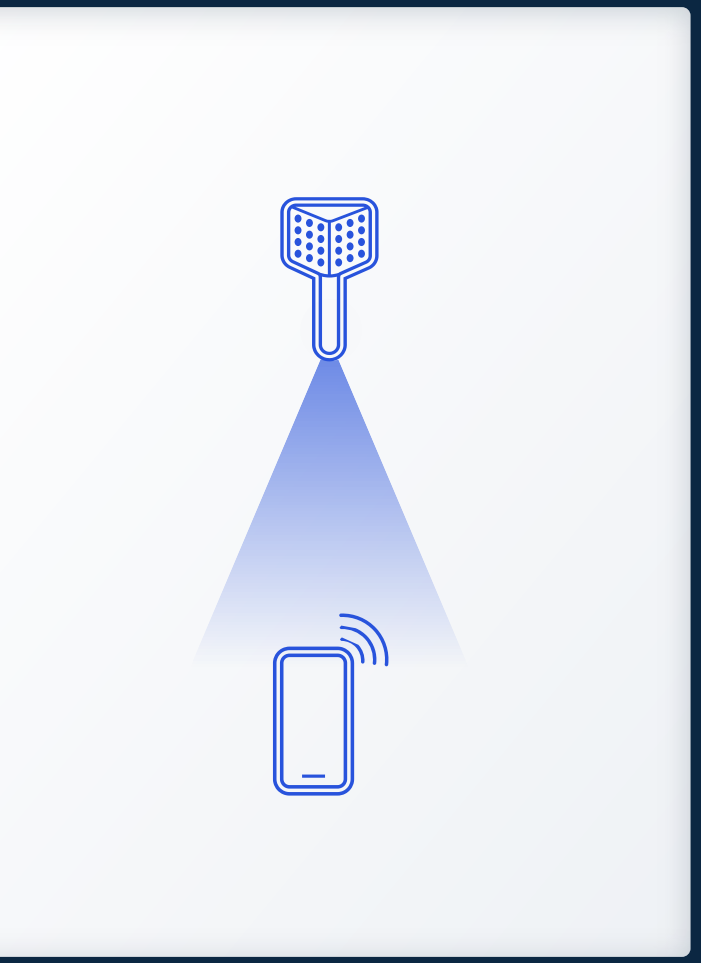
Reduced feedback overhead and transmission from device

Reduced transmission of reference signal from base station

Efficient beam tracking and beam search for higher performance

[1] Synchronization Signal Block / Channel State Information Reference Signal

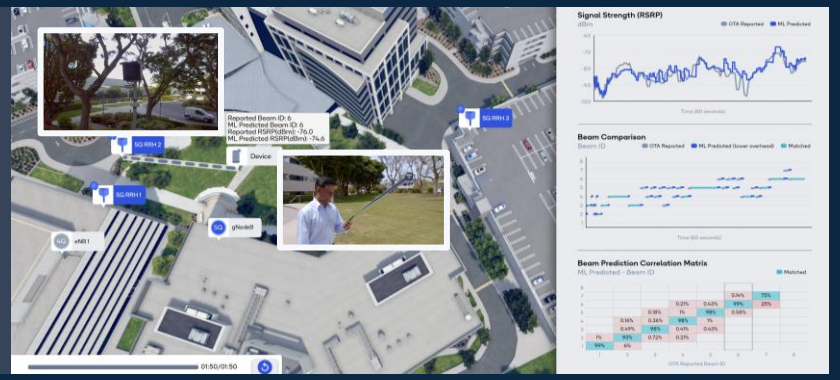
Using AI to enhance mmWave beam management



PAST DEMONSTRATIONS

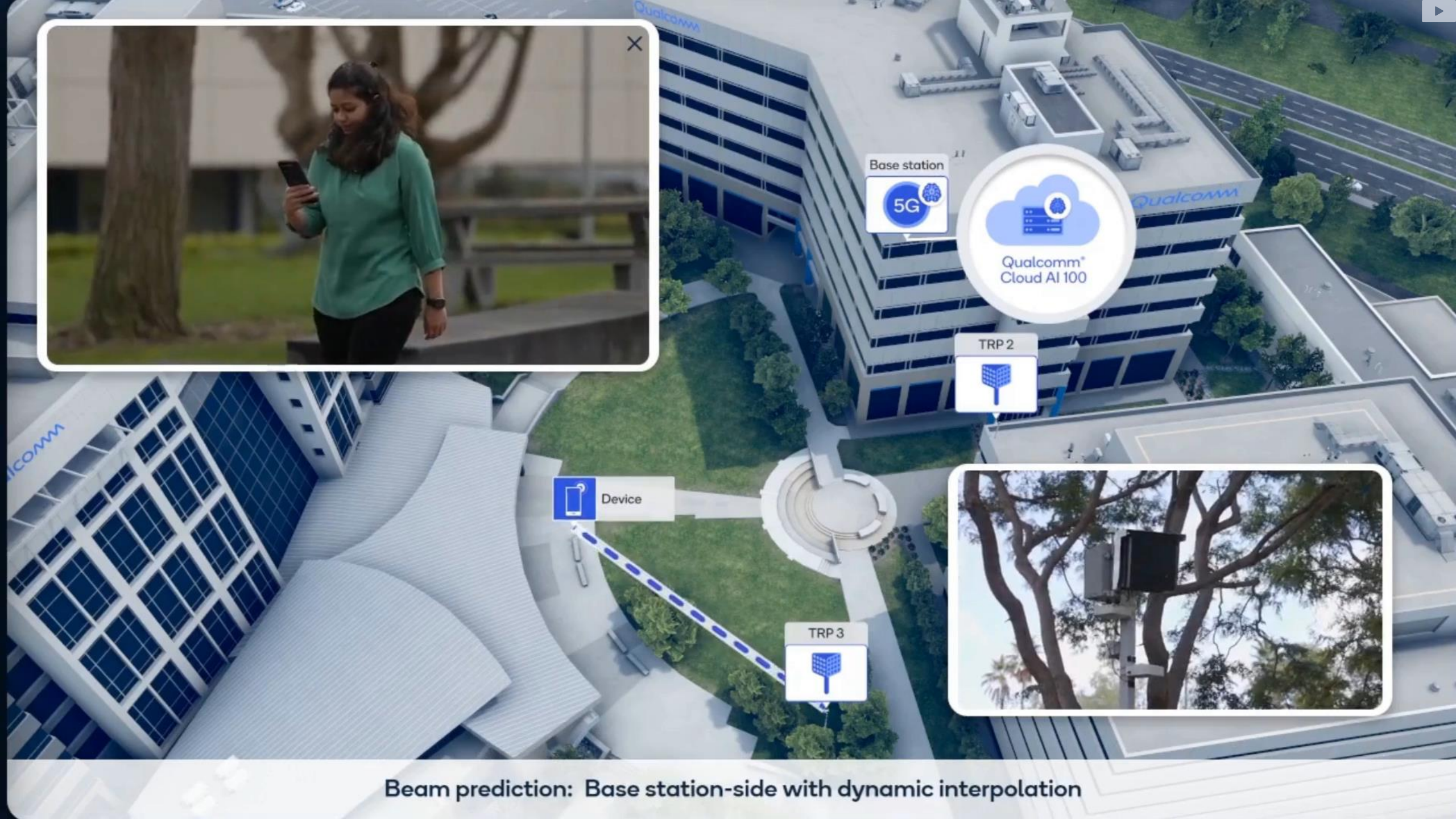
MWC’21: beam prediction

Implemented in our 28 GHz OTA test network machine learning based beam prediction that can increase system capacity and device battery life



MWC’22: gain quantification

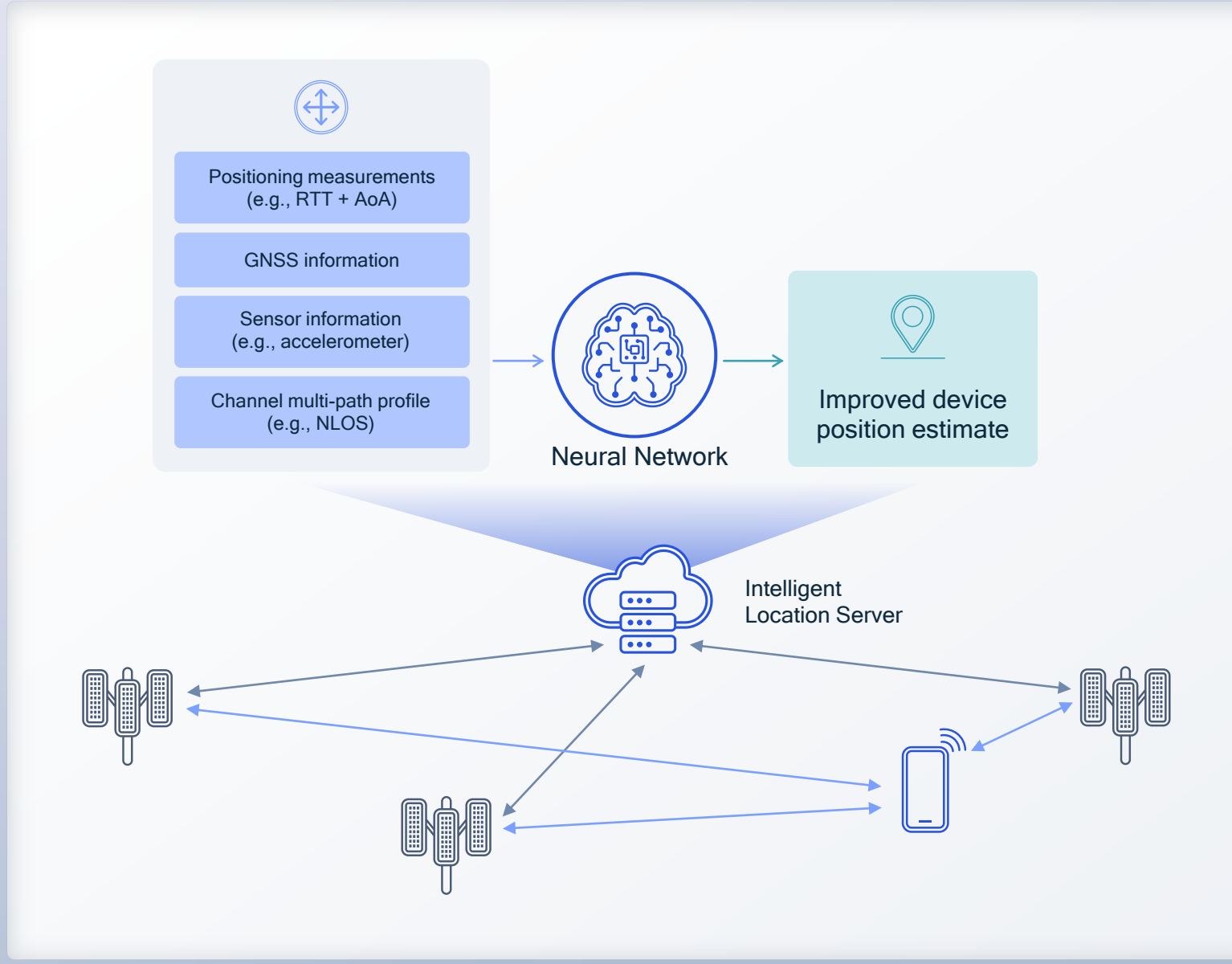
Expanded our OTA prototype of base station-side beam prediction, quantitatively showcased improved usable capacity and extended device battery life, utilizing Qualcomm® Cloud AI 100 platform and Snapdragon® Modem-RF system



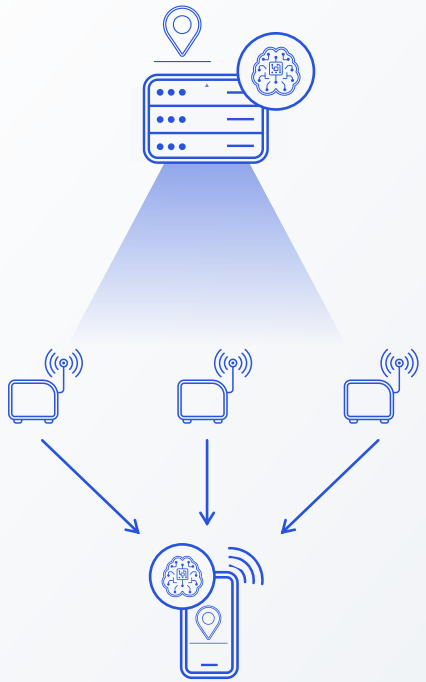
Beam prediction: Base station-side with dynamic interpolation

AI/ML for enhanced positioning and RF sensing performance

Supplemented by various assisting information, such as GNSS, multi-path profiles, and other sensors



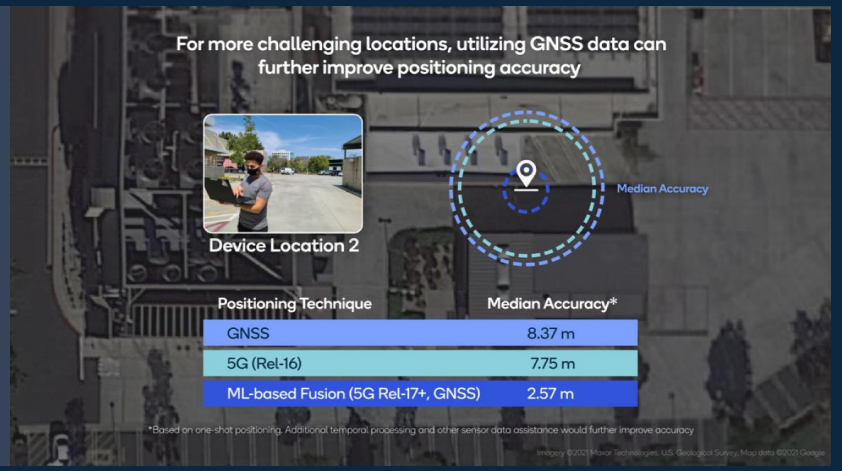
Using AI to enhance positioning and RF sensing accuracy



PAST DEMONSTRATIONS

MWC'21: 5G RTT+AoA

Implemented in our 3.5 GHz OTA test network machine learning based wide-area 5G positioning using RTT+AoA with GNSS sensor fusion



MWC'21: Wi-Fi indoor

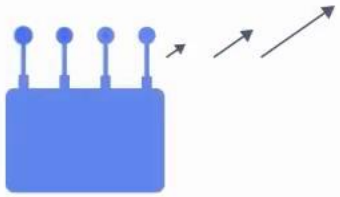
Leveraged unsupervised/weakly supervised learning – also applicable to 5G RF sensing (e.g., for positioning, motion and gesture detection)

Qualcomm | Intelligent Industrial Positioning

Over-the-air

Multipath

Numerous indoor surfaces generate multiple reflections



Obstruction

Non-line of sight (NLOS)

Variety of indoor obstructions block line of sight

Further evolving wireless AI in 5G Advanced

Potential work items for Release 19 and beyond

Realizing the key use cases studied in Release 18

Potential Release 19 work items

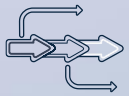
Two-sided CSI[1] with spatial/frequency-domain compression and time-domain prediction

Beam management with spatial-domain and time-domain prediction

Direct and AI-assisted positioning accuracy enhancements

Specifying protocol and signaling to enable wireless AI

Device capability signaling, including model and functionality identification



Life Cycle Management including Model ID and AI/ML functionality related procedures: e.g., configuration, activation, deactivation, switching, fallback

Data/measurement reporting as input to inference

Also exploring additional use cases on the path to 6G…

Such as inter-cell beam management, mobility management, new learning framework, and more…

[1] Channel State Information

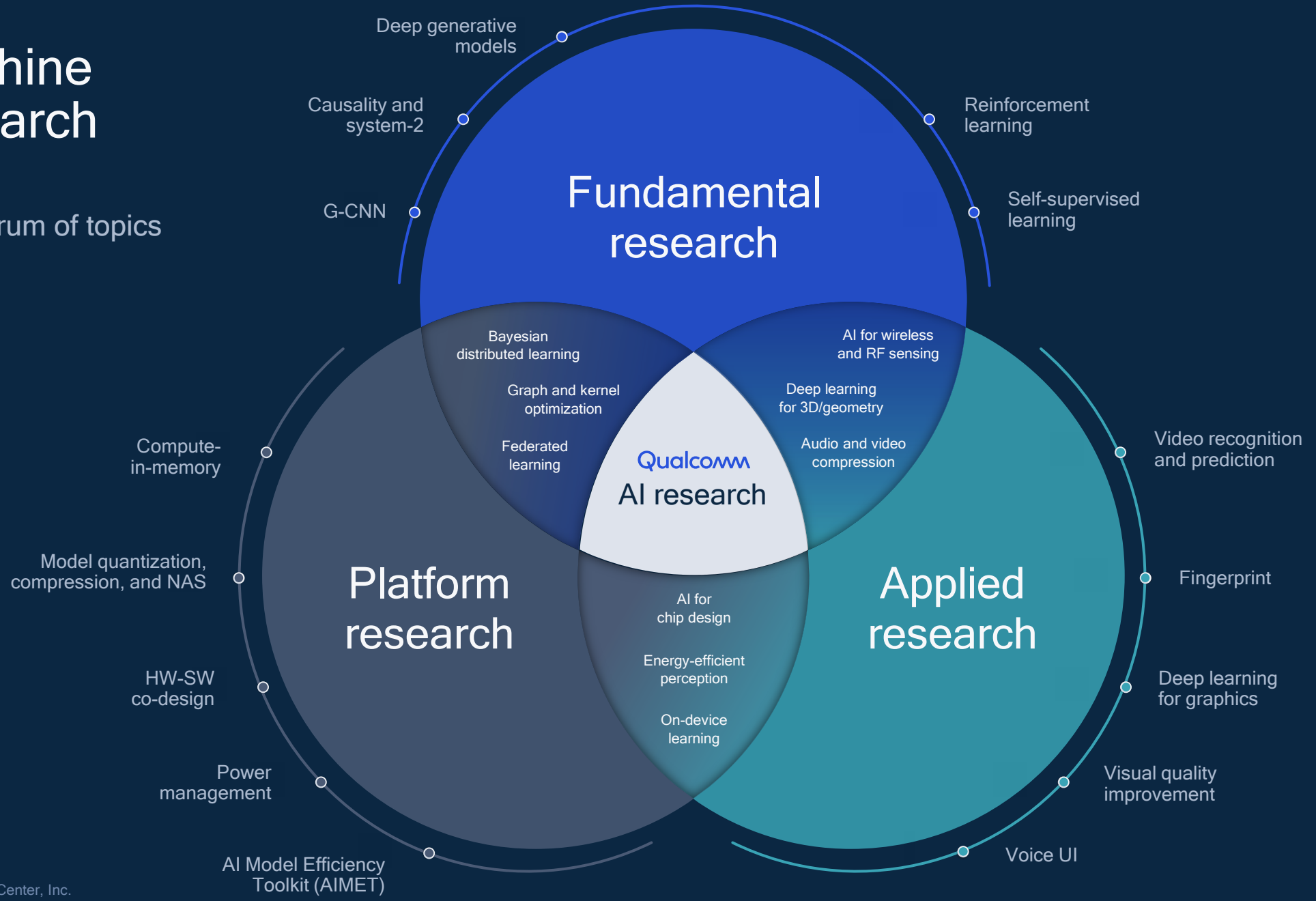
We are driving foundational wireless AI innovations leading towards an AI-native 6G

Leading wireless research for 5G Advanced and 6G

1. Drive cutting-edge foundational technology advancements and system innovations
2. Validate system designs via comprehensive simulations and over-the-air prototypes
3. Lead global technology standardization with broad ecosystem collaborations
4. Prepare for commercialization at scale in future products and services

Leading machine learning research for edge AI

across the entire spectrum of topics



AIMET is a product of Qualcomm Innovation Center, Inc.

Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc.

Key longer-term research vectors

enabling the path towards 6G

Key longer-term research vectors

enabling the path towards 6G

AI-native E2E communications

Data-driven communication and network design, with joint training, model sharing and distributed inference across networks and devices

Scalable network architecture

Disaggregation and virtualization at the connected intelligent edge, use of advanced topologies to address growing demand

Expanding into new spectrum bands

Expanding to THz, wide-area expansion to higher bands, new spectrum sharing paradigm, dynamic coordination with environmental awareness

Air interface innovations

Evolution of duplexing schemes, Giga-MIMO, mmWave evolution, reconfigurable intelligent surfaces, non-terrestrial communications, waveform/coding for MHz to THz, system energy efficiency

Merging of worlds

Physical, digital, virtual, immersive interactions taking human augmentation to next level via ubiquitous, low-power joint communication and sensing

Communications resiliency

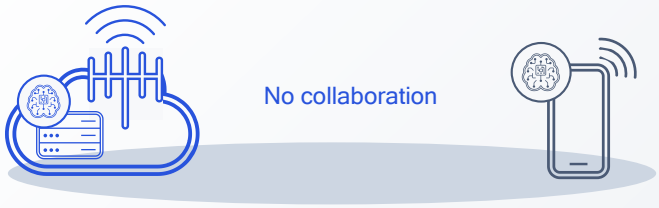
Multifaceted trust and configurable security, post quantum security, robust networks tolerant to failures and attacks

Evolving towards an AI-native wireless system

Multiple wireless AI/ML training and inference scenarios

Overlay AI/ML

INDEPENDENTLY AT THE DEVICE OR NETWORK



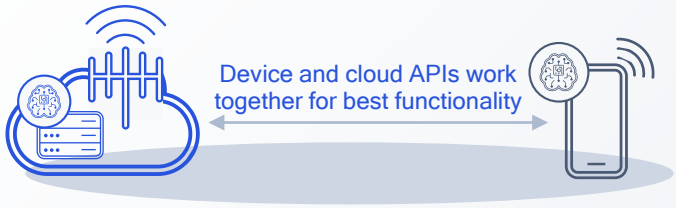
ML operates independently at the device and network as an optimization of existing functions

Proprietary ML procedures including model development and management

Proprietary and standardized data collection used as input to training

Cross-node AI/ML

COORDINATED BETWEEN DEVICE AND NETWORK



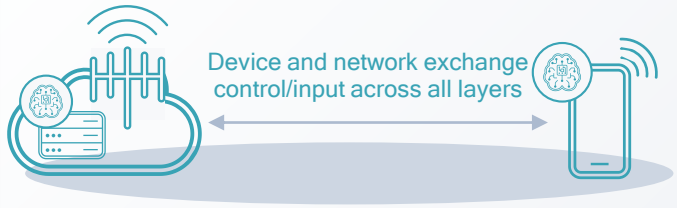
ML operates in a coordinated manner between the device and network

Proprietary and standardized ML procedures including model development and management

Further data collection used as input to training as well as monitoring

Native AI/ML

AT ALL DEVICE AND NETWORK LAYERS



ML operates autonomously between the device and network across all protocols and layers

Integrated ML procedures across to train performance and adapt to different environments

Data fusion for integrated dynamic ML lifecycle management

5G

5G ADVANCED

6G

Generative AI

Creating new and original content for innovative consumer and enterprise use cases
(e.g., text, images, video, audio, and other data)

SOME GOOD EXAMPLES…

Large language models for text generation
(e.g., ChatGPT, BARD)

Diffusion models for image generation
(e.g., Stable Diffusion, DALL-E2)

Stable Diffusion images generated on an Android phone powered by Snapdragon, with the prompt: "Super cute fluffy cat warrior in armor, photorealistic, 4K, ultra detailed, vray rendering."

World's first on-device demo of Stable Diffusion running on an Android phone

1B+ parameter generative AI model runs efficiently and interactively

Full-stack AI optimization to achieve sub-15 second latency for 20 inference steps

Enhanced privacy, security, reliability, and cost with on-device processing

Fast time-to-market enabled by Qualcomm AI Research and Qualcomm® AI Stack

Applying generative modeling to improve wireless communications system design

Wide applicability for Generative Modeling

Real-time use cases for air interface

Propagation channel
Beam management
Interference prediction
Scheduler optimization
Traffic source
Mobility enhancement

Link / system simulation

Deployment optimization

Positioning and sensing

Network and device optimization

Others…

Application examples

Channel rendering

Text description of image or semantic map

Diffusion model
(To generate channel information)

Channel sampled from a conditional distribution
P (h | conditioning from inputs, location)

Network / device prediction

Context in text, e.g., history of device reports and base station responses

Large language model
To learn link, beam, protocol languages

Next action for base station and/or device, sampled from a conditional distribution
P (next action | conditioning from inputs)

Our on-going wireless generative AI research areas

3D mapping and material learning

Foundation models (e.g., link and protocol level use cases, beam prediction, and others)

Neural channel rendering (e.g., map-based, ray tracer augmented, site-specific, and others)

Customized ML-based stochastic channel

Neural surrogate for base station scheduler and applications traffic

And others…

Driving the 5G Advanced technology evolution into 6G

Next technology leap for new capabilities and efficiencies

Foundational research, vision, requirements, etc.

6G

Rel-21 and beyond
New innovation platform

Historically 10 years between generations

A key enabler of the connected intelligent edge

Technology foundation for the next generation

5G

Rel-15
eMBB focus

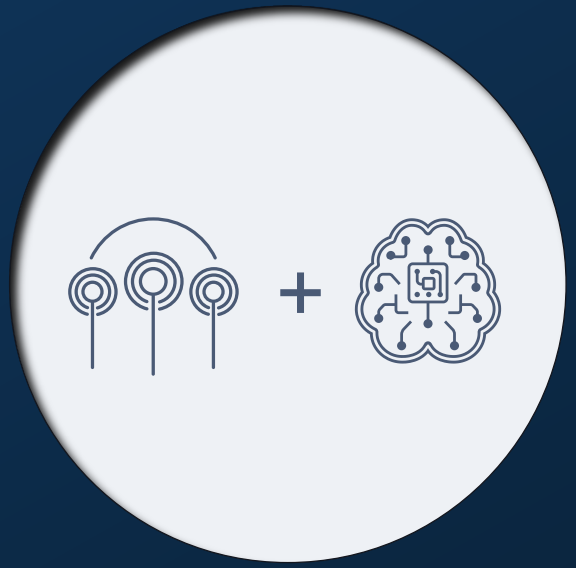
Rel-16 and 17
expanding to new industries

5G ADVANCED

Rel-18, 19. 20 and beyond
Continued 5G evolution and proliferation

Strong 5G momentum sets stage for global expansion


AI-enabled end-to-end communication
Expanding into new spectrum bands
Cellular air interface innovations
Precise positioning and RF sensing
Powering the metaverse
Private network innovations
Wide-area IoT evolution
Advanced automotive connectivity
Introduction
Introduction
Introduction
Introduction
Introduction
Introduction
Introduction
Introduction
Demo 1: Advanced ML-based mmWave beam management
Demo 1: Giga-MIMO for wide-area coverage in 7 to 16 GHz
Demo 1: Path to full duplex
Demo 1: Precise positioning everywhere
Demo 1: Boundless AR with dynamic distributed compute
Demo 1: Adaptive RAN operation for 5G PN
Demo 1: 5G IoT coverage extension with device mesh
Demo: Cloud-based VRU safety
Demo 2: Multi-vendor cross-node ML-based CSF
Demo 2: Sub-terahertz communication in 100+ GHz
Demo 2: Green networks with super-QAM
Demo 2: High-resolution sensing
Demo 2: 5G API for immersive applications
Demo 2: Intelligent scheduling for virtualized 5G PN
Demo 2: Narrowband positioning for 5G IoT
Watch all demos on YouTube
Building a stronger, more capable wireless system foundation
Qualcomm
Demo 3: Perception-assisted 5G for enhanced XR
Demo 3: Multi-AP joint transmission for Wi-Fi
Taking 5G to new, more diverse verticals and use cases



Towards an AI-native communications system design

A closer look at how AI can meaningfully improve wireless performance starting with 5G Advanced

The rise of AI brings a unique opportunity to revolutionize the future of wireless technology

AI plays an essential role in every part of the cellular system, transforming the way it evolves going forward

5G Advanced starts the era of wireless AI, focusing on use cases that can bring immediate benefits

Qualcomm

Qualcomm is driving foundational wireless AI innovations leading towards an AI-native 6G

Thank you

Qualcomm

Follow us on:

For more information, visit us at:

qualcomm.com & qualcomm.com/blog