Qualcomm

@QCOMResearch

San Diego, CA

September 10, 2024

How can we shape the future of mobile connectivity with 6G?

Fusing connectivity and AI to deliver bespoke experiences that evolve

Dr. Tingfang Ji
Vice President, Engineering
Qualcomm Technologies, Inc.

Dr. Gavin Horn
Senior Director, Engineering
Qualcomm Technologies, Inc.

Shaping the digital future

Powering the mobile revolution

Redefining connected processing

Enabling intelligent computing everywhere

Where are we in the cellular innovation cycle?

5G

Ramping volume and expanding to new use case

5G ADVANCED

Completing 1st standard —2nd phase of 5G innovations

6G

Aligning on vision, foundational research, timeline, requirements

INVENTION

Invent new technologies and e2e system architecture

VISION

Identify a problem or need; establish requirements

PROOF-OF-CONCEPT

Deliver end-to-end prototypes and impactful demonstrations

COMMERCIALIZATION

Engage with global network operators to deploy new features with standards-compliant infrastructure and devices

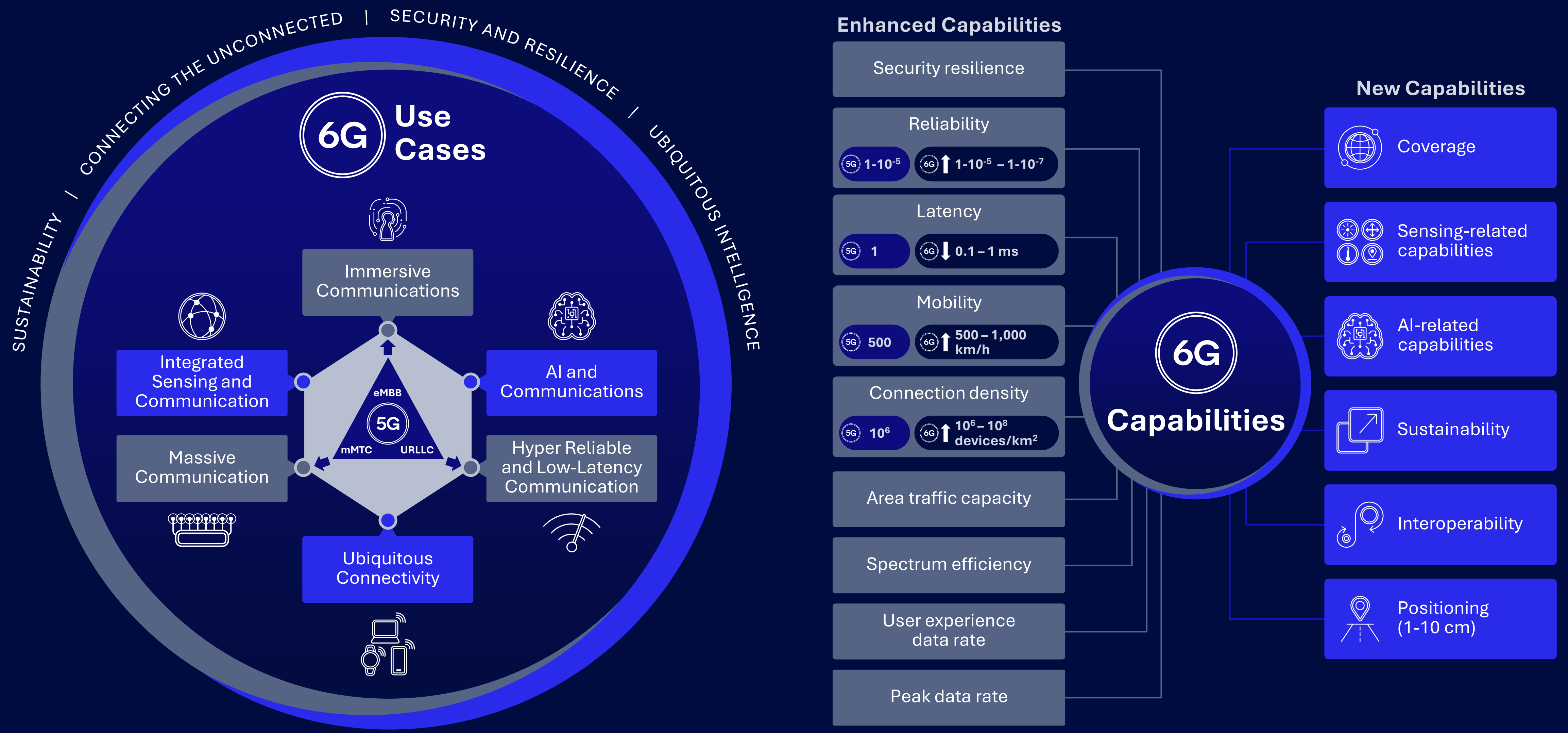
STANDARDIZATION

Drive e2e design with ecosystem and through standards process

TRIALS

Collaborate on OTA field trials that track 3GPP standardization and drive ecosystem towards rapid commercialization

6G vision from ITU-R — Usage scenarios and capabilities



Source: ITU-R's IMT towards 2030 and beyond (https://www.itu.int/en/ITU-R/study-groups/rsg5/rwp5d/imt-2030/Pages/default.aspx)

6G will support an unprecedented range of frequency bands

LOW BANDS
below 1 GHz (~20 MHz BW)

With 6G, lower frequencies with narrower bandwidths will provide even better long-range coverage.

MID BANDS
1 — 7 GHz (~100 MHZ BW)

Bandwidths up to ~100 MHz, shorter wavelengths, massive MIMO antennas and MU-MIMO enable high capacities

UPPER MID-BANDS
7 – 24 GHz (~500 MHz BW)

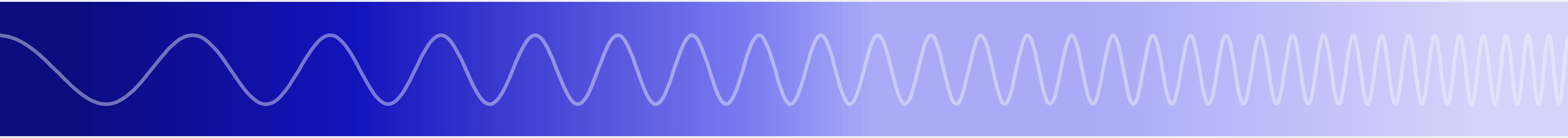
New 6G spectrum with bandwidths up to ~500 MHz bring additional wide-area capacity for communications and sensing

mmWAVE BANDS
24-71 GHz

6G spectrum bringing additional local-area capacity for communications and sensing

SUB-THZ
above 100 GHz

Higher frequencies with wide bandwidths provide excellent precision



Improving coverage and capacity in legacy bands and supporting new frequency bands for growth

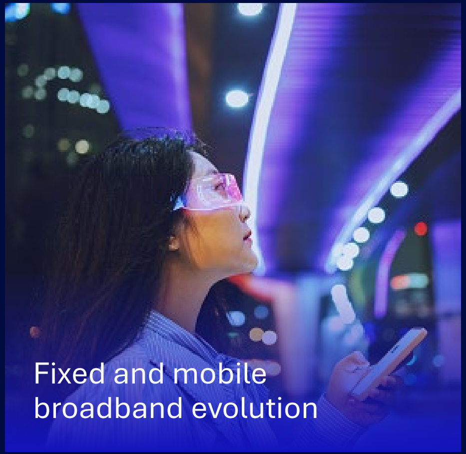
Leading the 5G Advanced evolution toward 6G

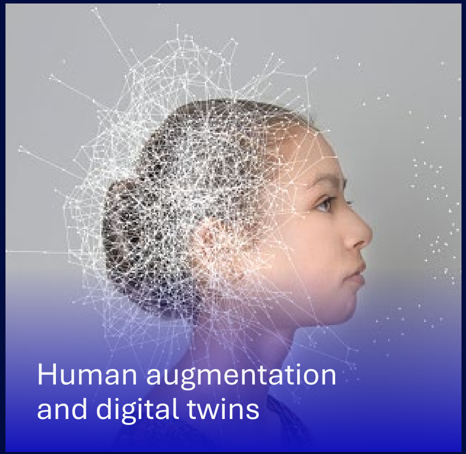
Global Momentum for 6G is growing

We are leading key discussions and working groups to promote early government investments in critical technologies

IMT-2030 defines next-gen mobile system requirements for 2030 and beyond

The standards body responsible for global 6G technology standardization





Propelling next-level experiences and innovative use cases in the new era of the connected intelligent edge for 2030 and beyond

Evolving wireless with AI

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

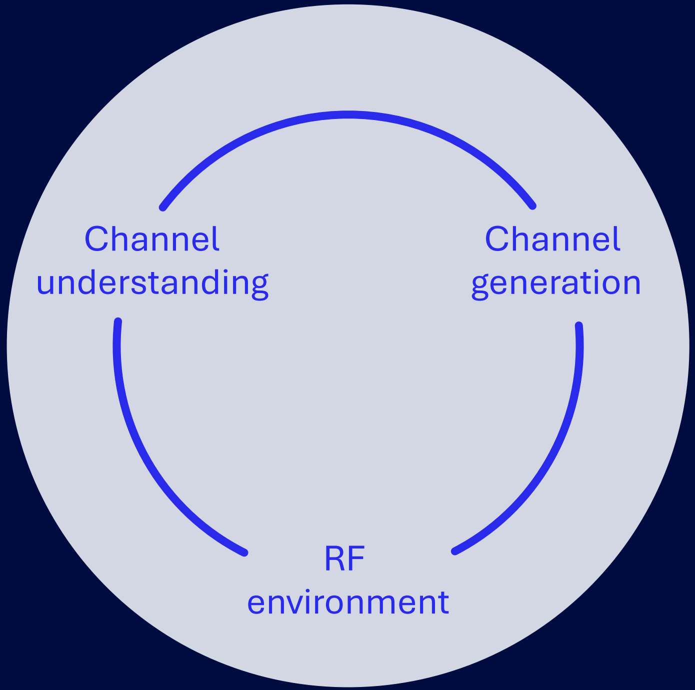
AI-enhanced wireless communications

AI
strengths

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

Using AI/ML to solve hard-to-model wireless problems

Example: AI/ML can improve channel modeling



Channel Modeling

Channel Generation

Synthetic channel generation, e.g., GAN[1]

$$p(x_{1:T}) = \prod_{t=1}^{T} p(x_t|x_{1:t-1})$$

Channel Prediction

Fading channel prediction, RSRP prediction[2]

$$p(x_{t+n}|x_{1:t})$$

Channel Inference

Inferring channel on other antennas, beams, frequencies, or locations[2]

Recover full x_t based on partial observation:

$$\tilde{x}_t\colon p(x_t|\tilde{x}_{1:t})$$

Channel Compression

Channel State Feedback[2,3]

Compress x_{t+1} based on prior

$$p(x_{t+1}|x_{1:t})$$

1 Eren Balevi, Akash Doshi, Ajil Jalal, Alexandros Dimakis, Jeffrey G. Andrews, “High Dimensional Channel Estimation Using Deep Generative Networks”, IEEE JSAC, 2021; 2 RP-213599, “New SI: Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR Air Interface”, 3GPP RAN Plenary, 2021; 3 Chao-Kai Wen, Wan-Ting Shih, Shi Jin, “Deep Learning for Massive MIMO CSI Feedback”, IEEE Wireless Comm Letters, 2018;

BUILDING ON THE

5G and 5G Advanced foundation

Targeting to expand wireless machine learning to the end-to-end system across network, RAN, device, and air interface

Network architecture enhancements

Allowing for machine learning to run over different HW/SW and future RAN function split to improve flexibility and efficiency

AI/ML procedure enhancements

Optimizing system for model management, training (e.g., federated and reinforcement learning), and inference

Data management enhancements

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

New and expanded use cases

Supporting traffic / mobility prediction, coverage / capacity optimization, massive MIMO, SON, CSI feedback, beam management, and other PHY/MAC and upper layer improvements

Qualcomm

MWCB 2024

Wireless AI Interoperability

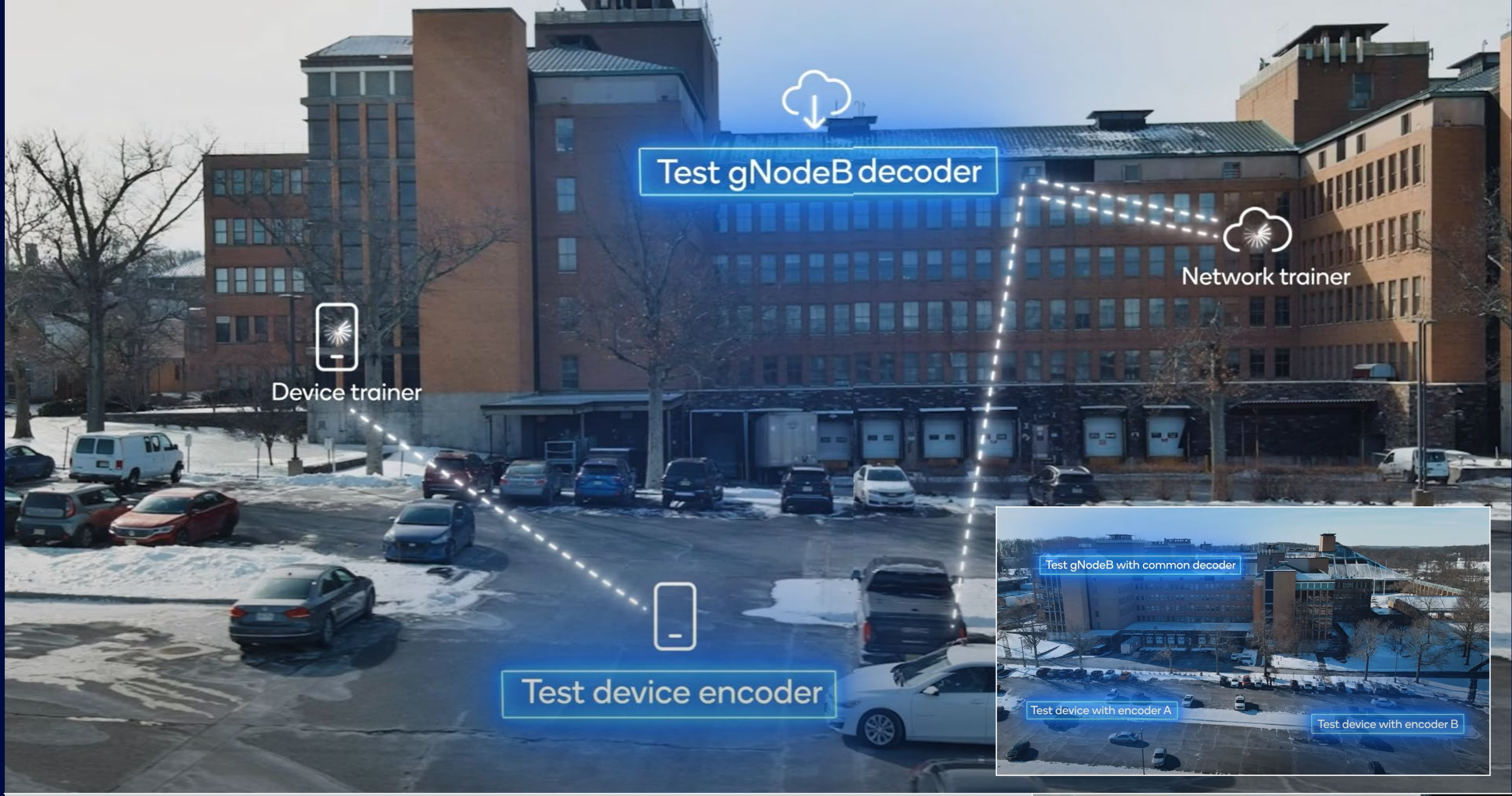
For multi-vendor system

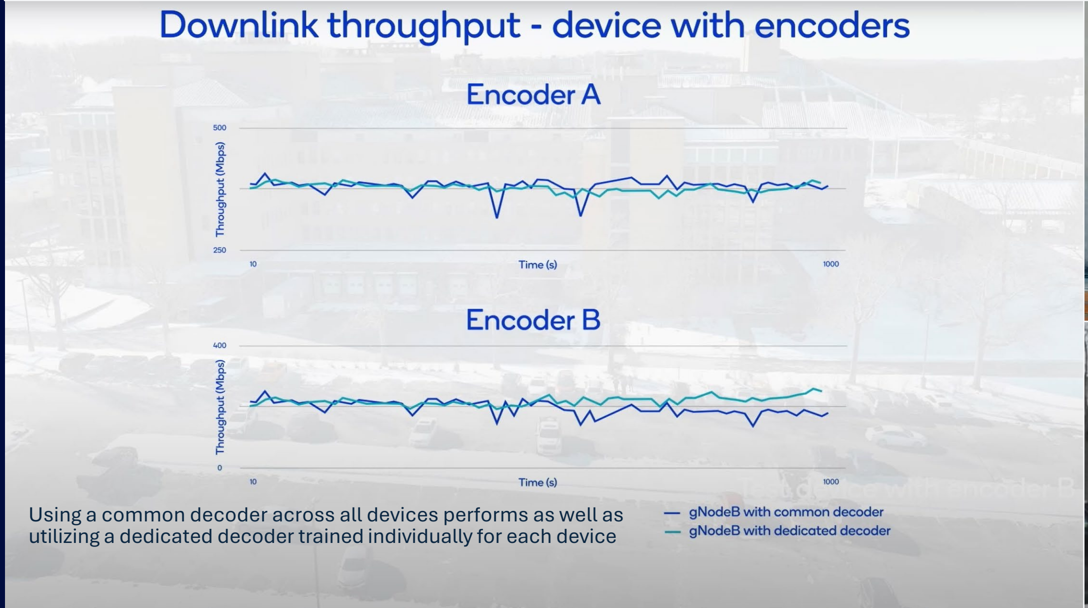
Close collaboration with Nokia Bell Labs on an over-the-air prototype of two-sided channel state information (CSI) feedback

Test network in Murray Hill, NJ, with Nokia infra and Snapdragon 5G Modem-RF

Sequential training enables data sharing but not neural network structures (models)

3GPP global standards compliant, potentially a part of 5G Advanced Rel-19+





Qualcomm

MWCB 2024

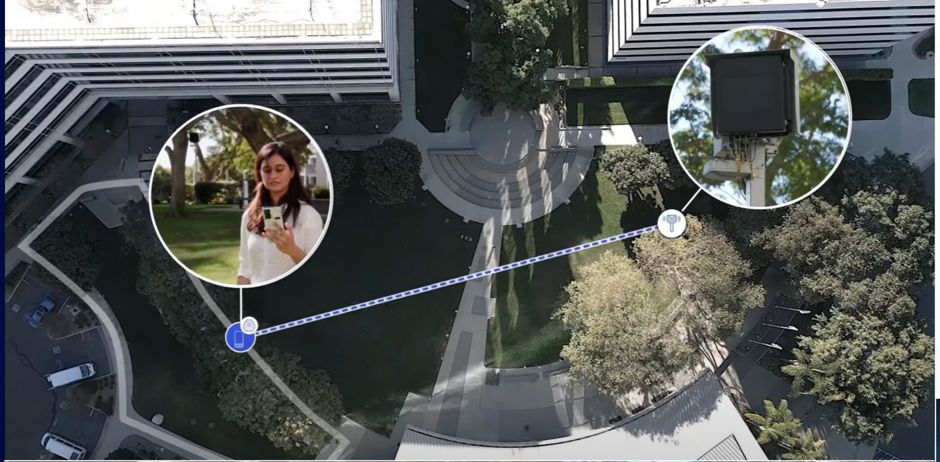
AI-enabled 5G mmWave Beam Management

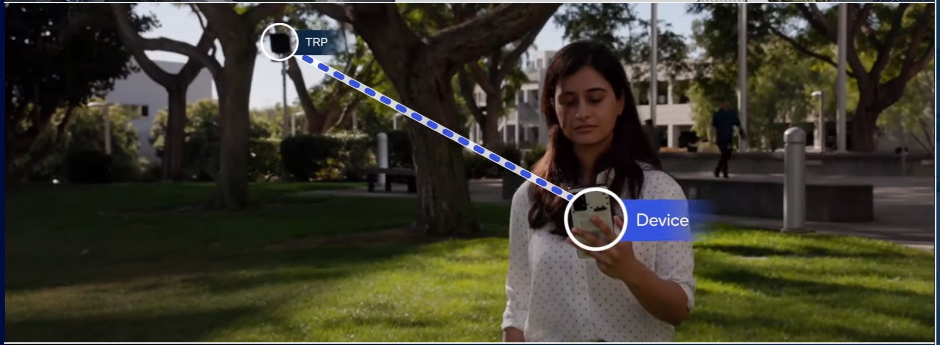
Time-domain and spatial-domain device beam prediction

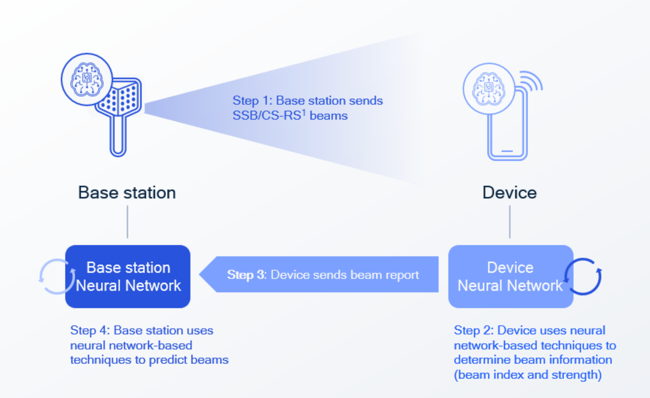
Reduced signaling overhead yields a more energy-efficient system design

Global standards-compliant design based on 3GPP Rel-18/19 projects

Over-the-air testbed operating in the 28 GHz band with up to 800 MHz bandwidth







A KEY PILLAR OF THE
5G ADVANCED ERA

Wireless AI

3 projects in Release 19

Study on AI/ML for Next-Gen Radio Access Network[1]

New use cases including network slicing and coverage and capacity optimization (CCO)

Continued studies on mobility optimization for NR-DC, split architecture support, enhanced energy saving, continuous MDT, and multi-hop device trajectory

Study on AI/ML to enhance 5G NR mobility[2]

Focusing on L3 device mobility, including RRM measurement & event prediction, device assistance information for network-side model, enhanced LCM, evaluation on testability, interoperability, impacts on RRM requirements and performance

1 RAN 3 led; 2 RAN 2 led; 3 RAN 1 led; 4 Continued study with corresponding checkpoints in RAN#105 (Sept '24)

Source: RP-234039 (AI/ML for NR air interface); RP-234054 (Study on AI/ML for NG-RAN); RP-234055 (Study on AI/M for mobility in NR)

Work on AI/ML Air Interface[3]

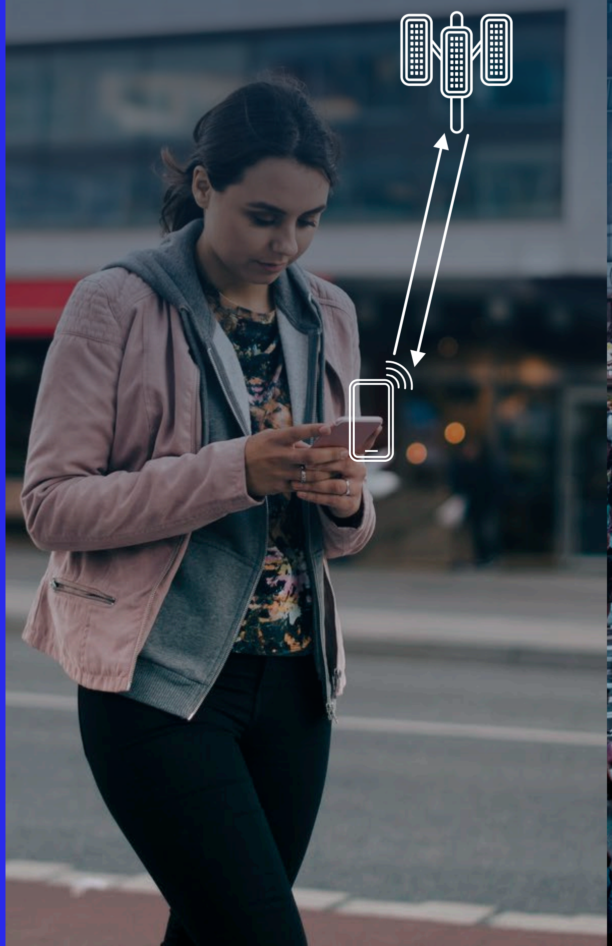
General Wireless AI Framework

Support Life Cycle Management (LCM) of one-sided (i.e., device or network) AI/ML models

Channel feedback[4]

Further study 2-sided CSI compression, 1-sided CSI prediction, model transfer/deliver, ...

Improve user downlink throughput and reduce uplink overhead



Beam management

Support device/network-sided beam prediction model in time/spatial domain

Reduce overhead, latency, and improve beam selection accuracy

Precise positioning

Support single-sided model for both AI-direct and AI-assisted positioning

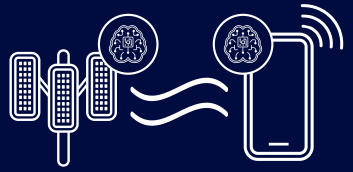
Improve positioning accuracy for different indoor/outdoor scenarios

AI-native 6G


RRC
SDAP
PDCP
RLC
MAC
PHY
AI/ML
Training and inference
6G protocol stack
6G Radio Access Network
AI-native 6G

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)

τ

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...

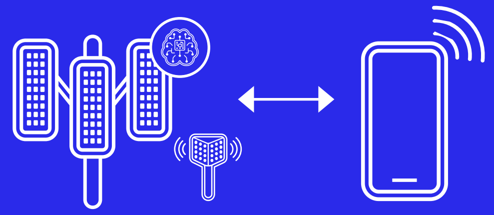
Machine learning can offer continuous wireless enhancements

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

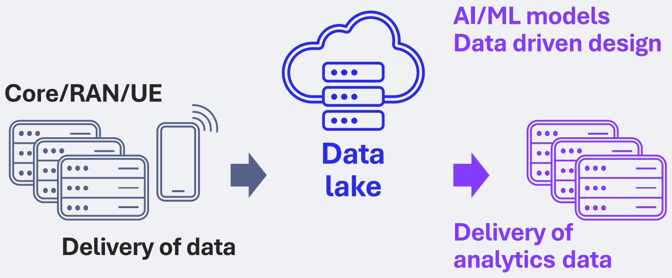
Data-driven system configuration provides end-to-end optimizations

Dynamic parameter adaptation based on fast machine learning algorithms

Neural network system design can customize to given wireless environment

Solve system-wide issues to support AI/ML in the 6G standard

Management of AI and ML, including model and life-cycle management across functions



Data collection

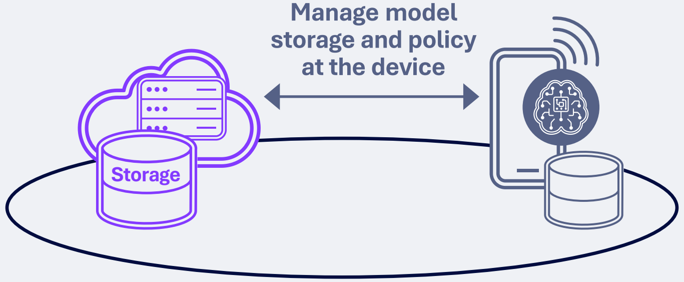
Collect data across different devices and network platforms into a common data lake

Continue to define data collection for new use cases at device and network

Make the data discoverable and accessible by SMO and other network entities

Use the data as input to:

1. Train AI/ML models
2. Device performance optimization
3. Data driven design

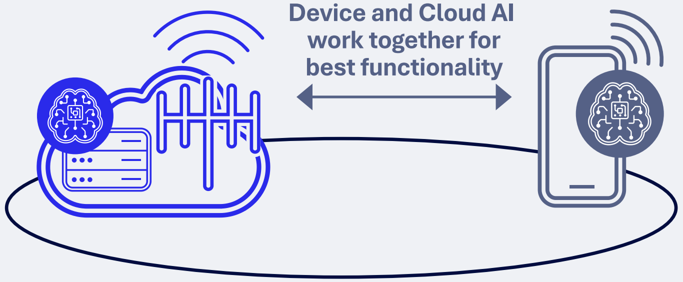


Device and network management

Manage storage and policy of AI/ML models at the device (and potentially network nodes)

Enable per UE and per RAN node AI/ML model provisioning across features (hosted by Qualcomm or 3rd party e.g., OEM or operator)

Support related procedures required to enable the device for AI/ML services, e.g., device configuration and policies



AI/ML management procedures

Joint Cloud/Core/RAN and device AI/ML functions to optimize connectivity and user experience

Device-side only or joint device-cloud AI/ML inference across different use cases

Leverages virtualization/containerization to scale and enable data driven device and network autonomy

DATA ANALYTICS | LIFE CYCLE MANAGEMENT | AI/ML SERVICES

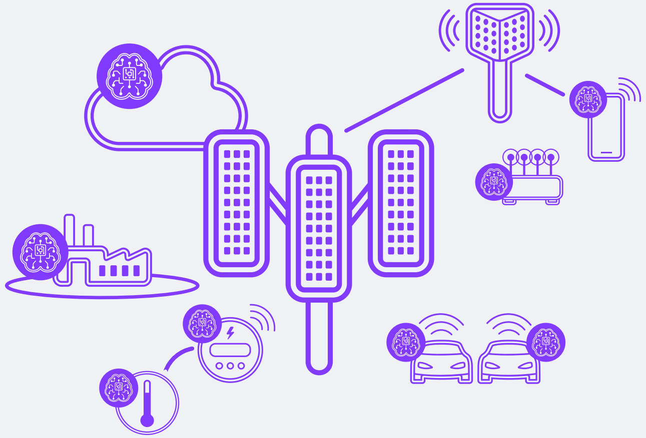
AI-native radio protocols – Why now?

Protocol adapts to the user experience instead of the user experience being defined by the protocol behavior

Paradigm shift

- Current configuration options in the user plane do not support a range of performance that allows the device to match QoS requirements
- Standard is defined to produce **predictable performance and simple compliance test** of protocol functions instead of allowing flexibility to adapt to various scenarios

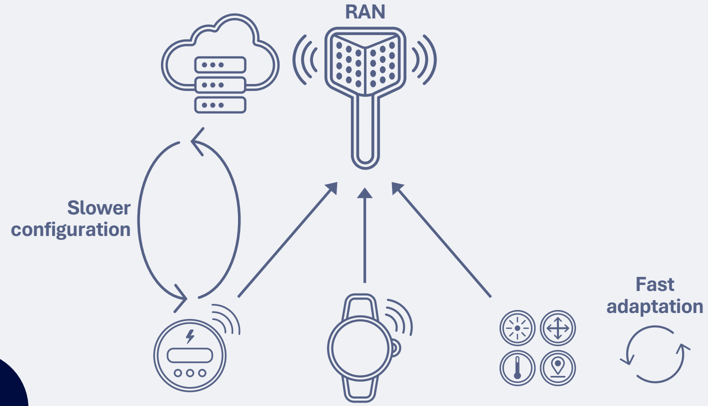
+



Leveraging AI capabilities

- Continual improvement through self-learning
- Data-driven development of **AI models responsive to user experiences and services**
- Dynamic **parameter adaptation based on fast machine learning algorithms** hosted at the device and assisted by the cloud

=

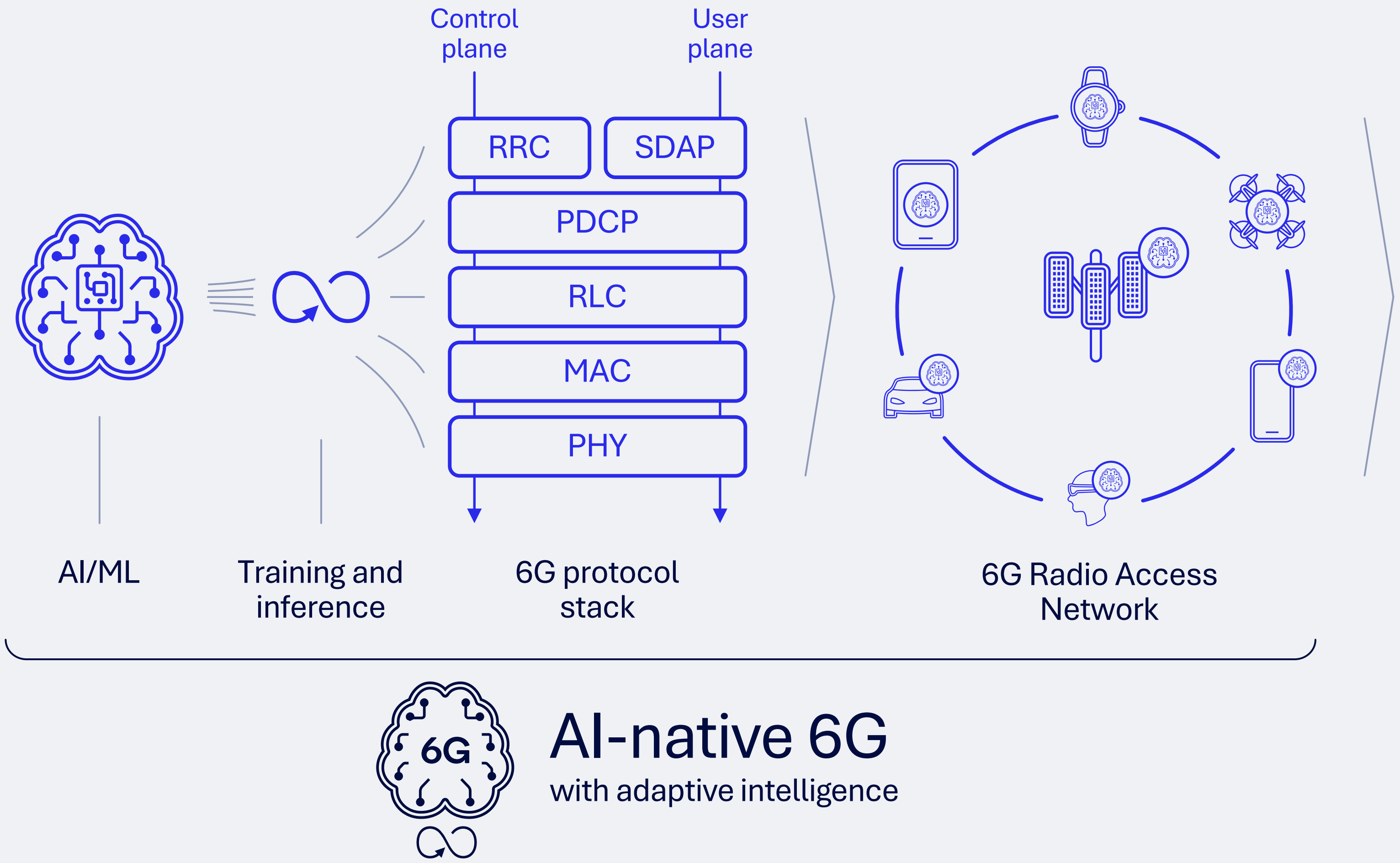


User experience optimization

- Slower network configuration to **define a range of behavior within which the device can autonomously adapt** based on current state
- Device's contextual awareness determines how to support user experience **based on real-time application requirements and local conditions**

AI-native 6G: Adaptive intelligence across the 6G protocol stack

Adaptivity is why AI and 6G are complementary



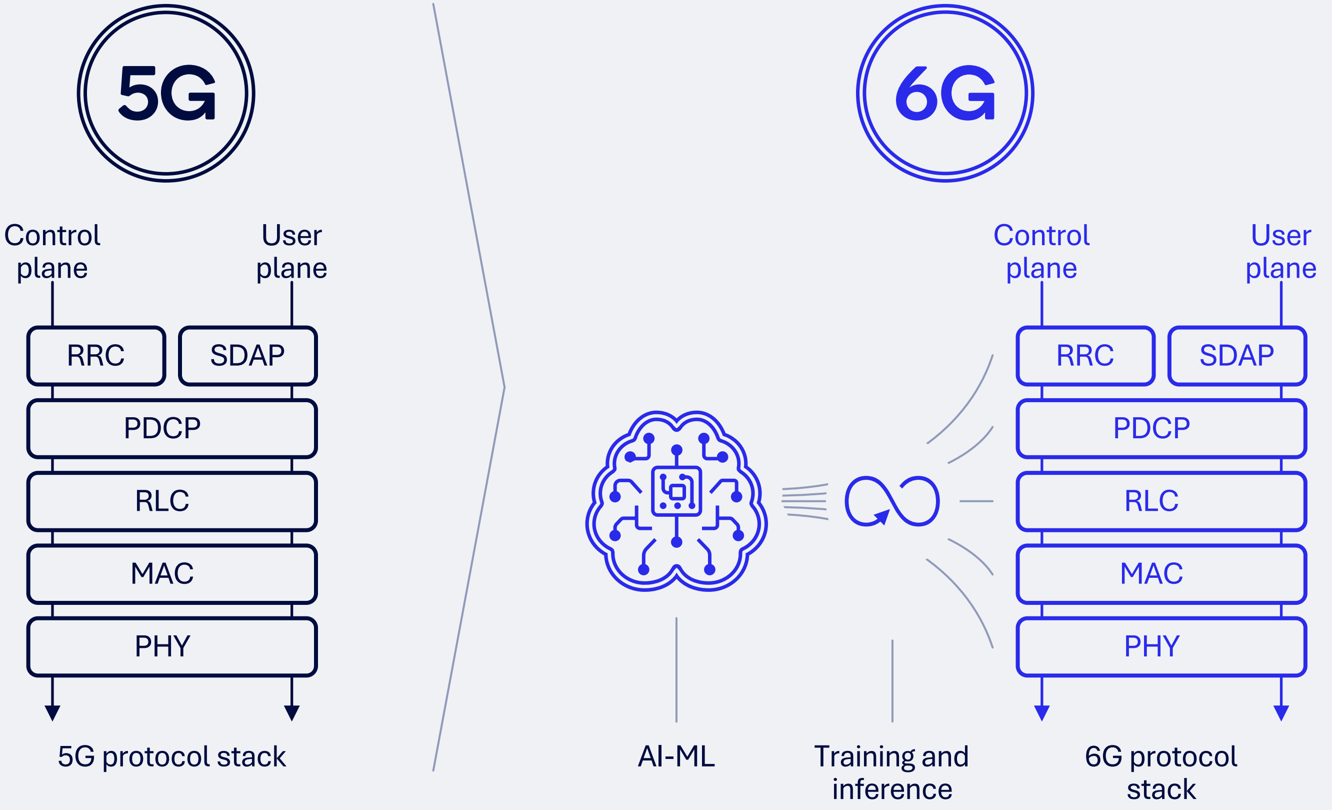
- Efficient resource utilization
- Robust service delivery
- User-specific adaptivity
- Application-aware adaptivity
- Fault-resilient network operation

Deliver performance-driven user experiences with adaptive intelligence

A network protocol that adapts to optimize user experiences instead of user experiences being defined by protocol behaviors

Network parameters are set to **static** values based on worst-case network requirements

Device behaviors are **fixed** across all scenarios for **behavior-driven** compliance



Network parameters are **adapted** intelligently within acceptable ranges under network supervision to meet key performance indicator (KPI) targets

Device behavior is **robust** for **performance-driven** compliance

A paradigm shift from conformance to adaptation and performance-driven optimization

Migrating from 5G to 6G with a focus on lower TCO

Migrating to 6G: Prioritize 6G Standalone for a range of benefits

Multi-RAT Spectrum Sharing (MRSS) and Dual-Stack are complementary tools for enhanced 6G experiences over wide areas

Provide wide area 6G coverage with MRSS

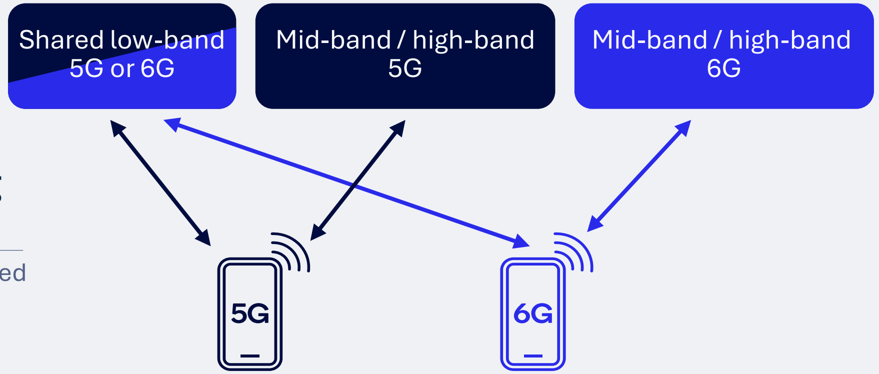
Enhance the robustness of 6G with MRSS for accessing a reliable low-band coverage anchor

Deploy 6G efficiently using CP-OFDMA-compatible 6G waveforms for MRSS in existing 5G bands

Combine existing carriers with new 6G carriers in higher frequencies for higher capacity

5G/6G Multi-RAT spectrum sharing (MRSS)

6G carriers can be deployed in 5G frequencies using CP-OFDMA-compatible waveforms



Improve user experiences with Dual-Stack

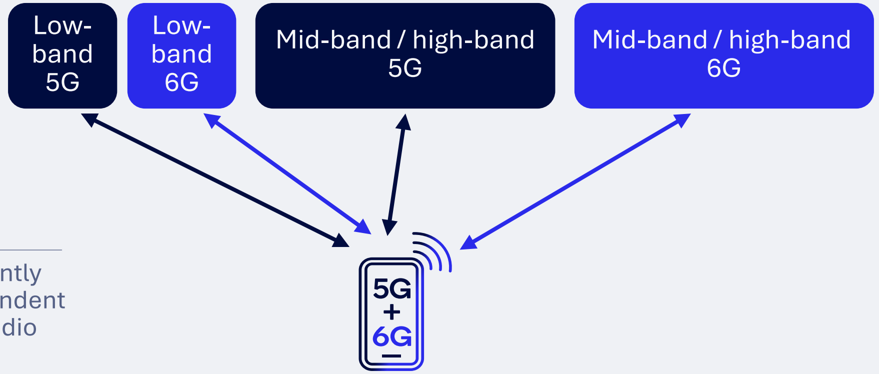
Simplify 6G introduction by eliminating 5G-6G RAN coordination with Dual-Stack

Move 5G-6G interworking complexity to the device to reduce the RAN transformation effort

Provide unprecedented user experiences by combining 5G and 6G network capacities

Dual-Stack

Device operates concurrently on 5G and 6G with independent transceivers for the two radio access networks



6G Standalone

Prioritizing infrastructure **reuse** and **extensibility** for fronthaul evolution

REUSE

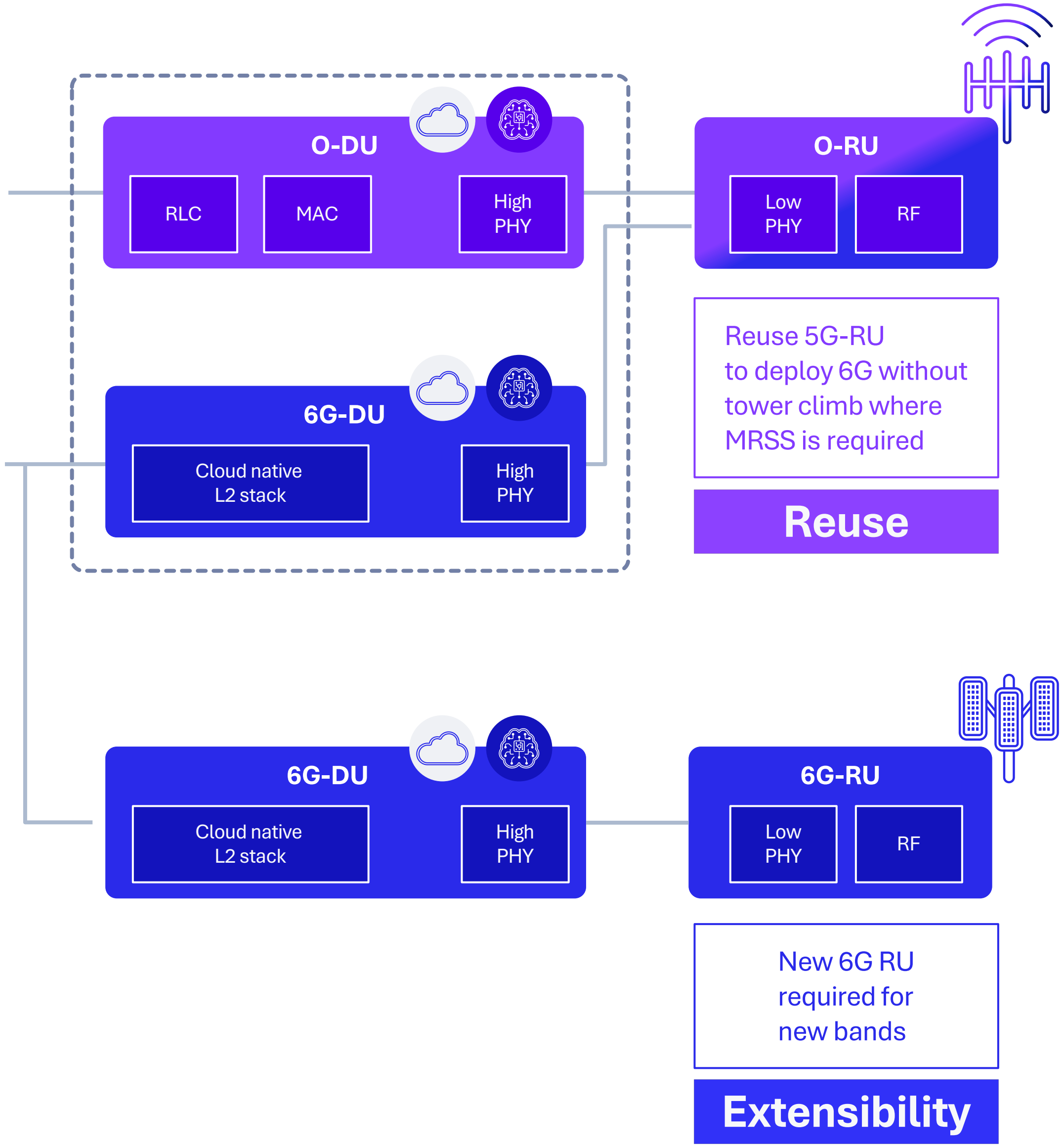
Migrate open fronthaul from 5G to 6G on existing bands

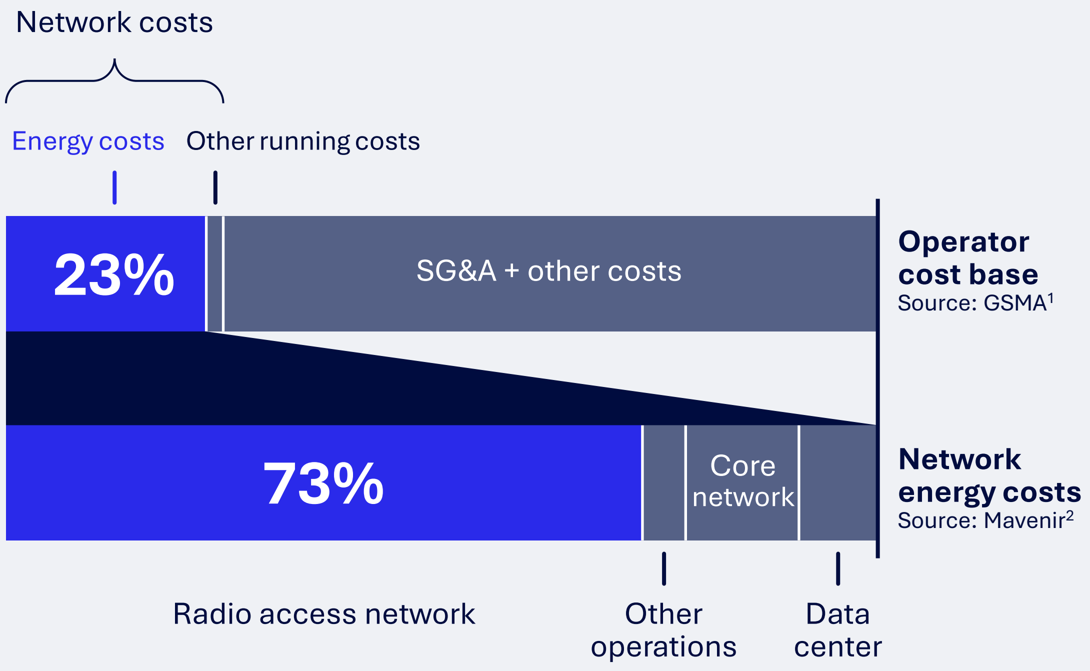
- Strong operator interest to keep Open Fronthaul for 6G
- Reuse existing open fronthaul and radio unit (RU) when migrating existing bands to 6G with MRSS

EXTENSIBILITY

Evolve open fronthaul for new 6G features

- OFH design areas under existing OFH architecture and framework
- **Layer-1 driven:** SBFD and FD, Enhanced MIMO processing of PUCCH, SRS-based DL precoding at RU, Dense UL detection for NOMA
- **Layer-2 driven:** Signaling simplifications, Topology aware scheduling for C-RAN
- **Service driven:** Sharing of RU by DUs belonging to different operators, as a TCO reduction and Multiplexing performance gain opportunity





Breakdown of the operator cost base

5G standard

Release 16 (2020)

- Wakeup signal (WUS)
- Other power saving techniques
- Low-power carrier aggregation control

Release 17 (2022)

- PDCCH skipping
- Search Space Set Group (SSSG) Switching
- Low-complexity RedCap devices
- Paging optimizations

5G-Advanced standard

Release 18 (2024)

- Adaptation of DTX/DRX
- Mobility and Paging Enhancement
- PDSCH transmission power adaptation
- Adaptation of antenna elements
- Inter-band carrier aggregation with SSB-less carriers
- Enhanced eDRX in inactive mode for RedCap devices

5G-Advanced standard

Release 19 (2025)

- Low-power wakeup signal and receiver (WUS/WUR)
- On-Demand SSB
- Ambient IoT
- Energy efficiency as a service

Release 20 (2027)

Potential additional improvements

6G standard

Release 21 (2029+)

The next generation of technology for new capabilities and efficiencies

Building on the 3GPP's long standing efforts to improve energy efficiency

1. https://data.gsmaintelligence.com/research/research/research-2020/5g-energy-efficiencies-green-is-the-new-black ,
2. https://www.mavenir.com/resources/a-holistic-study-of-power-consumption-and-energy-savings-strategies-for-open-vran-systems/

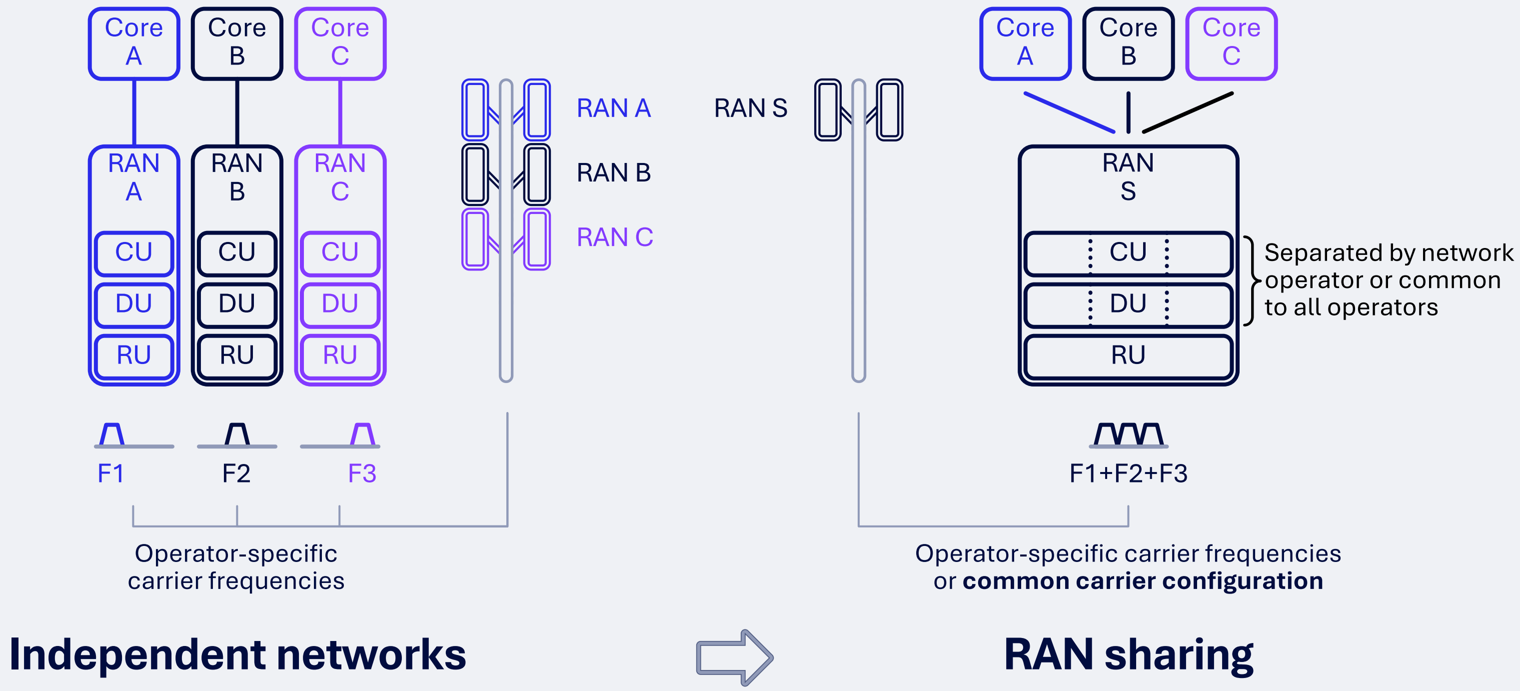
Reduce TCO and accelerate rollouts with advanced RAN sharing

Share some or all RAN components to reduce CAPEX and OPEX

Differentiate network services and user experiences with separate core networks

Continue with operator-specific spectrum or combine spectrum resources for joint scheduling over a common wide carrier (~ 500 MHz BW)

Reduce antenna tower loading and tower lease costs with fewer antennas by using the common carrier configuration



Other TCO reduction technologies:

AI-based network automation for continuous operational optimization

Non-terrestrial networks for energy- and cost- efficient rural coverage

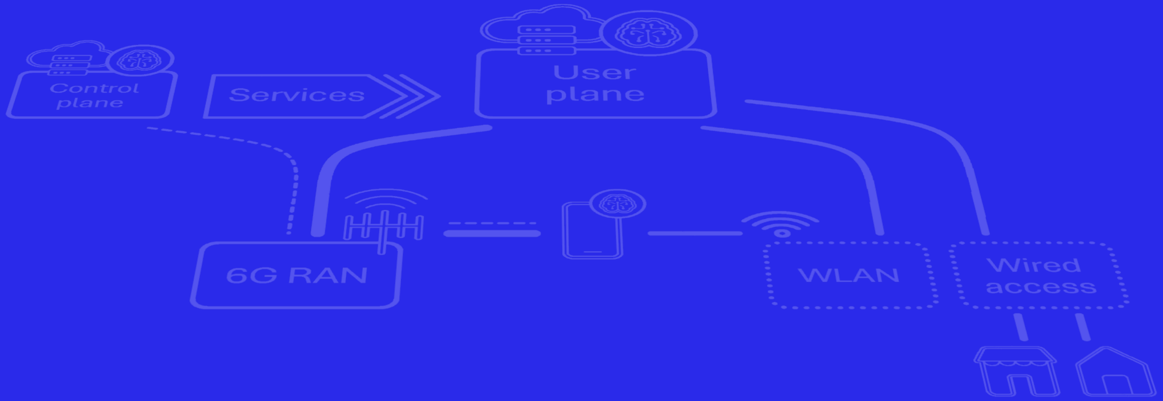
Green networks with energy-saving operational modes for lower OPEX

Examples of commercial RAN sharing:

4G/5G regional RAN sharing in Europe, Japan and Latin America

5G national RAN sharing in China

Rethinking the control plane



Emerging services and external technologies are driving deployments

Since 3G introduced data the network architecture and connectivity functions that have been deployed is largely unchanged

1G	2G	3G	4G	5G	6G
1980s	**1990s**	**2000s**	**2010s**	**2020s**	**2030s**
Analog voice	**Digital voice**	**Wireless Internet**	**Mobile broadband**	**Connected intelligent edge**	**Next-gen wireless**
AMPS, NMT, TACS	D-AMPS, GSM, IS-95 (CDMA)	CDMA2000/EV-DO WCDMA/HSPA+,	LTE, LTE Advanced, Gigabit LTE	5G New Radio	AI-native, new spectrum, RF sensing, and many more…

Adopting better network solutions

Adopting IP routing to support Internet + Move to COTS HW solutions for network boxes + Move to cloud to host Core and some RAN functions + **E2E cloud native Cloud services New HW for AI**

Support for emerging services

Data (SMS) + Smartphone with mobile broadband + IoT and other verticals + **More verticals AI/ML/Compute Sensing**

Key differences across Gs can be characterized by

- How the network is implemented
- Increased diversity of services OTT
- Evolution of connectivity features (e.g., security, mobility)

Separate user plane architecture helped scale the mobile Internet

3GPP adopted the Internet's IP-centric hourglass model for the user plane

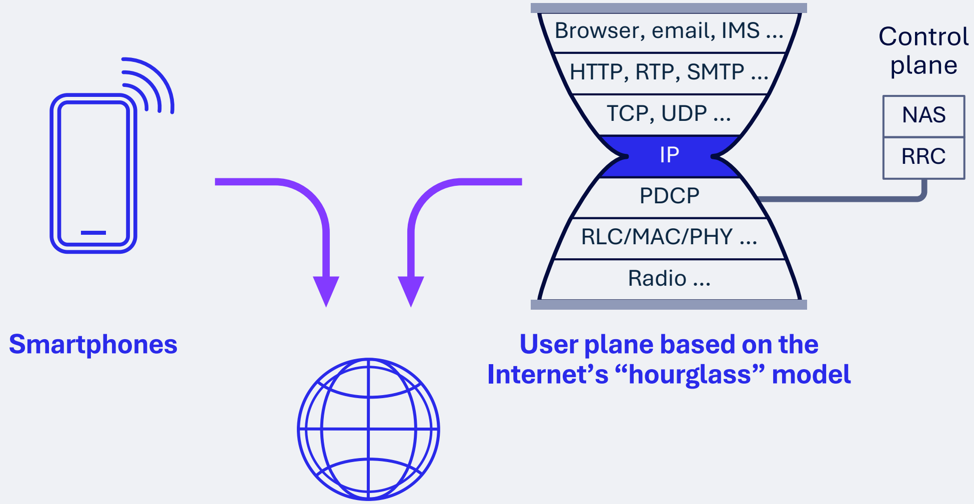
The Internet was designed to scale and serve a broad range of devices and services

The IP layer and the "hourglass" model were instrumental in facilitating the Internet's scale and versatility

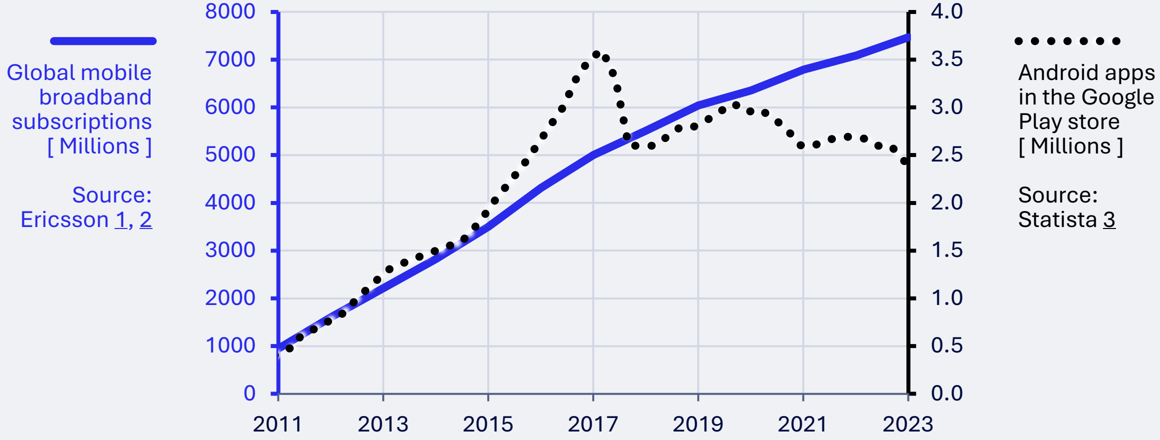
- Packet-switching
- Protocol flexibility above and below the IP layer

3GPP adopted the IP "hourglass" model for the user plane to connect mobile devices to the Internet

Mobile users enjoyed an unprecedented diversity of rapidly-evolving services due to the flexibility provided by the smartphone and the IP-native user plane



Rapid growth of the global mobile Internet



1: https://www.ericsson.com/en/reports-and-papers/mobility-report/mobility-visualizer?f=3&ft=1&r=1&t=9&s=4&u=1&y=2012,2023&c=2
2: https://www.ericsson.com/4a98ba/assets/local/reports-papers/mobility-report/documents/2011/traffic-and-market-data-report-q4-2011.pdf
3: https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/

The 3GPP control plane: Ever richer feature sets with limited adoption

4G and 5G continue to rely on a separate non-IP control plane as was first introduced in 3G

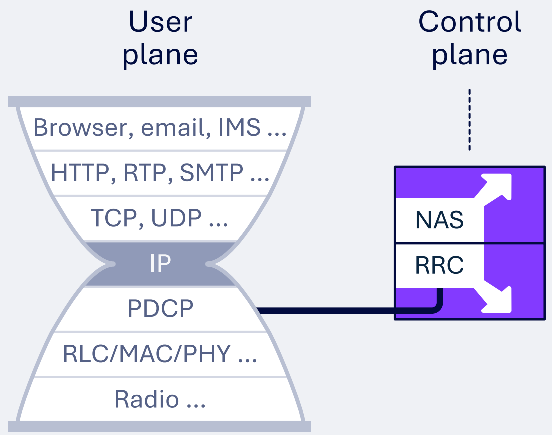
The non-access stratum (NAS) and radio resource control (RRC) control plane protocols have grown in complexity with more services in each 3GPP release

Control plane services have included:

- New horizontal features such as carrier aggregation and dual connectivity
- New verticals including Internet of Things (IoT) and Vehicle-to-Everything (V2X)
- Services hosted as part of the NAS/RRC control plane protocols, e.g., user location through NAS, or data collection through RRC for minimization of drive tests (MDTs)

Services over NAS or RRC require specialized knowledge for development, introduction, testing and maintenance compared to the more mainstream IP-based services

New control-plane services often impact various network entities, which affects adoption



Adopt a user plane-first design approach for 6G

Simplify 6G architecture and accelerate the introduction and operation of new services

User plane first design approach

Benefit from the scale and developer ecosystem for Internet services and protocols

Leverage the advantages of new cloud-based deployments, eliminating the need for additional hardware

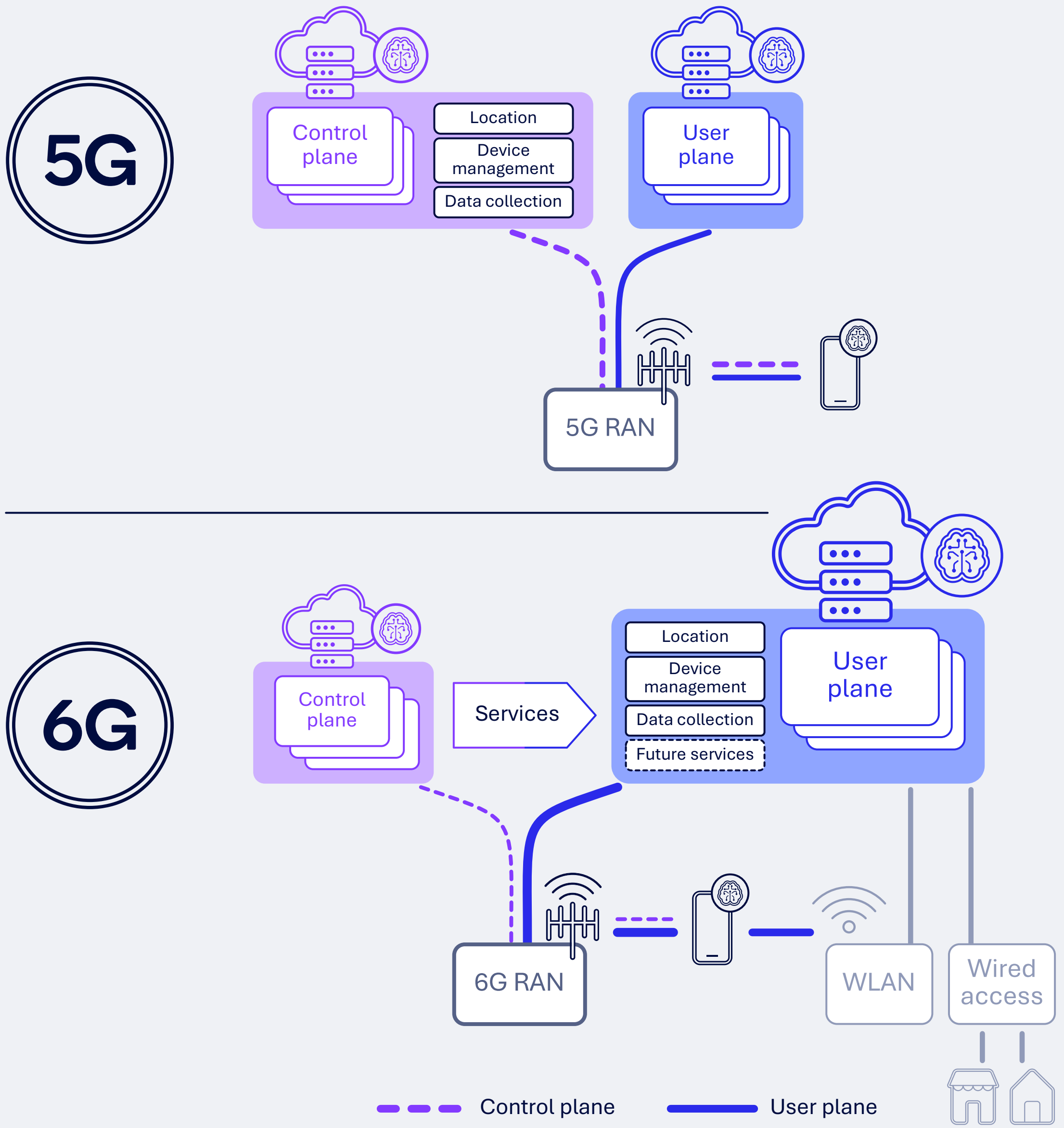
Develop "G-agnostic" services that can work seamlessly across different generations of cellular networks

Backport services to previous generations of mobile networks or cross-port them to other access technologies like Wi-Fi, fiber or cable

Lean control plane

Use the control plane only for aspects fundamental to maintaining connectivity and establishing data services, such as

- Security,
- Mobility, and
- Session management



Reuse and Replace protocols and interfaces

Reuse

External interfaces (Exposed to 3rd parties)

Common for 5G and 6G for service and data interworking

Data exposure (NEF) and IP functions (UPF) should be backward compatible

Internal Core interfaces (Service Based)

Reuse HTTP2/TCP
- Potentially evolve to HTTPS3/QUIC to support cloud platform evolution

Service discovery and authorization (NRF),
and analytics (NWDAF) common for 5G and 6G

Subscription and Policy management

Common for 5G and 6G as underlying functionality should not change

Replace

NAS protocol

Replace 5G NAS with a modular 6G NAS protocol with independent protocols for mobility management, session management, security, etc.
- Allows for more flexible and independent evolution of functions and introduction of new verticals
- Common transport of control plane signaling between UE and CN

Adopt a corresponding 6G AMF/SMF solution

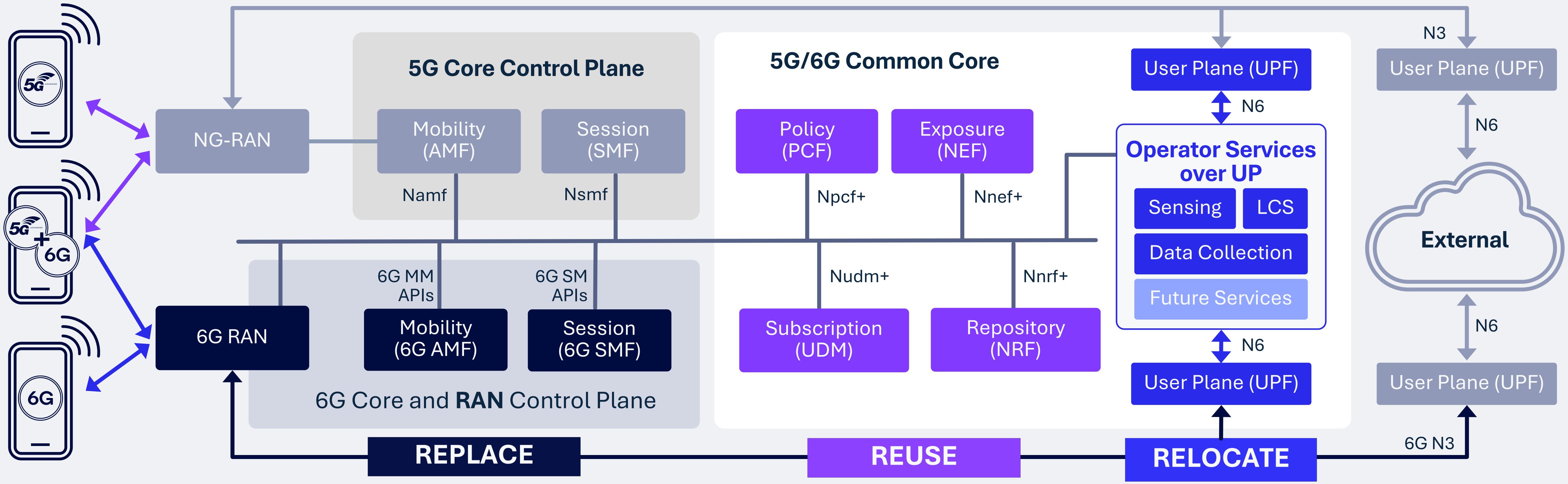
RAN to CN interfaces

Replace reference point interface with SBA based APIs

Allows for common orchestration of RAN and CN and allows RAN to benefit from cloud functionality

6G Core – **Relocate**, **Reuse** and **Replace** (System Architecture)

6GC evolved from 5GC network architecture*



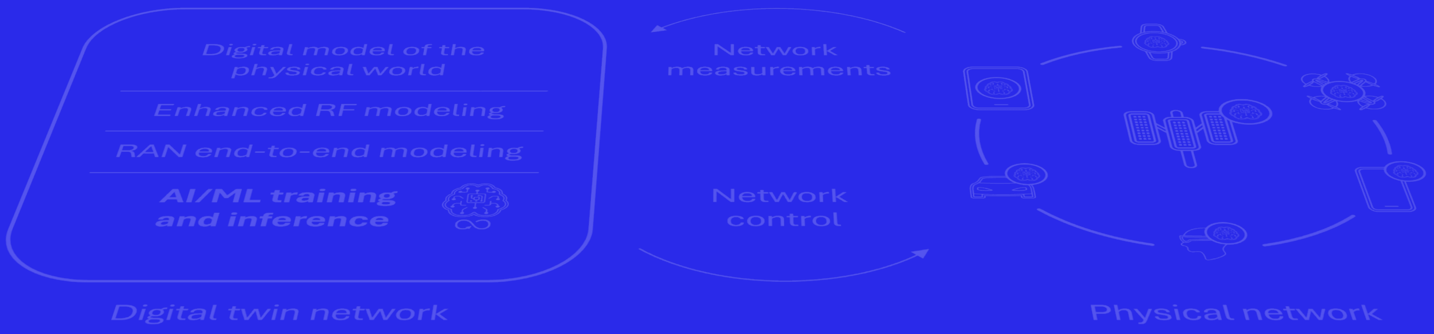
Evolve to a more service delivery -friendly 6G system while reusing as much of 5GS as possible

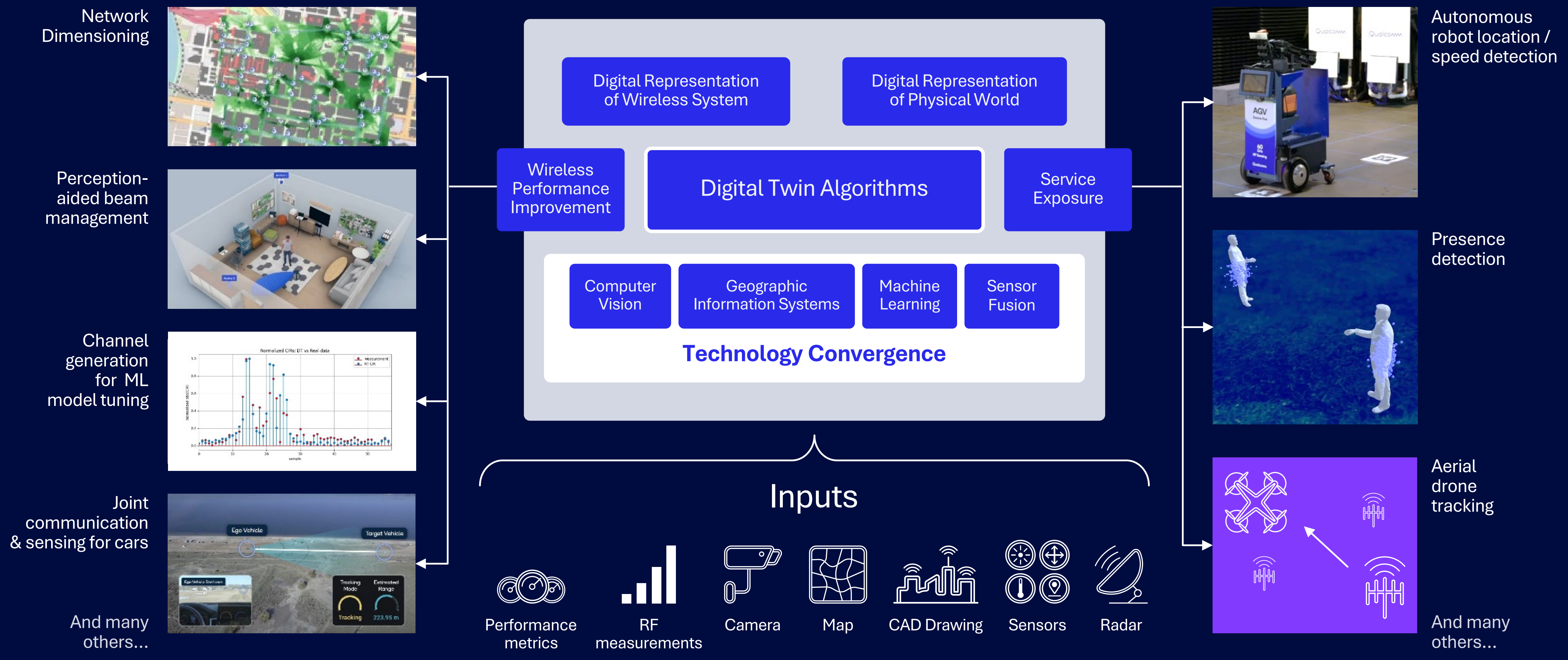
- **REPLACE** 5G NAS with new 6G NAS functions to support mobility, session management, etc.
- **RELOCATE** UE services to the user plane (where network is only a transport)
- **REUSE** API framework from 5G core for 6G core and RAN (based on 5GC service-based architecture)
- **REUSE** 5G network functions for policy subscription, user plane, etc.

Legend:
- 5G functions
- 6G function (**REPLACE**)
- 5G/6G common function (**REUSE**)
- 5G/6G User-plane function (**RELOCATE**)
- 5G User plane
- 6G User plane
- 5G/6G SBA
- Services over UP

* Note: Many additional 5GC network functions are not shown for clarity

Digital twin networks





Digital twin networks: Virtual representations of physical networks

Monitoring and performance optimization of its real-world counterpart

6G

Qualcomm

MWCB 2024

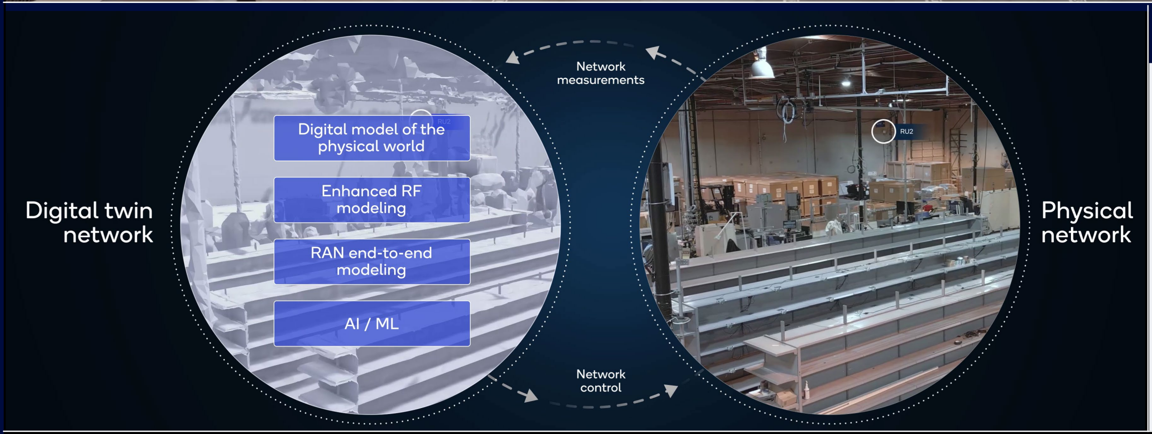
Digital Twin Network (DTN)

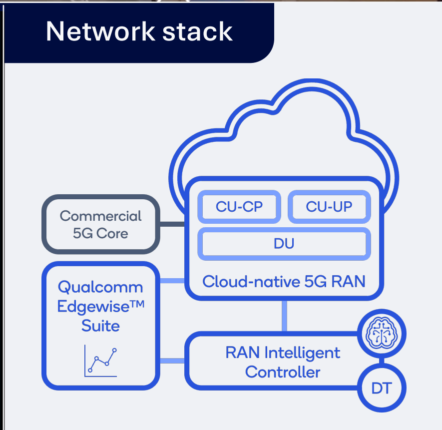
Converging expertise in wireless system modeling, computer vision and AI to create the high-fidelity DTN

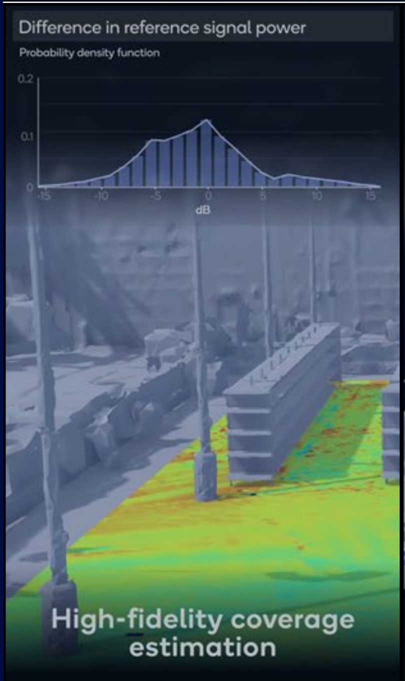
Generating synthetic data to addresS the data collection challenge from real world deployments

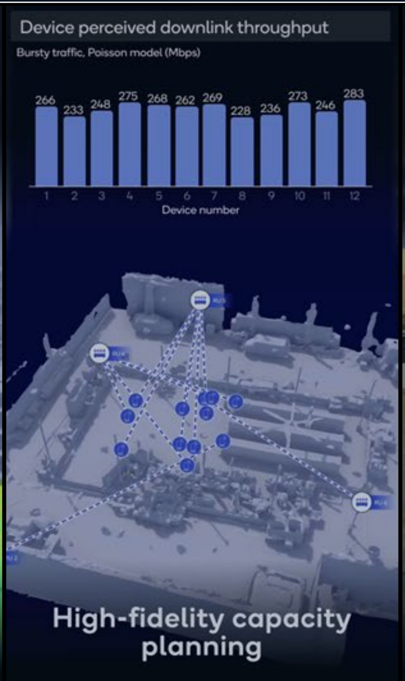
Sophisticated dynamic modeling of 5G RAN infrastructure

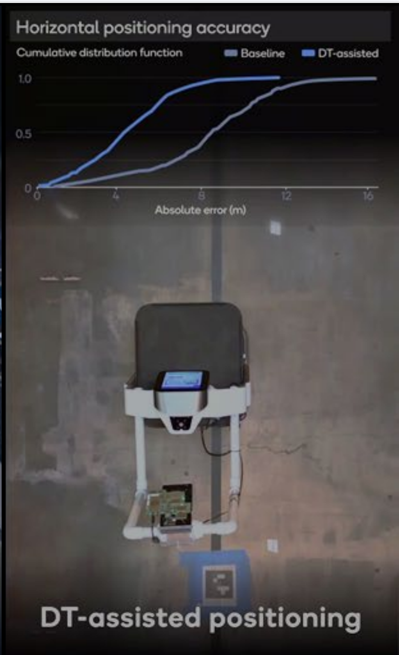
Over-the-air testbed operating in the 3.35 GHz band with cloud-native 5G RAN, RAN Intelligent Controller (RIC) and the Qualcomm Edgewise™ Suite











Our innovations on the path to 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

Snapdragon
X80 5G Modem-RF

Qualcomm®
5G AI Suite
Gen 3

AI-enhanced 5G Advanced user experience

Multi-antenna management to improve user experience

Contextually-aware QoS and latency improvements

60%* faster CPE service acquisition (mmWave)

10%* lower power in connected mode (mmWave)

Location accuracy improvement by 30%*

Best-cell selection time reduced by 20%*

30%* faster link acquisition

* Up to indicated numbers

Snapdragon 8 Gen 3

Titan of on-device intelligence

Generative AI powerhouse

up to **10B** parameters on device

up to **20** tokens per sec for 7B LLMs

AI

Qualcomm® Hexagon™ NPU

98% faster

40% more efficient

Stable Diffusion and ControlNet <1 sec fastest in the world

77GB/s, 4.8GHz LPDDR5x memory bandwidth

Qualcomm AI Stack

Pytorch ExecuTorch and optimized models

First to support **multi-modal gen AI models**

On-device personalization with Qualcomm sensing hub

World's 1st to support speculative decoding

4nm process node

Qualcomm® Kryo™ CPU

3.3GHz CPU max frequency

30% performance

20% efficiency

Security

Truepic with C2PA

Strongbox protection

Dual Always-Sensing ISPs

Gaming

25% performance

25% efficiency

240 FPS

Unreal Engine 5 with Lumen lighting system

Global illumination Next-gen lighting effects

Camera

Photo expansion

Video object eraser

Night vision for video capture

Dolby HDR photo capture

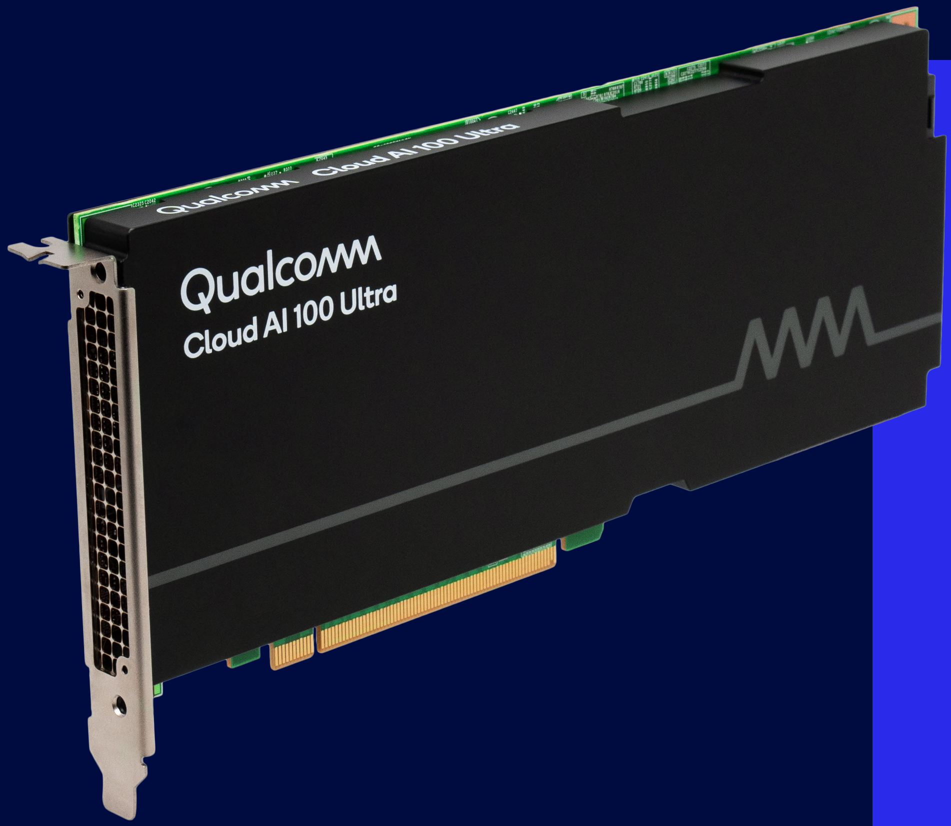
Sound

24bit, 96KHz Lossless

Qualcomm XPAN

PERFORMANCE AT SCALE

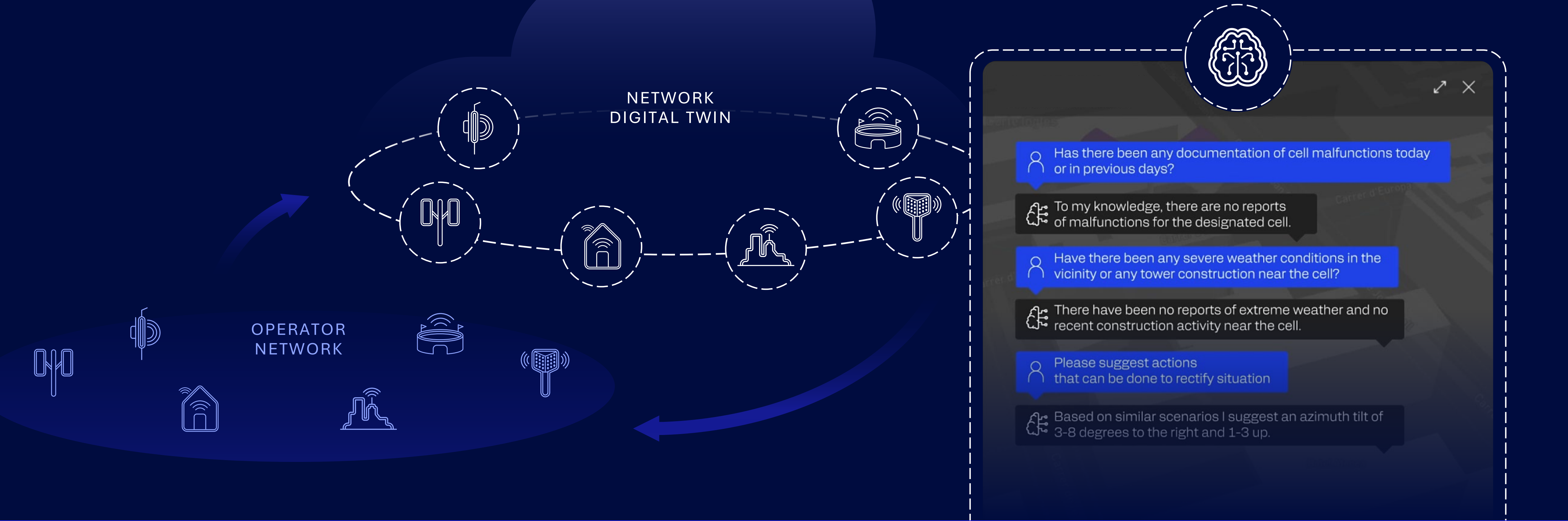
Introducing Qualcomm Cloud AI 100 Ultra



A performance- and cost-optimized AI inference solution, purpose-designed for Generative AI and large language models (LLMs).

- Best Perf/TCO$
- 100B Gen AI models on a single card
- Software tools for frictionless porting of pre-trained models
- 8x larger models within a single server
- Fully programmable and with support for recent AI techniques and data formats

Vision for gen AI-augmented and autonomous networks



Intelligent monitoring and management	On-the-fly modeling	Proactive alerts	Programming AI-assistants	'Level-3' autonomous networks

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

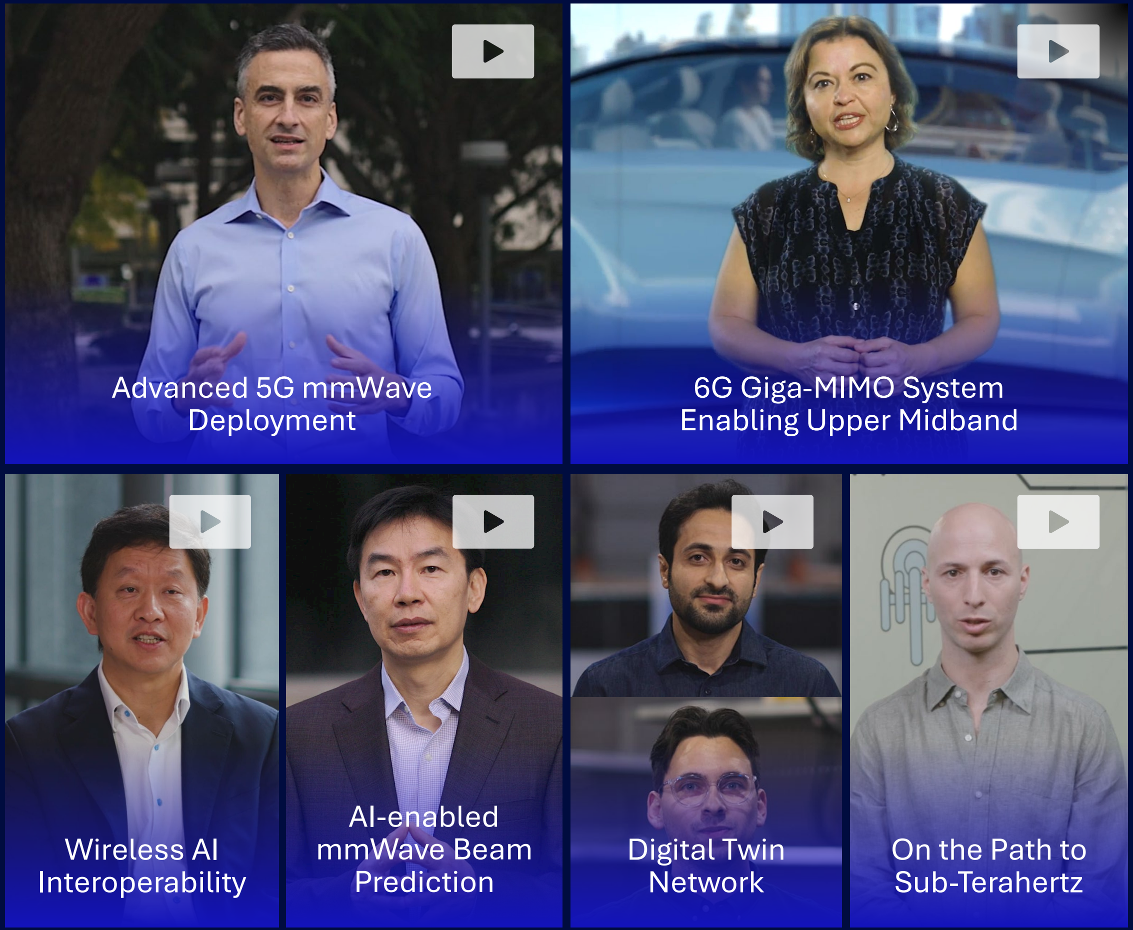
Physical, digital, virtual, immersive interactions taking human augmentation to next level via ubiquitous, low-power joint communication and sensing

Communications resiliency

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

Qualcomm Showcasing the latest technology innovations on the path to 6G

FOUNDATIONAL WIRELESS INNOVATIONS



5G BEYOND MOBILE BROADBAND

Other demonstrations: Super-QAM | Enhanced Link Adaptation | Enabling Subband Full Duplex | …

Sign up for the **What's Next in Wireless Tech** newsletter!

Thank you

Follow us on:

For more information, visit us at qualcomm.com & qualcomm.com/blog