MOBILE TEST SUMMIT 2025
VIRTUAL CONFERENCE CONNECTING THE WORLD

MOBILE TEST SUMMIT 2025
Virtual conference connecting the World

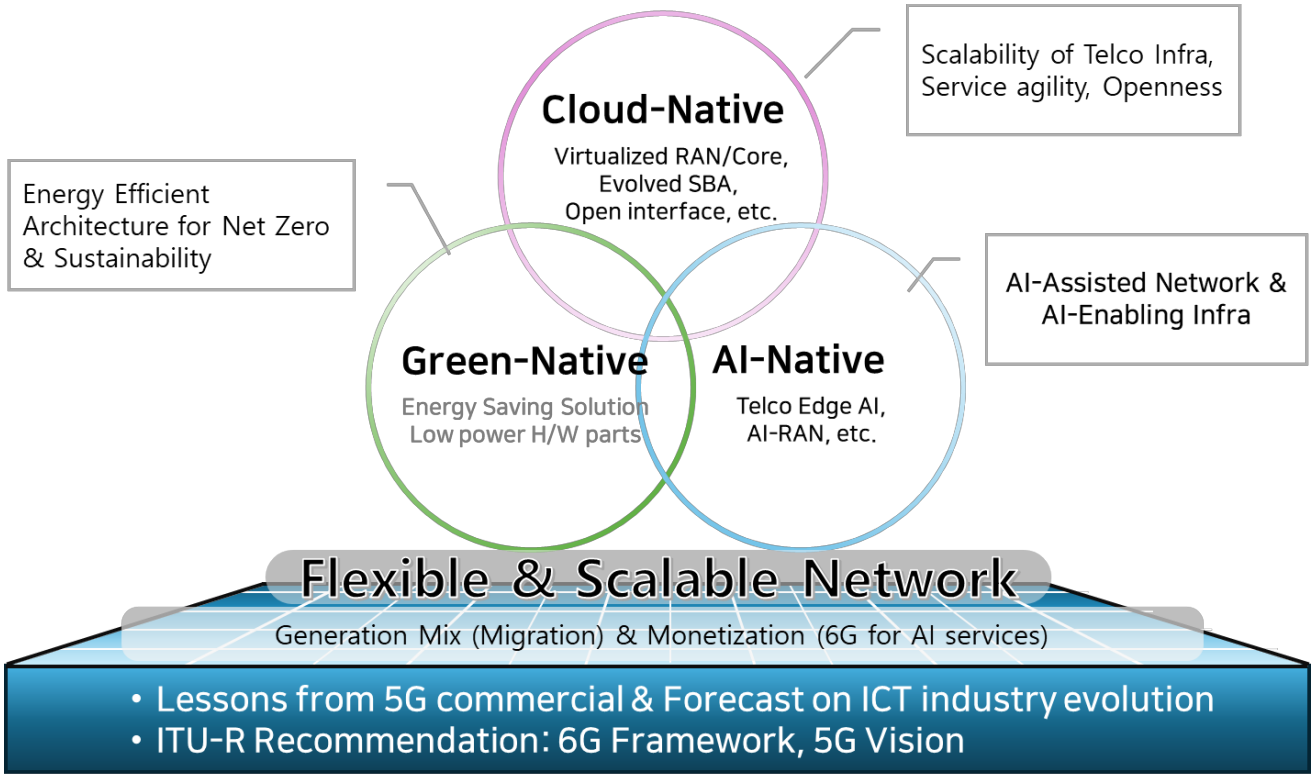
UNLOCKING THE POTENTIAL OF AI-RAN: A PATHWAY TOWARDS FUTURE NETWORK INTELLIGENCE

TONY Q.S. QUEK, IEEE Fellow, WWRF Fellow, Fellow of Academy of Eng. Singapore
ASSOCIATE PROVOST (AI & DIGITAL INNOVATION), SUTD
DIRECTOR, FUTURE COMMUNICATIONS R&D PROGRAMME
CHENG TSANG MAN CHAIR PROFESSOR AND FULL PROFESSOR

ROHDE&SCHWARZ
Make ideas real

6G Vision – SK Telecom

Evolution towards flexible & scalable network as the foundation for Cloud/Green/AI-native



[Key Considerations in Designing an Overall 6G Architecture]

※ Source: SK Telecom 6G White Paper "View on Future AI Telco Infrastructure," Oct. 2024, download : https://bit.ly/4eYS0t1

Source: SK telcom

AI-RAN Vision – Shaping Future of Telecom

Transforming Traditional RAN into Intelligent, Adaptive, and Revenue-Generating Platforms

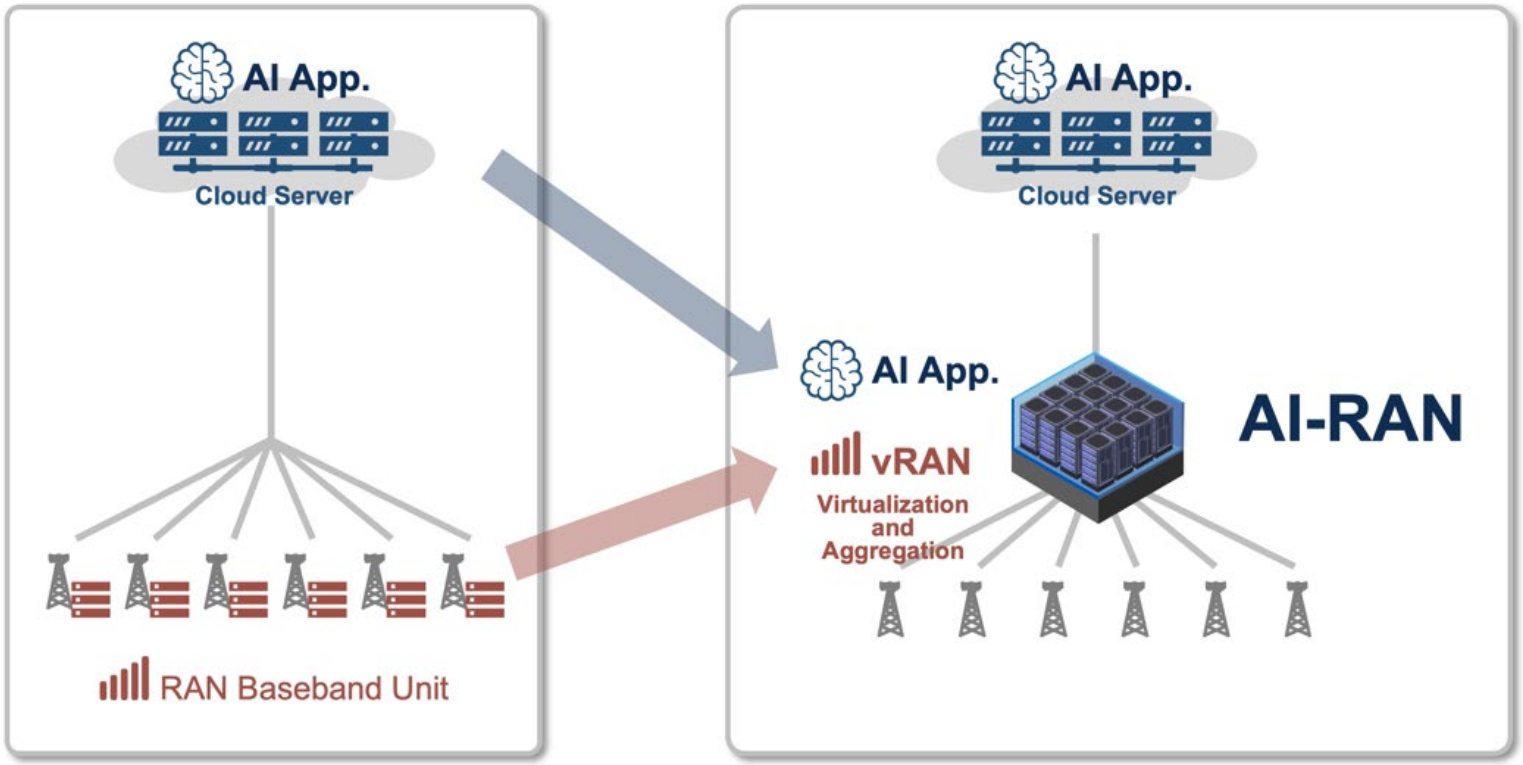
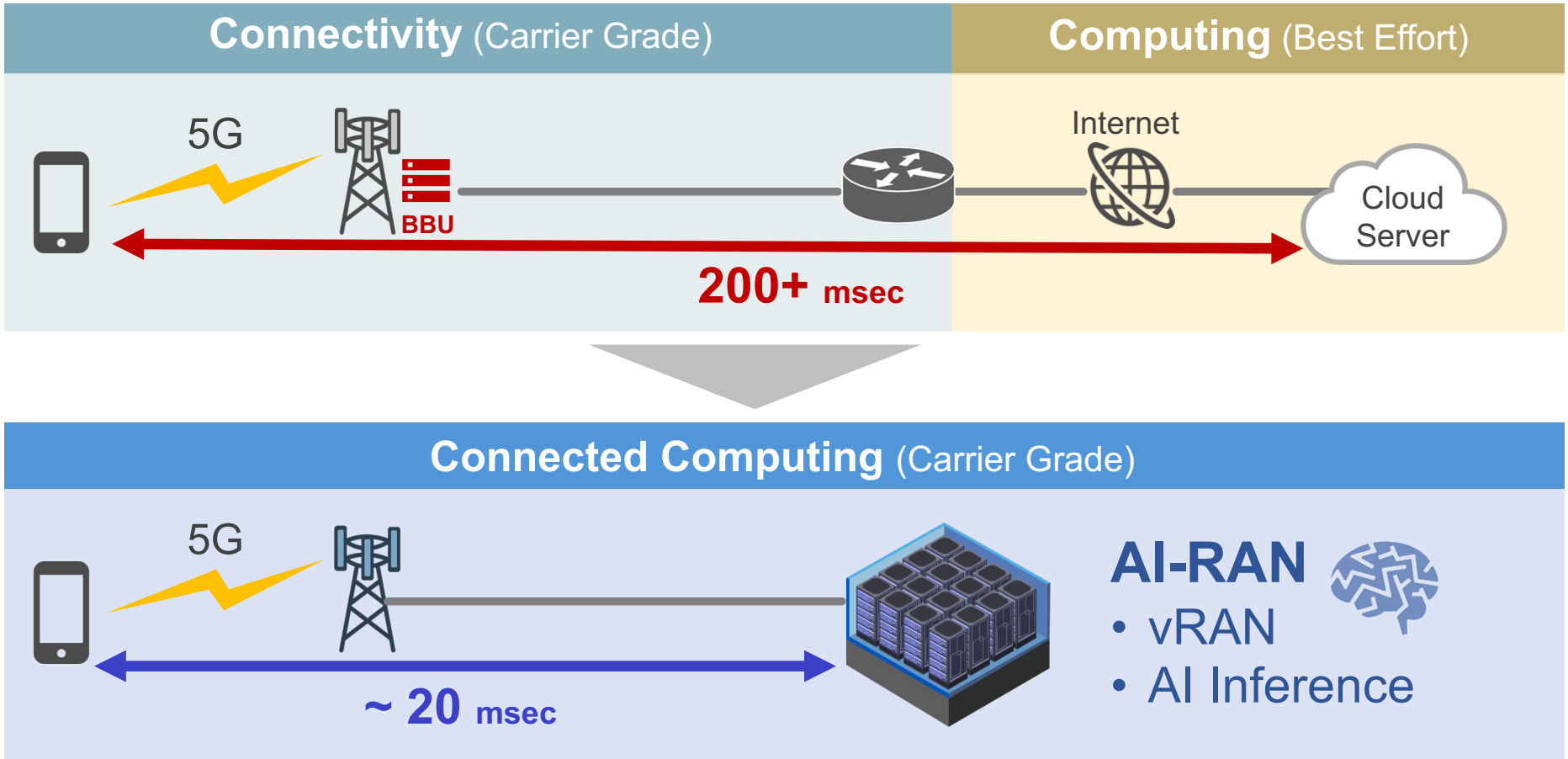


Figure 1. AI-RAN: AI and RAN integration

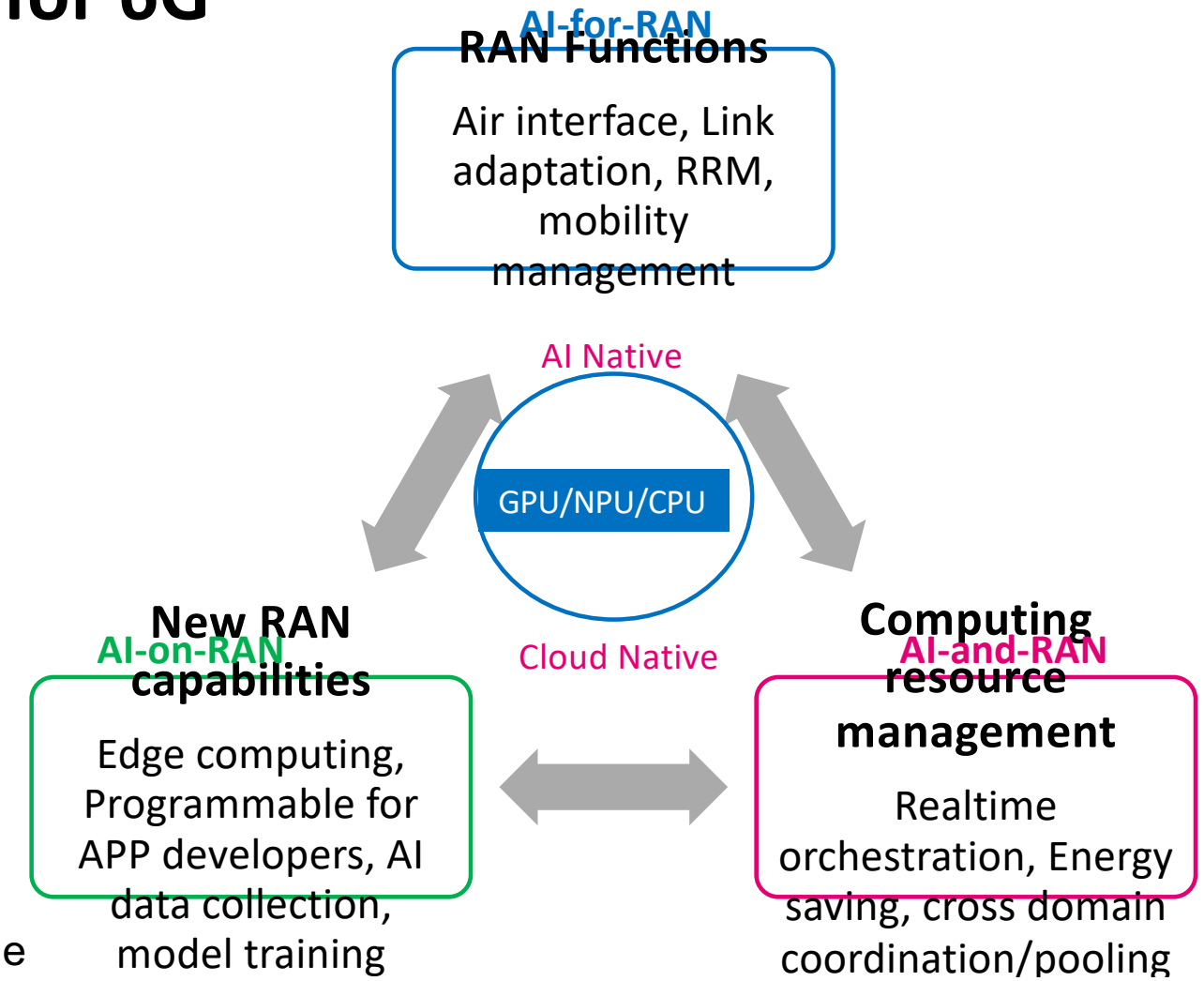
Source: Softbank AI-RAN White Paper

Integration of Connectivity and Computing



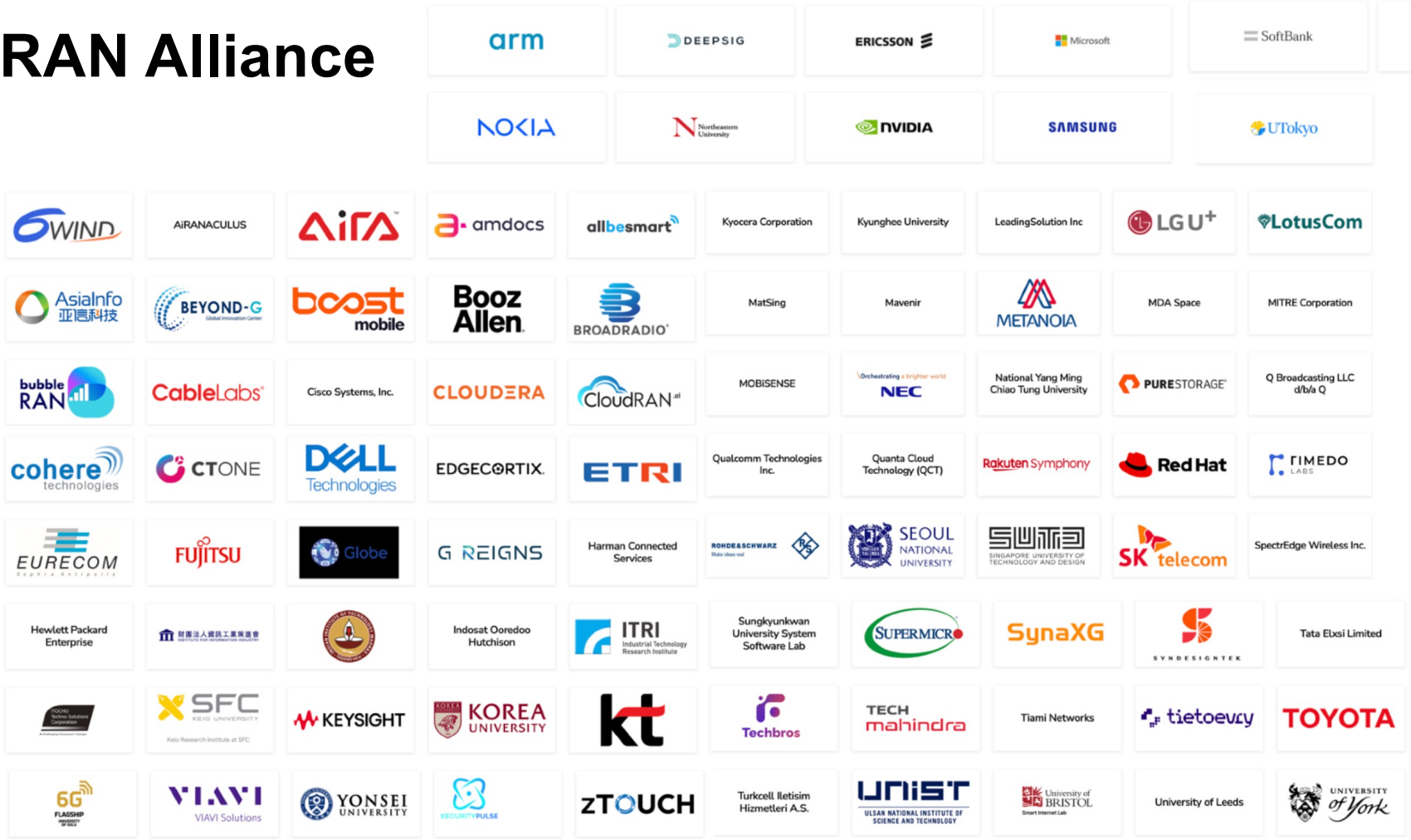
Source: AI-RAN Alliance

AI-RAN and Integrated Communication and Computing for 6G



Source: T-Mobile

AI-RAN Alliance



AI-RAN – Transforming RAN with AI

AI-and-RAN	AI-on-RAN	AI-for-RAN
AI and RAN sharing the same infrastructure	AI applications enabled by RAN	AI for the enhancement of RAN
Asset Utilization	New Applications	Spectral Efficiency

Source: AI-RAN Alliance

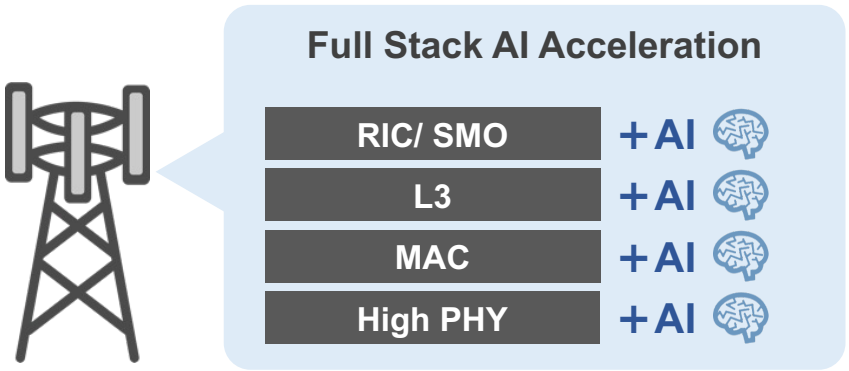
WG1: AI-for-RAN

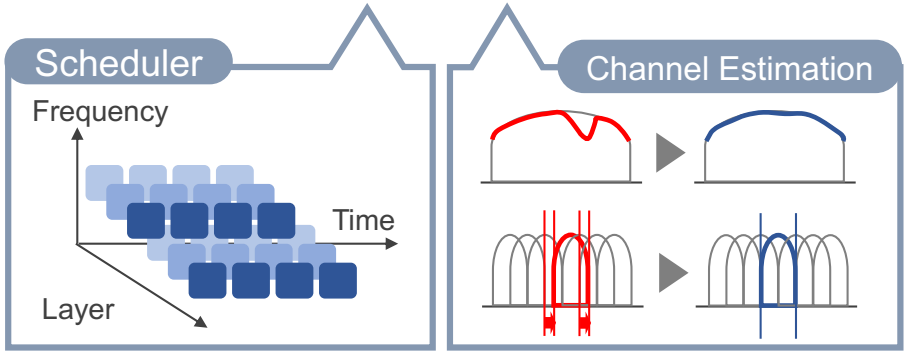
- **Improving RAN Performance with AI**
- Identify and explore AI applications in the RAN for improving its performance (spectral, energy, processing and system efficiency)
- Focus on AI-native RAN as the network evolves towards 6G with AI (potentially) at every layer
- Leverage and advance AI/ML efforts in 3GPP standards and related industry initiatives

Source: AI-RAN Alliance

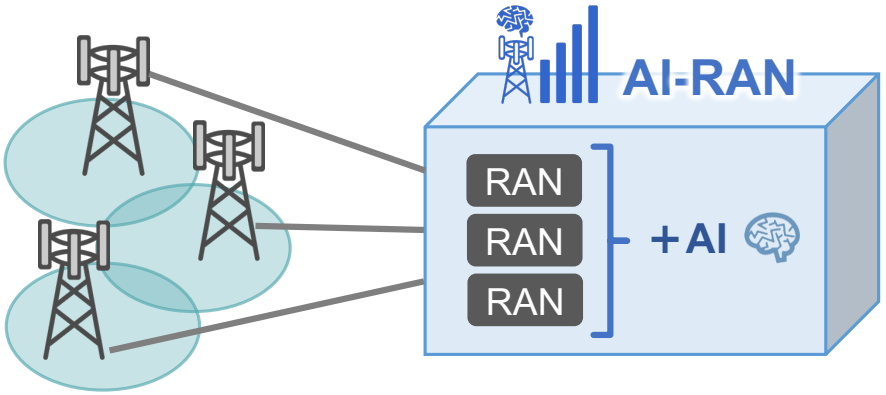
AI-for-RAN: Enhancement of RAN with AI

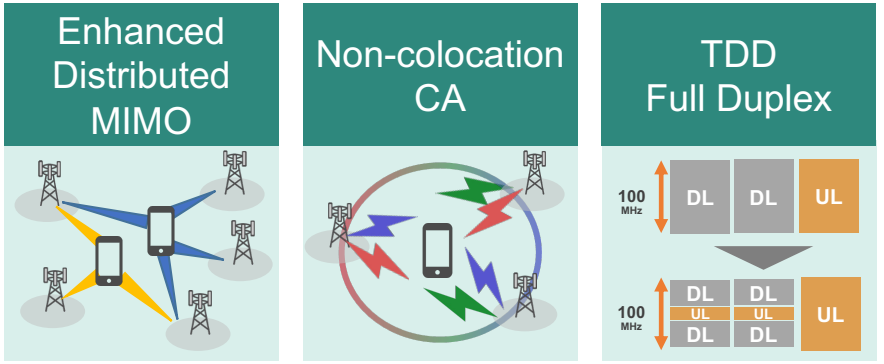
Individual Cell





Inter-Cell Coordination





Source: AI-RAN Alliance

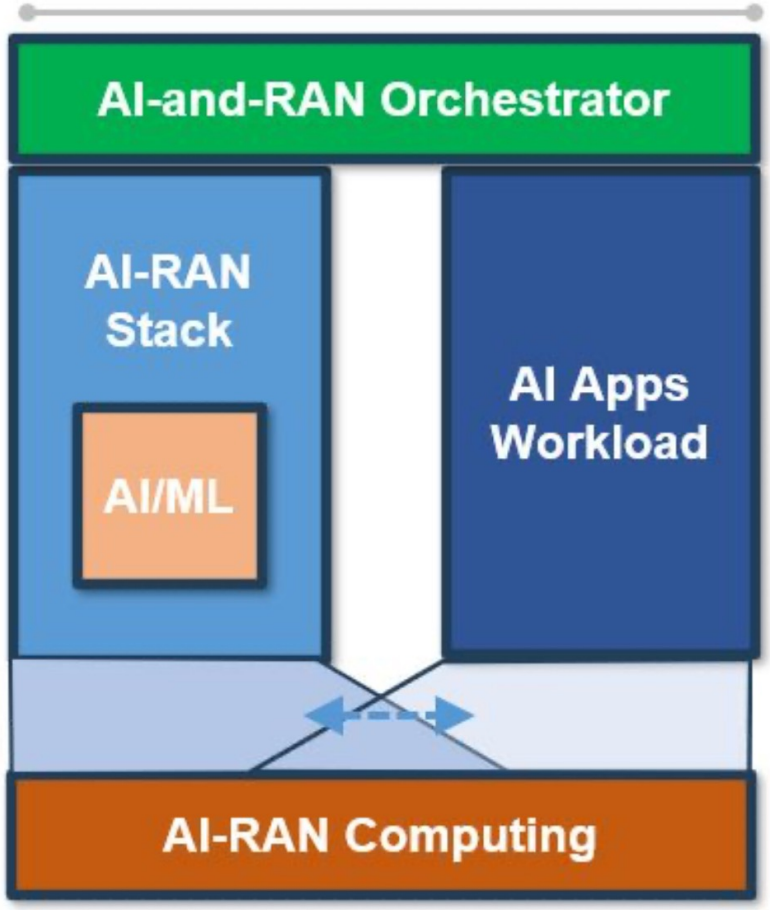
WG2: AI-and-RAN

- **Sharing equipment between RAN and AI**
- Explore the use of converged computer-and-communications infrastructure to run RAN and AI (e.g., GenAI) workloads simultaneously
- Aim to increase platform utilization for RAN workloads while creating monetization opportunities by running AI on the same compute cluster

Source: AI-RAN Alliance

AI-and-RAN: Comparison with MEC

- **Integration with AI infrastructure**: AI-and-RAN integrates **AI-specialized infrastructure optimized for both RAN and AI computations**, whereas MEC typically relies on generic computing infrastructure.
- **Scalability and Flexibility**: AI-and-RAN offer greater scalability and flexibility in **supporting both RAN and AI workloads**, as well as in dynamically managing resources between the two different types of workloads.
- **Security and Data Localization**: AI-and-RAN place a stronger emphasis on **data security and localization**, particularly for AI applications requiring sensitive data to be processed close to the data source.



Source: AI-RAN Alliance

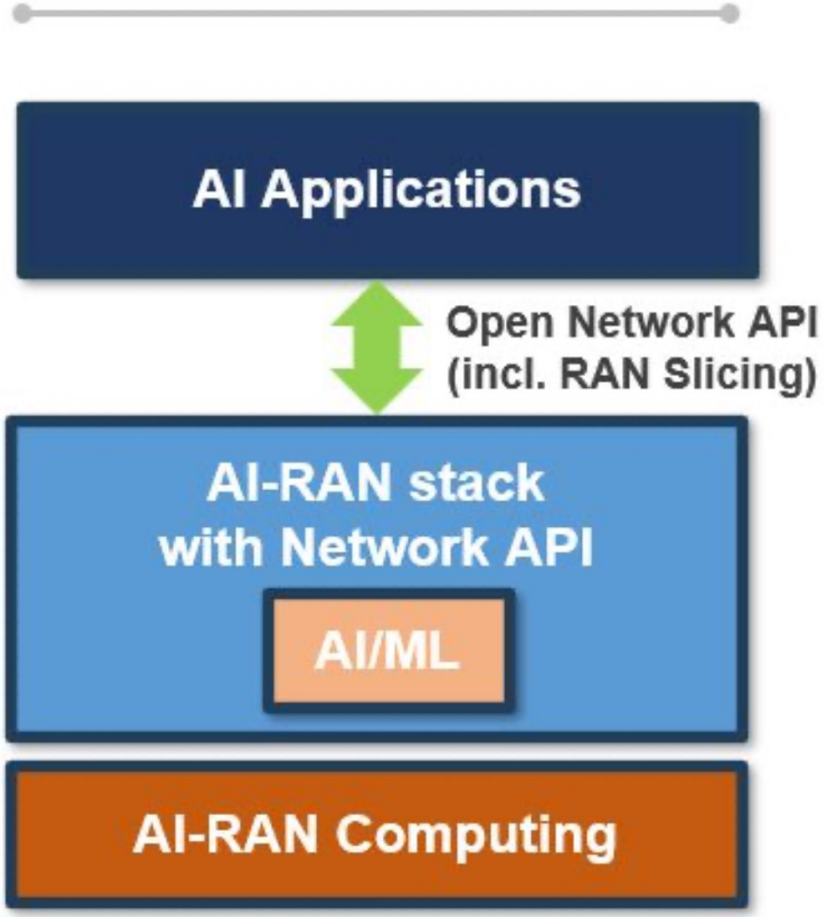
WG3: AI-on-RAN

- **New AI Applications enabled by RAN**
- RAN for AI, exploring how AI-applications like edge AI and real-time inferencing can be run on RAN infrastructure.
- Goal is to ensure that RAN can meet the needs of low-latency AI applications, especially, for the most prominent examples/use cases (e.g., autonomous driving, smart cities, etc...)

Source: AI-RAN Alliance

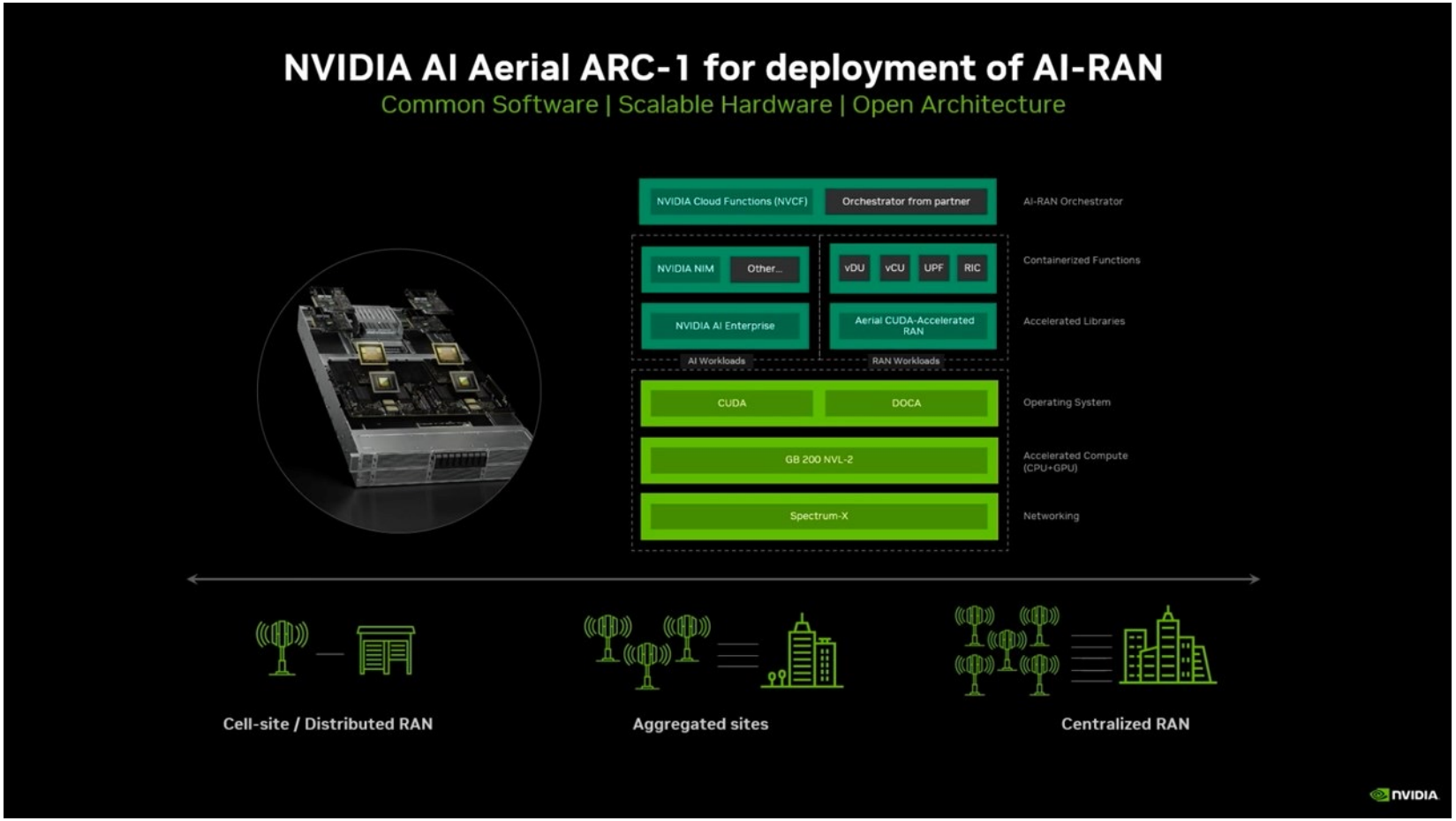
AI-on-RAN: Bridging AI Applications with AI-RAN

- **RAN network API:** Serve as the **critical interface between AI applications and the RAN system**, facilitating real-time interaction between the two. It will provide essential metrics and capabilities, allowing AI applications to adapt dynamically to changing network conditions.
- **Optimizing RAN Performance: Determine the optimal resource** like RAN slice for AI applications. Eg, low-latency slice for real-time AI applications and high-throughput slice for data-intensive AI applications.
- **Monitoring RAN Performance**: **Monitor delay, throughput and reliability** to optimize deployment of AI applications. Eg, model splitting.



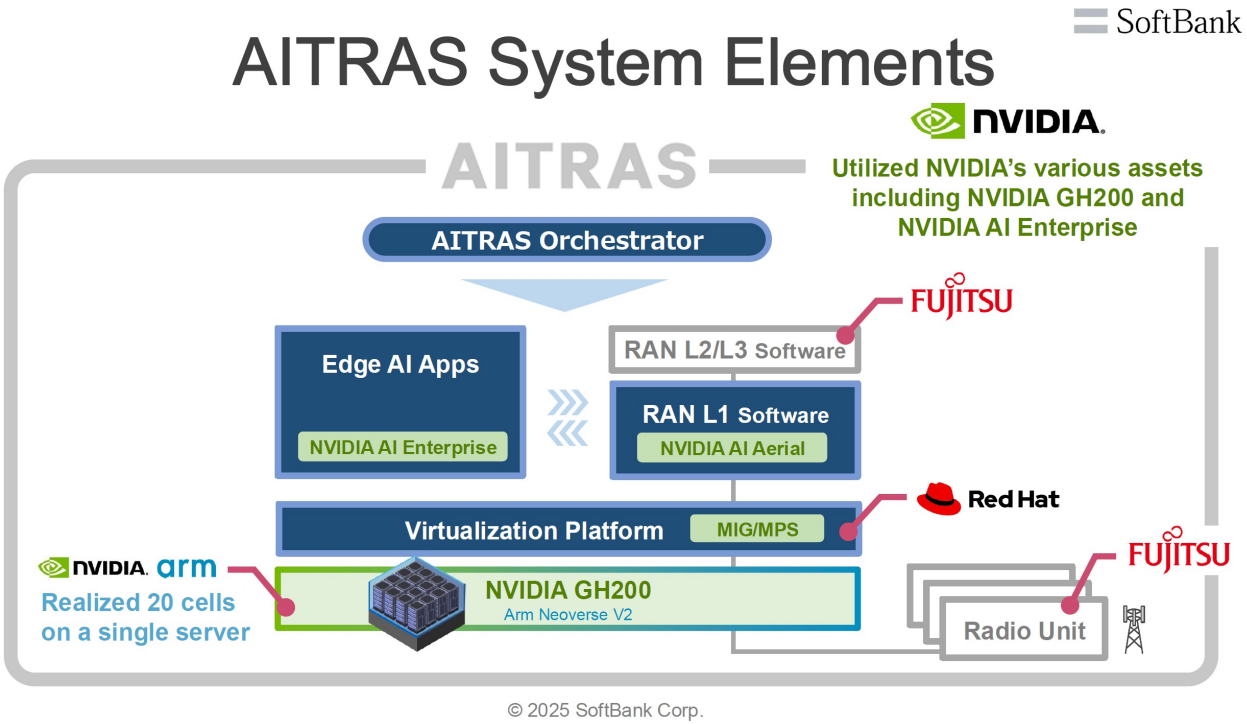
Source: AI-RAN Alliance

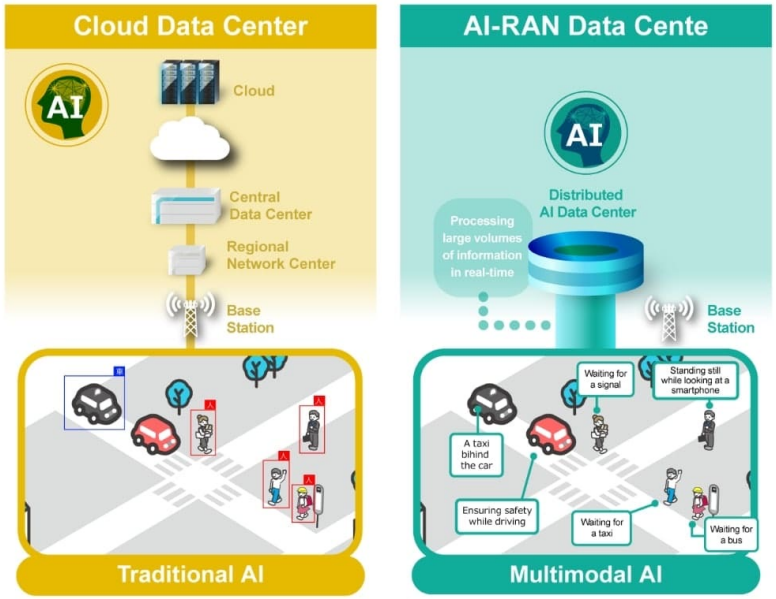
Nvidia – Leading the AI-RAN Movement



Source: Nvidia

Softbank AI-RAN Deployment



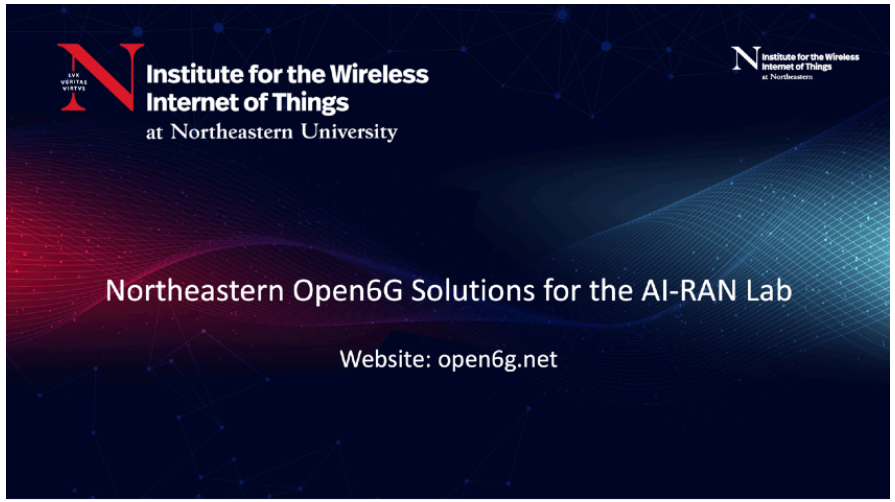


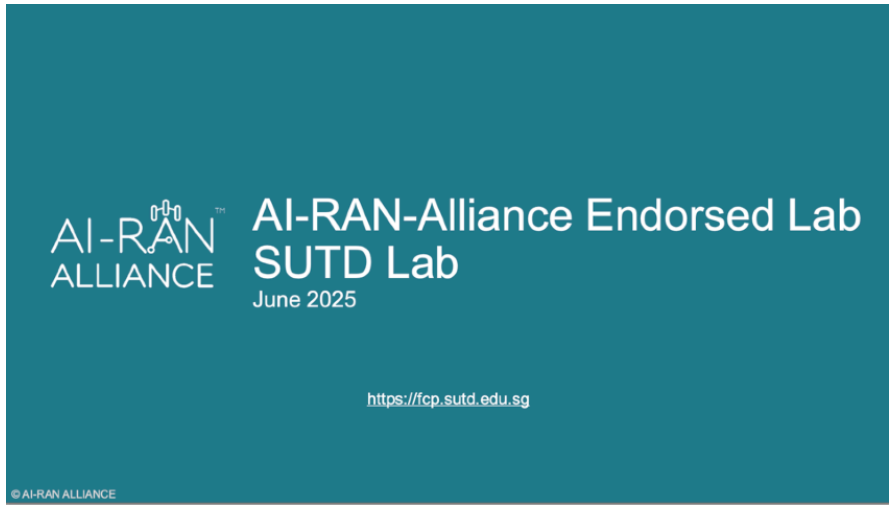
AI and image recognition have long been capable of detecting people and vehicles

Multimodal AI enables situational awareness The movement of numerous vehicles can reveal ongoing events in a city

Source: Softbank

AI-RAN Alliance Labs





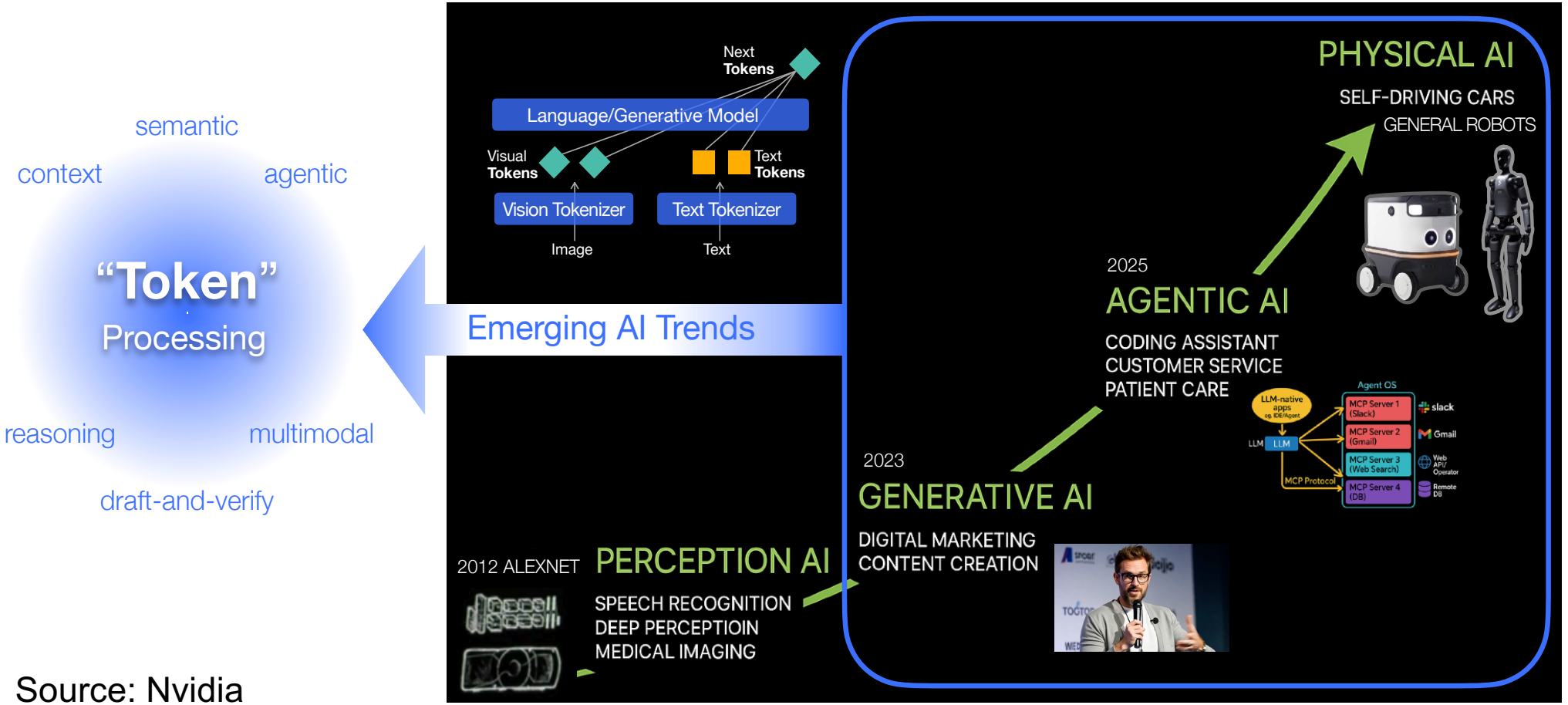
FCP
FUTURE COMMUNICATIONS
R&D PROGRAMME

SUTD
SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

AI-RAN for Token Communications

AI Trends – Token-based Processing

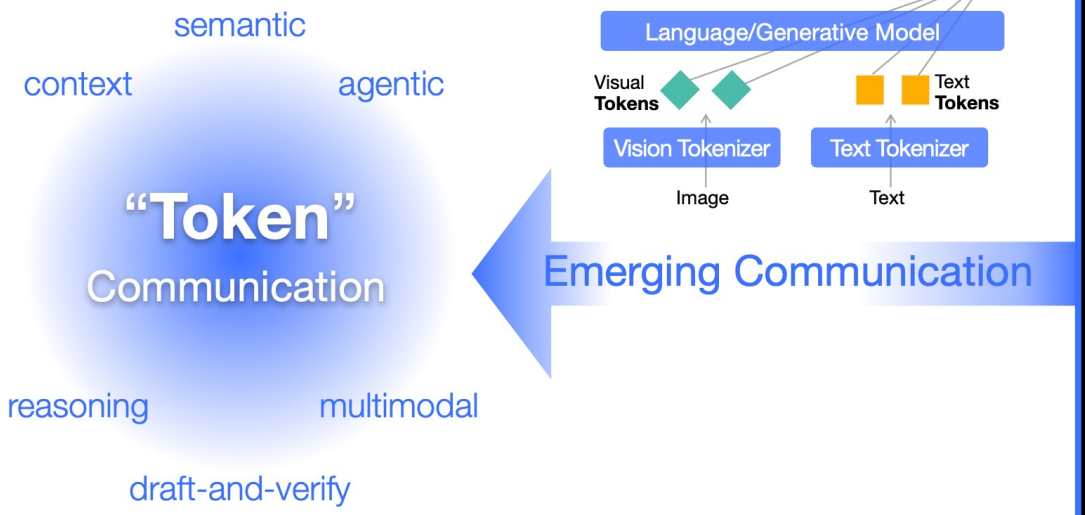
Tokens are the fundamental units of processing in **generative** & **large language model-based applications**.

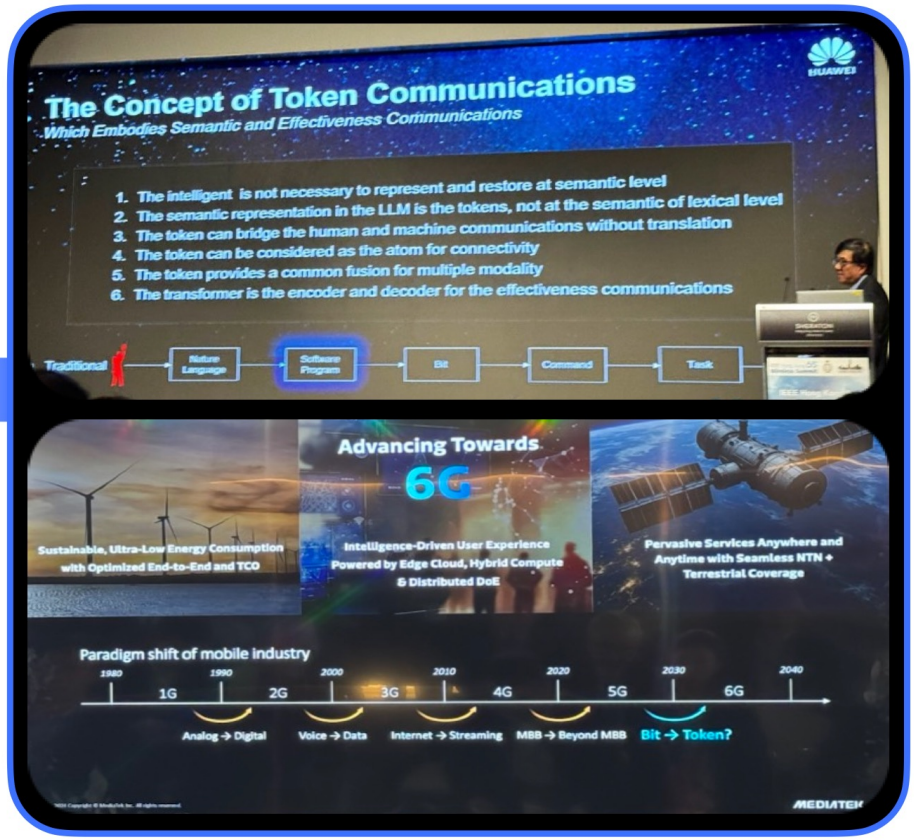


Source: Nvidia

AI Trends – Token-based Communication

To address the shift from bits to tokens, **token-aware** communication & **token-level** multiple access have emerged.



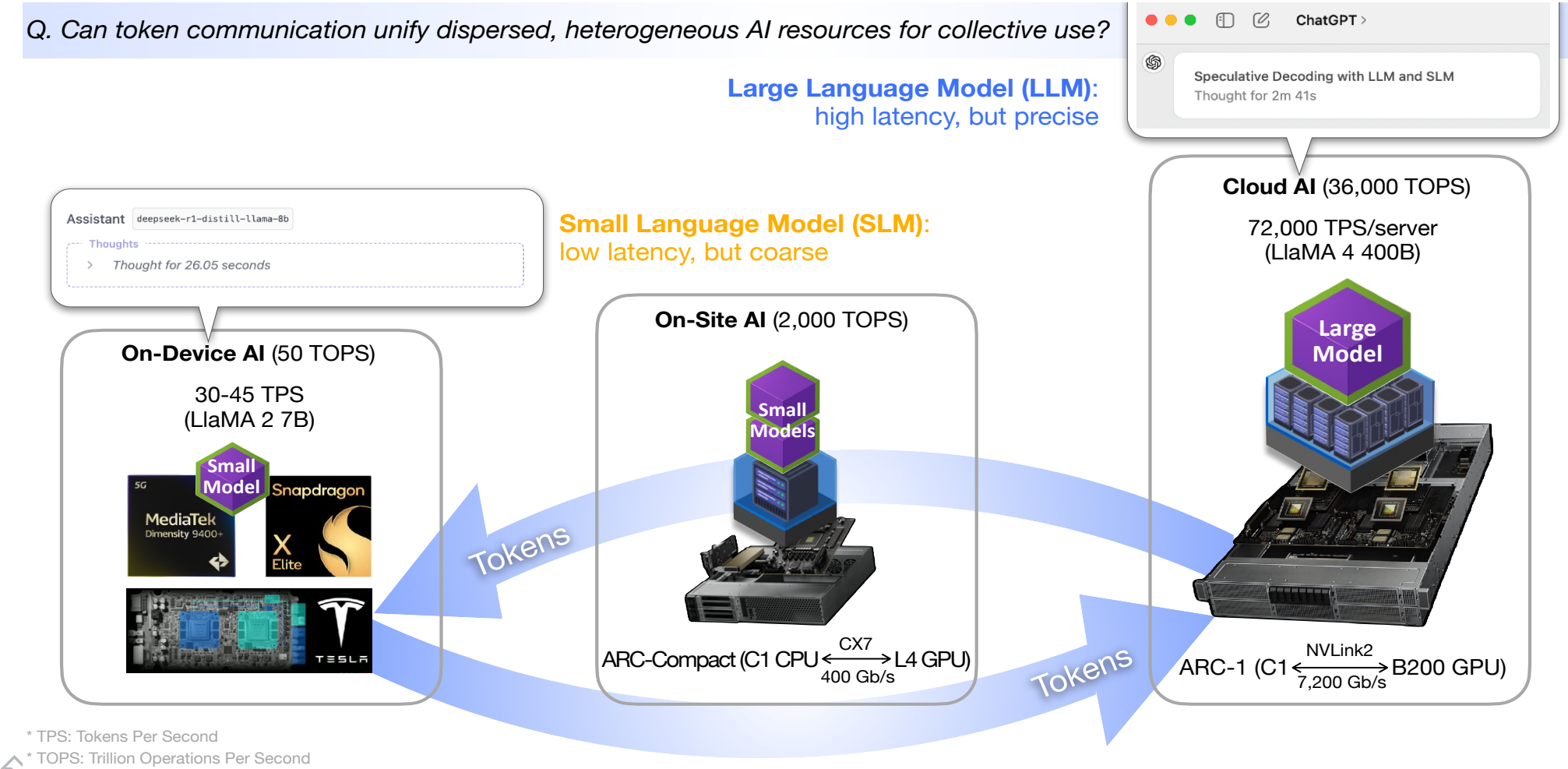


Source: Huawei, MediaTek

Source: Huawei, MediaTek

Challenges – Distributed & Heterogenous AI

Q. Can token communication unify dispersed, heterogeneous AI resources for collective use?

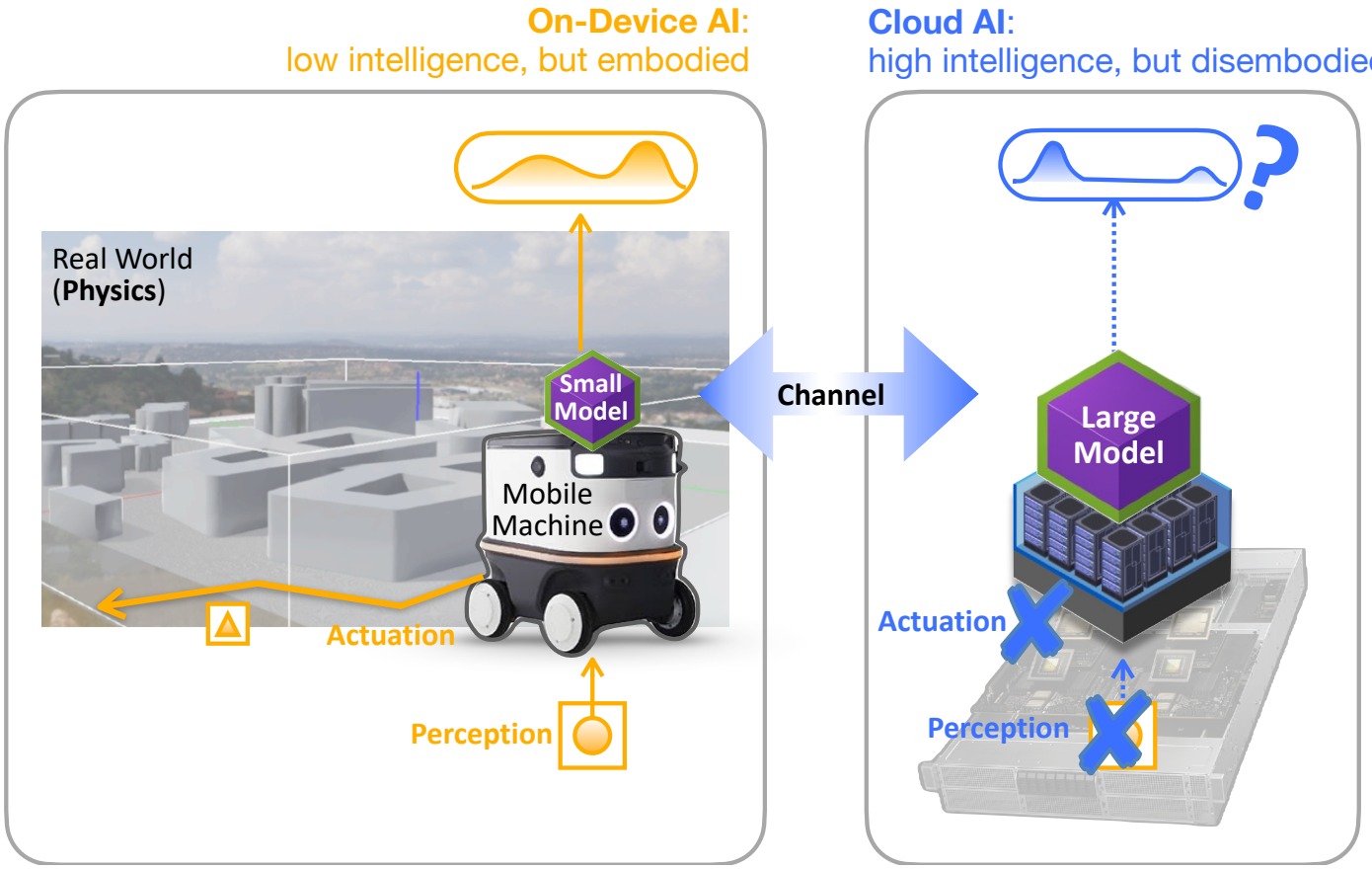


* TPS: Tokens Per Second
* TOPS: Trillion Operations Per Second

Source: Nvidia, MediaTek, Qualcomm, Tesla

Challenges – Disembodied Edge/Cloud AI

Q. Can embodied on-device AI communicate tokens with disembodied cloud AI?

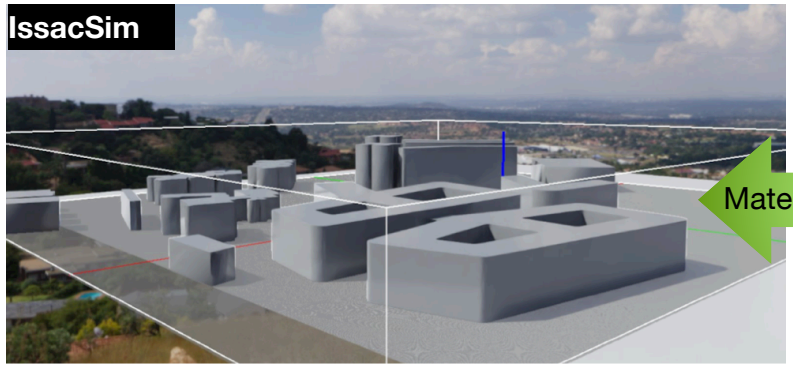


Enabling communication with embodied Cloud AI requires:

- **Perception**
- **Actuation**
- **Physics**
- **Channel**

Challenges – Twinned AI

Q. How can we synchronize digital-physical worlds, despite limited access to high-fidelity environmental information?



Material information mismatch

Robot HW mismatch

- **Fusion of real world and synthetic data**
- **Natural and artificial cognition**

Future Communications R&D Programme (FCP)

Future Communications R&D Programme (FCP) hosted by SUTD

SUTD integrates design, AI and technology into a holistic, interdisciplinary education and research experience. This unique approach encourages our students to push the boundaries of innovating solutions to real-world problems.

FCP was set up in May 2021 with S$68.7M initial investment from the government

- Through research grant calls: Build up Singapore's communications and connectivity R&D capabilities.
- Establish international stature by partnering and participating with established countries, companies and committees.
- Through translation grant calls: Build/grow companies to keep pace with developing/translating innovative applications and services.

FCP received S$20M additional investment from the government in 2025 to continue capability building and grow collaborations.

THE FCP TEAM AT SUTD

LEADERSHIP

Prof Tony Quek
FCP Director

Prof Chen Binbin
FCP Deputy Director
(Industry)

Prof Park Jihong
FCP Deputy Director
(Research)

ADMIN TEAM

Ms Dawn Chia
Admin Team Lead

Ms Li Jiayan
Deputy Manager

Ms Stacey Zhang
Deputy Manager

TECHNICAL TEAM

Dr Ngo Van Mao
Technical Team Lead

Mr Yeo Siow Long
Communications Testbed Engineer

Mr Liang Xian Loong
Communications Testbed Engineer

Mr Joshua Dean Samjaya
Research Assistant

Mr Hariz Yet
Communications Testbed Engineer

Mr Nguyen Thanh Tam
GPU Testbed Engineer

Mr Ngo Van Tuan
Software Engineer

Dr Wang Peng
Research Fellow

Mr Vishal Choudhary
Research Associate

Dr Gu Zhouyou
Research Fellow

Mr Lim Yi Shen
Research Assistant

FCCLab Testbed: Key Features – **AI-Native Design**

AI-enabled / Software-defined: deployable in virtualized environment & leverage AI for intelligent orchestration

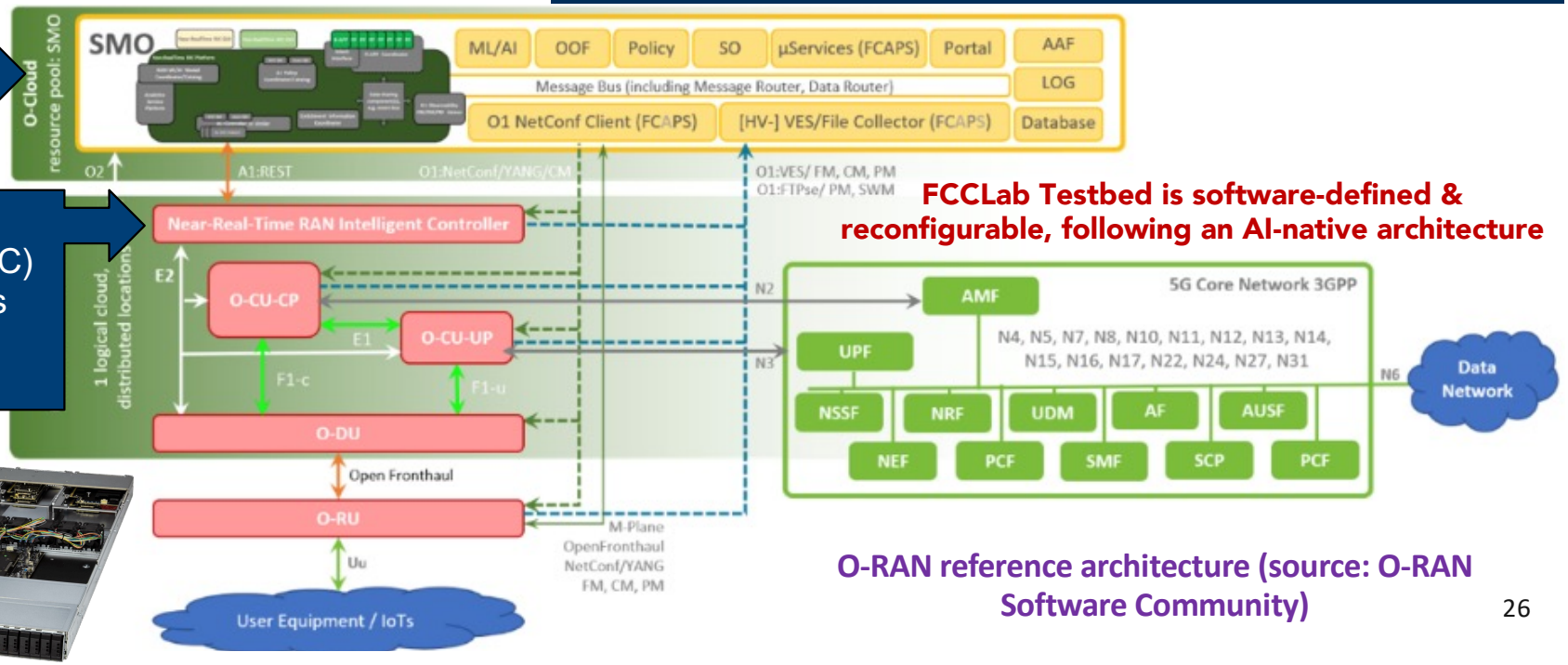
Reconfigurable: customizable for diverse use cases & adapt to changing environment / user needs

Success stories (thanks to our AI-native design):

- FCCLab is the only team from Asia in the main stage of RIC Forum 2024 (organized by US NTIA), presenting its AI-based rApp for mitigating network interference
- FCCLab is among the first AI-RAN Alliance labs in the world

AI-based SMO can help operators automatically optimize their networks

xApps and rApps running on RAN Intelligent Controller (RIC) allow AI / software companies to accelerate the innovation in the telco ecosystem



FCCLab Testbed is software-defined & reconfigurable, following an AI-native architecture

O-RAN reference architecture (source: O-RAN Software Community)

Beyond 5G, Programmable and AI RAN Innovation

The Future Communications Connectivity Lab (FCCLab) @SUTD is an all-in-one store that provides full services in:

Research & Development

Innovation with 5G, 6G, programmable and AI RAN

Test & Integration

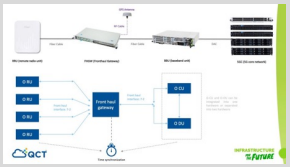
Certification, verification, and interoperability

System Integration

System design with AI and technologies maximizes performance and business success

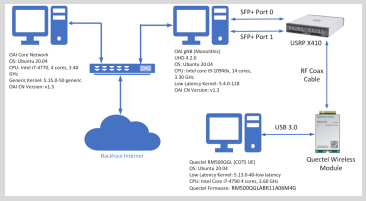
Testbeds and Platforms

Programmable and AI RAN



Single-Vendor 5G Open RAN Network

Multi-Vendor 5G Open RAN Networks



Open Source Programmable 5G Network

Application Verticals

Drone Arena

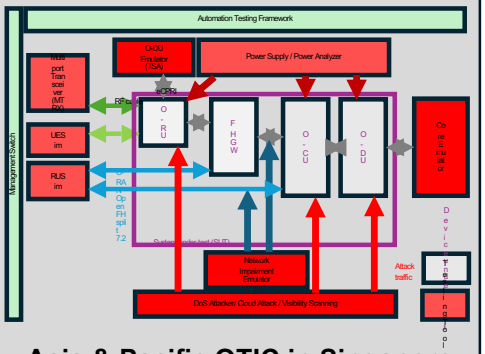
Cyber-Physical Campus

Metaverse Testbed

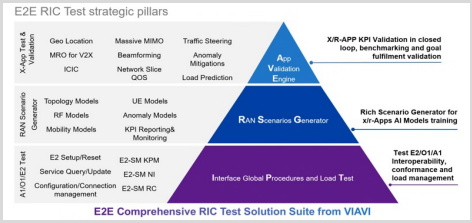
Testing & Integration

E2E - TESTED | CERTIFIED | INTEROPERABLE

O-RAN Certification and Badging

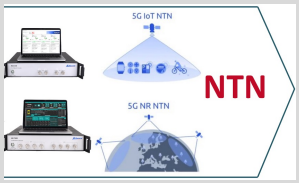


Asia & Pacific OTIC in Singapore

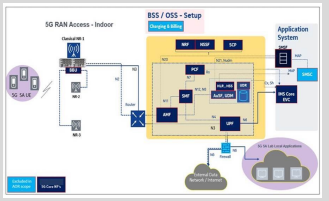


RIC/xApp/rApp Testing Capabilities

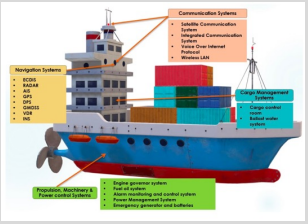
Emulation & Cyber Range



Non-Terrestrial-Network Testbed



Nokia 5G Cyber-Range Testbed



Maritime Shipboard OT Testbed

FCC Lab & AI-RAN Lab Equipment

5G Core

OAI 5G Core Network

Open5GS Core Network

Fraunhofer open5gcore

Keysight Core Sim

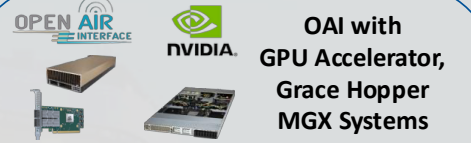
QCT

QCT 5G Core

Ataya 5G Core

CU, DU

OAI with GPU Accelerator, Grace Hopper MGX Systems



OAI, srsRAN with USRP on Bare Metal Server

Synergy O-CU & O-DU

SynaXG BBU

HTC BBU

QCT BBU

All-in-One

mmWave gNB

SC-250

Fronthaul

Keysight TSA

PTP and SyncE Provider and Tester

Paragon Neo

Falcon Switch PTP Grandmaster

Dell PTP Switch

NE3 Network Emulator

Packet Capture Appliance

RU

LiteOn O-RU

Delta O-RU

Foxconn 鴻海科技集團

Compal O-RU

SERA O-RU

WNC O-RU

X410, X310, B210

Indoor O-RU

Outdoor O-RU

RuSIM

Channel

Adjustable RF Attenuator J720x

MTRX

PropSIM

RF shield box

OTA

FR1: 3400 – 3450 MHz

FR2: 25900 – 26300 MHz

SUTD Campus

Aerial Arena

UE

5G-native Drone

Humanoid robot

Quadruped robot dog

AR/VR Goggle

5G 4K Camera (Pegatron Nura4K)

5G Cell Phone

TEMS Pocket

Wheeled Robots (AMR, Limo ROS)

Pegatron FR2 CPE

PEGATRON 5G

5G Dongle (Pegatron, APAL)

UeSIM

VIAVI TM500

Server for RIC x/rAPP

RIC Test

Vector signal generator

Spectrum Analyzer

Power Analyzer

ALifecom NE7500/ NE6000

CMX500

CMW500


FCP
FUTURE COMMUNICATIONS
R&D PROGRAMME


SINGAPORE UNIVERSITY OF
TECHNOLOGY AND DESIGN

Thank you!