

Briefing on AI-RAN

New Technology Development

Announcement

SoftBank Corp.

Speaker

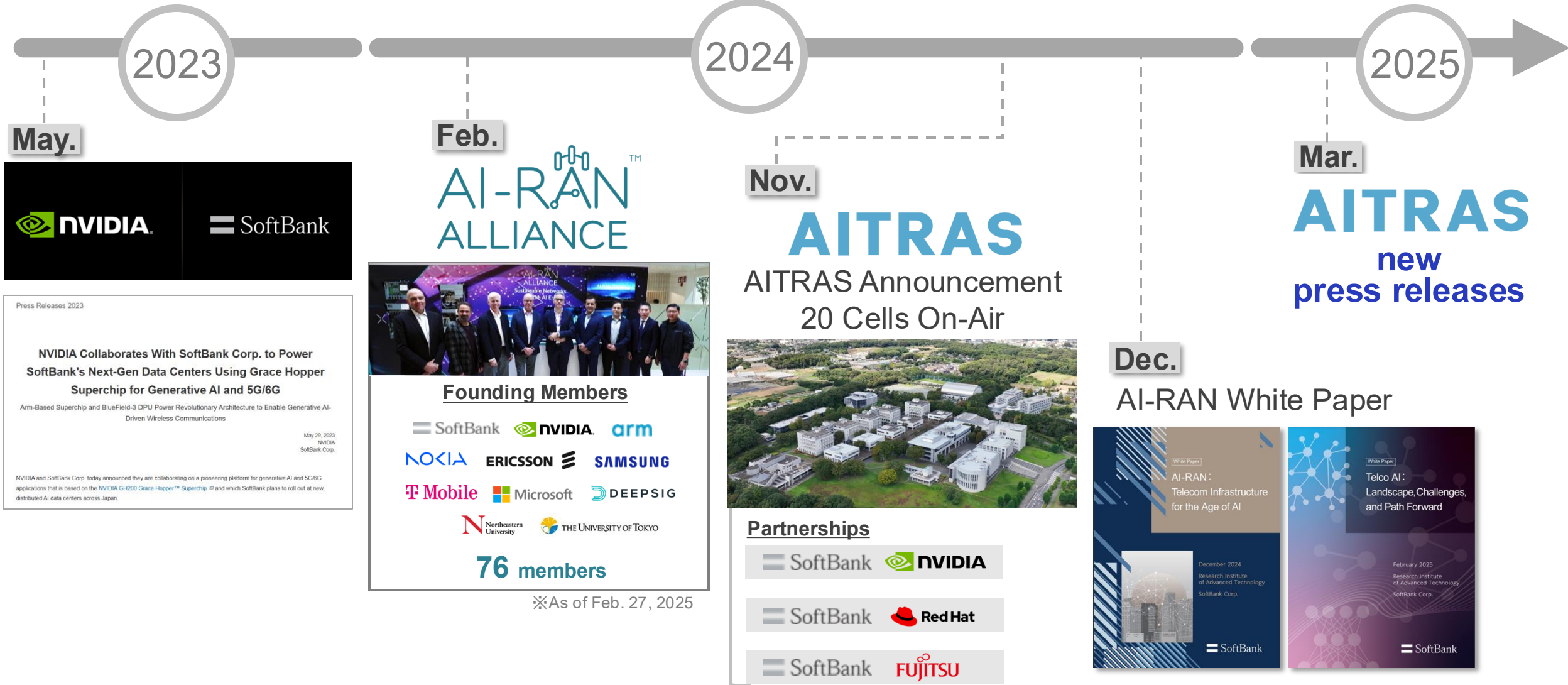


Ryuji Wakikawa

VP, Head of Research Institute of Advanced Technology,
SoftBank Corp.

SoftBank's Initiatives for AI-RAN

SoftBank's Initiatives for AI-RAN



AITRAS

SoftBank's original product
based on the AI-RAN concept

AITRAS System Elements

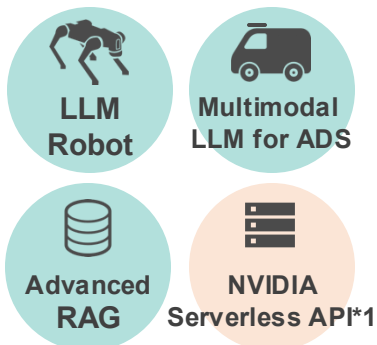
: SoftBank Developed

AITRAS

AITRAS Orchestrator

Manage & control software

Provide Edge AI Service Menu



Edge AI Apps

NVIDIA AI Enterprise*1

RAN L2/L3 Software

RAN L1 Software

NVIDIA AI Aerial

Advanced radio signal processing software

Virtualization Platform

MIG/MPS

Built & implemented virtualized infrastructure

NVIDIA GH200*2
Arm Neoverse V2

Radio Unit



*1: Serverless API powered by NVIDIA AI Enterprise

*2: NVIDIA GH200 Grace Hopper Superchip

AITRAS System Elements



Utilized NVIDIA's various assets including NVIDIA GH200 and NVIDIA AI Enterprise

AITRAS Orchestrator

Edge AI Apps

NVIDIA AI Enterprise

RAN L2/L3 Software

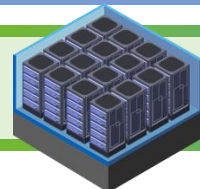
RAN L1 Software

NVIDIA AI Aerial

Virtualization Platform

MIG/MPS

 **NVIDIA.** **arm**
Realized 20 cells on a single server



NVIDIA GH200
Arm Neoverse V2



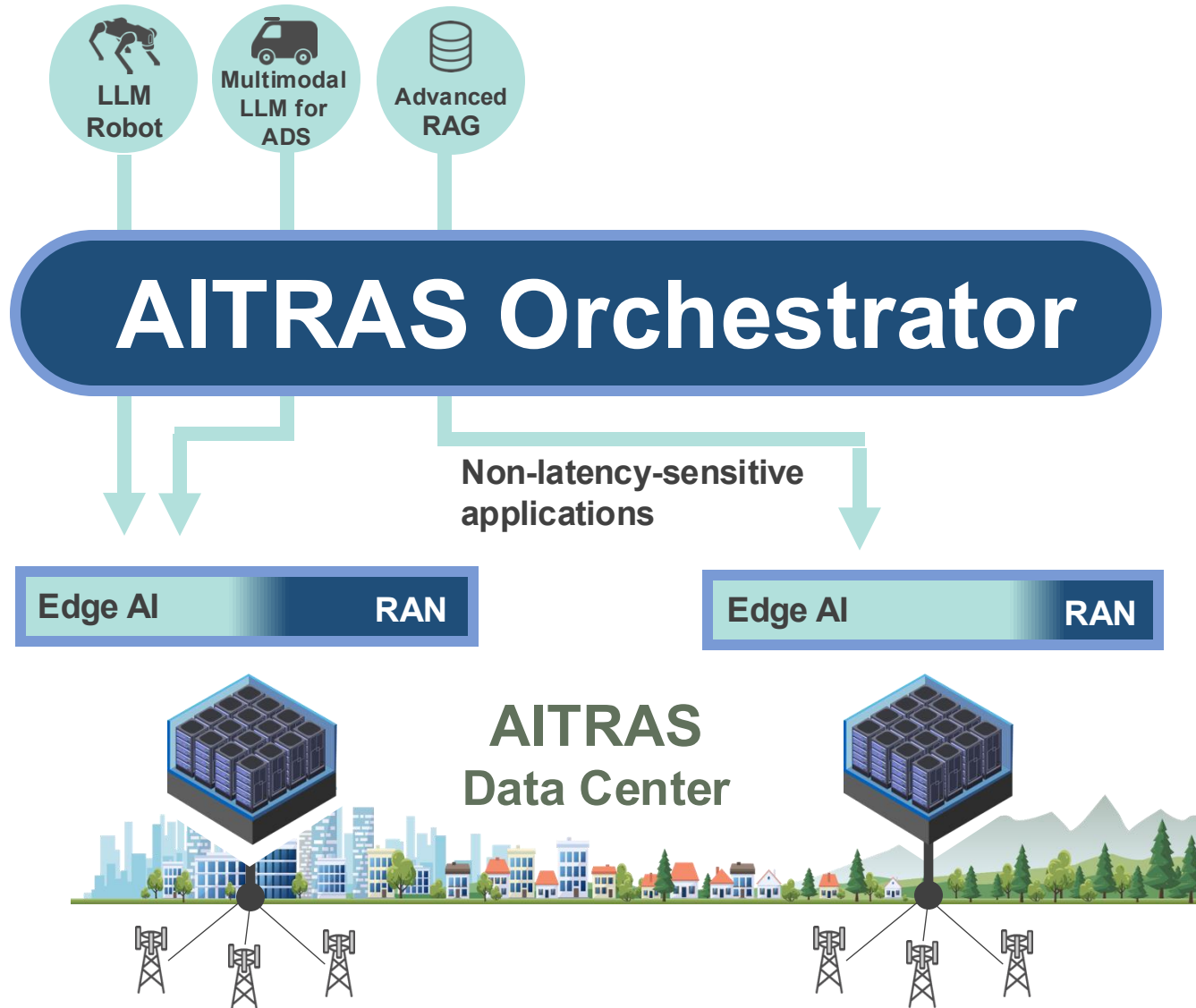
 **Red Hat**



Radio Unit

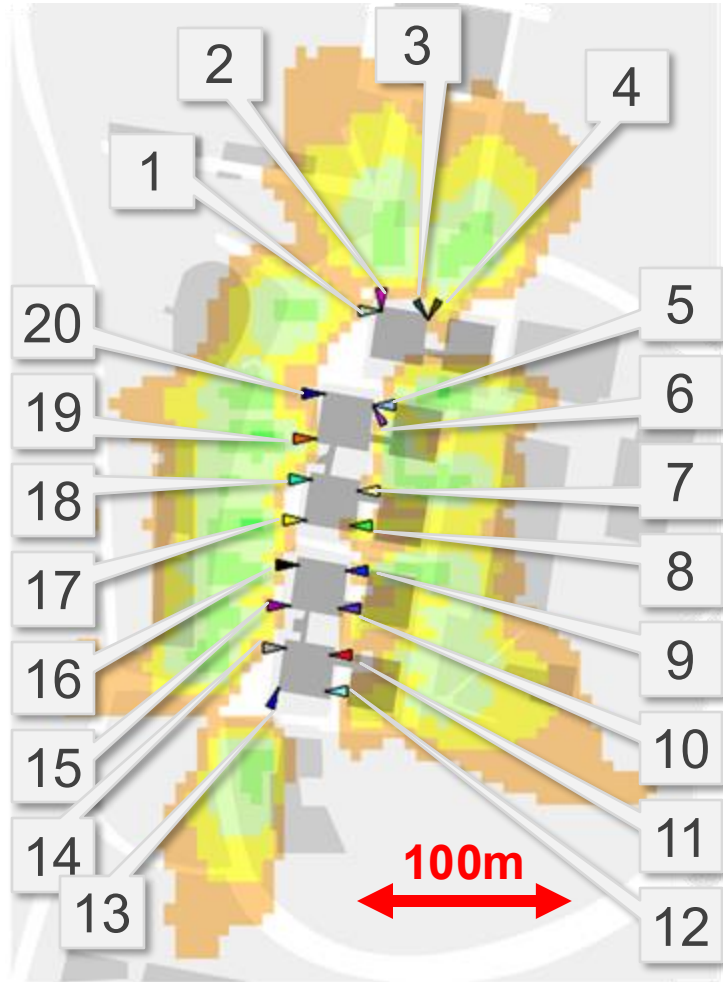


AITRAS Orchestrator



- Automatic resource allocation by AI
- Dynamic changes of server roles
- NVIDIA Serverless API, NVIDIA AI Enterprise Supported

AITRAS 20 Cells On-Air



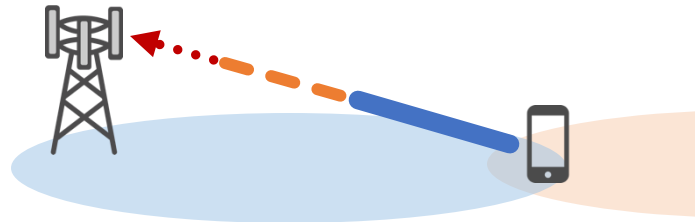
Announcing New Developments

①

SoftBank Corp. Demonstrates RAN Performance Enhancement with AI Technology

Developing AI-for-RAN Technologies

UL Signal Processing



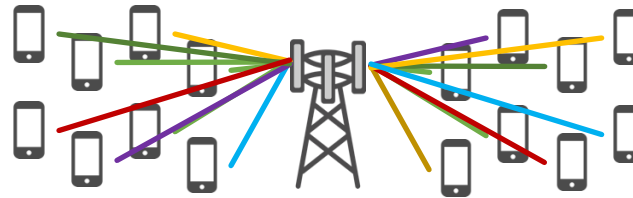
Low signal in noisy areas and cell edges affects channel error

AITRAS



Channel
Interpolation

MU-MIMO



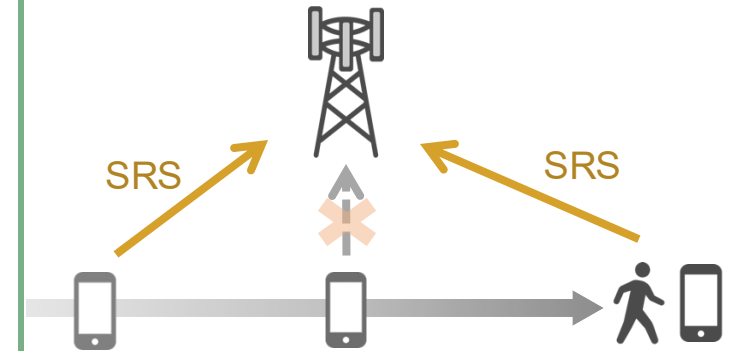
Massive matrix calculations for device pairing

AITRAS



User Pairing
Optimization

Beamforming



More devices cause wider SRS intervals

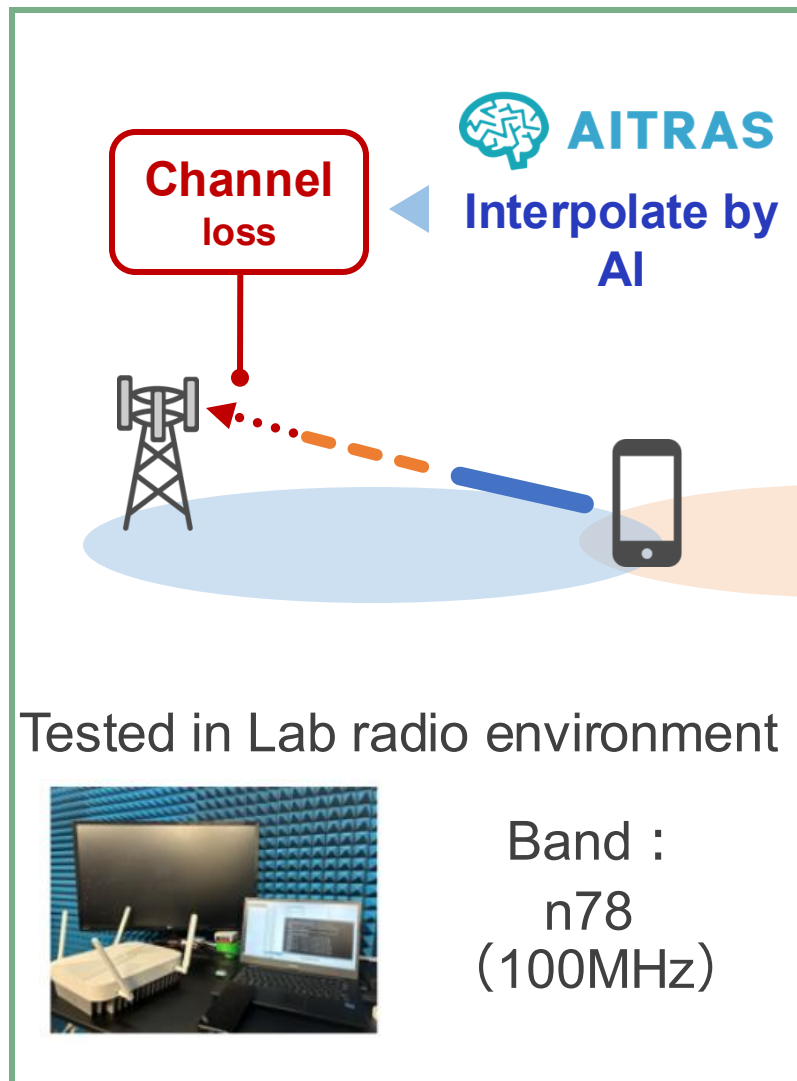
SRS : Sounding Reference Signal

AITRAS



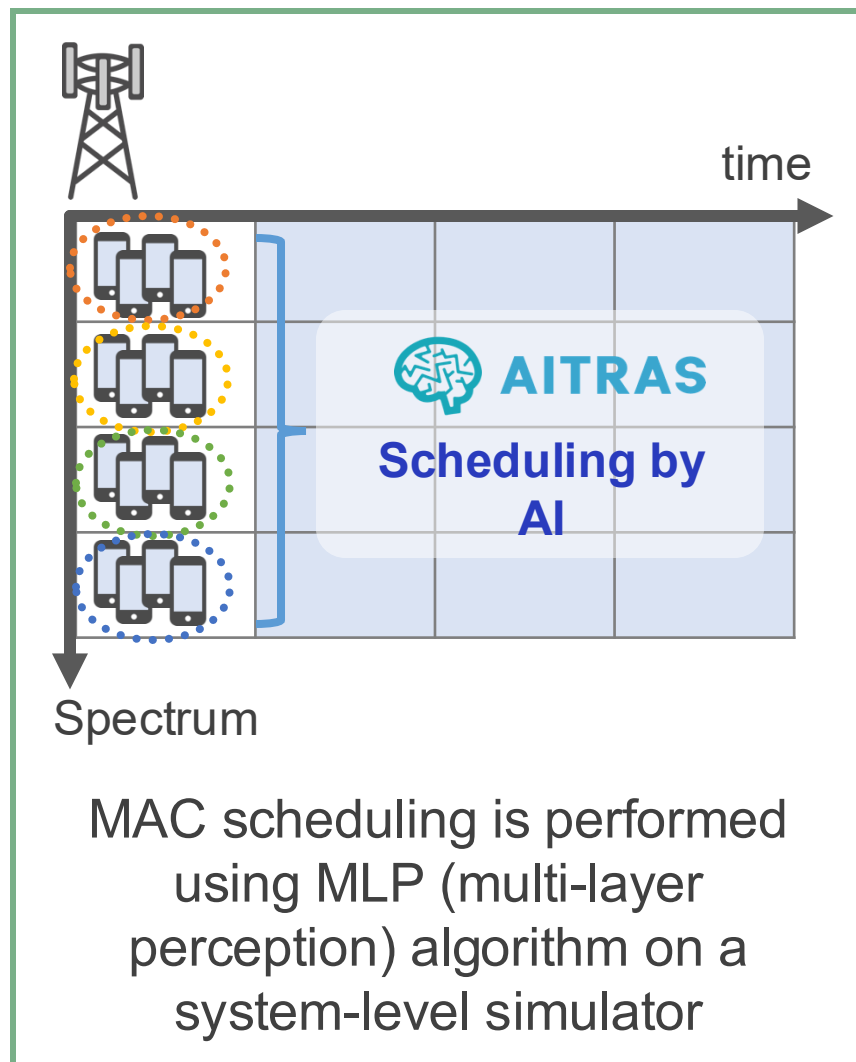
SRS
Prediction

UL Channel Interpolation



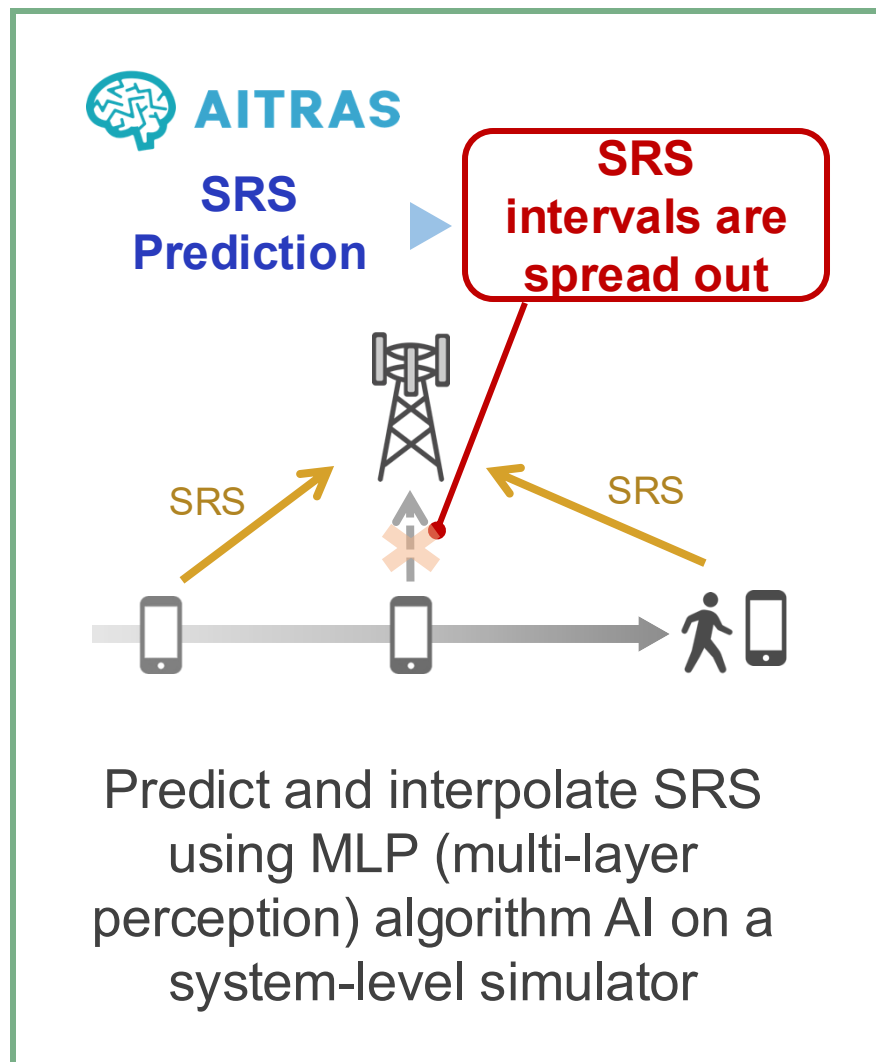
**Throughput improves
by approx. 20%**

MU-MIMO : User Pairing Optimization



**Cell throughput improves
by approx. 9%**

Beamforming : SRS Prediction



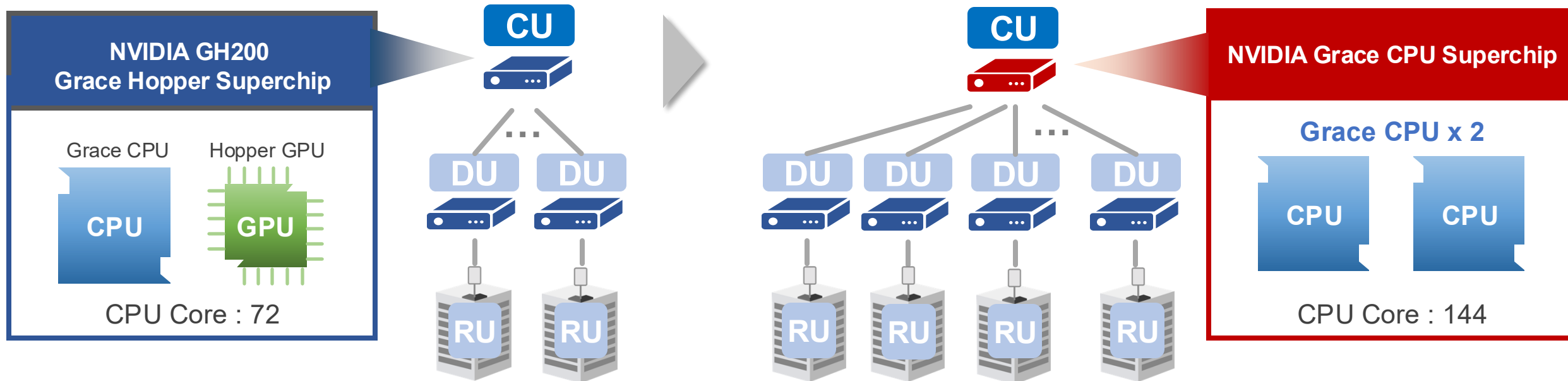
**User throughput improves
by approx. 13%**

②

SoftBank Corp.'s "AITRAS" Implements
Centralized and Distributed AI-RAN
Architectures
Using Arm-based NVIDIA Platform

Evolution of AITRAS①

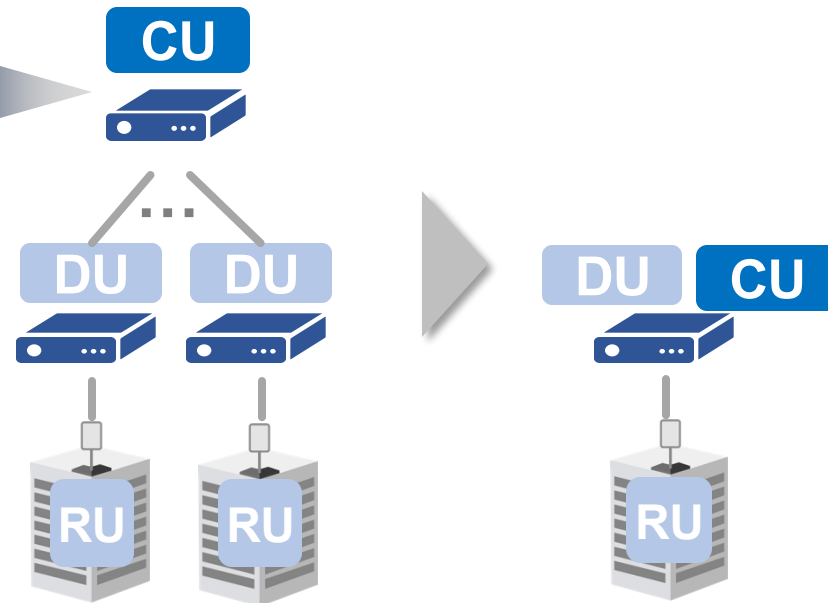
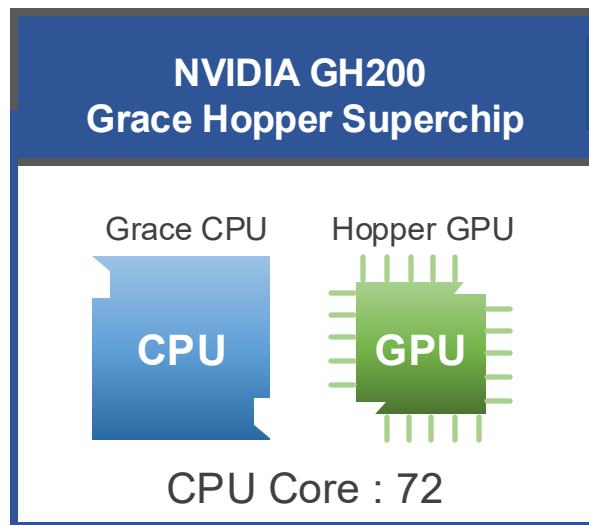
CU Implementation for NVIDIA Grace CPU Superchip Server



CU capacity is approx. **doubled**
by NVIDIA Grace CPU

Evolution of AITRAS②

Implement both DU and CU on a single NVIDIA GH200 Server

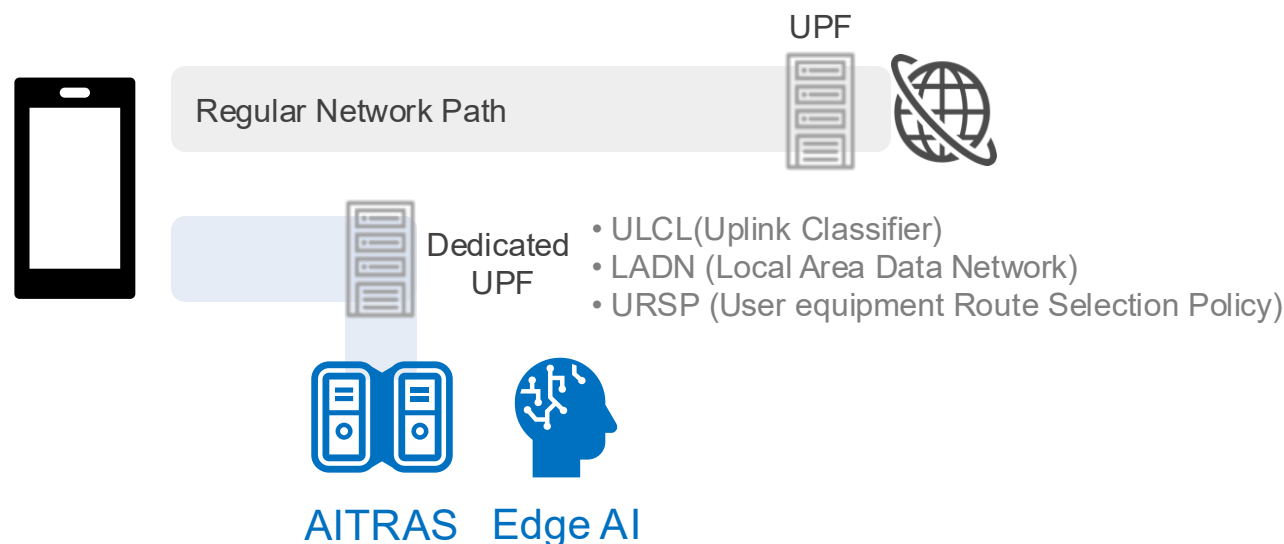


**Deploy D-RAN
configured AITRAS**
in areas with increasing
local AI demand such as
enterprise facilities

③

SoftBank Corp. Develops
Secure Access Feature to Edge AI Servers
on “AITRAS”
Utilizing Local Breakout Technology

Edge Routing for AI-on-RAN Applications

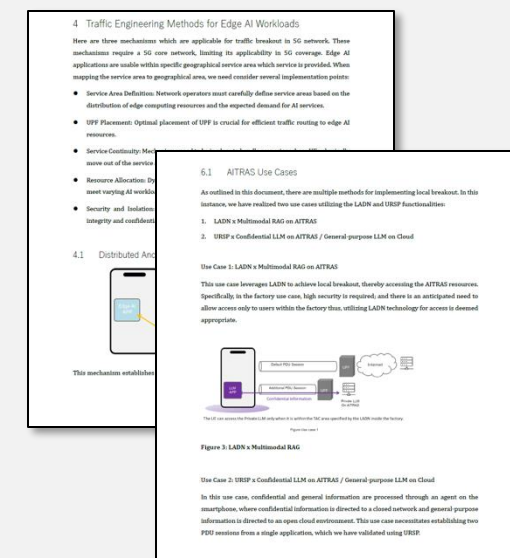
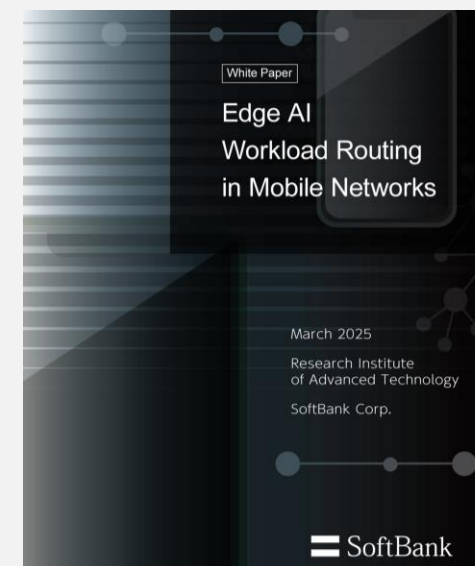


Routing technology enabling dynamic path selection at the network edge

- Smart switching between internet and local breakout paths
- Expanded deployment options for AI-on-RAN

White Paper Available

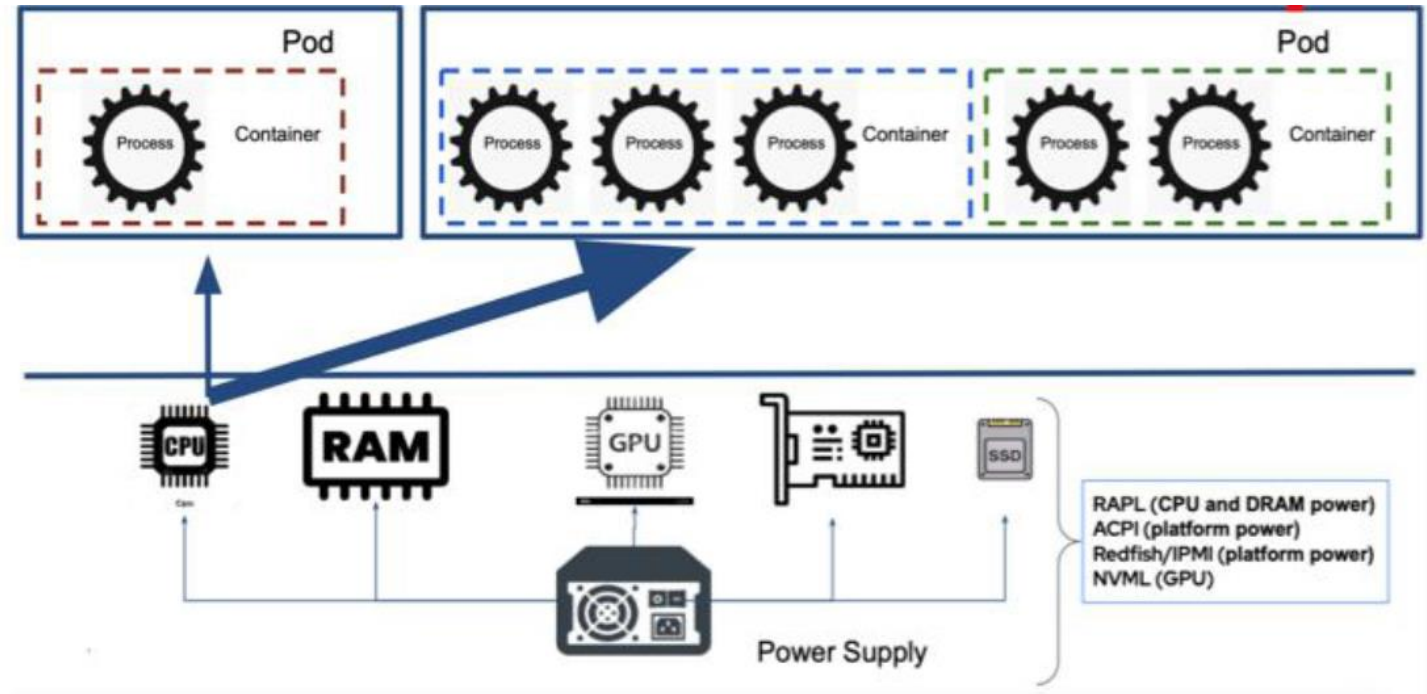
“Edge AI Workload Routing in Mobile Networks”



④

SoftBank Corp. and Red Hat Develop Solution to Optimize Power Consumption in AI-RAN Data Centers

Red Hat Kepler Project

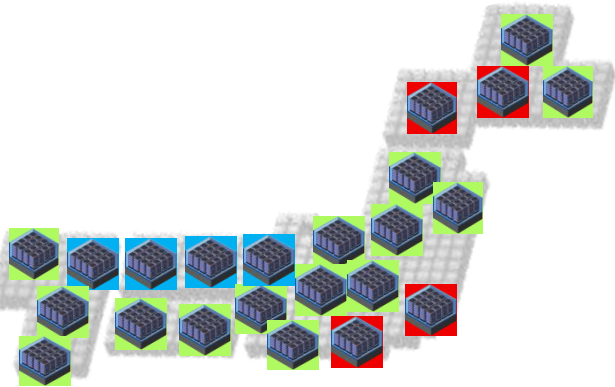


Red Hat's Kepler project:

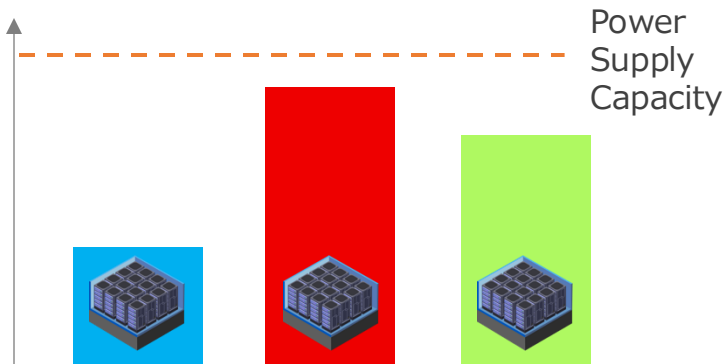
An open-source initiative that helps **measure and optimize energy consumption of the Pod on Red Hat OpenShift**

AITRAS Orchestrator x Red Hat Kepler

Regional differences in power supply and demand



Power utilization rate



AITRAS Orchestrator

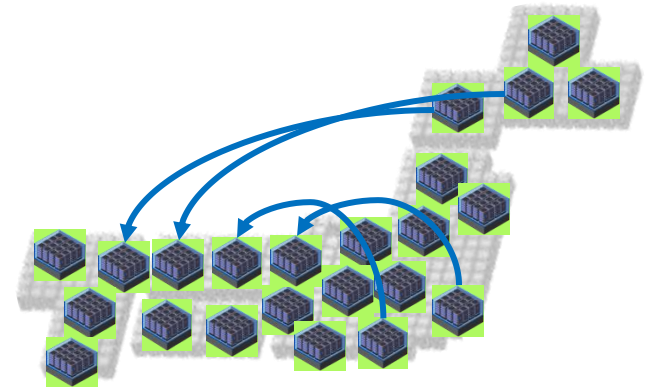
Deploy pods on OpenShift to optimal DC based on:

- ✓ Location
- ✓ Latency requirements
- ✓ Resource availability
- ✓ **+ Power Consumption**

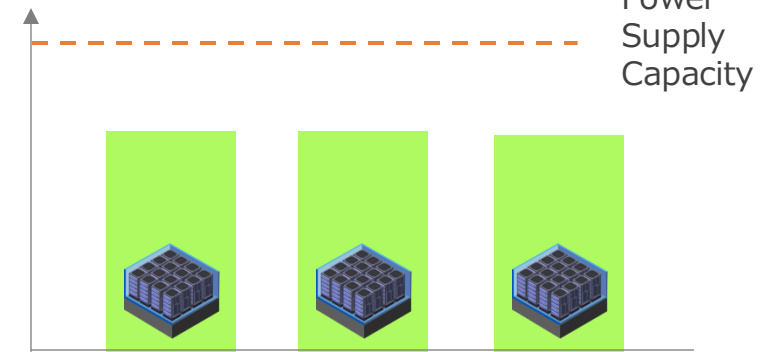
Red Hat Kepler

Predict pods power usage

Predict pods power usage and optimize placement



Power utilization rate



AITRAS Orchestrator x Red Hat Kepler

Cluster 1

Power supply capacity

10,000 J

0 J

Cluster 2

Power supply capacity

3,000 J

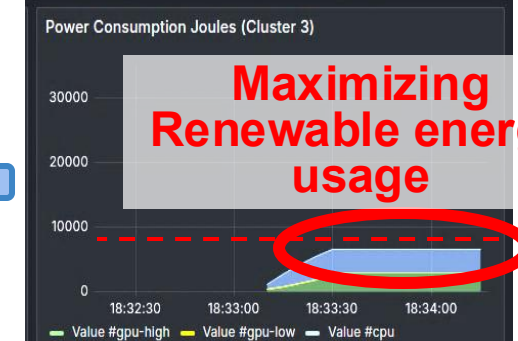
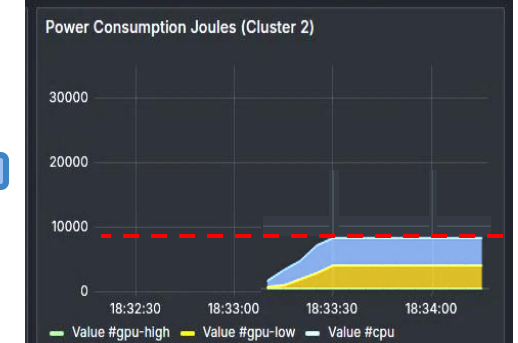
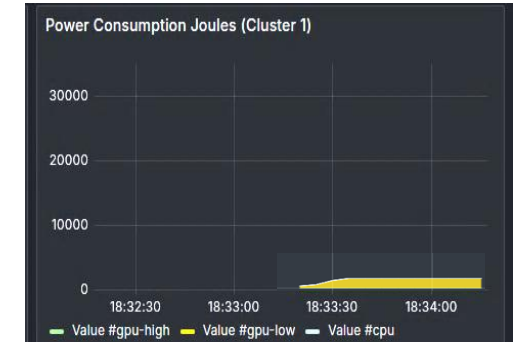
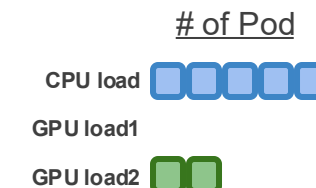
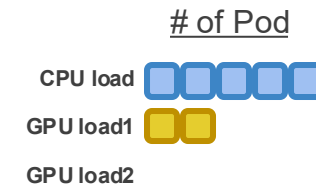
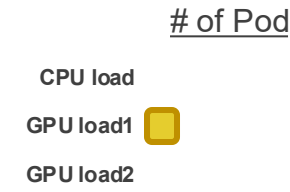
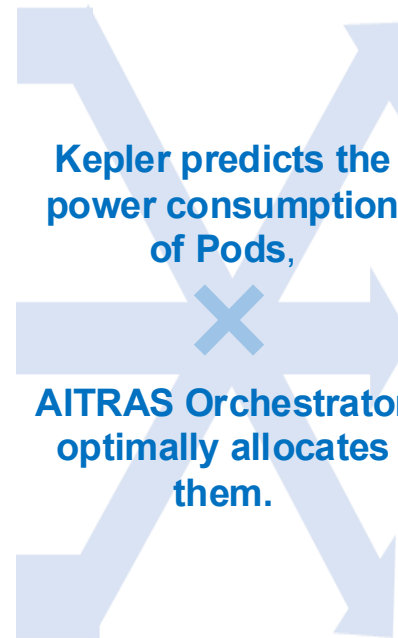
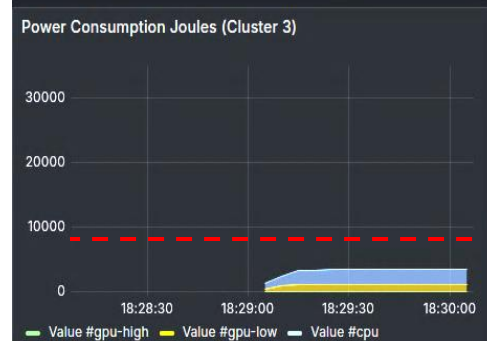
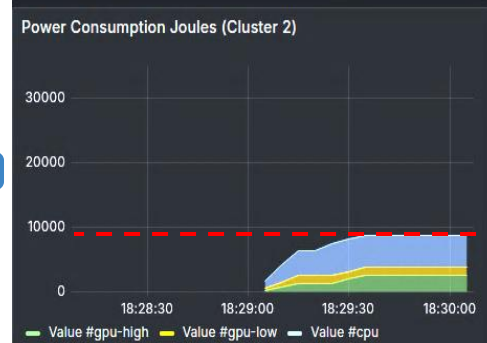
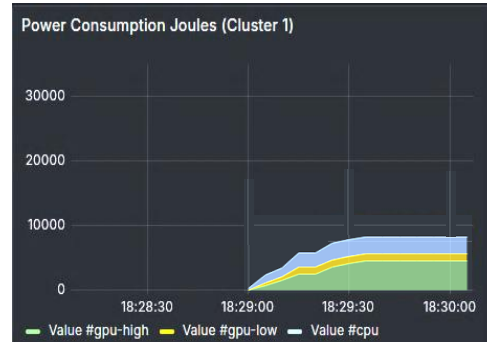
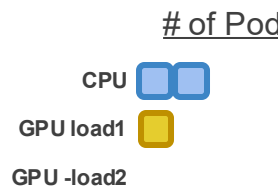
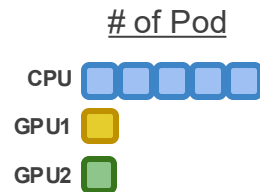
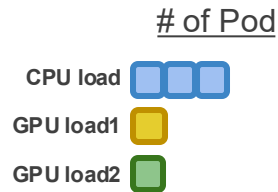
8,000 J

Cluster 3

Power supply capacity

3,000 J

8,000 J

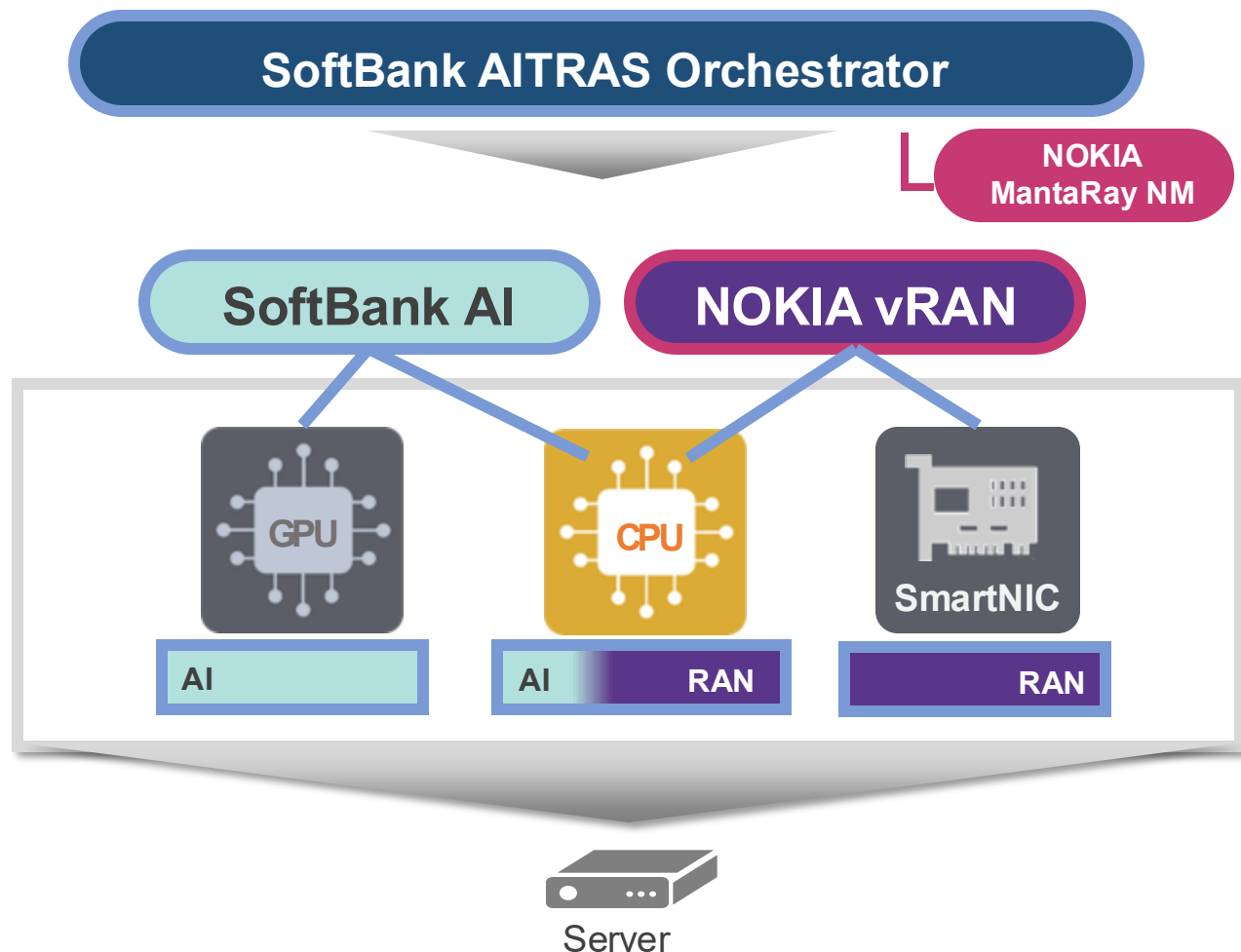


Maximizing Renewable energy usage

⑤

SoftBank Corp. and Nokia Achieve
AI and vRAN Coexistence
with Automated Optimal Resource Allocation
on a Single Server

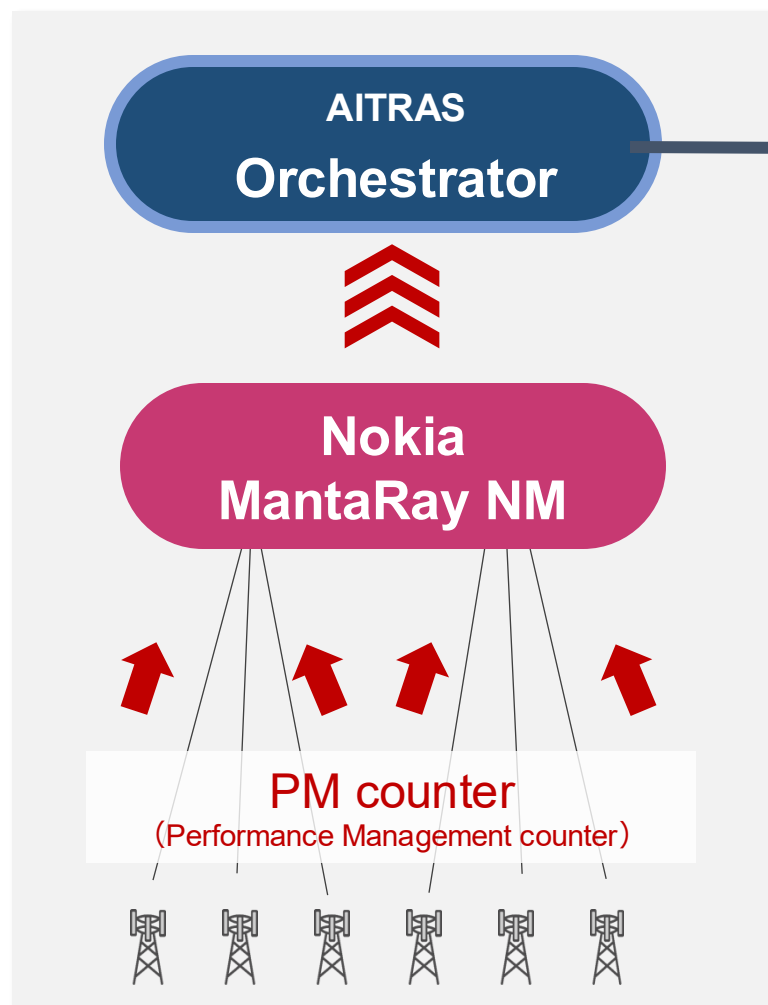
AI and vRAN Coexist on a Server



- SmartNIC that handles part of vDU processing extended to support GPU servers
- AITRAS Orchestrator optimizes resource allocation

SoftBank and Nokia enabled AI-and-RAN utilizing Nokia's vRAN software.

AITRAS Orchestrator x Nokia MantaRay NM



- Predict vRAN traffic behavior
- Optimize AI and vRAN resource allocation

