SoftBank

AI-RANの新規技術開発発表
事前説明会

説明者

船吉 秀人

ソフトバンク株式会社

先端技術研究所 先端無線統括部

統括部長

ソフトバンク
AI-RANの取り組み

ソフトバンクのAI-RANの取り組み

2023 2024 2025

5月.

NVIDIA SoftBank

Press Releases 2023

NVIDIA Collaborates With SoftBank Corp. to Power SoftBank's Next-Gen Data Centers Using Grace Hopper Superchip for Generative AI and 5G/6G

Arm-Based Superchip and BlueField-3 DPU Power Revolutionary Architecture to Enable Generative AI-Driven Wireless Communications

May 29, 2023
NVIDIA
SoftBank Corp.

NVIDIA and SoftBank Corp. today announced they are collaborating on a pioneering platform for generative AI and 5G/6G applications that is based on the NVIDIA GH200 Grace Hopper™ Superchip and which SoftBank plans to roll out at new, distributed AI data centers across Japan.

2月

AI-RAN ALLIANCE™

Founding Members

SoftBank NVIDIA arm
NOKIA ERICSSON SAMSUNG
T-Mobile Microsoft DEEPSIG
Northeastern University THE UNIVERSITY OF TOKYO

76 社

※2025/2/27時点

11月

AITRAS

AITRAS発表
20セルオンエア

パートナーシップ

SoftBank NVIDIA
SoftBank Red Hat
SoftBank FUJITSU

12月

AI-RAN White Paper

3月

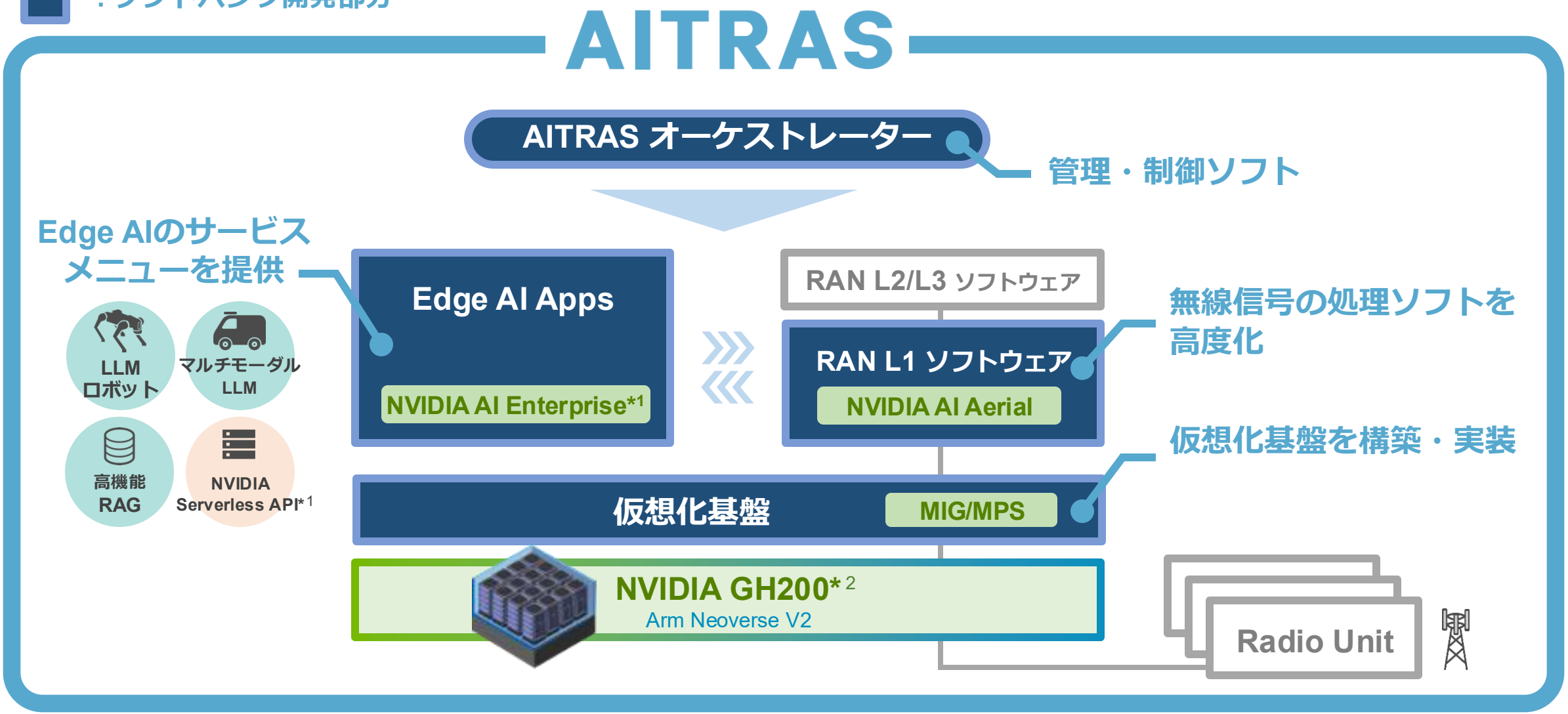
AITRAS
新規発表

AITRAS

AI-RANコンセプトにもとづいた
ソフトバンクオリジナルのプロダクト

AITRASのシステム構成

：ソフトバンク開発部分



*1:Serverless API powered by NVIDIA AI Enterprise
*2: NVIDIA GH200 Grace Hopper Superchip

AITRAS のシステム構成

AITRAS

NVIDIA GH200や
NVIDIA AI Enterpriseなど
さまざまなアセットを活用

AITRAS オーケストレーター

Edge AI Apps

NVIDIA AI Enterprise

RAN L2/L3 ソフトウェア

RAN L1 ソフトウェア

NVIDIA AI Aerial

仮想化基盤

MIG/MPS

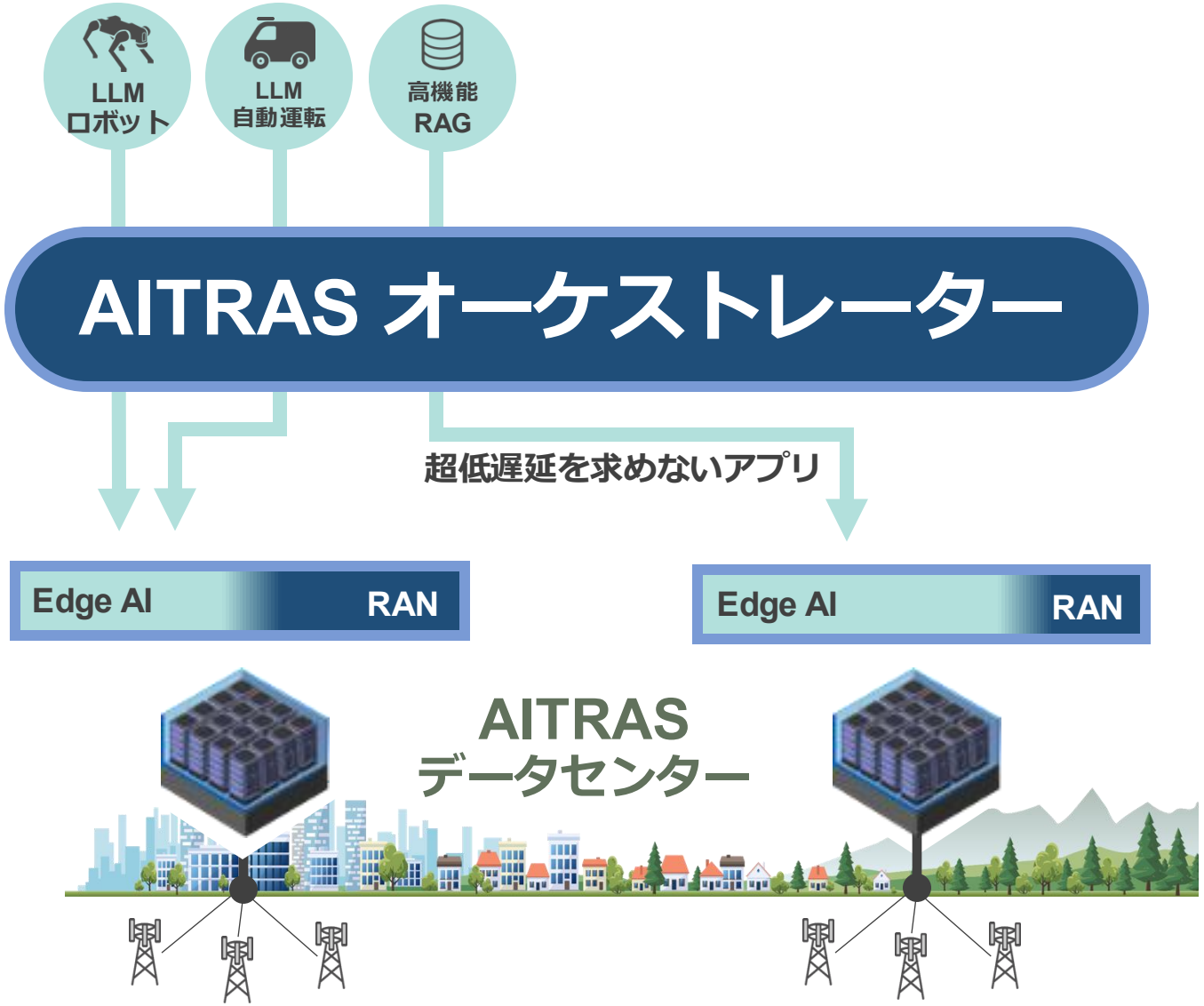
NVIDIA GH200

Arm Neoverse V2

1サーバーで
20セルを実現

Radio Unit

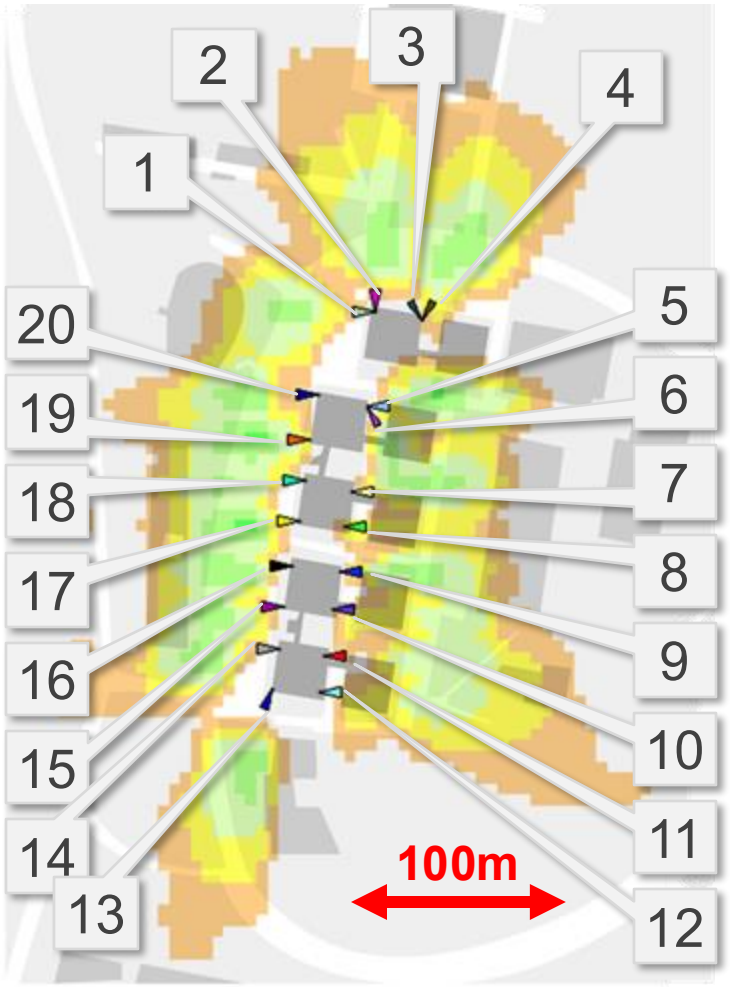
AITRAS オーケストレーター



- AIによる自動リソース割当
- サーバーの役割変更
- NVIDIA Serverless API,
  NVIDIA AI Enterprise 対応

*Serverless API powered by NVIDIA AI Enterprise

AITRAS 20 セルをオンエア



AI-RANに関する新規開発の発表

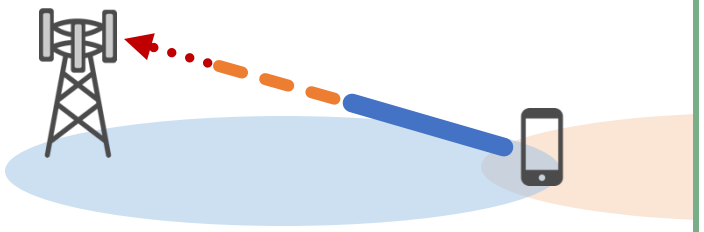
①

AI技術によるRANの性能向上効果を実証

AI for RAN技術を開発

SoftBank

UL Signal Processing

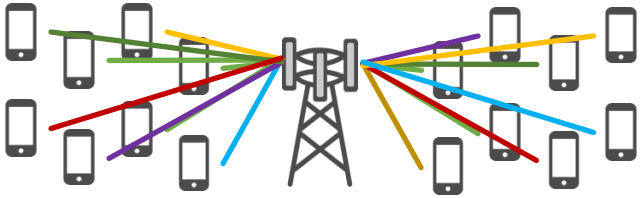


ノイズが多いエリアやセル
エッジでは受信感度が低く
チャネルエラーに影響

AITRAS

チャネル補間

MU-MIMO

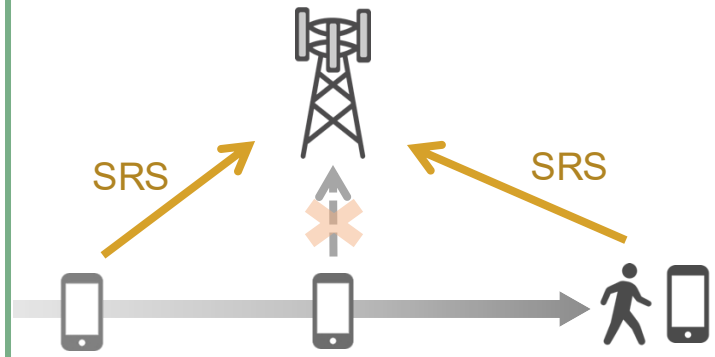


端末のペアリングに関する
膨大なマトリクス計算が発生

AITRAS

ユーザーペアリングの最適化

Beamforming



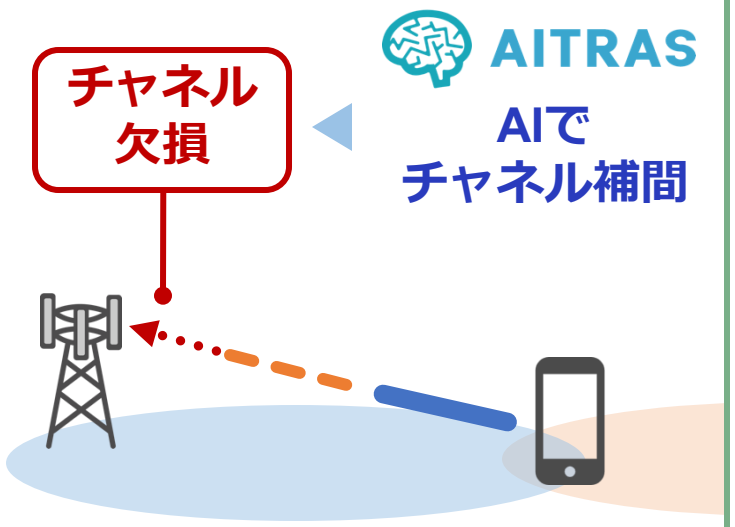
接続端末が増えると
SRSの送信間隔が大きくなる

SRS：Sounding Reference Signal
サウンディング参照信号

AITRAS

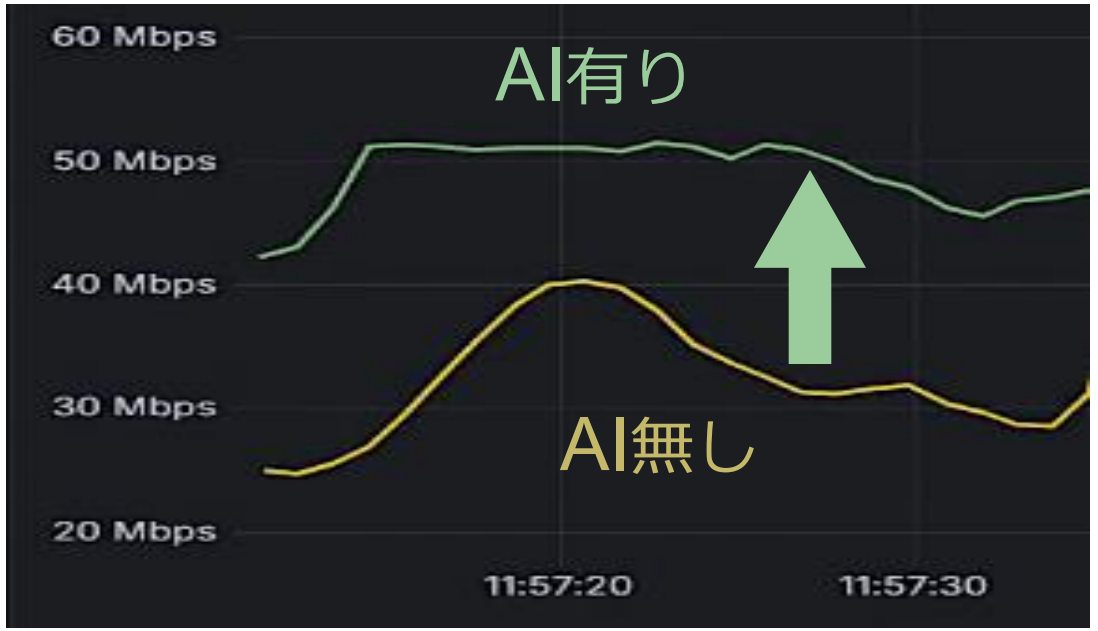
SRS 予測

ULチャネル補間



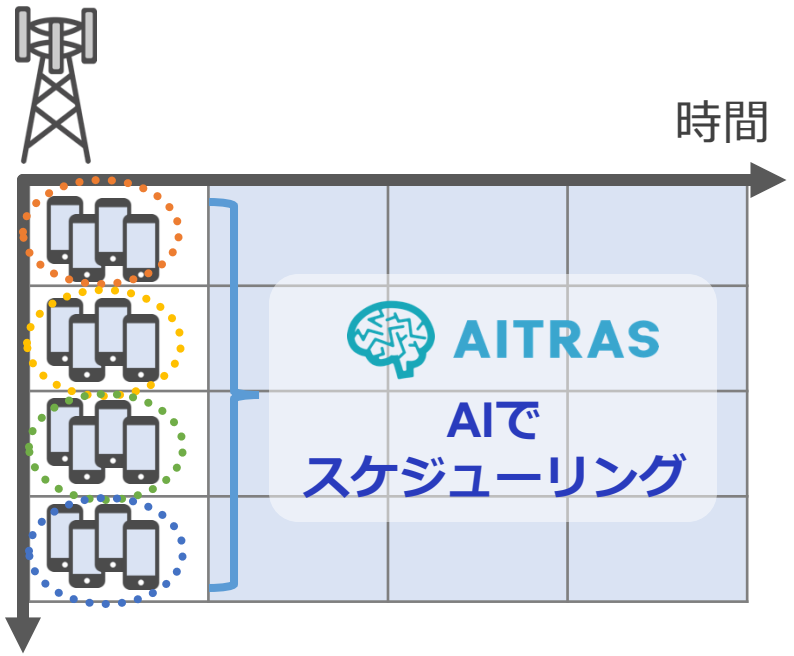
ラボの無線環境にて実証

周波数帯域：
n78
（100MHz）



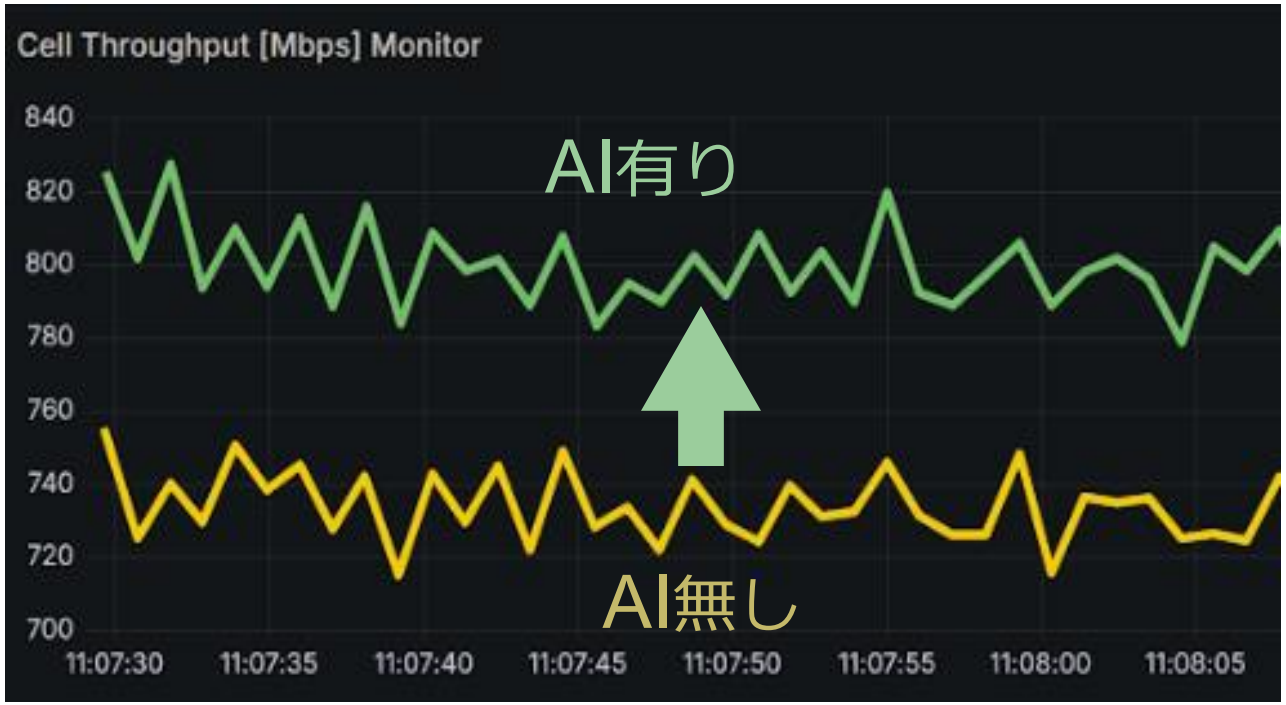
スループットが
約20%向上

MU-MIMO：ユーザーペアリングの最適化



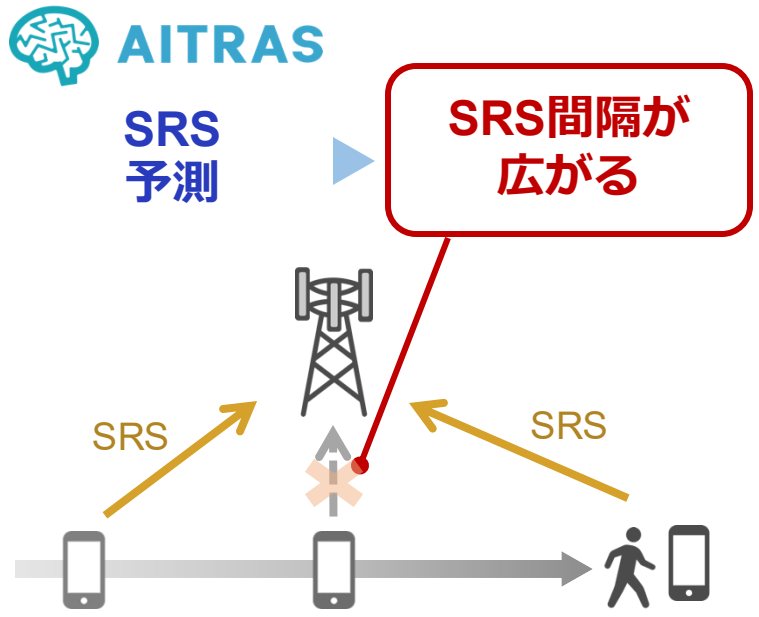
システムレベルシミュレーター上で
MLPアルゴリズムを用いて
MACスケジューリングを実施

MLP：Multilayer Perceptron
多層パーセプトロン

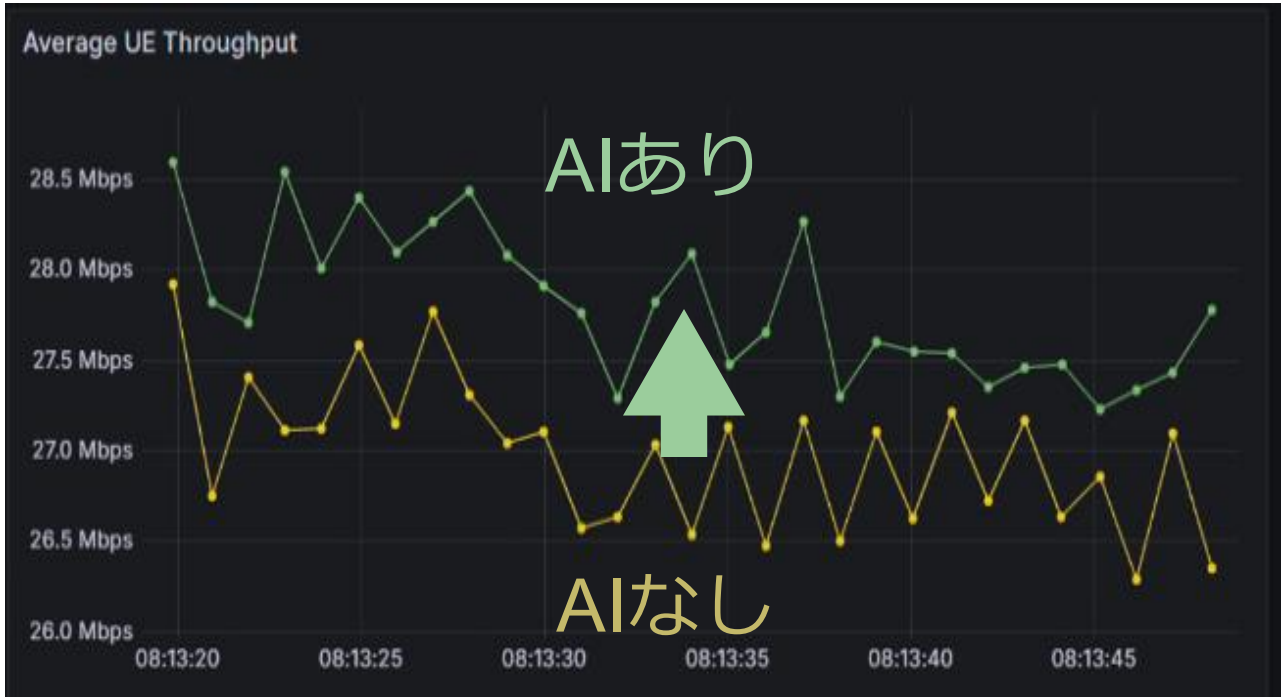


セルスループットが
約9%向上

Beamforming：SRS 予測



システムレベル
シミュレーター上で、
MLPアルゴリズムのAI を使用し
SRSを予測・補間



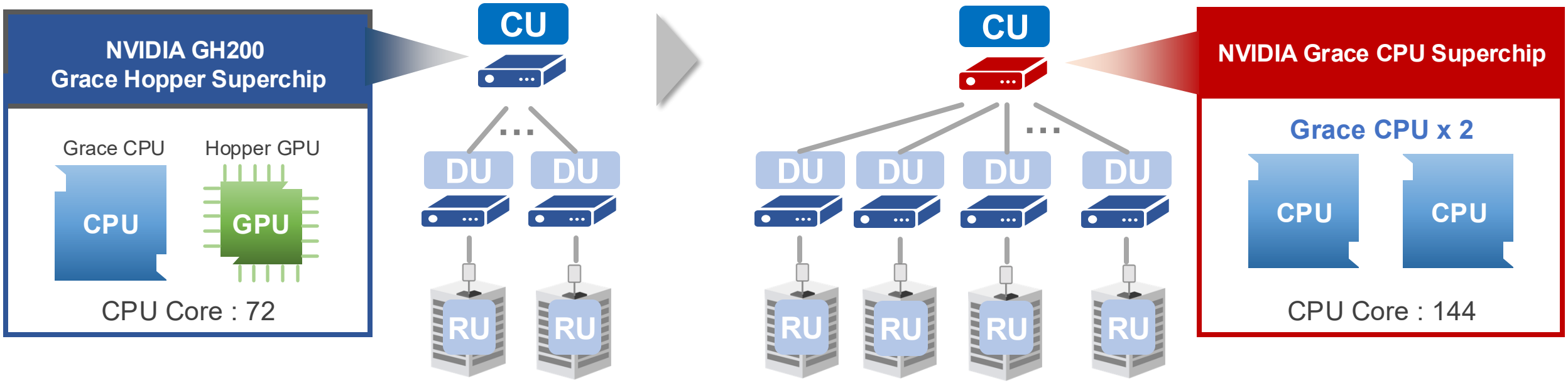
ユーザースループットが
約13%向上

②

「AITRAS」、
ArmベースのNVIDIAプラットフォームを活用した
C-RANとD-RANの
AI-RANアーキテクチャー実装の完了について

AITRASの進化①

NVIDIA Grace CPU Superchip ServerへのCU実装

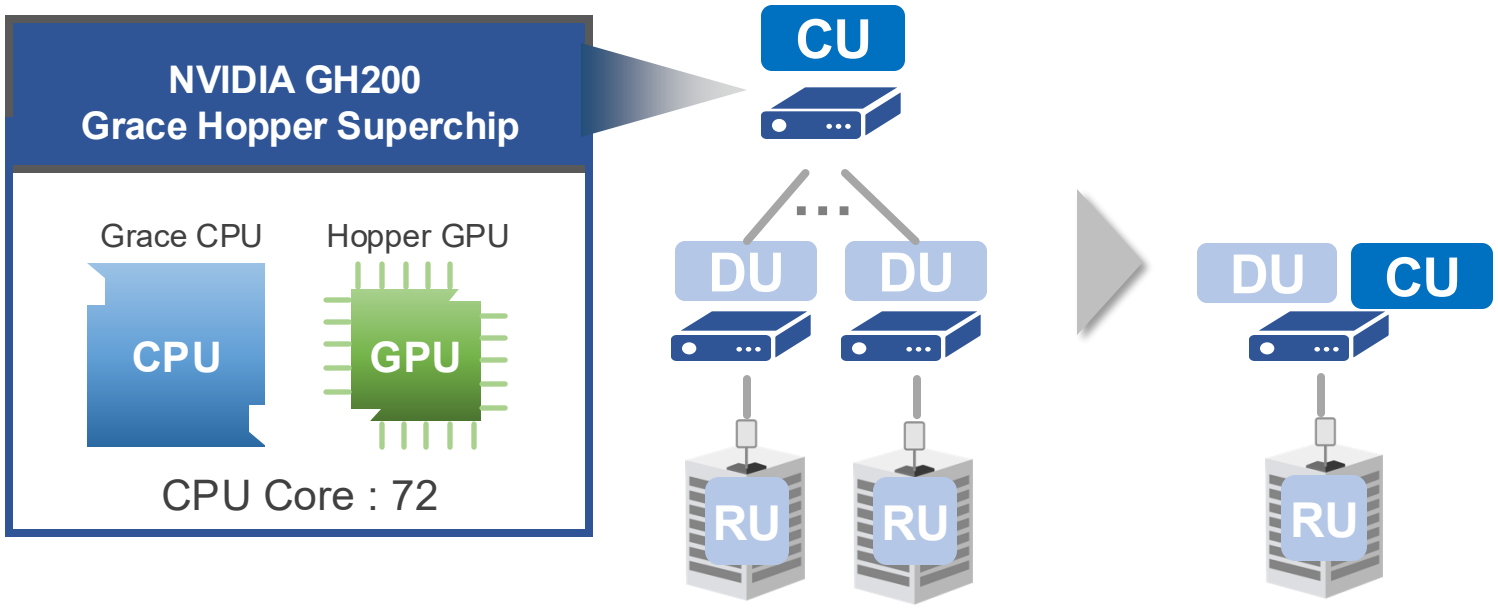


NVIDIA Grace CPUによって

CUの収容率はおよそ2倍に

AITRASの進化②

1台のNVIDIA GH200 ServerへDU、CUの両方を実装



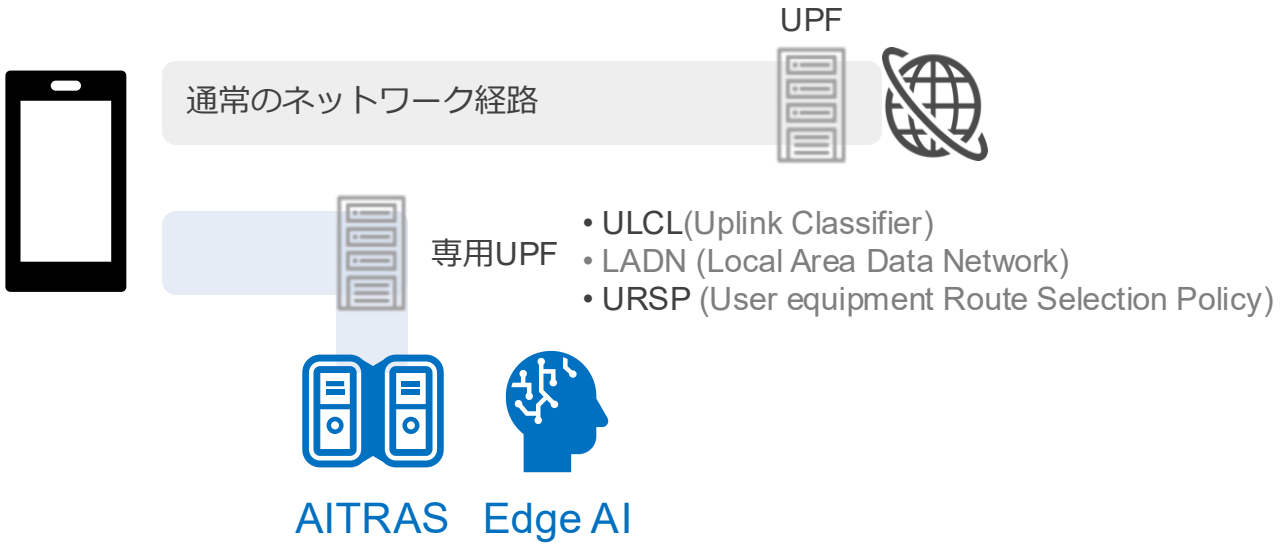
エンタープライズ施設
等で局所的にAI需要が
高まるエリアへ

**D-RAN構成の
AITRASを展開可能に**

③

ローカルブレイクアウト技術を活用し、「AITRAS」上のエッジAIサーバーへセキュアにアクセスする機能を開発

AI on RAN に向けた エッジルーティング技術の開発

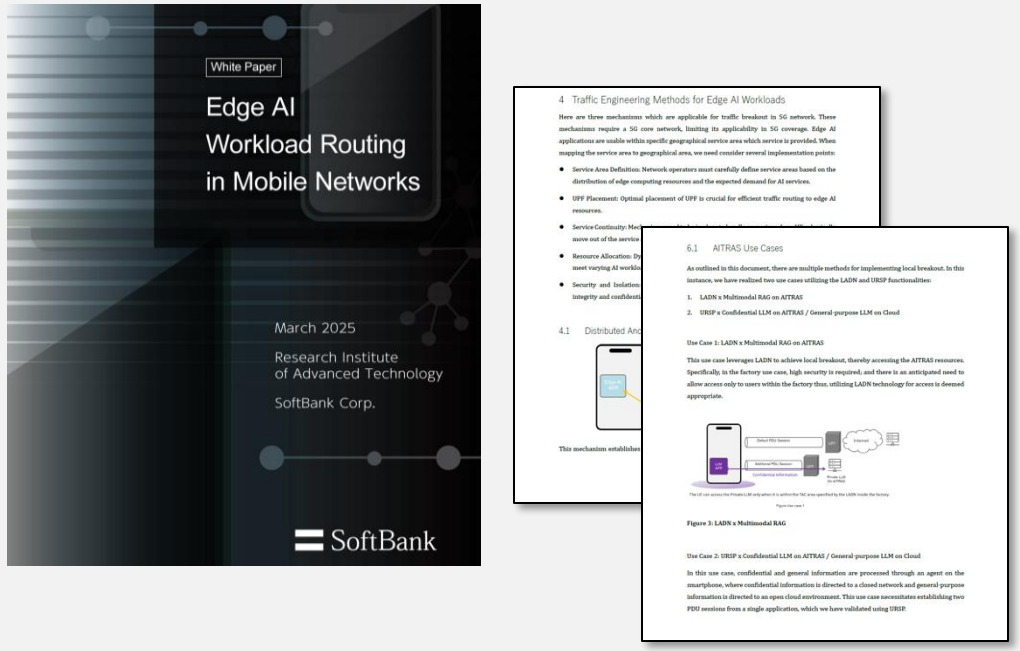


ネットワークエッジでの動的経路選択を可能にするルーティング技術

- Internetとローカルブレイクアウト経路間のスマートな切り替え
- AI on RANの展開オプションを拡大

White Paper 公開

“Edge AI Workload Routing in Mobile Networks”



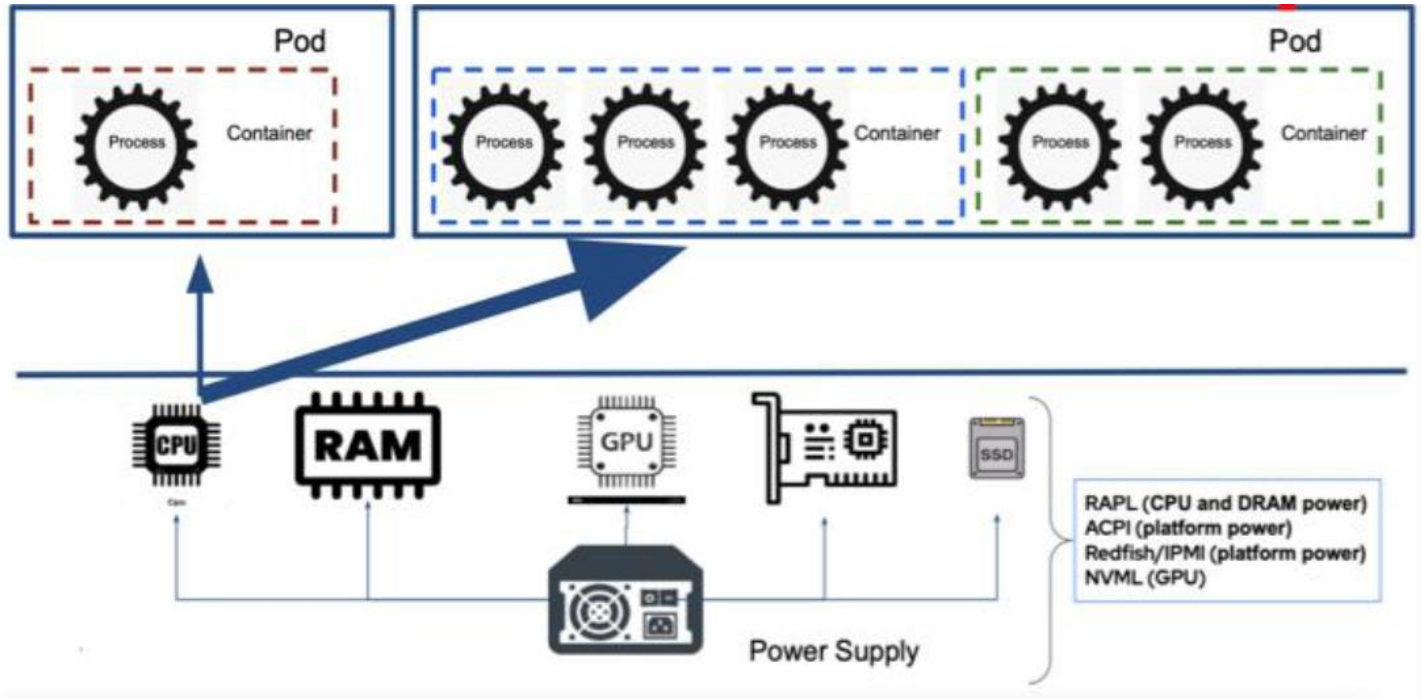
④

ソフトバンクとレッドハット、AI-RANのデータセンターにおける消費電力を最適化するソリューションを開発

Red Hat Kepler プロジェクト

Red Hat

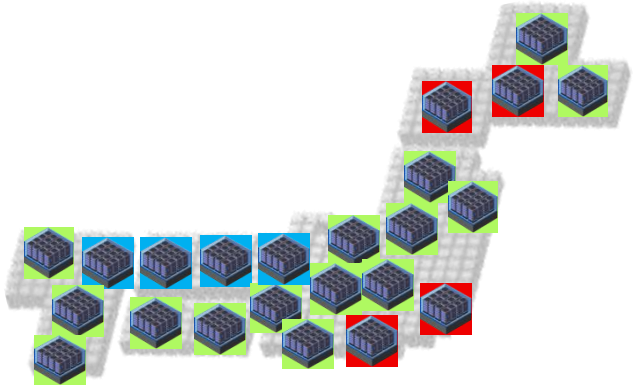
KEPLER

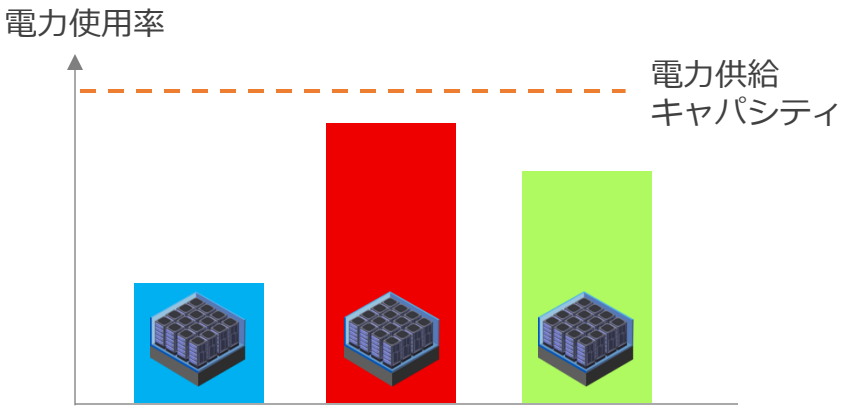


Red Hat Kepler プロジェクト：
Red Hat OpenShift上のPodのエネルギー消費を追跡し、
測定および最適化を行うオープンソースの取り組み

AITRAS オーケストレーター × Red Hat Kepler

電力の需要と供給の地域差





AITRAS オーケストレーター

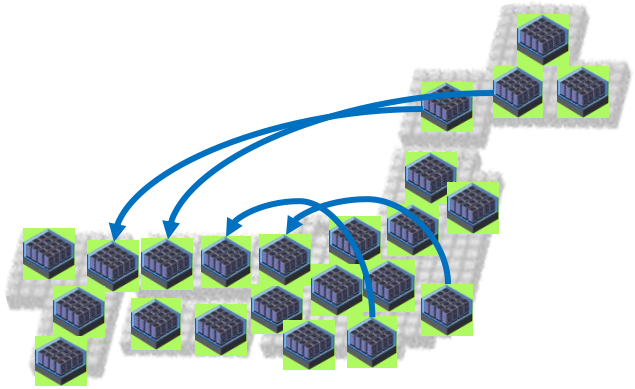
以下を考慮し、最適なDCにOpenShift上のPodをデプロイ

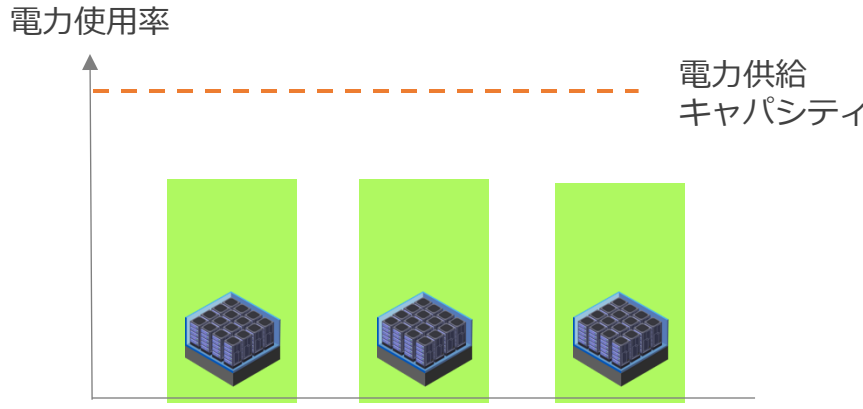
- ✓ 場所
- ✓ 遅延の要求
- ✓ リソース状況
- ✓ + 電力消費

Red Hat Kepler

Podの電力使用を予測

Podの電力使用料を予測し、配置を最適化





AITRASオーケストレーター x Red Hat Kepler

Before

After

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

# of Pod

CPU load

GPU load1

GPU load2

# of Pod

CPU

GPU1

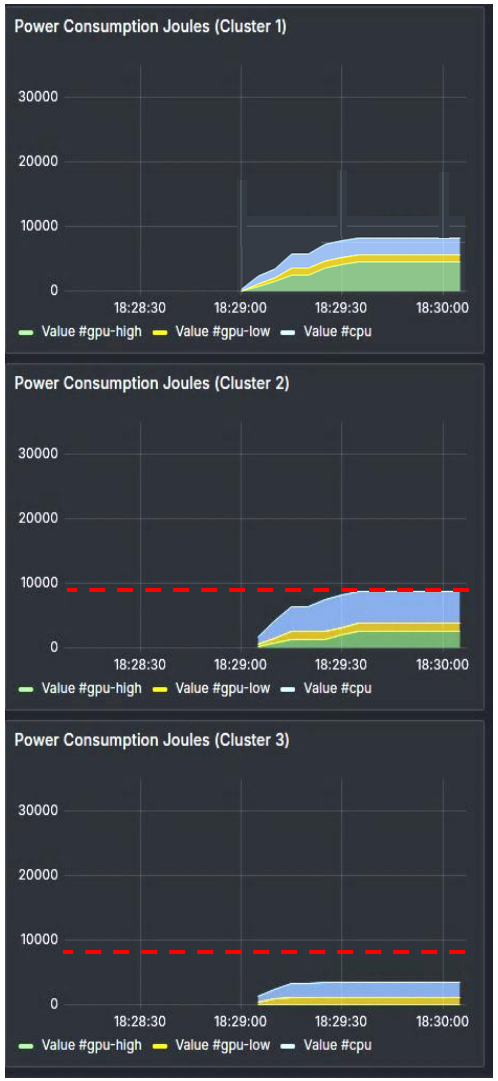
GPU2

# of Pod

CPU

GPU load1

GPU -load2



Kepler がPodsの電力消費を予測

AITRAS オーケストレーターがPodsを最適に配置

# of Pod

CPU load

GPU load1

GPU load2

# of Pod

CPU load

GPU load1

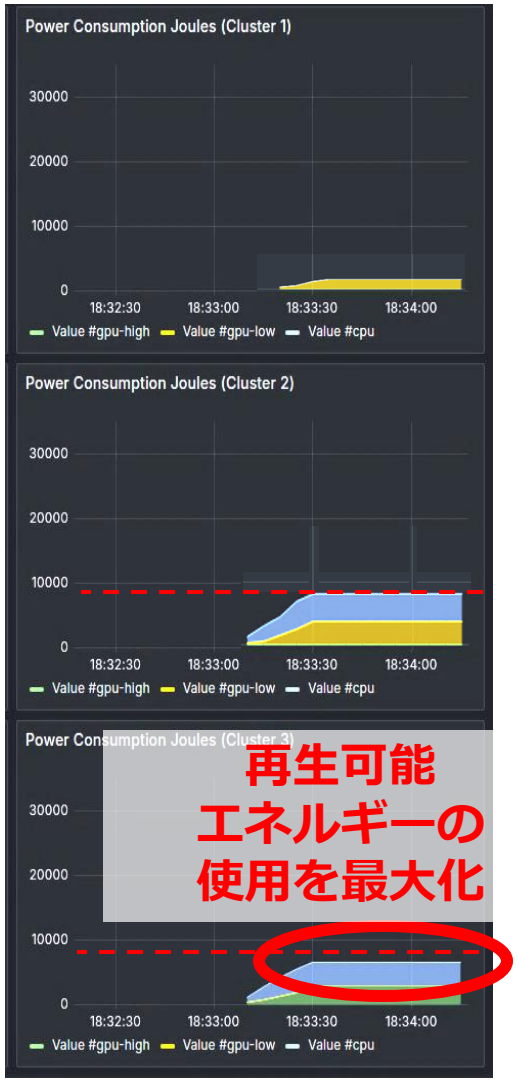
GPU load2

# of Pod

CPU load

GPU load1

GPU load2

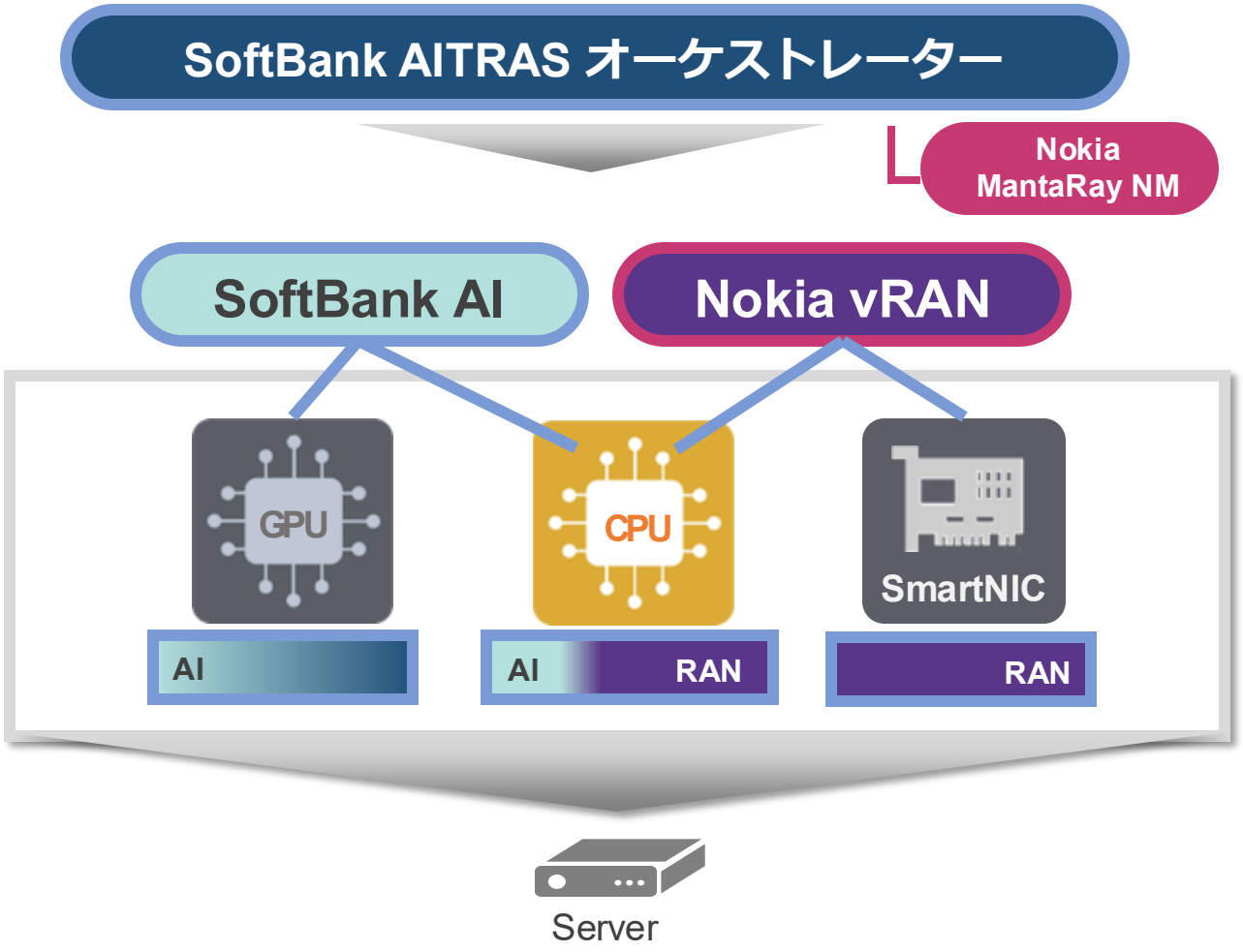


再生可能エネルギーの使用を最大化

⑤

ソフトバンクとノキア、1台のサーバー上でAIとvRANの共存と最適なリソース割り当ての自動化を実現

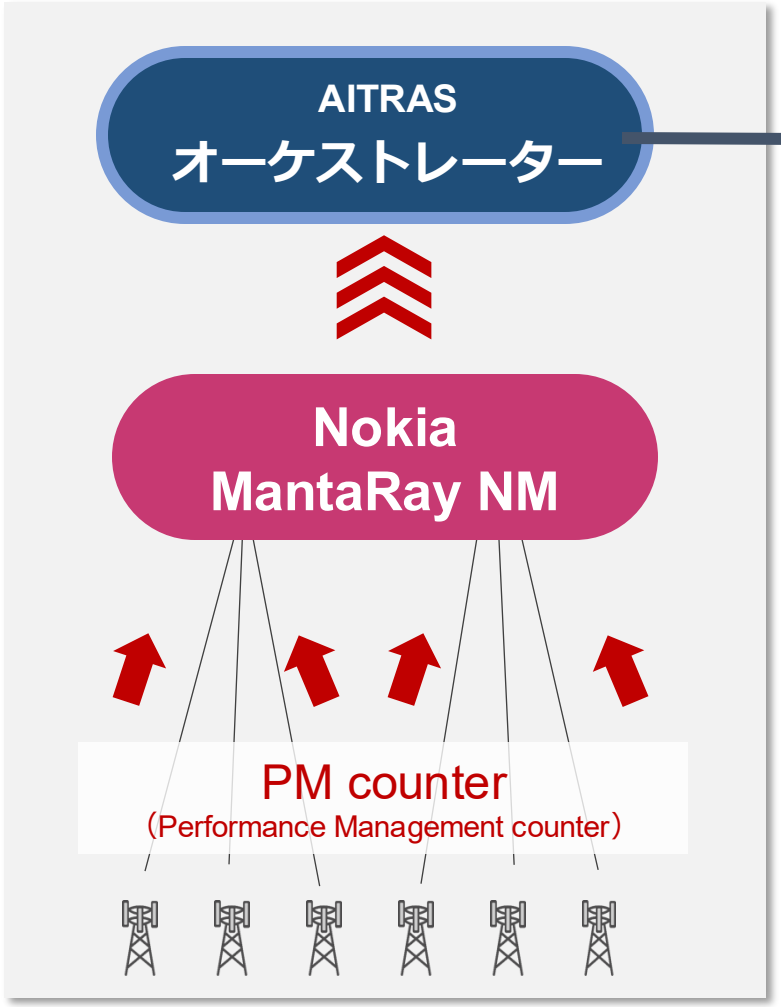
AIとvRANが1台のサーバー上で共存



- vDU処理の一部を担当するSmartNICがGPUサーバーをサポートするよう拡張
- AITRASオーケストレーターがリソース配分を最適化

NokiaのvRANソフトウェアを利用してAI and RANを実現

AITRAS オーケストレーター × Nokia MantaRay NM



- vRANトラフィックを予測
- AIとvRANのリソース配分を最適化