

SoftBank R&D

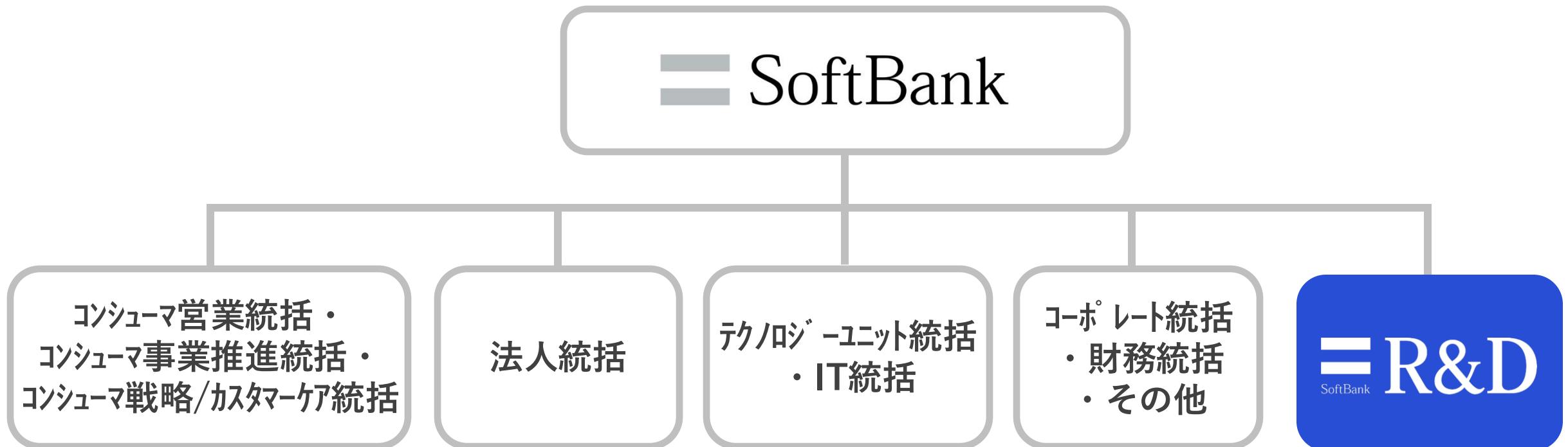
ソフトバンクのAIインフラ戦略

2025.12.11

ソフトバンク株式会社
執行役員 兼 先端技術研究所 所長
湧川 隆次

先端技術研究所について

通信のさらなる進化や先端技術を活用した新規ビジネスのためのR&Dを中心に活動

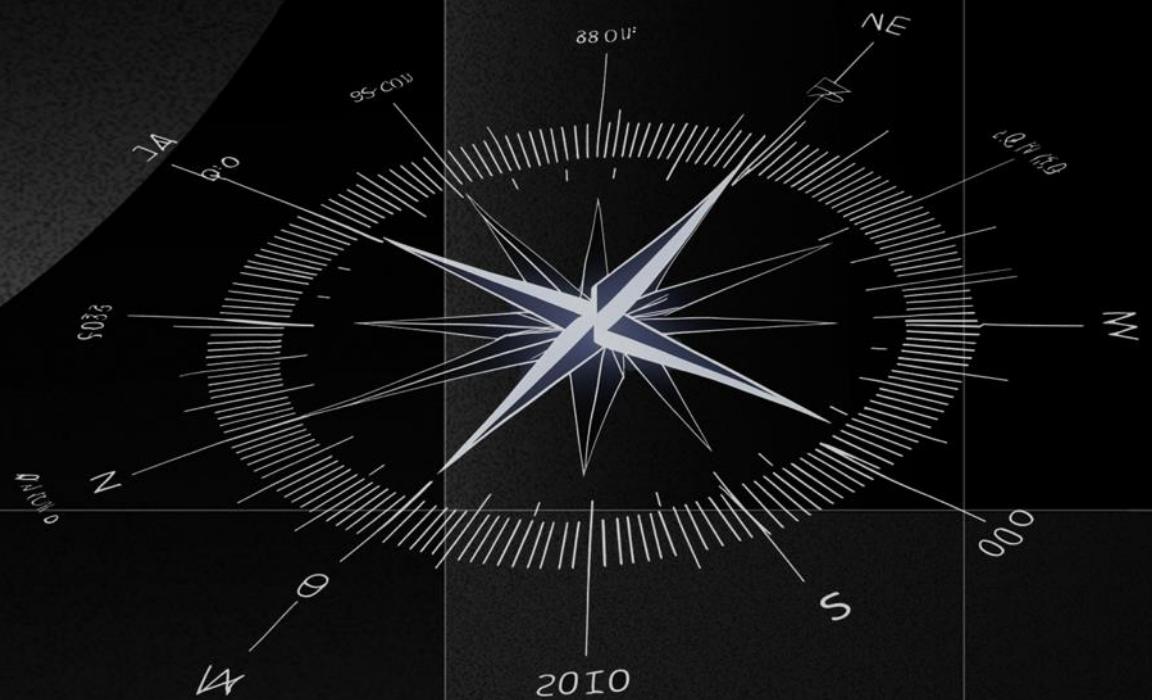


先端技術研究所
2022年4月新設

SoftBank R&D

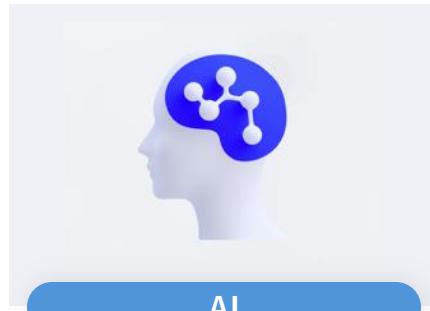
“技術は羅針盤”

技術に裏付けられた
ビジョンをつくる

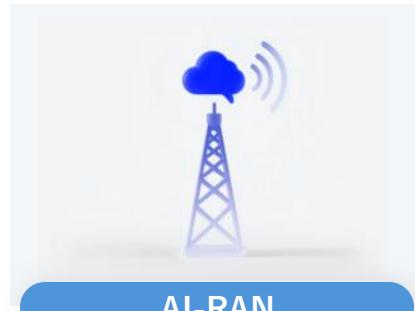


先端技術研究所の活動領域

70以上のプロジェクトを推進中



AI



AI-RAN



Large Telecom Model



6G · 5G Advanced



コアネットワーク



光無線



テラヘルツ通信



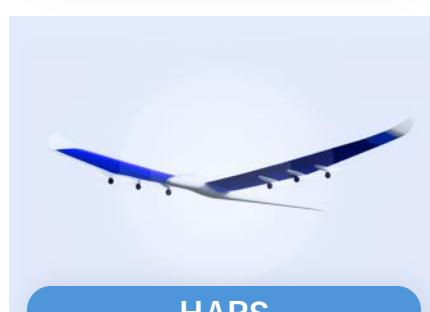
センシング



自動運転



量子暗号・量子技術



HAPS



次世代バッテリー



HAPS Alliance
HIGH ALTITUDE PLATFORM STATION

ソフトバンクのAI関連活動

SoftBank

AI Data Center



AI-RAN:AITRAS



2024年8月

4x4
on-air



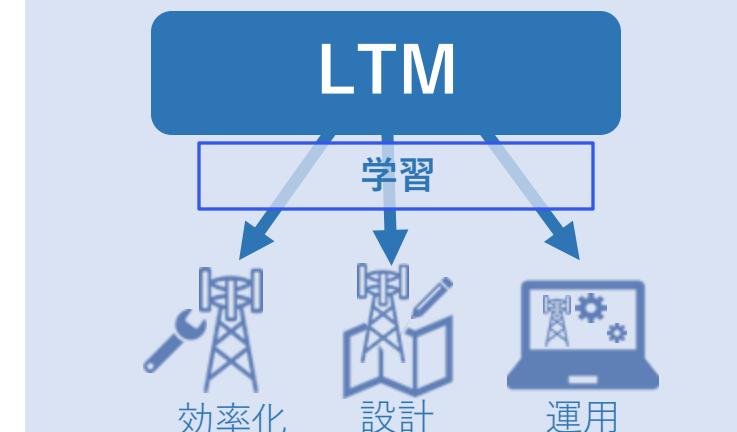
2025年10月

Massive MIMO
16 Layers
on-air

LTM: Large Telecom Model



継続学習
学習データ(~10TB)
・SB NW設定.
・SB KPI データ



インフラの構造的課題を抜本的に解決

現在



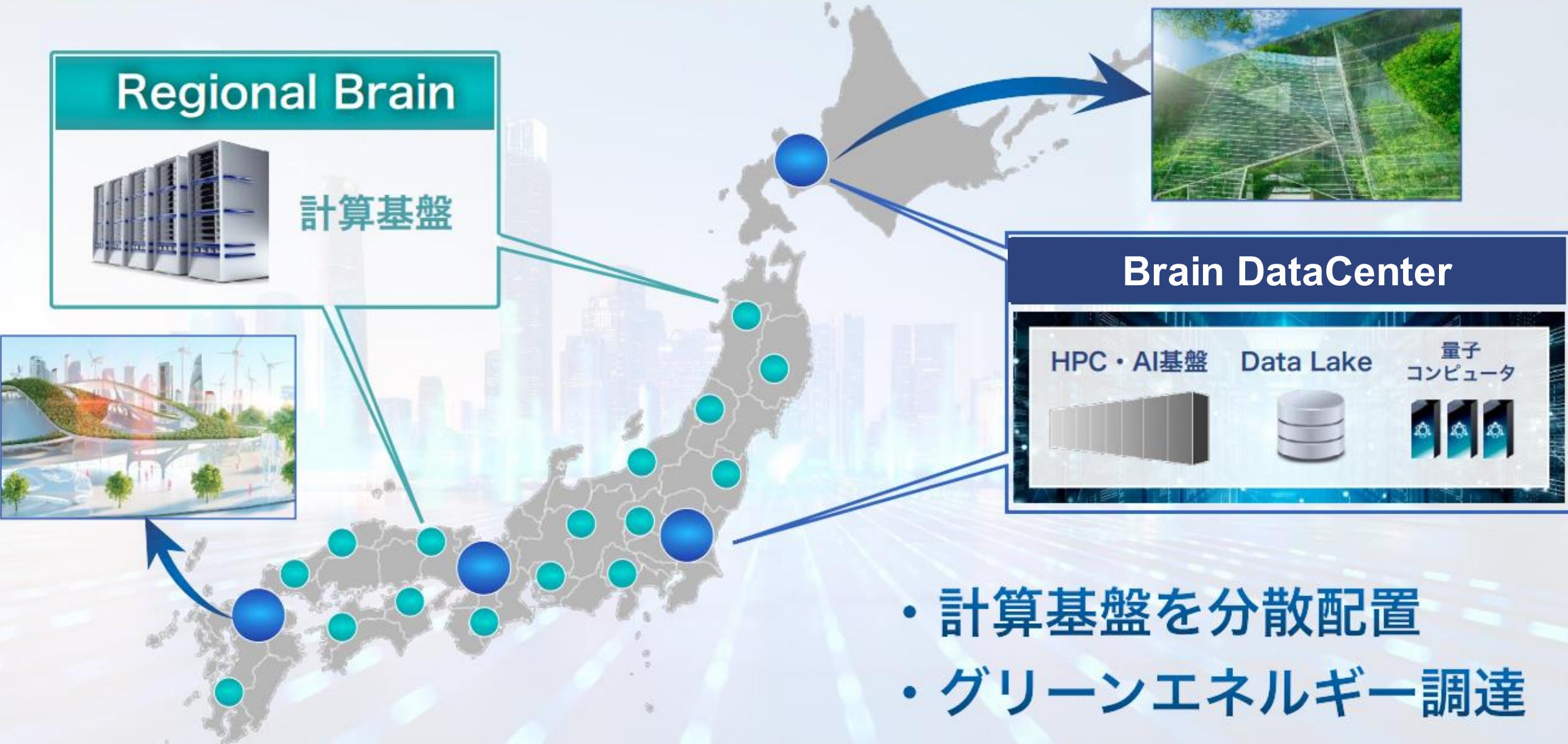
データ処理/電力消費が
都市部に集中

次世代社会インフラ



データ処理/電力を
地方分散処理

分散型AIデータセンター



AIコンピューティング基盤

4,000基超のNVIDIA Blackwell GPUを搭載



「NVIDIA Blackwell GPU」を搭載した
「NVIDIA DGX SuperPOD」として、
世界最大のAI計算基盤を構築

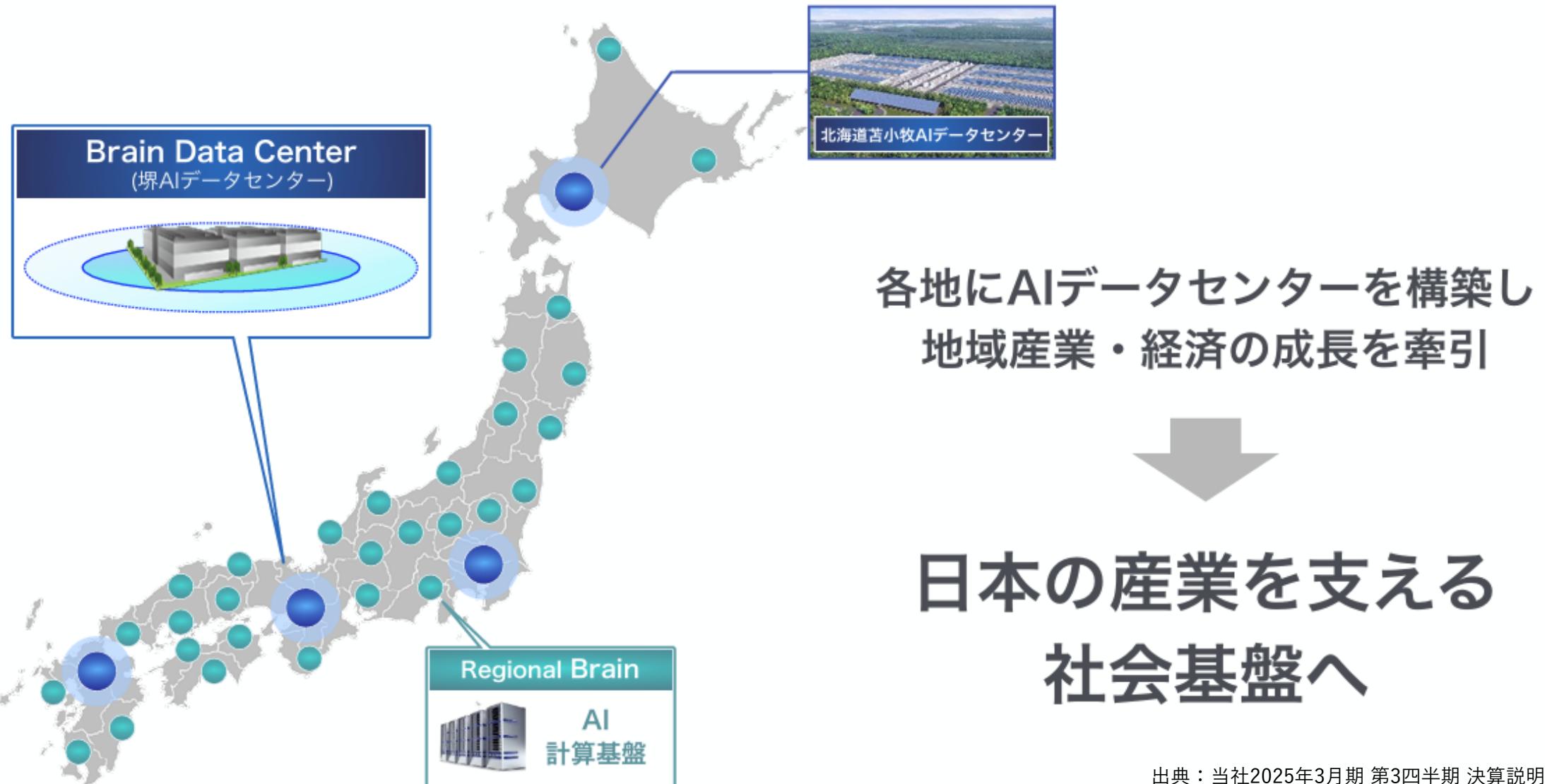
～4,000基超の「NVIDIA Blackwell GPU」の整備が完了～

2025年7月23日
ソフトバンク株式会社

ソフトバンク株式会社（以下「ソフトバンク」）は、現在拡張を進めているAI（人工知能）計算基盤において、4,000基超の「NVIDIA Blackwell GPU」を搭載した、「NVIDIA DGX B200」システムで構成された「NVIDIA DGX SuperPOD」の構築を2025年7月22日に完了しました。このAI計算基盤は、「NVIDIA DGX B200」システムで構成された「NVIDIA DGX SuperPOD」として世界最大※1の計算能力を実現しています。また、この計算基盤は、AI処理に必要な大量のデータを高速かつ安定してやり取りできるよう、NVIDIAの最新ネットワーク技術「Quantum-2 InfiniBand」を採用しています。さらに、企業向けのAI開発や運用に必要なツール群「NVIDIA AI Enterprise」にも対応しており、AIの開発や運用をより効率的かつ安全に行える環境を提供します。これにより、ソフトバンクが構築したAI計算基盤全体のGPU（Graphics Processing Unit）は1万基超となり、計算処理能力の合計は13.7EFLOPS（エクサフロップス※2）となりました。

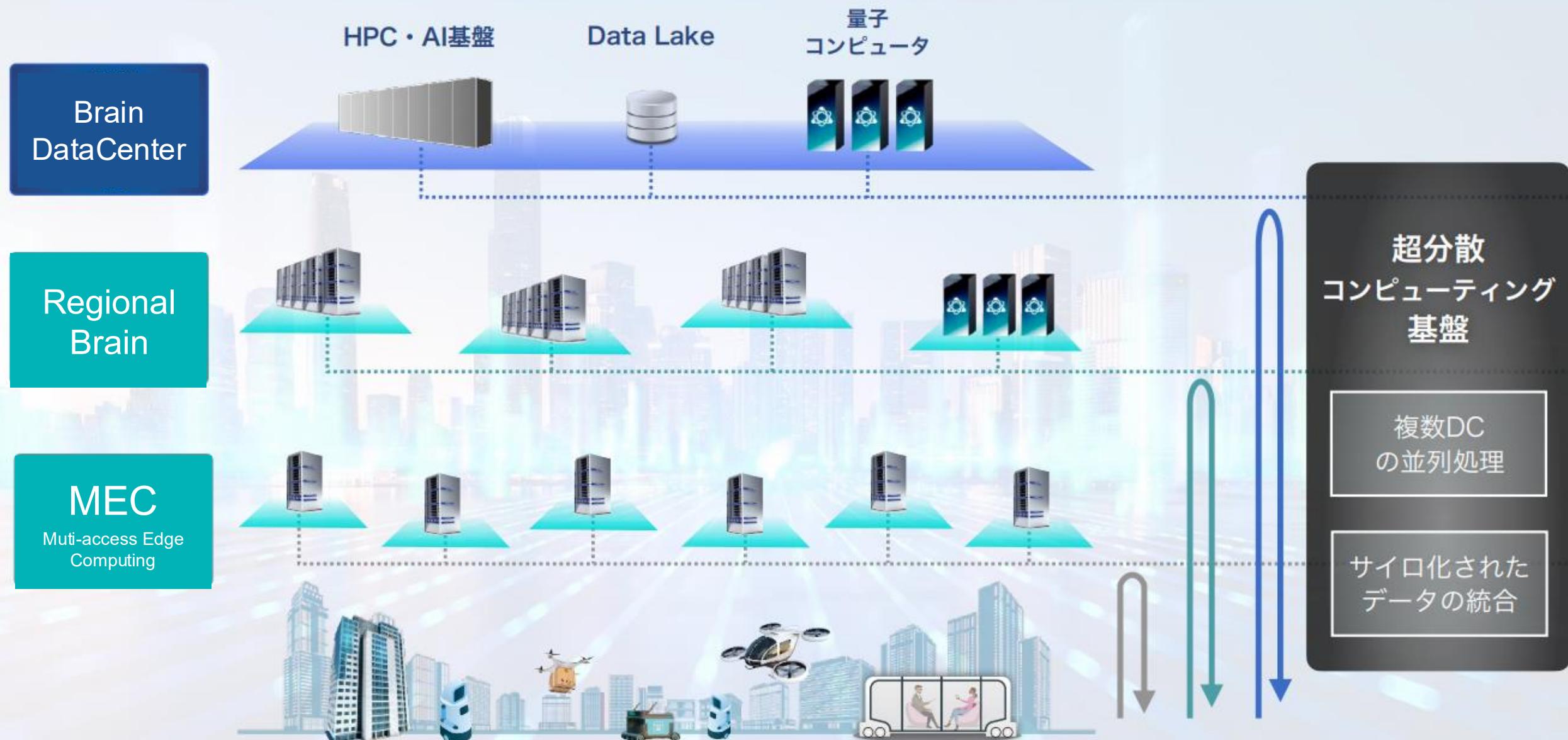
出典：2025/7/23 当社プレスリリース
“「NVIDIA Blackwell GPU」を搭載した「NVIDIA DGX SuperPOD」として、世界最大のAI計算基盤を構築”

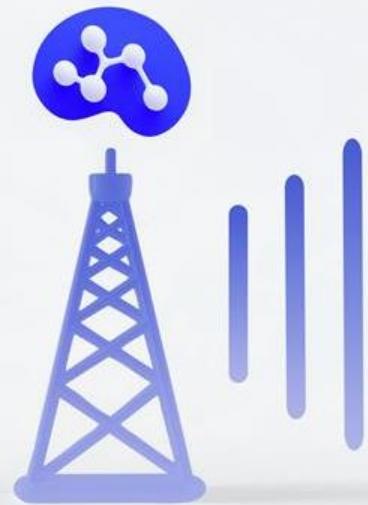
AIデータセンターを中心とした「産業集積地」構想



出典：当社2025年3月期 第3四半期 決算説明会

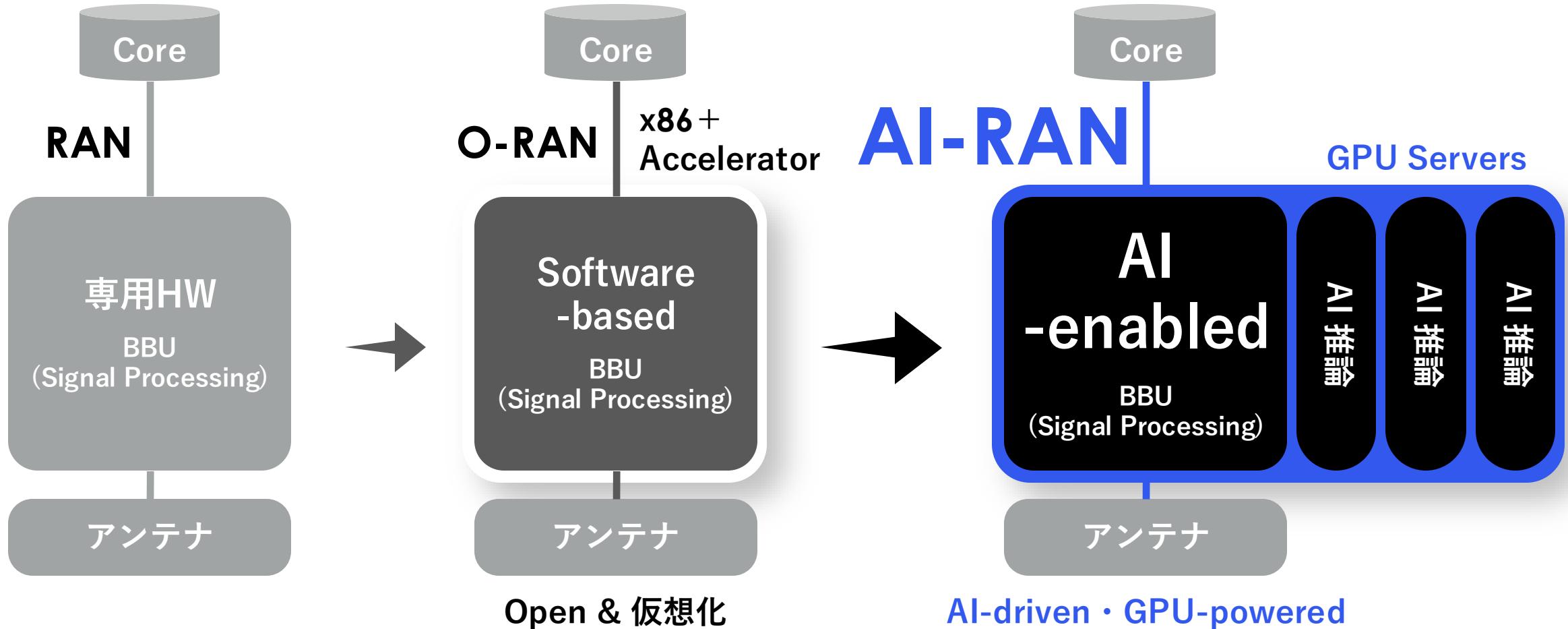
分散型AIデータセンターの配置





AI-RAN

AI-RANとは？



AI-RAN解説動画

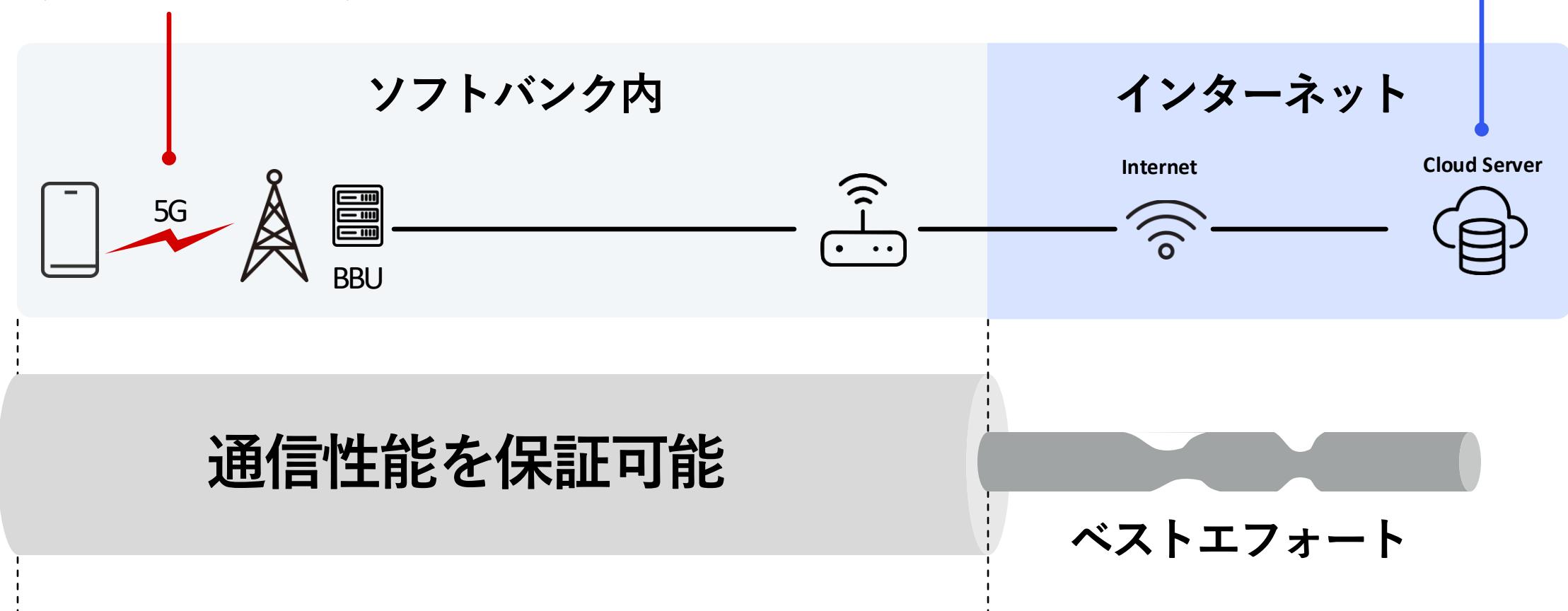


<https://www.youtube.com/watch?v=4cY2b09xz0Q>

AIサービスを使用するには

5Gで高速化

(赤い箇所だけが速くなる)



キャリアの強みを最大限に



AI-RANで進化するAIサービス



低遅延

高帯域通信

高セキュリティ

分散コンピューティング

AI推論



ロボットとの
リアルタイムなやりとり



手術補助



自律型ロボット
意思決定



機密情報
LLM RAG

NVIDIAとのGPUに関する取り組み

2019年からGPU-vRANに関する検証を共同で実施

GPU vRAN参戦



衝撃の「NVIDIA、通信業界に“再参入”」。KDDIとSBが採用表明した“ソフトウェア化する5G基地局”とは

…「EGX Edge Supercomputing Platform」という新しい“エッジコンピューティング”と呼ばれる最新技術向けのサーバー製品群を発表した。また、これを活用した5G向けの通信機器を、スウェーデンのエリクソン、米国のレッドハットと協業して提供していくと明らかにした。

しかも、これらの製品は、日本の通信キャリアであるKDDI、ソフトバンクが評価を開始していると明らかにした。…

出典：[TECH INSIDERの記事](#)を基に作成

GPU vRAN検証



NVIDIAのGPUを活用した5G仮想基地局の技術検証を実施

→処理速度や消費電力などで優れた性能を確認

ソフトバンク株式会社 (以下「ソフトバンク」) は、NVIDIA Corporationと協力し、ご当地GPU (Graphics Processing Unit) を活用して世界初となる「NVIDIA Maxine」(以下「TSM-AI-HW」) の技術検証を実施しました。

これは、大規模なデータを扱ううえで、高い処理速度や低消費電力を要する「TSM-AI-HW」が導入されました。具体的には、NVIDIAのGPUによってAI処理を実現する「Tensor Core」や、NVIDIAの独自開発した「Multi-Input Multi-Output」が組み込まれたことで、通信機器のデータ処理を最適化する機能が実現されました。これにより、通信機器のデータ処理をオフロードすることができ、データの処理速度を大幅に向上させました。また、NVIDIAのGPUによってAI処理を実現する機能が組み込まれたことで、通信機器のデータ処理を最適化する機能が実現されました。これにより、通信機器のデータ処理をオフロードすることができ、データの処理速度を大幅に向上させました。

今回の検証では、「NVIDIA Aerial」(以下「NVIDIA Aerial」) の性能を検証しました。検証内容は、より柔軟なアーキテクチャを実現するため、NVIDIAのGPUによってAI処理を実現する機能を搭載したデータセンターのデータ処理を実現するパラメータです。このパラメータは、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。パラメータは、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

「NVIDIA Aerial」は、O-RAN AllianceTMで定めた標準規格に基づいており、5Gの通信機器の構造や機能において柔軟性を持ったパラメータです。このパラメータは、データセンターの構造によってデータの種類を考慮することなく、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

GPUによるAI処理によるデータセンターの構造や機能を実現するため、AIによるデータ処理を必要とするデータセンターを構成するため、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

MEC AI検証



5GとMEC、「NVIDIA Maxine」を活用して映像を“超解像”する実証実験に成功

→高品質なビデオ品質を実現

ソフトバンク株式会社 (以下「ソフトバンク」) は、MEC (Multi-access Edge Computing) とNVIDIA Corporation (以下「NVIDIA」) が実施する「NVIDIA Maxine」(以下「TSM-AI-HW」) の技術検証を実施しました。

これは、大規模なデータを扱ううえで、高い処理速度や低消費電力を要する「TSM-AI-HW」が導入されました。具体的には、NVIDIAのGPUによってAI処理を実現する「Tensor Core」や、NVIDIAの独自開発した「Multi-Input Multi-Output」が組み込まれたことで、通信機器のデータ処理を最適化する機能が実現されました。

具体的な検証内容は、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

また、NVIDIAのGPUによってAI処理を実現する機能を搭載したデータセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

「NVIDIA Aerial」は、O-RAN AllianceTMで定めた標準規格に基づいており、5Gの通信機器の構造や機能において柔軟性を持ったパラメータです。このパラメータは、データセンターの構造によってデータの種類を考慮することなく、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

「NVIDIA Aerial」は、O-RAN AllianceTMで定めた標準規格に基づいており、5Gの通信機器の構造や機能において柔軟性を持ったパラメータです。このパラメータは、データセンターの構造によってデータの種類を考慮することなく、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

AI-on-5G Lab開設



完全仮想化されたプライベート5Gの商用化に向けて研究施設「AI-on-5G Lab.」をNVIDIAと合同で開設

→NVIDIAとMECが融合したプライベート5Gの商用化を加速

ソフトバンク株式会社 (以下「ソフトバンク」) は、MEC (Multi-access Edge Computing) とNVIDIA Corporation (以下「NVIDIA」) が実施する「NVIDIA Maxine」(以下「TSM-AI-HW」) の技術検証を実施しました。

これは、大規模なデータを扱ううえで、高い処理速度や低消費電力を要する「TSM-AI-HW」が導入されました。具体的には、NVIDIAのGPUによってAI処理を実現する「Tensor Core」や、NVIDIAの独自開発した「Multi-Input Multi-Output」が組み込まれたことで、通信機器のデータ処理を最適化する機能が実現されました。

具体的な検証内容は、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

また、NVIDIAのGPUによってAI処理を実現する機能を搭載したデータセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

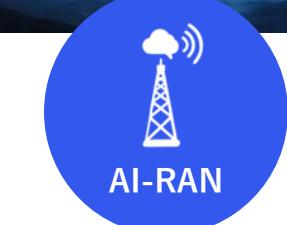
「NVIDIA Aerial」は、O-RAN AllianceTMで定めた標準規格に基づいており、5Gの通信機器の構造や機能において柔軟性を持ったパラメータです。このパラメータは、データセンターの構造によってデータの種類を考慮することなく、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

「NVIDIA Aerial」は、O-RAN AllianceTMで定めた標準規格に基づいており、5Gの通信機器の構造や機能において柔軟性を持ったパラメータです。このパラメータは、データセンターの構造によってデータの種類を考慮することなく、NVIDIAのGPUによってAI処理を実現する機能を実現するため、データセンターの構造や、データの種類によって異なるため、複数のパラメータを用いて検証を行いました。

AI-RAN Alliance 設立



AI-RAN ALLIANCETM



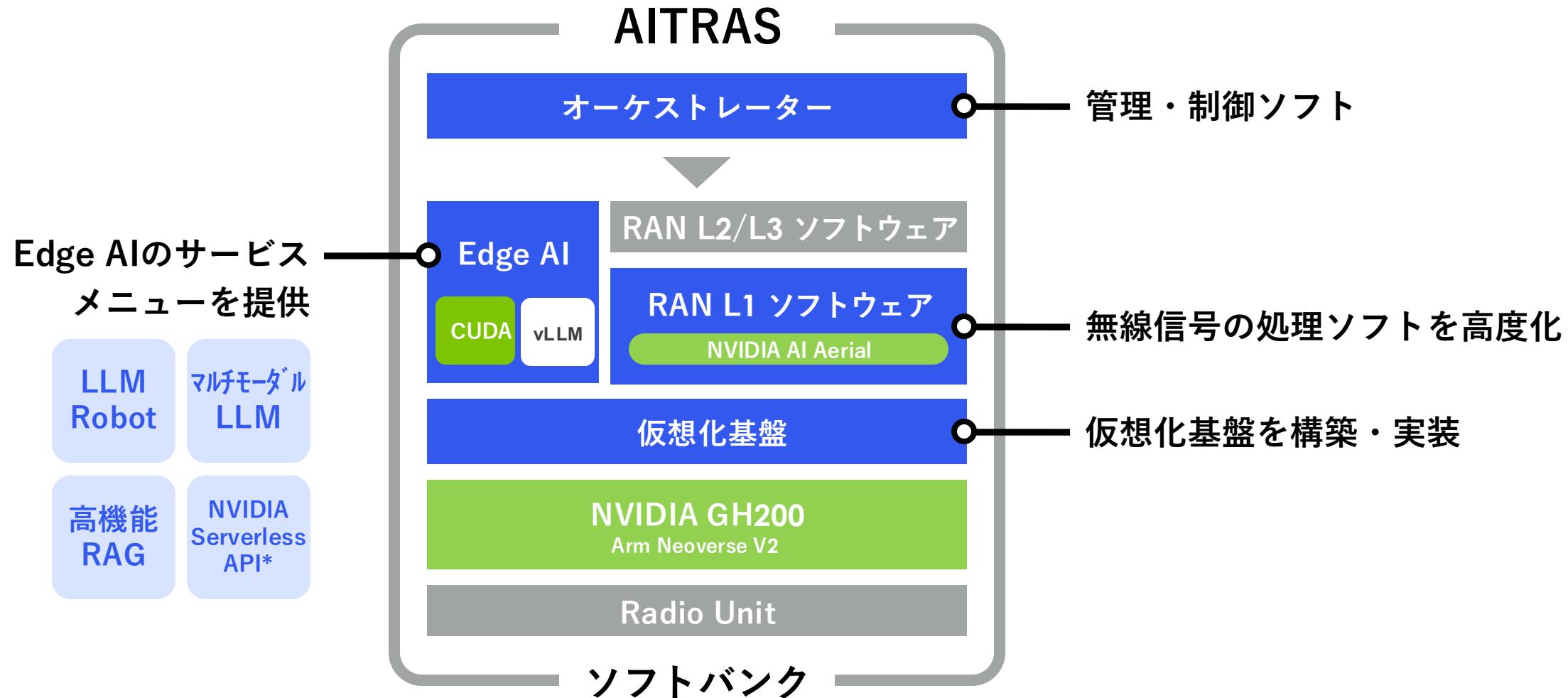
ア イ ト ラ ス

AITRAS

AI-RANコンセプトに基づいた
ソフトバンクオリジナルプロダクト

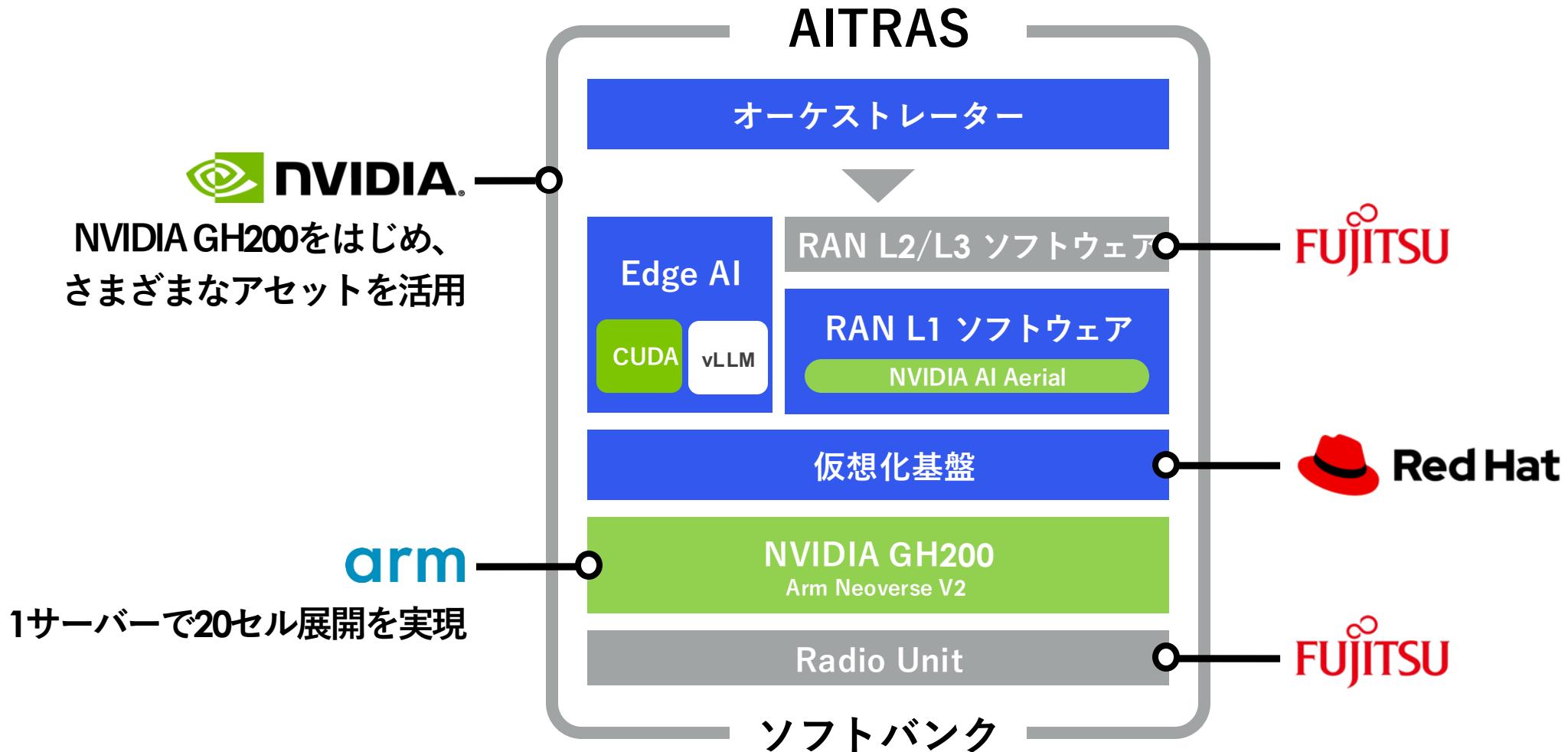
AITRASのシステム構成

ソフトバンクが開発した要素技術

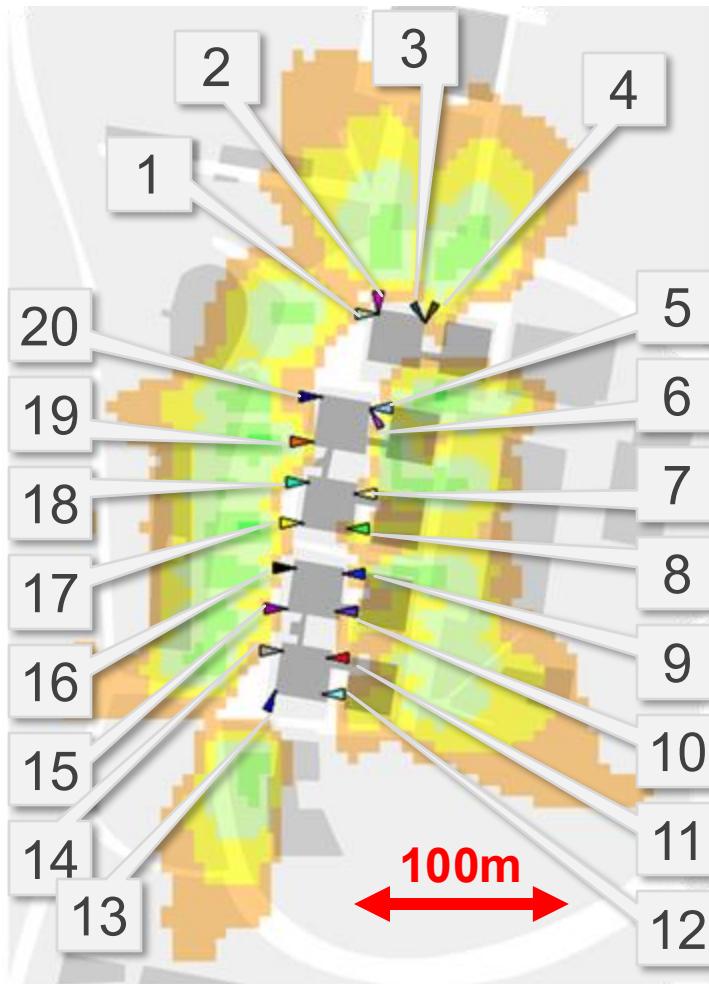


AITRASのシステム構成

パートナーとのコラボによる高性能化



2024年11月@SFC AITRAS 20セルをオンエア(実験局)

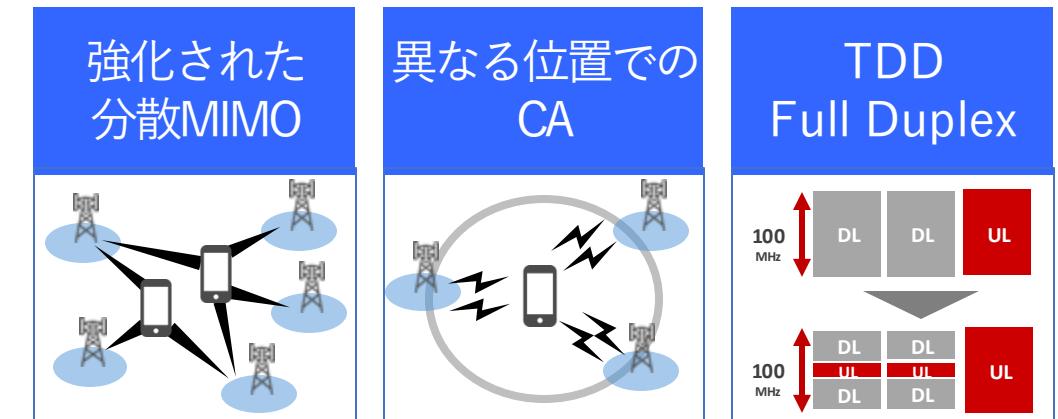
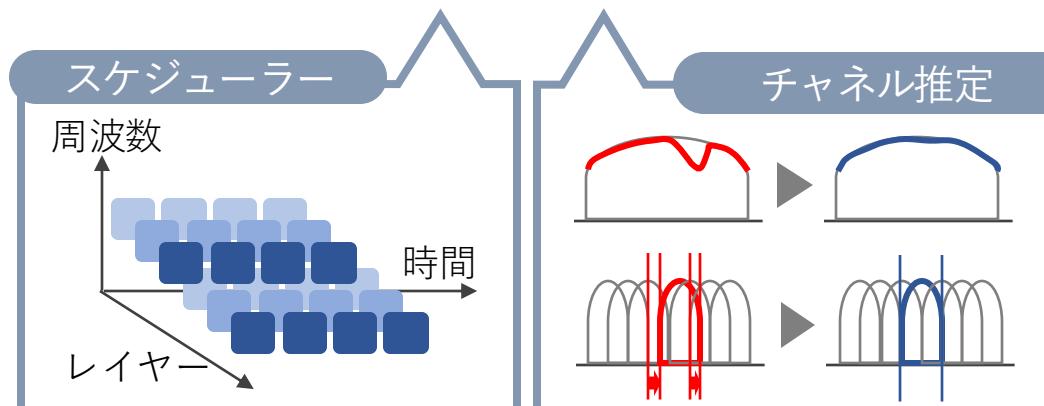
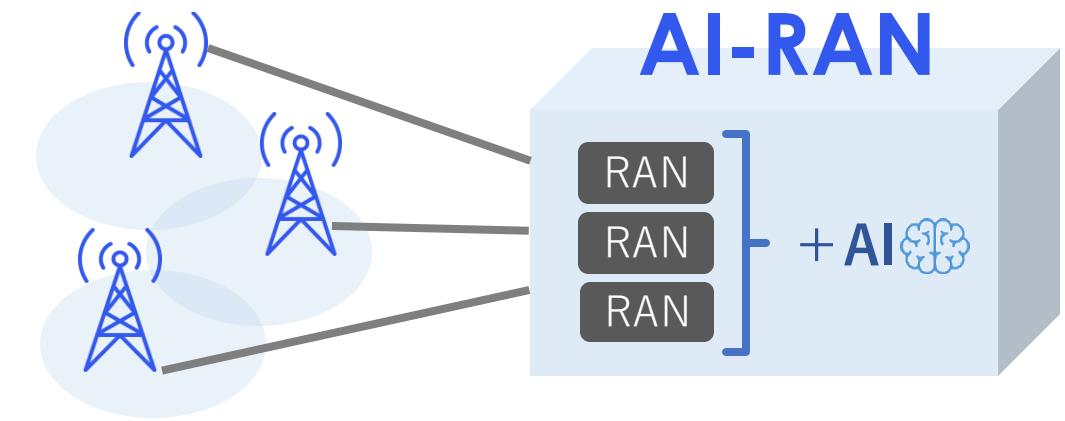


AIによって無線リソース効率最大化へ

AIによる単独セルの最適化

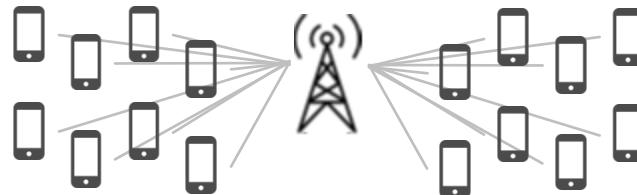


AIによるセル間の連携・最適化



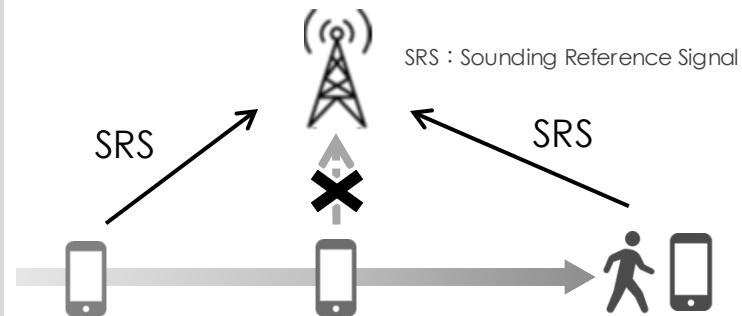
AI無線開発実績

MU-MIMO



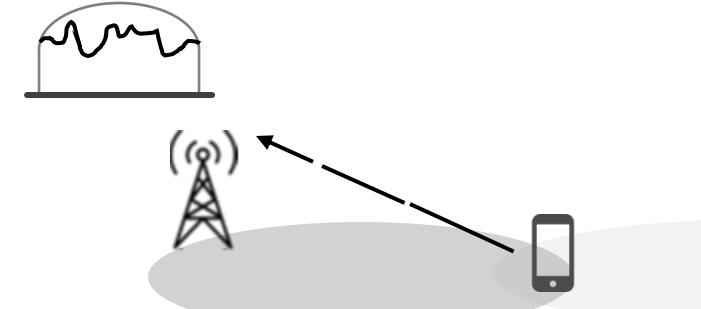
多数・多種デバイス接続による
非効率なリソース割当

ビームフォーミング



多数デバイス接続による
ビームフォーミング性能劣化

UL信号処理



干渉による
ULチャネル劣化

MACスケジューラー

スループット向上
9%

SRS補完

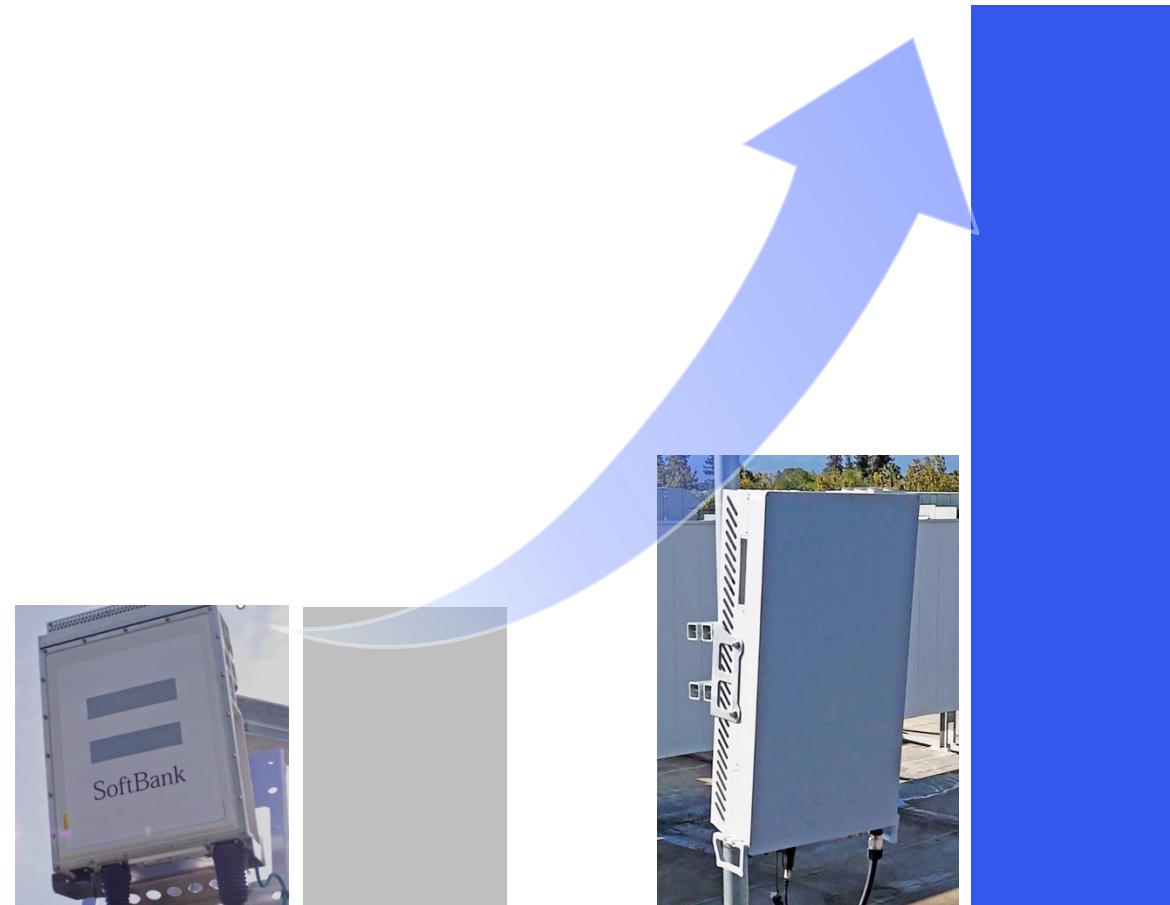
スループット向上
13%

ULチャネル補完 with Transformer

スループット向上
30%

Massive MIMO 16 Layer On-Air

サンタクララNvidia本社にて実証

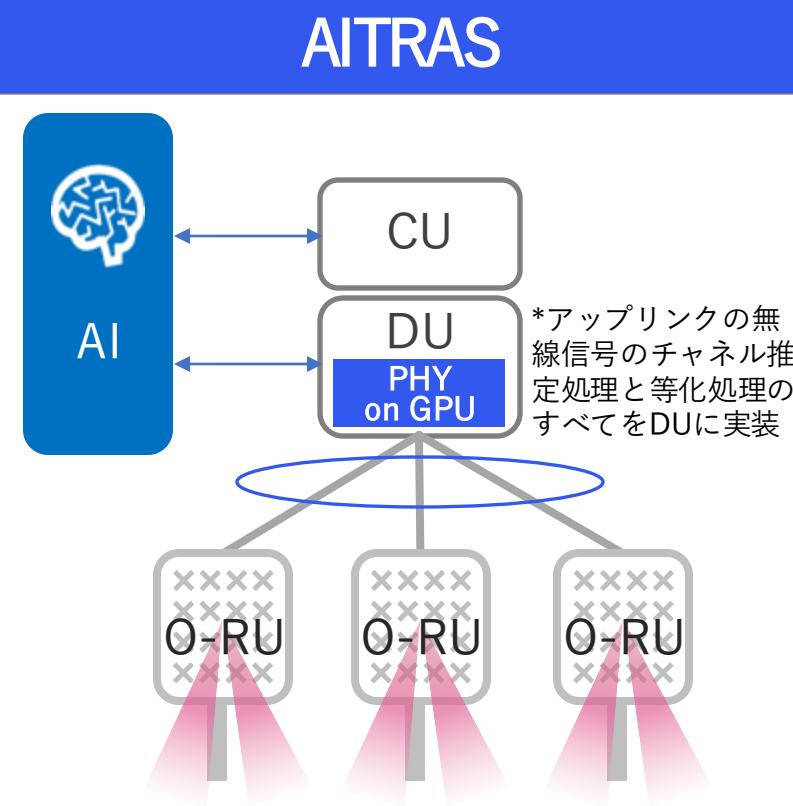
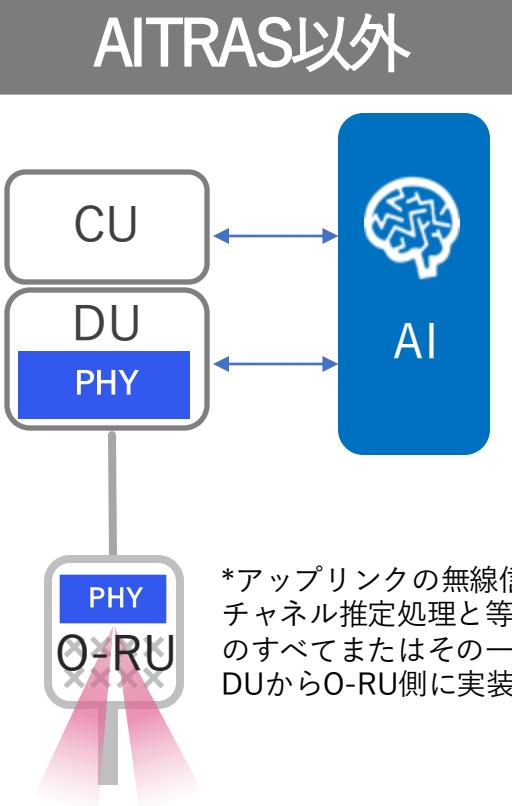


SU-MIMO

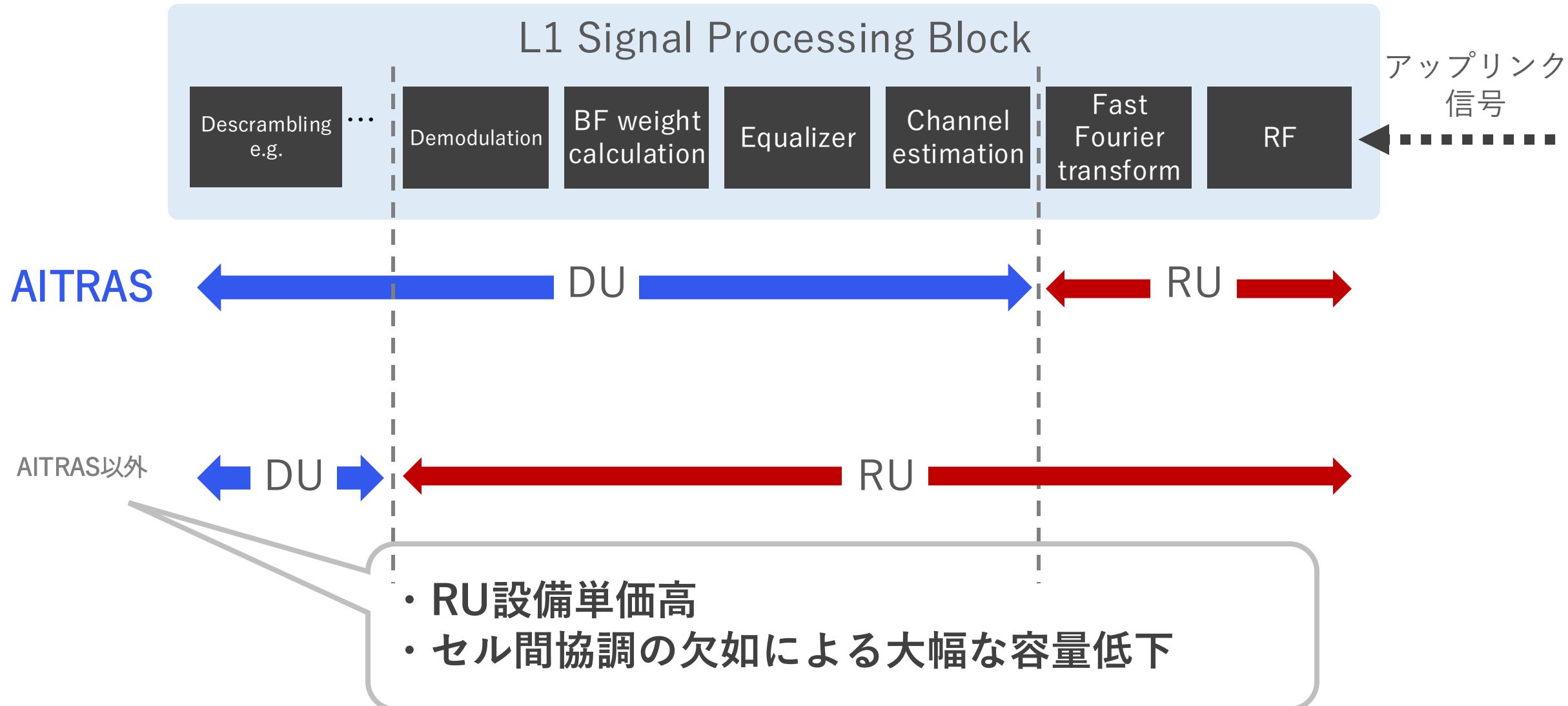
MU-MIMO

スループット
3倍向上

AITRASシステム構成

AITRAS		AITRAS以外
システム構成 O-RAN Split Option 7.2x	 <p>*アップリンクの無線信号のチャネル推定処理と等化処理のすべてをDUに実装</p>	 <p>*アップリンクの無線信号のチャネル推定処理と等化処理のすべてまたはその一部をDUからO-RU側に実装</p>
AIで利用可能な無線品質最適化に必要な情報	完全	限定的/局所的
アップリンクチャネル補間による無線品質の改善	高	中～低
ビームフォーミングの最適化	高	中～低

Massive MIMOをソフトウェアで完全実装



AI-RAN™
ALLIANCE



The logo graphic for AI-RAN consists of the letters 'A', 'I', 'R', 'A', 'N' in a teal sans-serif font. The letter 'R' is stylized with a central vertical stroke and two diagonal lines extending from its top and bottom. Above the 'R', there is a graphic element resembling a neural network or a signal path, composed of two horizontal teal lines with circular endpoints, which meet at a central teal circle. To the right of the 'N', there is a small teal 'TM' symbol.

2024年2月AI-RAN アライアンス設立

SoftBank

共同での課題克服



Chair:
Dr. Alex Jinsung Choi,
SoftBank

- ✓ AI-RANの活用と高性能化
- ✓ 新たな収益機会の創出
- ✓ AI-RANエコシステムの構築
- ✓ 研究開発活動を通じたパフォーマンス向上

3つのWorking Group

AI-for-RAN



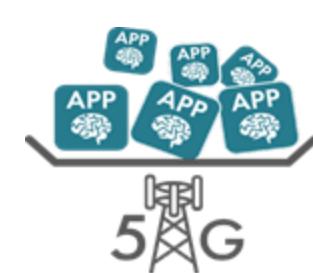
周波数の利用効率

AI-and-RAN



設備の利用効率

AI-on-RAN



新規サービス

AI-RAN Alliance Members: 100社を超えた

SoftBank

Founding Members (11)



General Members (104)

- 6WIND
- A10 Systems Inc. d/b/a AiRANACULUS
- Aira Technology
- AirHop Communications
- Ajou University
- Allbesmart Lda
- Altera Corporation
- Amdocs
- Amini AI
- Arizona State University
- AsiaInfo Technologies (H.K.) Limited
- Battelle Memorial Institute
- Beyond-G Global Innovation Center, Hanyang University, Seoul, Korea
- Bloxtel Inc.
- Boost Mobile Network
- Booz Allen Hamilton Inc.
- BROADRADIO INTERNATIONAL PTE. LTD.
- BubbleRAN
- CableLabs
- Capgemini America Inc
- Cisco Systems, Inc.
- Cloudera
- Cloudly, Inc.
- CLOUDRAN.AI PTE.LTD.
- Cohere Technologies
- Concordia University
- CTOne Inc.
- Dell Technologies
- EdgeCortix Inc.
- Ennoia Technologies, Inc.
- Eridan
- ETRI
- EURECOM
- F5, Inc.
- Fujitsu Limited
- G Reigns
- GlobalLogic Inc.
- Globe Telecom
- Harman Connected Services
- Hermes Network
- Hewlett Packard Enterprise
- Indian Institute of Technology Jodhpur
- Indian Institute of Technology Madras
- Indosat Ooredoo Hutchison
- Institute for Information Industry
- Itochu Techno Solutions Corporation
- ITRI
- Keio Research Institute at SFC, Keio University
- Keysight Technologies
- King's College London
- Korea University
- KT
- Kyunghee University
- LeadingSolution Inc
- LG Uplus
- LITE-ON TECHNOLOGY CORPORATION
- Matsing
- MediaTek Inc.
- Metanoia Communications Inc.
- MITRE Corporation
- Myrtle.ai
- National Yang Ming Chiao Tung University
- NEC Corporation
- NIA (National Information-society Agency)
- Paradise Mobile
- Poznan University of Technology
- Pure Storage
- Q Broadcasting LLC d/b/a Q
- Qualcomm Technologies Inc.
- Quanta Cloud Technology (QCT)
- Queen Mary University of London
- Rakuten Symphony
- Red Hat. K.K.
- RIMEDO SP. Z O.O. (Rimedo Labs)
- Rohde Schwarz GmbH Co KG
- Seoul National University
- Siemens Digital Industries Software
- Singapore University of Technology and Design
- SK Telecom
- SpectrEdge Wireless Inc
- Sungkyunkwan University System Software Lab

+ pending applications



AI-RAN ユースケース

2025/12/1 安川電機との協業発表

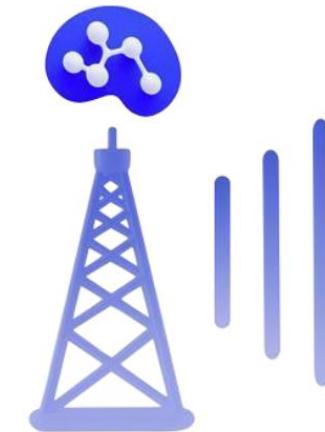
≡ SoftBank

YASKAWA

≡ SoftBank



AIロボティクス



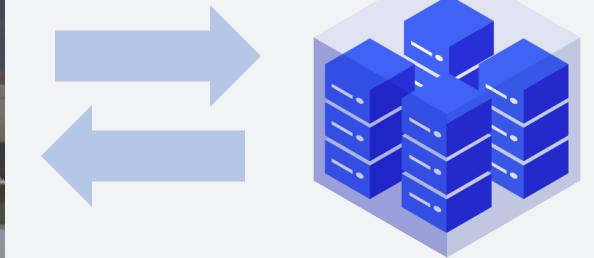
AI-RAN

AI-RANによるスマートビルの実現



フィジカルAIデモ

iREX 2025にてAI-RANとロボット、ビルOSが連携したデモを展示



キャリア品質の通信 × エッジAI



新サービス

新たなアーキテクチャー

すべてのエンタープライズへ新たな機会を提供



SoftBank