Wireless World Research Forum (WWRF) 2025

J.J. Vegas Olmos
Principal Engineer
Research Program Coordinator

Spoiler Alert and Agenda

- Working title of the session "Breakthrough Technologies Enabling 6G including usable Security and Privacy"
  - Breakthrough = state-of-the-art
  - Enabling = should be usable for X
  - 6G = really big field
  - Security = very large field
  - Privacy = very large field

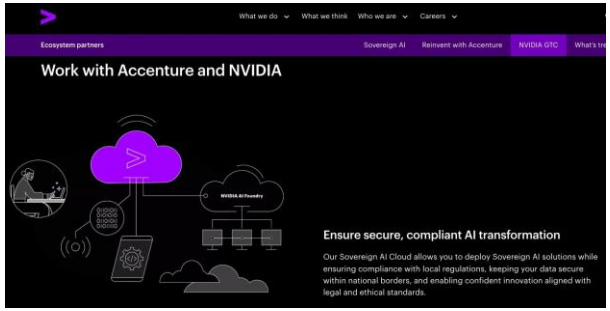
- It will be hard to please everyone in terms of content matching the title of the session

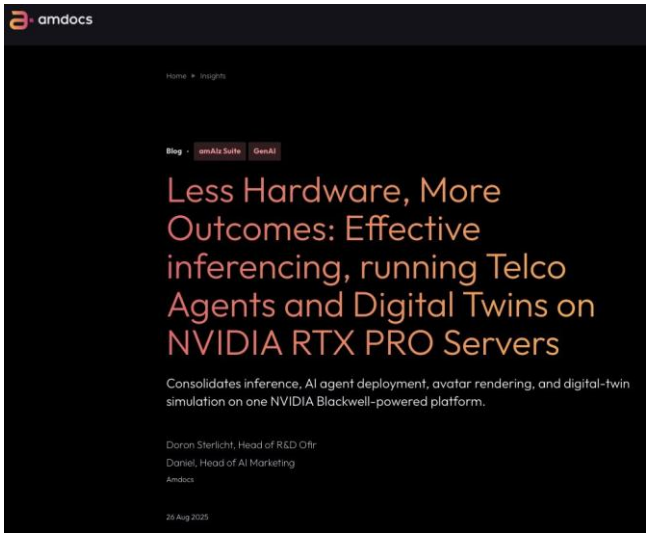
- Agenda
  - What is actually the telco sector?
  - AERIAL platform
  - Technologies

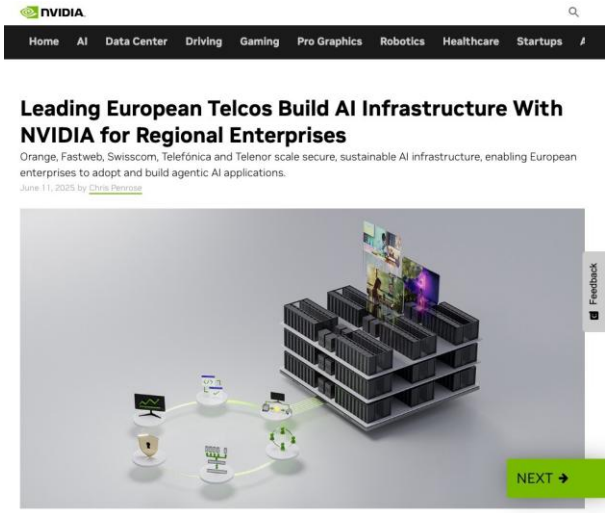
- Aprox 10-12 minutes

Recent Telco News

Telco is no longer telco, Telco is Data centers + AI + services + infrastructure…







NVIDIA

Reinventing Telcos With AI

Next Wave of Telco Transformation – from the data center to the edge

Gen AI For Operations

Save costs by improving Customer Experience, Network Operations and Employee Productivity

AI Factories

Generate new revenue stream by enabling AI infra/platform to government, enterprises, and startups

AI Enabled RAN

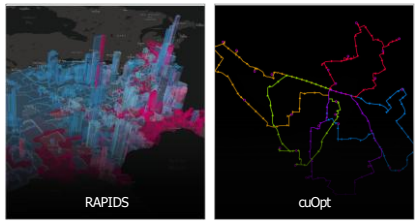
Obtain SW Defined + High Performance + AI-RAN

AI Infrastructure

Maximize utilization by accelerating/offload NFV Infrastructure and Applications

Telco means something different for each stakeholder

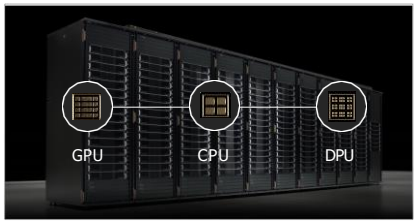
AT&T



Accelerated Data Processing

Technician Route Optimization

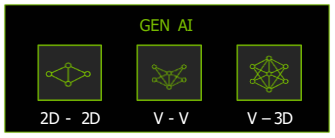
Accelerated Network Infrastructure

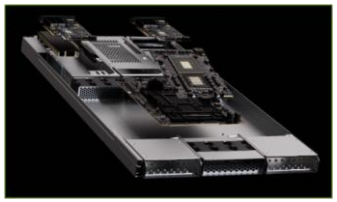


Energy & Cost Savings

Accelerated Networking | RAPIDS & cuOpt

SoftBank NVIDIA





Grace-Hopper Aerial Server

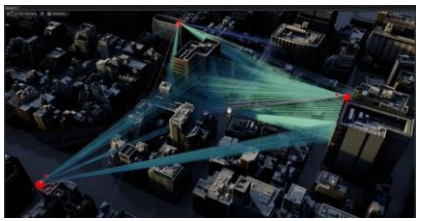
Software-Defined 5G vRAN

MGX Grace Hopper, NVIDIA Aerial 5G

ERICSSON HEAVY.AI

Digital Twin of the Network

Aerial Omniverse Digital Twin



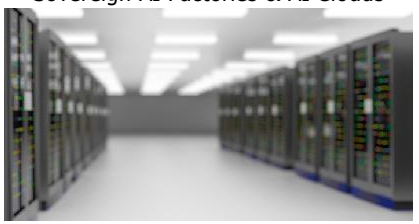
Network Simulation

Design and Test 5G & 6G Networks

NOTRAFFIC

AI at the Edge — Traffic Management to Reduce Emissions

Sovereign AI Factories & AI Clouds

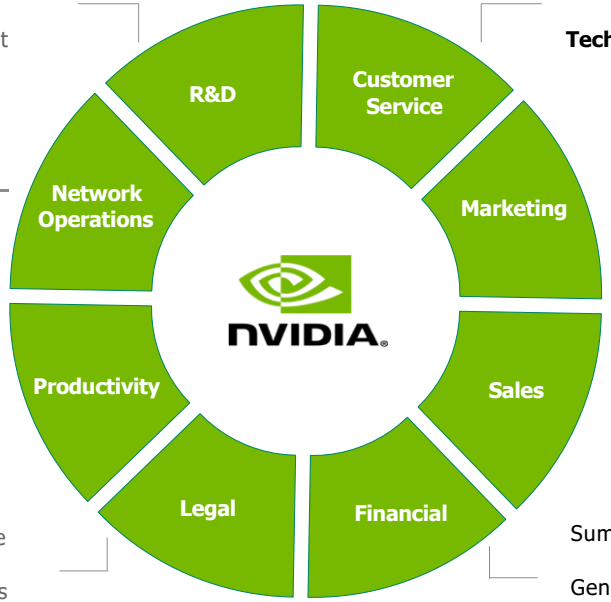


AI Infrastructure & Services

Edge Computing | Sovereign AI Factories

And it is evolving very rapidly

For example : LLMs for Ops functions across the board, especially in Customer Service, Network Operations, and Productivity



Market Research & Competitor Analysis

Internal Research Assistant

Virtual Assistant & Self Service Portal

Technical Support Services

Automated Infrastructure Documentation

Data Mining for Network Log Analysis

Configuring Network Management Systems

Field Technician Operation

Personalized email campaigns

Content Generation

Onboarding of New Employees

"How-to" for Employees

Email summaries, PPT templates

Lead Generation by Content

Automated Deal Flow

Summarization across multiple platforms

Summarization of Contracts / Red Line

Generating Contracts for Business Units

Summarization of Reports

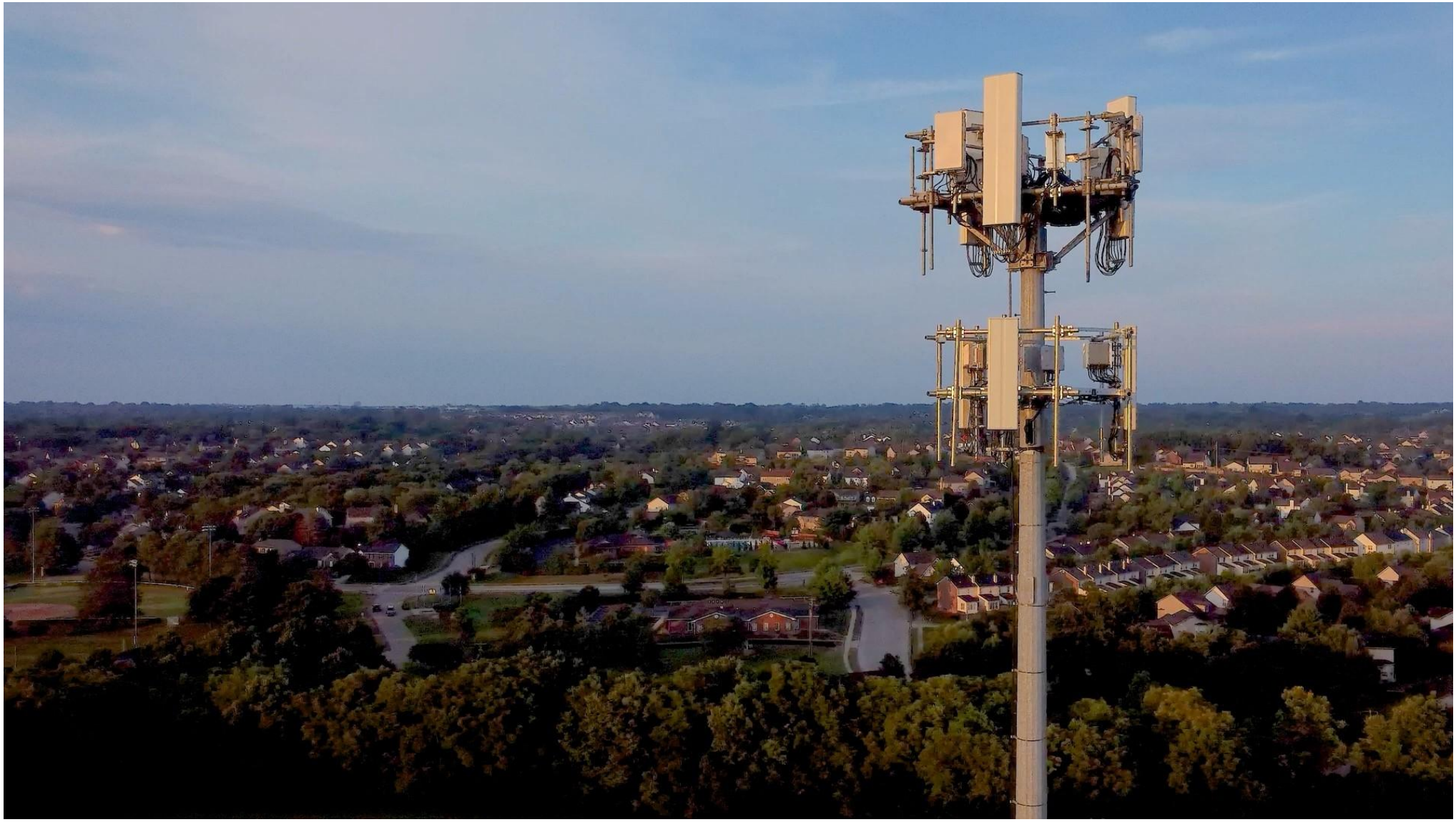
Generating Reports

Improve Customer Engagement

Improve Productivity and Reduce Risk

Reduce Costs via Operational Efficiencies

Faster Time to Market

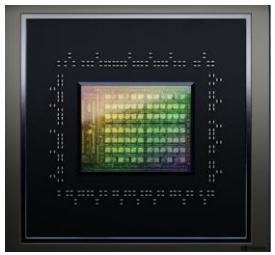


AERIAL platform

Our approach to 5G/6G

Virtualize it –Virtualize it all !

Enhanced Performance and Energy Savings for vRAN, Packet Core, and Telco Edge Applications
Baseband : all virtual, Wireless : HW front-end



Grace C1

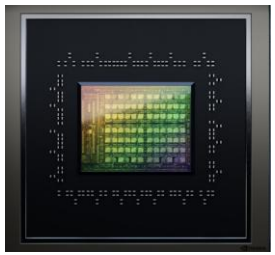
L1/L2+ Stack

NVIDIA Networking:

ConnectX-6 or 7 Adapter

Enabling eCPRI, PTP, packet filtering, time stamping, ordering, MACSEC

vRAN
DU



Grace C1

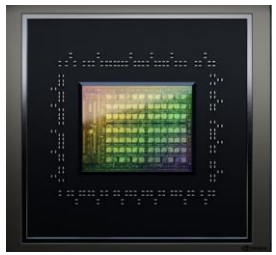
AMF, SMF, UDM, PCF

NVIDIA Networking:

ConnectX-6 or 7 Adapter

Accelerating UPF, dUPF, VNF, IPSEC, TLS , Encap/Decap

5G Packet Core
UPF, AMF, SMF, PCF, UDM, etc.



Grace C1

Compute Edge Applications

NVIDIA Networking:

BlueField-3 DPU

Offloading VNFs & enabling isolation

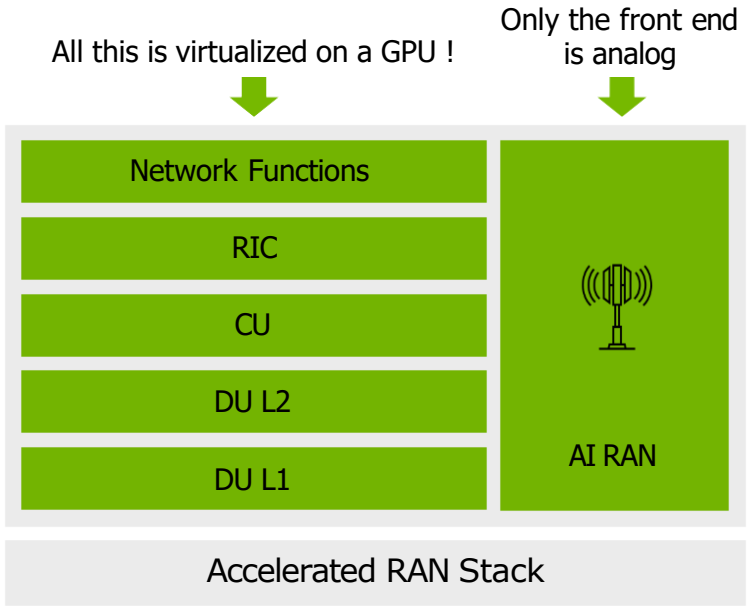
ConnectX-6 or 7 Adapter

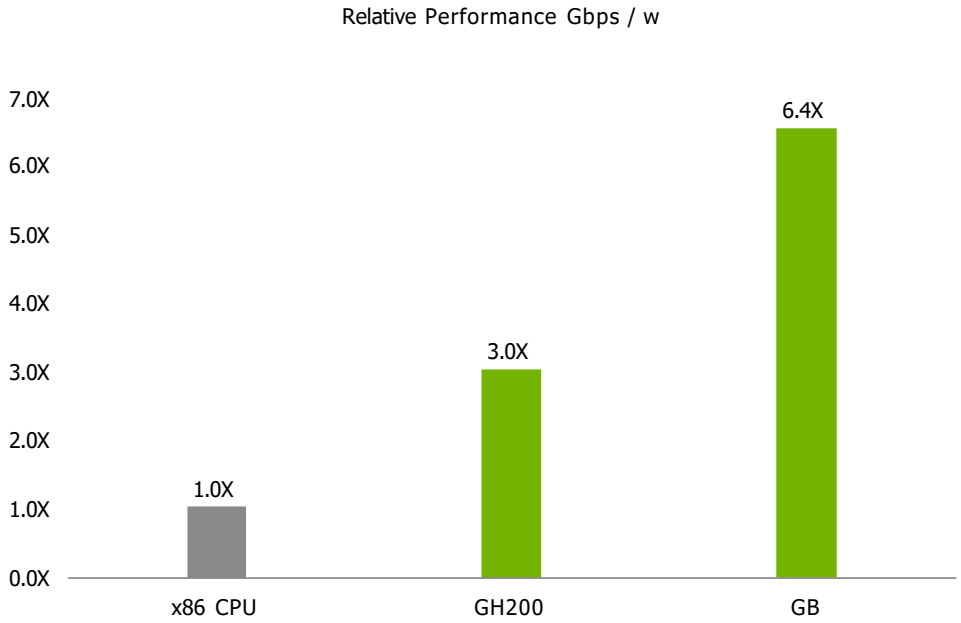
Accelerating VNFs

CDN, MEC, Regional DC
VNF, Routing, Security, Streaming

Aerial CUDA-Accelerated RAN

Improve TCO/ROI and Boost Spectral Efficiency

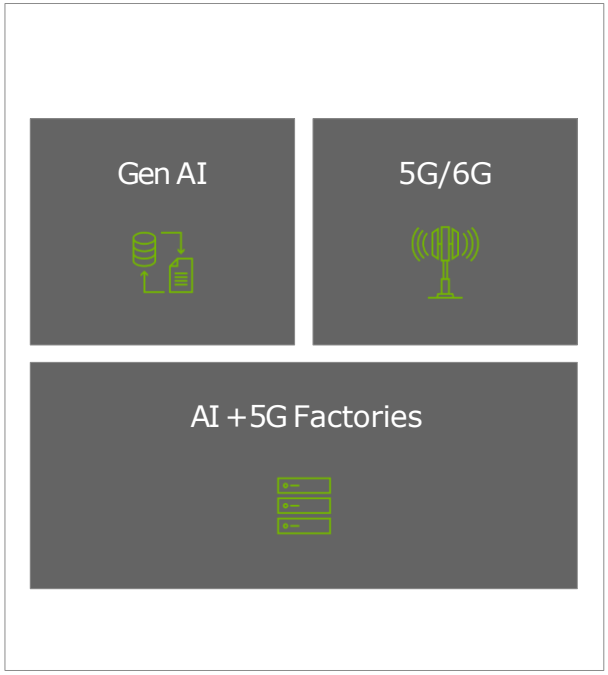




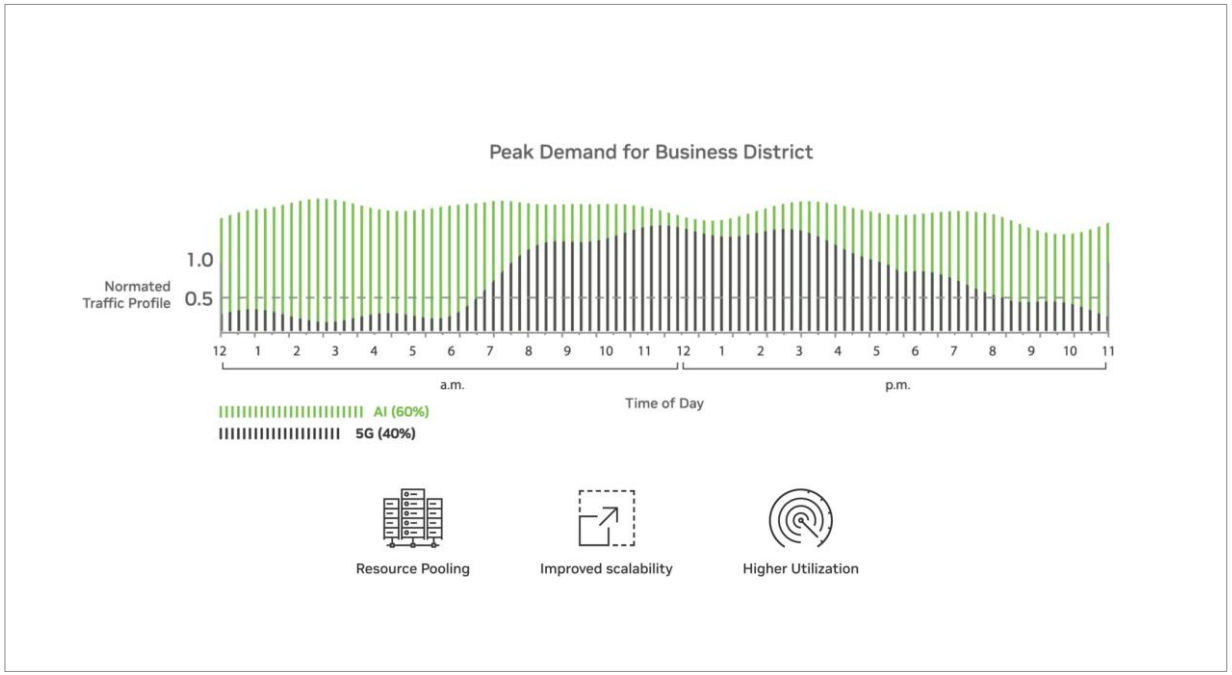
~2x TCO Gains in Spectral Efficiency on GPU

AI and 5G On The Same Infrastructure

Maximizing Data Center Utilization by Co-Hosting 5G RAN and Generative AI



Efficiency | Utilization | Monetization

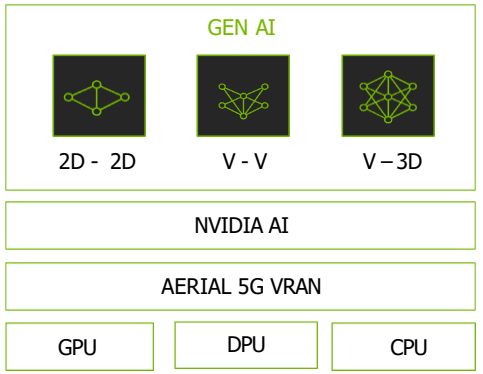


Dynamically Allocate Resources as Demand Changes Throughout the Day

Reinvent 5G Data Centers With Grace Hopper

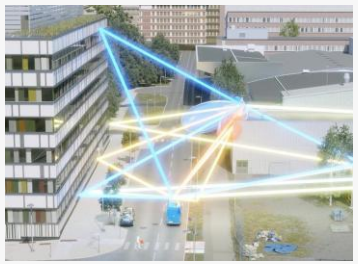
Software Defined 5G | Cloud Native | Gen AI

- One cloud platform for 5G and gen AI
- 2.5X power efficiency over state-of-the-art 5G vRAN
- Highest performance 5G vRAN
- AI-powered wireless communications, 6G Ready
- New monetization opportunities



NVIDIA Aerial: Driving Innovation From 5G to 6G

Three Platform Tools for Commercialization and Research

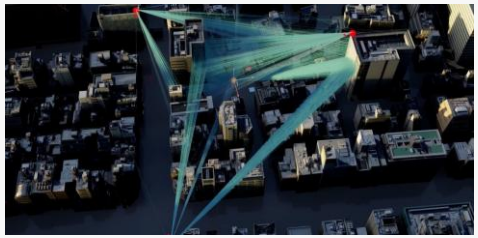


Aerial CUDA-Accelerated RAN

Building Commercial-Grade Networks

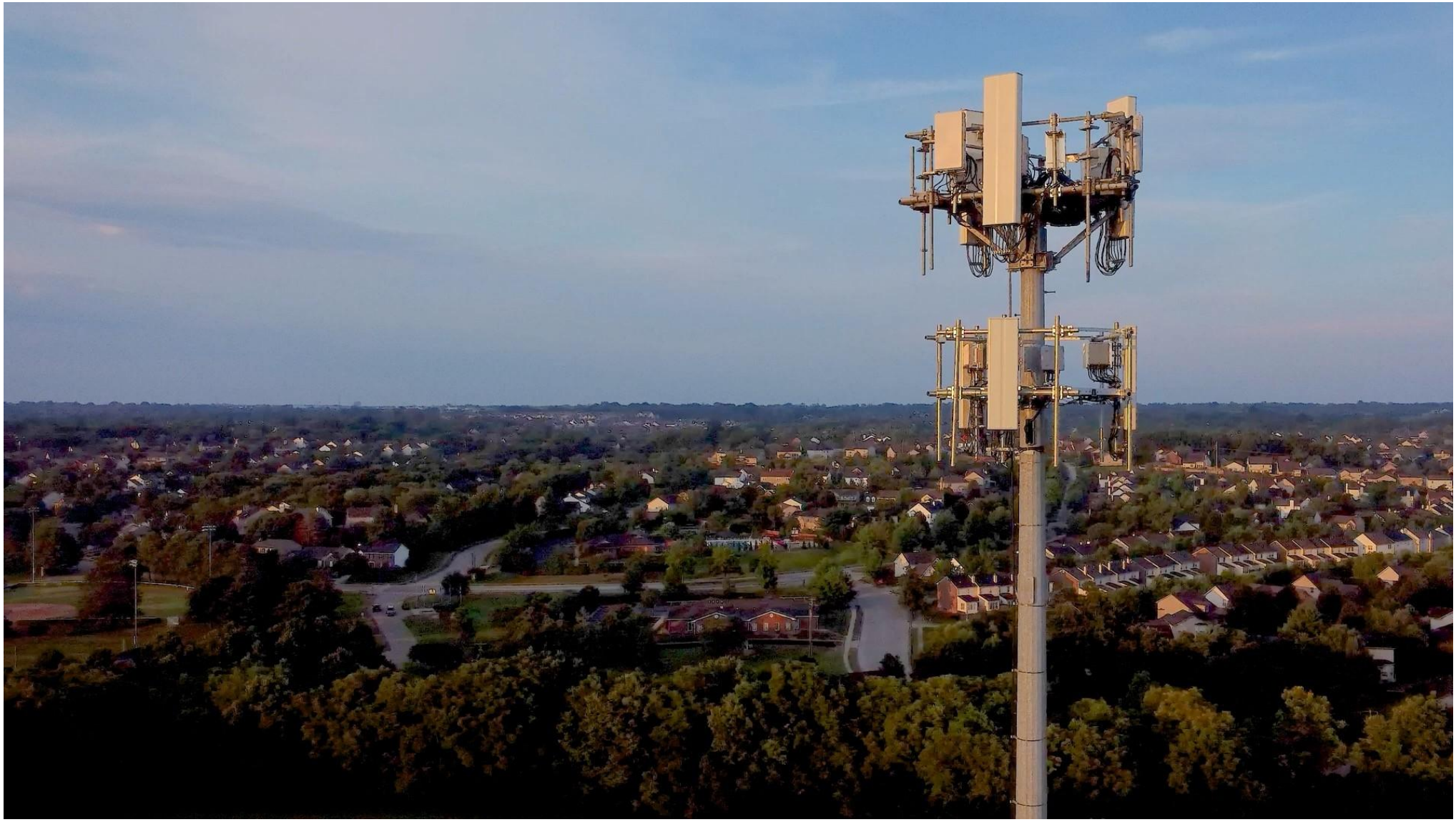
Aerial AI Radio Frameworks

Enabling AI/ML Training and Inferencing for the RAN



Aerial Omniverse Digital Twin

System-Level Simulator for 5G and 6G Research

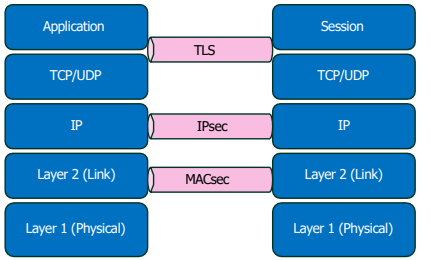


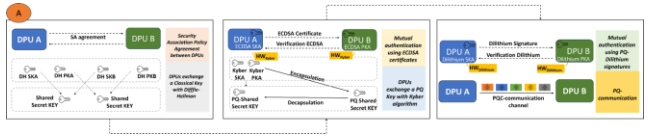
Enabling Technologies

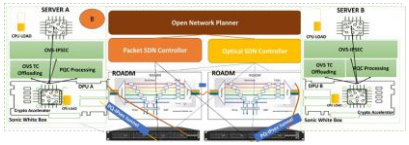
Two examples – that essentially provide security and privacy

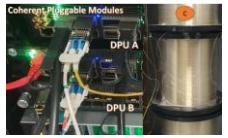
PQC is already standardized and you use it already without knowing it

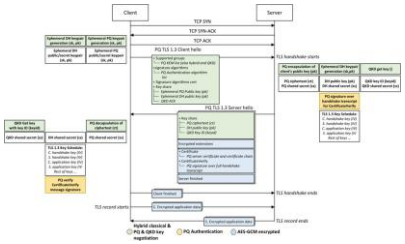
- PQC already integrated by many hyperscalers
  - If you use chrome and connect to a google service, you are using PQC in the TLS
- We have demonstrated operation of all PQC algorithms in network interfaces cards
  - CPU operation
  - Hardware offloads (CPU accelerators)
  - Hardware offloads (GPU accelerators)



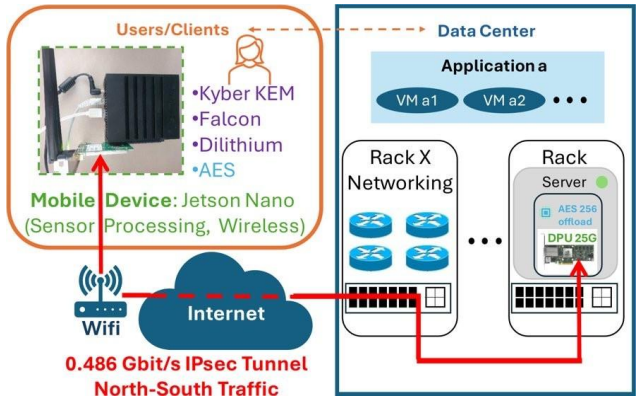








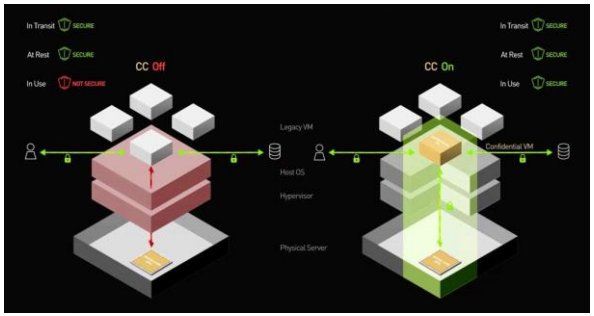
- Integration too with QKD



- Since AERIAL is virtualized, effectively we can use PQC for 5G/6G
  - The PQC algos are offloaded
  - No extra latency
  - Full line-rate capacity

Confidential Computing

Protect data while being processed (full protection in transit, at rest and in processing)



- Confidential computing uses trusted execution environments (TEEs)
  - **Data-in-use protection** → Keeps data encrypted in memory and only decrypted inside the enclave.
  - **Attestation** → Verifies that code running inside the enclave hasn't been tampered with.
  - **Isolation** → Prevents other applications, the OS, or even cloud admins from accessing the enclave.

- The RAN is actually a small data center
  - There are different GPU-modes:
    - **CC-Off**: normal mode.
    - **CC-On**: all confidentiality protection features enabled (firewalls, disabling performance counters that could leak info).
    - **CC-DevTools**: a mode that's like CC-On in many respects but enables development tools (e.g. profiling) thus relaxing some protections for debugging.
  - **Software / Open-Source Tools**
    - **nvTrust**: An open-source repo by NVIDIA that contains "utilities & tools, open-source code, and SDKs leveraged when using NVIDIA solutions in trusted environments (confidential computing)."
    - **GPU Operator with Confidential Containers and Kata**: For cloud-native / Kubernetes settings, NVIDIA is enabling confidential containers. These combine GPU isolation plus CPU TEEs (Intel TDX, AMD SEV/SEV-SNP) so that the whole stack is protected.
  - **Attestation & TEE support via CPU + GPU combination**
    - To support a truly confidential compute environment, NVIDIA's model includes requiring CPU virtualization/encryption TEE tech (e.g. **AMD SEV-SNP, Intel TDX**) so that the VM itself (and its memory) is protected. Then the GPU (H100) can be used within that VM to also enforce confidentiality.


NVIDIA
Thanks