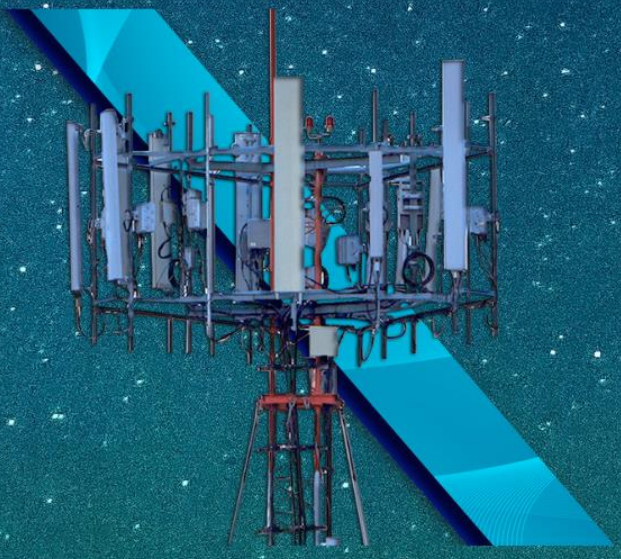
DIPL.ING.(FH)KLAUS ROCK

# HTTP-QuSS

## HTTP - QUANTUM SPEED AND SECURITY

February 14, 2022

## WI-FI | 3-4-5G TOWER SEGMENT

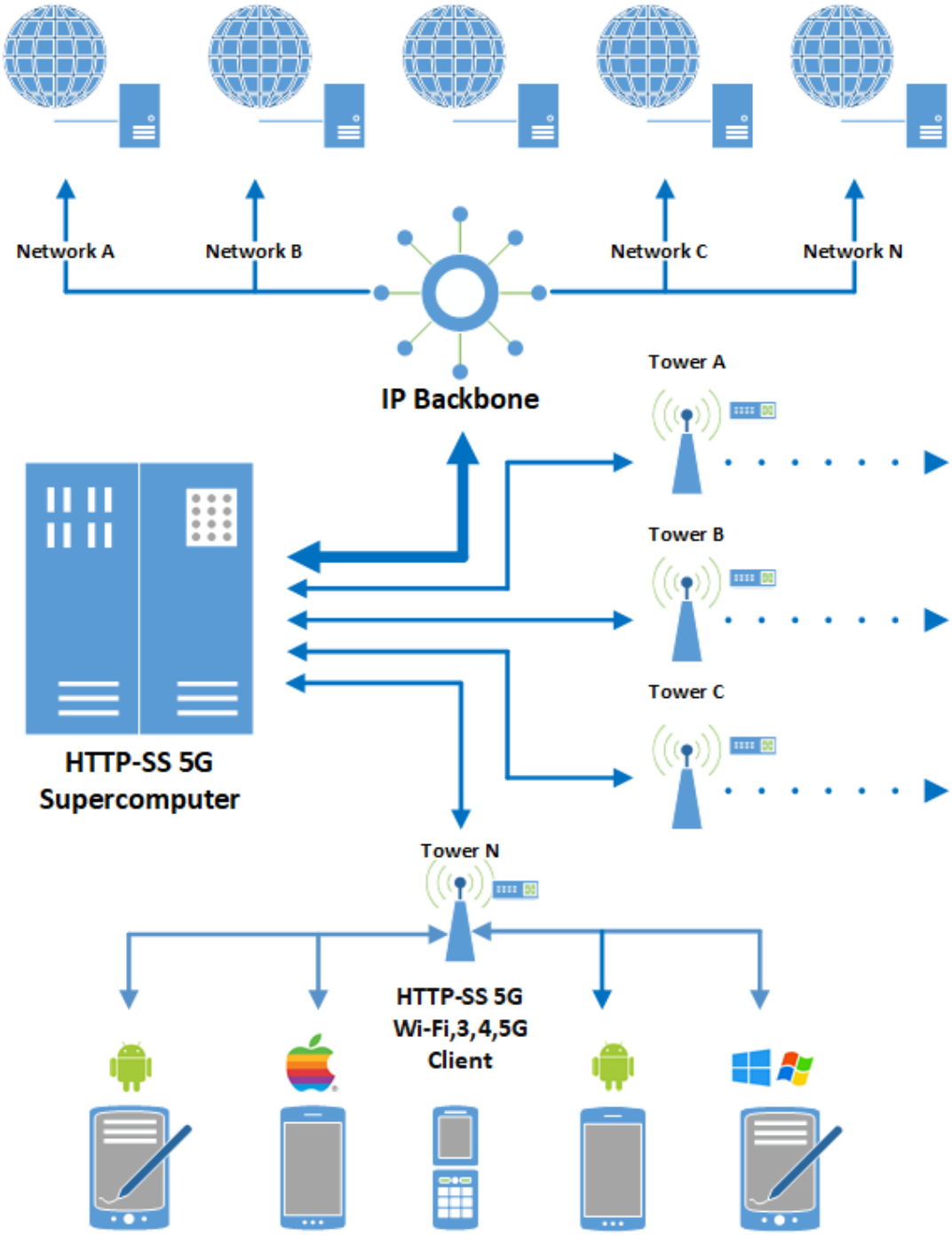


ROCK TECHNOLOGIES
Bonhoefferstr. 37 | 73432 Aalen | Germany | +49-7367-9222-958

# What is HTTP-QuSS

**HTTP-QuSS** consists of an **Embedded Supercomputer** of the newest Generation stationed at an IP Backbone Site and an **AI Edge Server** for **Wi-Fi, 3/4/5G Towers**.

The **HTTP-QuSS Supercomputer** eliminates the Bandwidth destroying Latency Issue caused by Protocol Handshakes, Long Distances, many Hops, Data Losses, congested Networks by many Users etc. and therefore existing Wi-Fi and 4G Networks are able to provide 5G Requirements like low Latencies and Gbit/s Bandwidths. 5G Networks on the other Hand can even fulfil their own claimed Features when accessing the normal Internet and not only connected Devices within a Basic Cell.



# Unique Selling Proposition

- **Total Elimination of the TCP Latency Problem and related Bandwidth Losses Within Wi-Fi, 3,4,5G Communication Networks**
  - ✓ Through a new **HTTP 5G S**ingle **S**tream Architecture and Technology
  - ✓ By using the next Generation of Embedded Server Supercomputer Hardware
  - ✓ By Artificial Intelligence supported Data Transmission and highly Parallelized Process Chains
  - ✓ Integrated Dynamic Bandwidth Shaping and Slicing for max Bandwidth Usage
- **Highly reduced Data Transmission for all WEB Objects and Files**
  - ✓ By AI supported WEB Object Push and File Descriptor Delta Data Algorithm
  - ✓ And 1 Round Trip Protocol Handshake
  - ✓ and therefore 90 % reduced secure Data Transmission
  - ✓ All WEB- and TCP Applications are supported
- **5G Performance and Latency for existing 4G and Wi-Fi Towers**
  - ✓ With its own highly efficient Process and Processor Management
  - ✓ Integrated parallel Processing Architecture without Site Effects
  - ✓ Smart parallel Process Chains with highly efficient Inter Communication
  - ✓ And much more …
- **Quantum Secure Cyber Security**
  - ✓ Through keyless 2 Level Data Encryption
- **AI Edge Tower Hardware and Software Solution**
  - ✓ No need of special Hardware on Client Site
  - ✓ No need of special Browser or Proxy Settings on Client Site
- **Fast and transparent Wi-Fi, 3,4,5G Tower Installation and Integration**
  - ✓ No need of Reconfigurations in existing Infrastructure
  - ✓ Fully transparent for User
  - ✓ Supports all Devices PC''s, Laptops, Smartphones (Android, iOS), DSL- and VSAT Routers etc.etc.

## Advantages for Telecoms

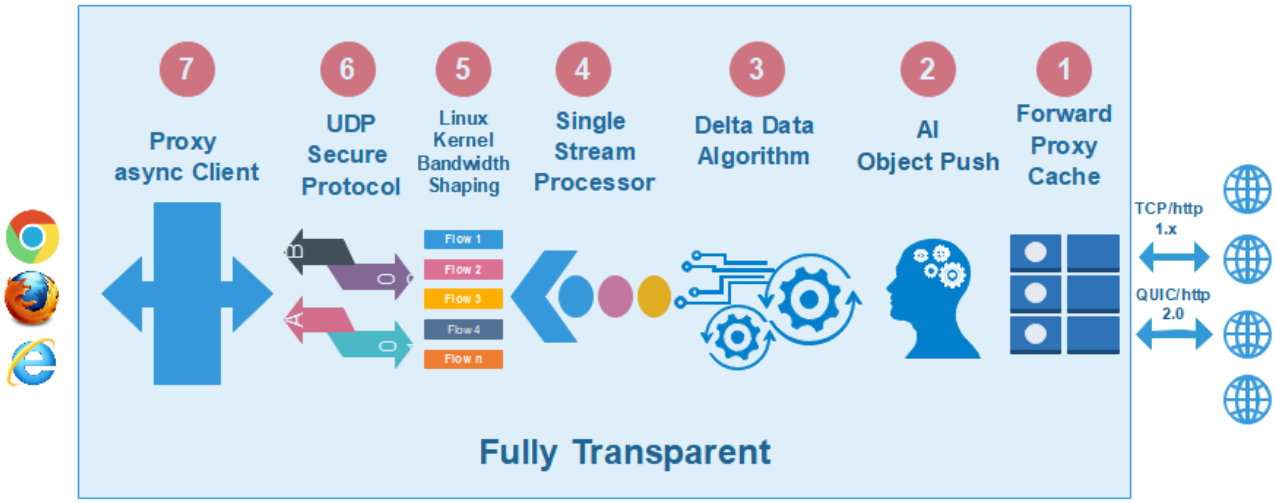
- Best use of limited Resources (Bandwidths)
- More efficient and faster Data Transmission
  - ✓ Faster WEB Page Loading
  - ✓ Significantly better Performance
- Transparent and fast Integration into existing Infrastructure
- Cost Savings
- Higher Customer Satisfaction
- New Customers, Competitive Displacement
- Image Profit

## Advantages for End-user

- Can double his available Bandwidth.
- Can reduce his monthly Costs dramatically because Data Volume will be reduced by **90 %**
- Much better WEB Page Load Time and App Performance
- 5G Performance even in Wi-Fi and 4G Networks
- In Satellite, WLAN / Wi-Fi, and Mobile Networks for no Latency Issues
- Much better Performance for Cloud Applications
- Fast Cloud Data Backup
- Broadband Availability, even in rural Areas
- No special App Software is needed
- Zero Client Installation Scenario

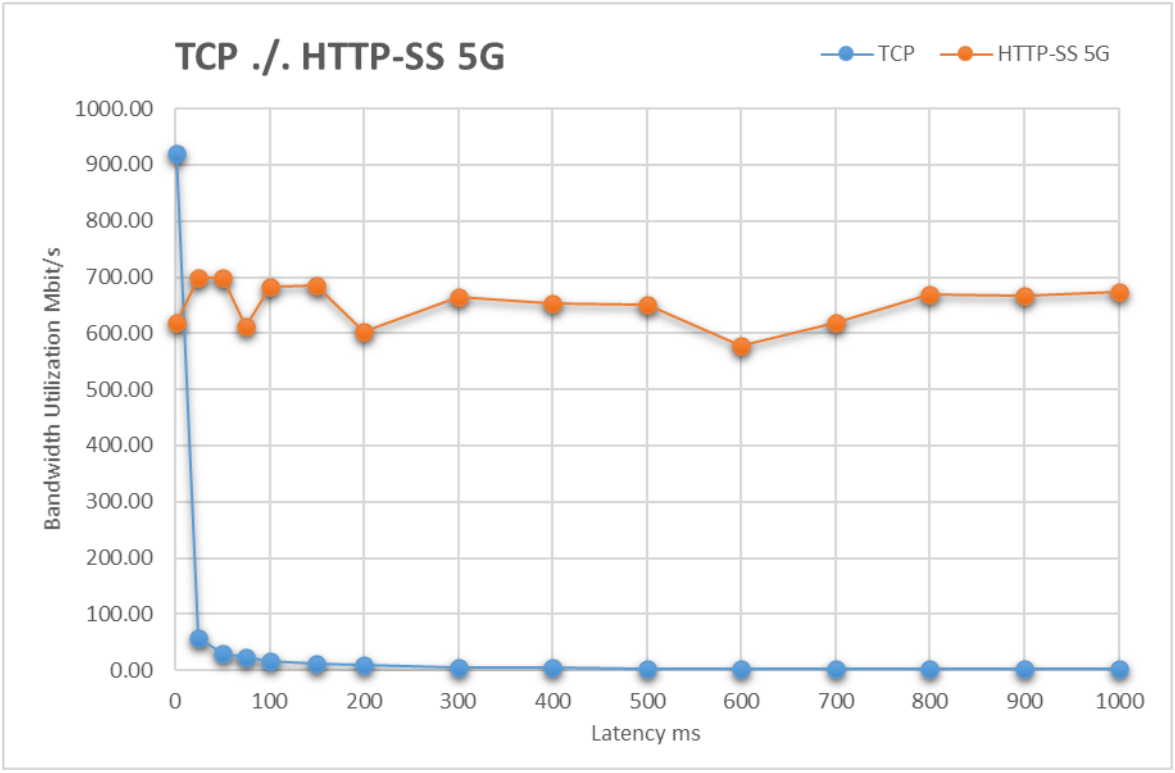
# Transparent Wi-Fi, 3,4,5G Network Integration

- No need to change Network Infrastructure by fully **transparent Tower Integration**
- Supports **all common Browsers**

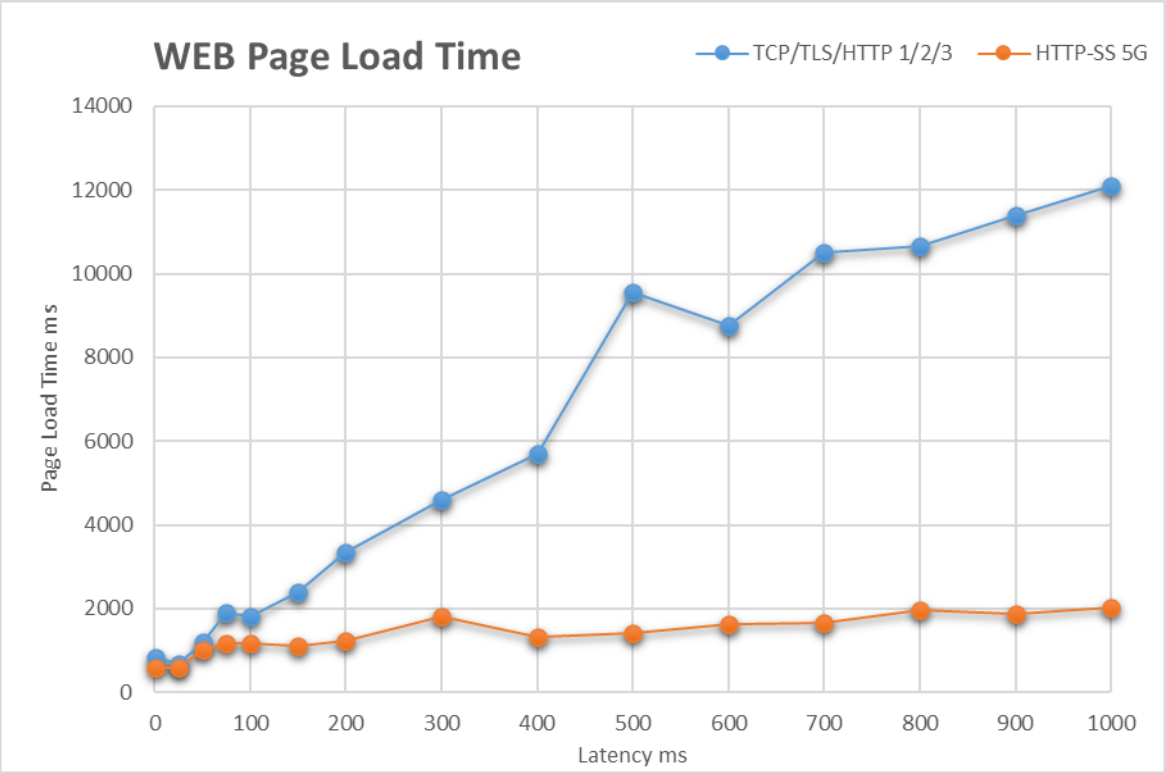



# No Latency to Bandwidth Dependency

- **No Dependency** between **R**ound **T**rip **T**ime and **TCP Bandwidth**
- **Breakthrough** in higher bandwidths Regions even at very long Round Trips



# Faster WEB Object Transfer Time



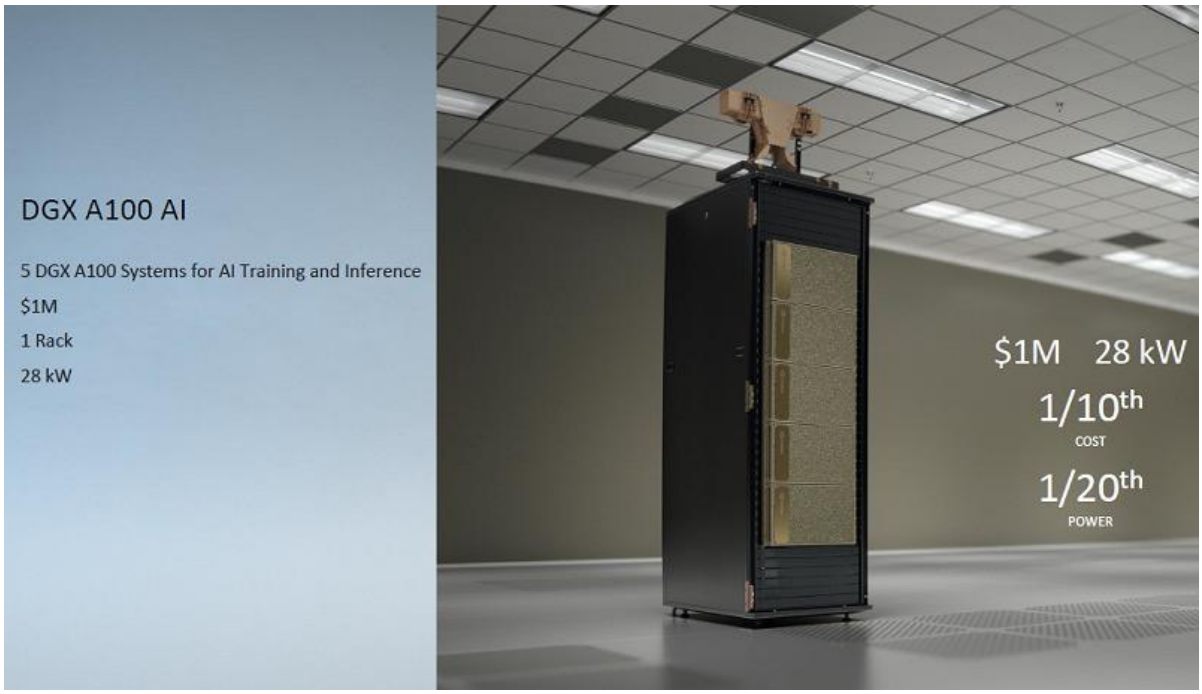

**WEB Object Transfer Time** nearly constant even at very high Round Trips

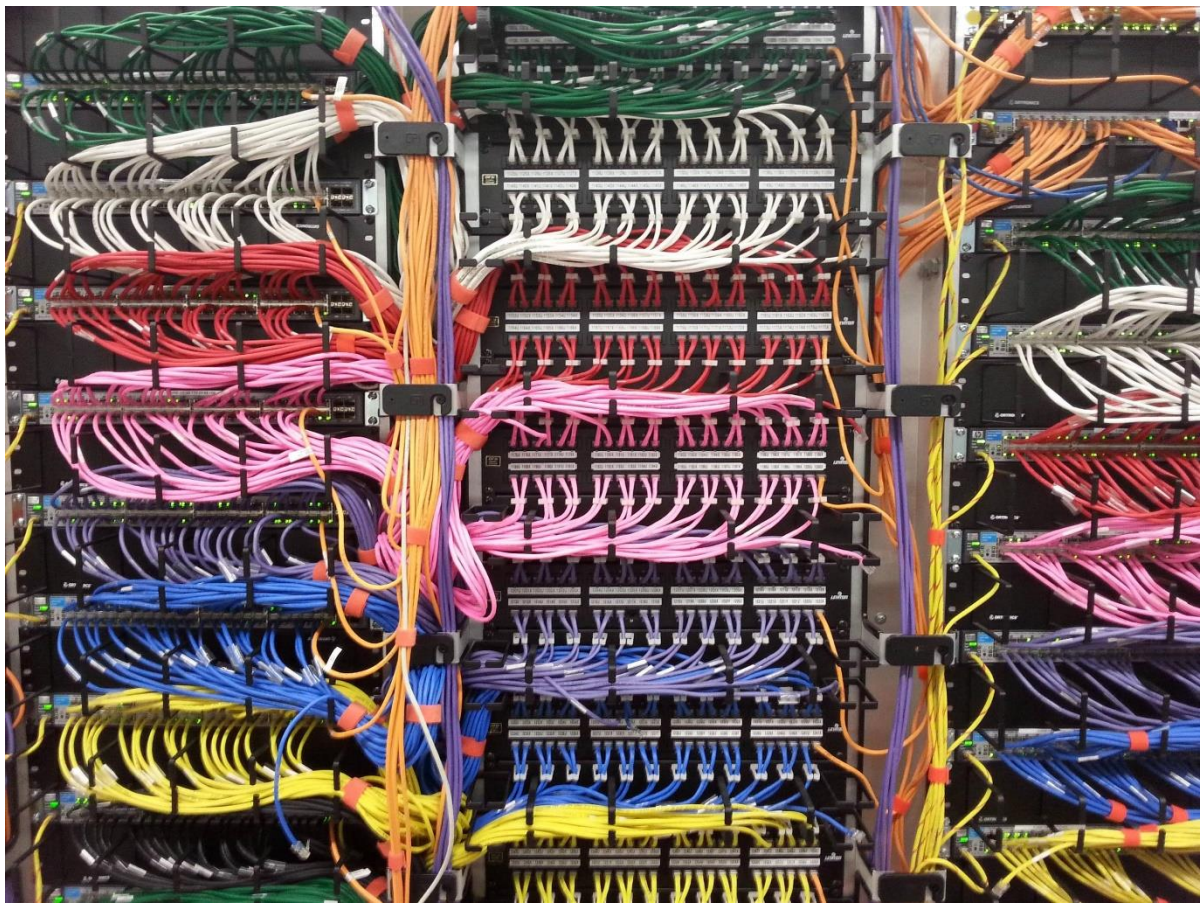
- **50 %** better Performance within RTT Range 50 – 100 ms
- Up To **500 %** better Performance in very High Latency IP Networks
- **No visible Difference** of Browser Page Load Time within higher or very high Latency Networks
- Even much better Performance in very High Bandwidth Regions **over 50 Mbit/s**

| RTT / ms | Transfer Time / s | | Performance |
|---|---|---|---|
| | HTTP v. TCP | HTTP-SS uncompressed* | |
| 50 | 1,16 | 1,07 | 8,41% |
| 100 | 1,95 | 1,098 | 77,60% |
| 150 | 2,734 | 1,19 | 129,75% |
| 200 | 3,533 | 1,216 | 190,54% |
| 250 | 3,535 | 1,272 | 177,91% |
| 300 | 4,169 | 1,32 | 215,83% |
| 350 | 4,794 | 1,364 | 251,47% |
| 400 | 5,536 | 1,41 | 292,62% |
| 450 | 6,166 | 1,465 | 320,89% |
| 500 | 6,817 | 1,524 | 347,31% |
| 550 | 7,475 | 1,542 | 384,76% |
| 600 | 8,099 | 1,621 | 399,63% |
| 650 | 8,728 | 1,665 | 424,20% |
| 700 | 9,384 | 1,715 | 447,17% |
| 750 | 10,045 | 1,774 | 466,23% |
| 800 | 10,686 | 1,817 | 488,11% |
| 850 | 11,31 | 1,874 | 503,52% |
| 900 | 11,963 | 1,921 | 522,75% |
| 950 | 12,63 | 1,986 | 535,95% |
| 1000 | 13,276 | 2,033 | 553,03% |

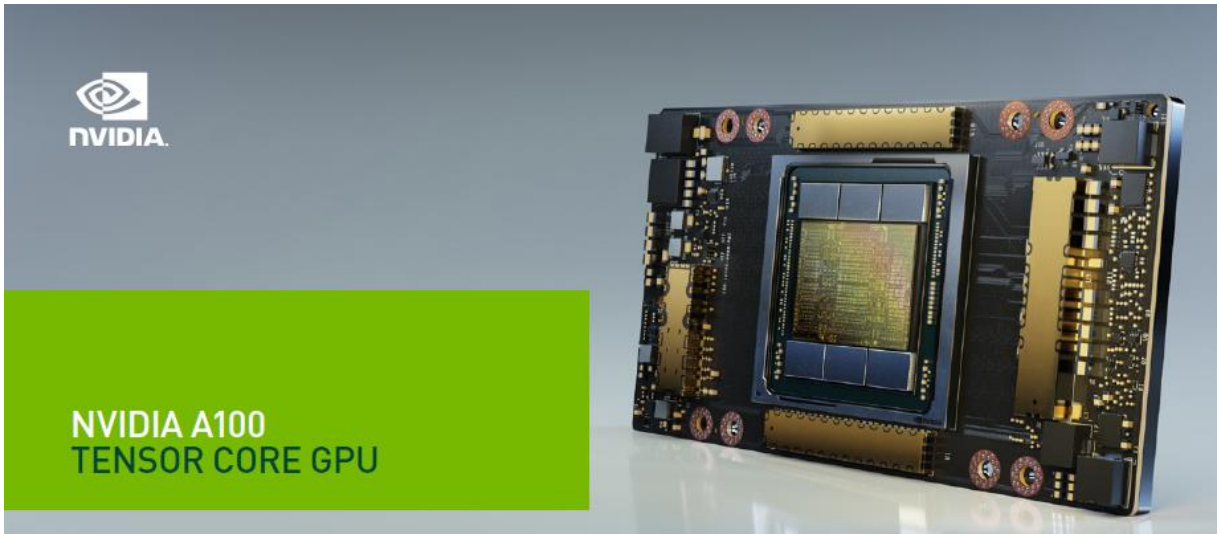
## NVIDIA A-100 as HTTP-QuSS IP-Backbone Supercomputer

The Internet backbone may be defined by the principal data routes between large, strategically interconnected computer networks and core routers of the Internet. These data routes are hosted by commercial, government, academic and other high-capacity network centers, as well as the Internet exchange points and network access points, that exchange Internet traffic between the countries, continents, and across the oceans.



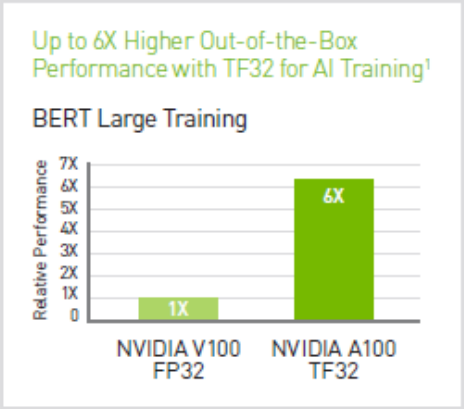



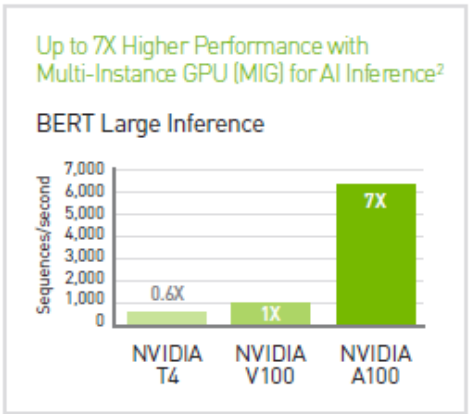
# NVIDIA A-100 Embedded Supercomputer

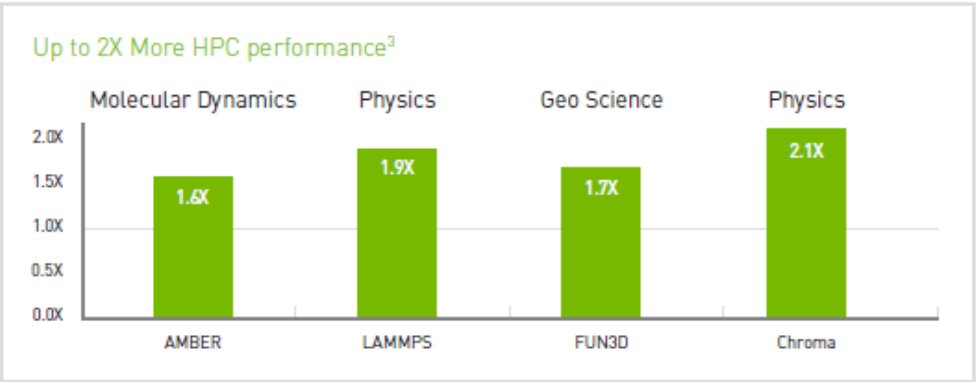



## Unprecedented Acceleration at Every Scale

The NVIDIA A100 Tensor Core GPU delivers unprecedented acceleration at every scale for AI, data analytics, and HPC to tackle the world's toughest computing challenges. As the engine of the NVIDIA data center platform, A100 can efficiently scale up to thousands of GPUs or, using new Multi-Instance GPU (MIG) technology, can be partitioned into seven isolated GPU instances to accelerate workloads of all sizes. A100's third-generation Tensor Core technology now accelerates more levels of precision for diverse workloads, speeding time to insight as well as time to market.



# SYSTEM SPECIFICATIONS (PEAK PERFORMANCE)

| | NVIDIA A100 for NVIDIA HGX™ | NVIDIA A100 for PCIe |
|---|---|---|
| GPU Architecture | NVIDIA Ampere | |
| Double-Precision Performance | FP64: 9.7 TFLOPS<br>FP64 Tensor Core: 19.5 TFLOPS | |
| Single-Precision Performance | FP32: 19.5 TFLOPS<br>Tensor Float 32 (TF32): 156 TFLOPS \| 312 TFLOPS* | |
| Half-Precision Performance | 312 TFLOPS \| 624 TFLOPS* | |
| Bfloat16 | 312 TFLOPS \| 624 TFLOPS* | |
| Integer Performance | INT8: 624 TOPS \| 1,248 TOPS*<br>INT4: 1,248 TOPS \| 2,496 TOPS* | |
| GPU Memory | 40 GB HBM2 | |
| Memory Bandwidth | 1.6 TB/sec | |
| Error-Correcting Code | Yes | |
| Interconnect Interface | PCIe Gen4: 64 GB/sec Third generation NVIDIA® NVLink®: 600 GB/sec** | PCIe Gen4: 64 GB/sec Third generation NVIDIA® NVLink®: 600 GB/sec** |
| Form Factor | 4/8 SXM GPUs in NVIDIA HGX™ A100 | PCIe |
| Multi-Instance GPU (MIG) | Up to 7 GPU instances | |
| Max Power Consumption | 400 W | 250 W |
| Delivered Performance for Top Apps | 100% | 90% |
| Thermal Solution | Passive | |
| Compute APIs | CUDA®, DirectCompute, OpenCL™, OpenACC® | |

* Structural sparsity enabled

** SXM GPUs via HGX A100 server boards; PCIe GPUs via NVLink Bridge for up to 2 GPUs

## GROUNDBREAKING INNOVATIONS

### NVIDIA AMPERE ARCHITECTURE

A100 accelerates workloads big and small. Whether using MIG to partition an A100 GPU into smaller instances, or NVLink to connect multiple GPUs to accelerate large-scale workloads, A100 can readily handle different-sized acceleration needs, from the smallest job to the biggest multi-node workload. A100's versatility means IT managers can maximize the utility of every GPU in their data center around the clock.

## THIRD-GENERATION TENSOR CORES

A100 delivers 312 teraFLOPS (TFLOPS) of deep learning performance. That's 20X Tensor FLOPS for deep learning training and 20X Tensor TOPS for deep learning inference compared to NVIDIA Volta™ GPUs.

## NEXT-GENERATION NVLINK

NVIDIA NVLink in A100 delivers 2X higher throughput compared to the previous generation. When combined with NVIDIA NVSwitch™, up to 16 A100 GPUs can be interconnected at up to 600 gigabytes per second (GB/sec) to unleash the highest application performance possible on a single server. NVLink is available in A100 SXM GPUs via HGX A100 server boards and in PCIe GPUs via an NVLink Bridge for up to 2 GPUs.

## MULTI-INSTANCE GPU (MIG)

An A100 GPU can be partitioned into as many as seven GPU instances, fully isolated at the hardware level with their own high-bandwidth memory, cache, and compute cores. MIG gives developers access to breakthrough acceleration for all their applications, and IT administrators can offer right-sized GPU acceleration for every job, optimizing utilization and expanding access to every user and application.

## HBM2

With 40 gigabytes (GB) of high-bandwidth memory (HBM2), A100 delivers improved raw bandwidth of 1.6TB/sec, as well as higher dynamic random-access memory (DRAM) utilization efficiency at 95 percent. A100 delivers 1.7X higher memory bandwidth over the previous generation.

## STRUCTURAL SPARSITY

AI networks are big, having millions to billions of parameters. Not all these parameters are needed for accurate predictions, and some can be converted to zeros to make the models "sparse" without compromising accuracy. Tensor Cores in A100 can provide up to 2X higher performance for sparse models. While the sparsity feature more readily benefits AI inference, it can also improve the performance of model training.

The NVIDIA A100 Tensor Core GPU is the flagship product of the NVIDIA data center platform for deep learning, HPC, and data analytics. The platform accelerates over 700 HPC applications and every major deep learning framework. It's available everywhere, from desktops to servers to cloud services, delivering both dramatic performance gains and cost-saving opportunities.

# NVIDIA EGX-Ready Servers integrated into Wi-Fi, 3/4/5G Towers

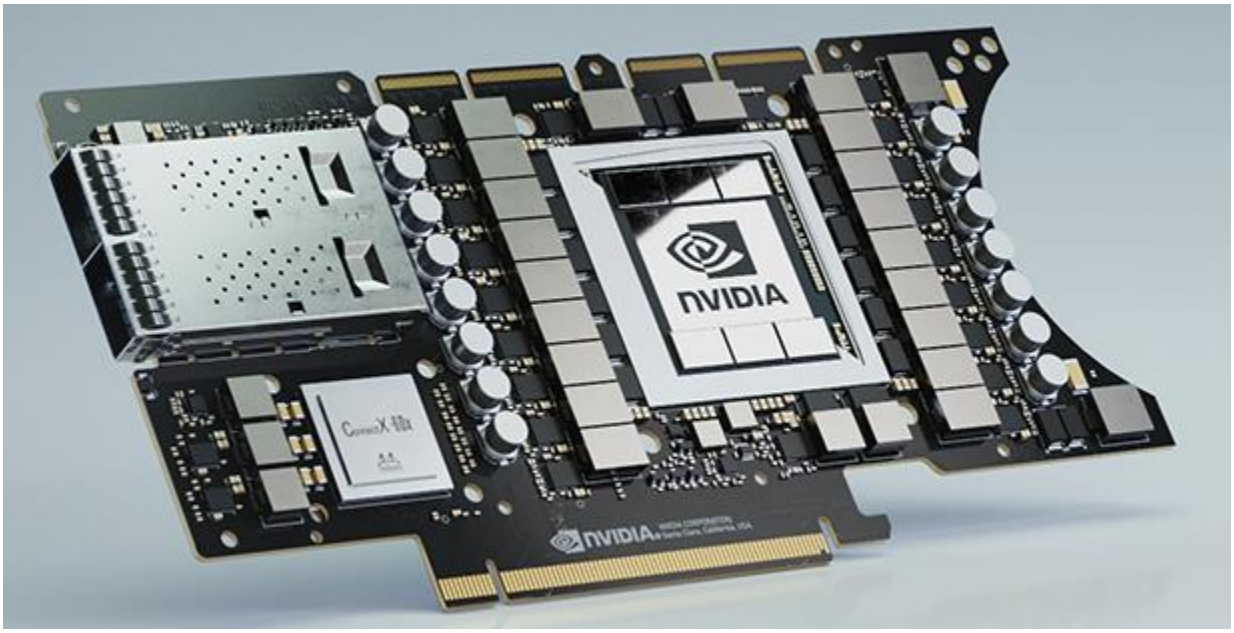
## NVIDIA Jetson Family

The EGX hardware portfolio starts with the power-efficient NVIDIA® Jetson Family, which includes the small but mighty Jetson Nano™ and Xavier™ NX providing between 0.5 to 21 trillion operations per second (TOPS) for tasks such as image recognition and sensor fusion. And it scales all the way to a full rack of NVIDIA T4 servers, delivering more than 10,000 TOPS to serve hundreds of users with real-time speech recognition and other complex AI experiences.


## NVIDIA EGX A100

More Powerful, Secure Edge AI

Is Part of the NVIDIA EGX™ platform, NVIDIA EGX™ A100 combines the powerful performance of the NVIDIA Ampere architecture with the disruptive new HTTP-QuSS security and latency elimination technology of the NVIDIA Mellanox® ConnectX-6 Dx SmartNIC. for Wi-Fi, 3/4/5G Towers.



**NVIDIA EGX Server Installation in Tower Base Transceiver Station**