TECHNICAL OVERVIEW

GPU-ACCELERATED SIGNAL PROCESSING

Delivering Real-Time, Low-Latency, and High-Performance Computing with the NVIDIA Aerial SDK

A FIREHOSE OF SENSOR DATA AND SPEED-OF-LIGHT COMPUTE

With high-throughput and low-latency demands, signal processing applications—from software-defined radio and communications systems to speech processing and beyond—have traditionally relied on special accelerators like FPGAs and ASICs to deliver real-time performance. Programming these devices, however, has remained a huge challenge, and the development-to-deployment cycle is prone to restarts. Additionally, as the signal processing community extends into applications of artificial intelligence and machine learning for intelligent networks, anomaly detection, and spectrum awareness, the seamless connection to software frameworks like PyTorch, **NVIDIA RAPIDS™**, and TensorFlow is critical.

5G, in particular, ushers in a new era in wireless communications that delivers more than 10X lower latency and 1,000X the bandwidth when compared to previous 3rd Generation Partnership Project (3GPP) standards: all while supporting millions of connected devices per square kilometer. As developers and decision-makers look toward the future, NVIDIA GPUs lead the way with a focus on fast input/output (I/O) handling, programmability, compute performance, and enablement of both AI training and inference for signal processing workloads. Whether at the edge, in a data center, or in the cloud, the **NVIDIA Aerial™ SDK** delivers a collection of state-of-the-art signal processing solutions for Python, CUDA®, and C++ developers alike.

NVIDIA AERIAL: BIGGER, STRONGER, MORE PRODUCTIVE

In 2019, NVIDIA announced Aerial—a software package targeted to the 5G signal processing surrounding virtual radio access networks (vRAN). Aerial has since grown beyond 5G and now encompasses the full spectrum of GPU-accelerated tools used as building blocks and solutions to signal processing applications.

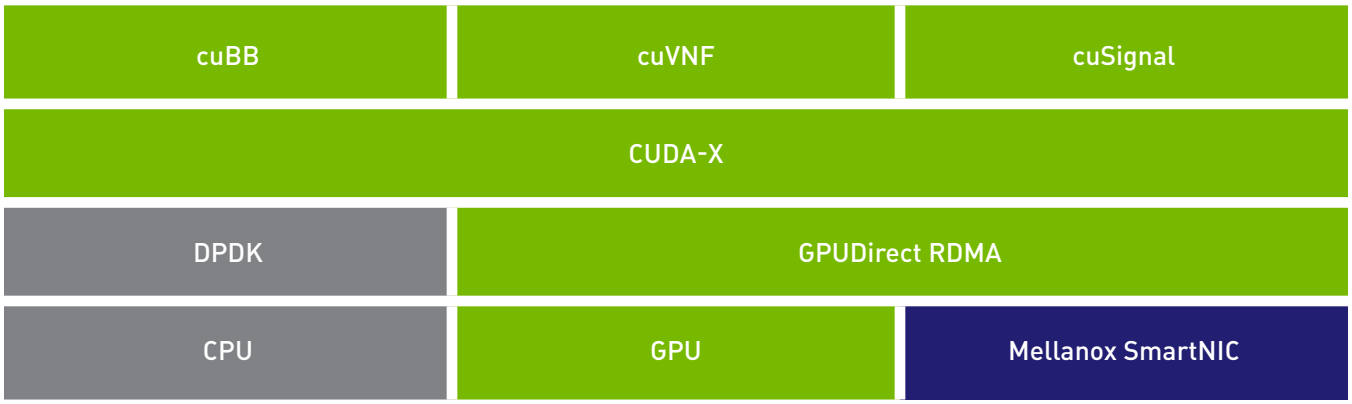

Figure 1. Aerial signal processing implementation stack

CORE AERIAL SDK LIBRARIES

CUDA Baseband (cuBB)
cuBB provides a fully accelerated 5G signal processing pipeline, including cuPHY, that delivers unprecedented throughput and efficiency by keeping all the physical-layer processing within the GPU's high-bandwidth memory. cuBB's functions include multi-cell support, channel estimation and equalization, beamforming, modulation and demodulation up to 256-quadrature amplitude modulation (256-QAM), and an optimized low-density parity check (LDPC) decoder to improve performance and reduce pipeline latency. Aerial is an O-RAN 7.2 fronthaul compliant 5G vRAN deployment SDK.

CUDA Virtual Network Functions (cuVNF)
cuVNF provides optimized I/O and packet placement with the GPU-accelerated Data Plan Development Kit (GPU-DPDK), sending 5G and other high-rate radio-over-network packets directly to GPU memory with a GPUDirect® remote direct memory access (RDMA)-capable network interface card like the Mellanox ConnectX-6 Dx. Additionally, GPU-DPDK provides capability for header-data splits, sending required metadata to the CPU while retaining the data in the GPU for processing. cuVNF is essential for satisfying the low-latency and high-throughput demands of 5G and other high-rate radio-over-network packet processing and can be extended to other applications and signal types like VITA 49.

cuSignal
cuSignal is a software library that brings GPU-accelerated core signal processing functions (convolution, spectrum estimation, filtering) to the Python programmer. By leveraging and extending the popular SciPy Signal application programming interface (API), cuSignal enables end-to-end GPU acceleration of a signal processing workflow—from online signal processing or preprocessing to machine learning training and inferencing, delivering 100X speedup over CPU with minimal code changes.

CUDA-X Signal Processing Libraries
Developers working at the C++ and CUDA layer can build their own signal processing routines by leveraging CUDA-X™ libraries like **cuFFT** for fast Fourier transforms, **cuBLAS** for general linear algebra, **CUTLASS** for optimized linear algebra, and **cuSOLVER** for solving linear systems. Building solutions at this level grants the developer the most flexibility and potentially optimum performance.

RECOMMENDED HARDWARE

For cuSignal

- NVIDIA Maxwell™ or newer GPU is required with support ranging from the Jetson Nano™ embedded GPU to the NVIDIA A100 Tensor Core GPU.

For cuBB/cuVNF

- NVIDIA V100 Tensor Core or newer GPU is required and must be paired with a Mellanox ConnectX-6 Dx SmartNIC, which includes many 5T-for-5G features, including O-RAN fronthaul conformance. For optimal performance, a PLX PCIe switch should be used to host the Mellanox SmartNIC and the GPU rather than directly connecting to the CPU's PCIe root complex.

ADVANTAGES OF GPU-ACCELERATED SIGNAL PROCESSING

Ease of Development and Deployment on a Variety of GPUs
Prototype and deploy high-performance applications in Python, C++, or native CUDA using the Aerial SDK on cloud to edge GPU platforms

Smart NIC Security
Boost security with the SmartNIC's 5T-for-5G features, including open radio access network (O-RAN) fronthaul conformance, Transport Layer Security (TLS) inline crypto acceleration, and hardware root trust.

Taming the I/O Beast
Enable direct memory access to NVIDIA GPUs using cuVNF and an NVIDIA® Mellanox® ConnectX-6 Dx SmartNIC with up to 100 gigabits per second (Gbps) throughput.

Connection to AI/ML Platforms
Gain insight into your signal's data and train networks to identify modulation types, increase signal-to-noise ratio, and detect anomalous behavior with cuSignal.

INDUSTRY USE CASES

Federal

- Radio frequency spectrum awareness and collaboration
- Anomaly detection in wireless communication systems
- Increased compute power for radar and sonar phased-array applications
- Time-series analysis of sensor data and cyber-threat hunting
- Low-latency packet processing

Telecommunications

- Efficient bandwidth utilization, beamforming, and channel estimation
- High-throughput packet processing
- Low-latency data pipelines
- Private and enterprise 5G deployments
- Edge data center Cloud RAN (C-RAN)

WHAT USERS ARE SAYING

"Accelerated computing technologies brought to the table by NVIDIA GPUs with CUDA, Aerial SDK, and the cuSignal Python library for DSP developers are transforming the digital signal processing community. With NVIDIA's Aerial software libraries, we are now able to instantaneously survey GHz of the radio frequency (RF) spectrum to detect and localize malicious RF emissions."

– **Bill Urrego,** Systems Engineer, the MITRE Corporation, the developer of Photon, a GPU-accelerated digital signal processing (DSP) platform

"cuSignal has been a tremendous enabler for Expedition Technology's portfolio of wireless machine learning programs, dramatically accelerating our capability development timelines and system performance."

– **Greg Harrison,** CTO, Expedition Technology

"The GPU is becoming one of the most efficient processors for signal processing, and cuSignal is paving the way."

– **John Ferguson,** CEO, Deepwave Digital

Learn More

NVIDIA Aerial SDK (including cuBB and cuVNF): **developer.nvidia.com/aerial-sdk**
cuSignal: **github.com/rapidsai/cusignal**
CUDA-X: **developer.nvidia.com/gpu-accelerated-libraries**

NVIDIA