

REPORT ON NVIDIA CUDA-X

INTRODUCTION:

NVIDIA CUDA-X is a collection of GPU-accelerated libraries, tools, and technologies built on top of NVIDIA CUDA, that delivers higher performance when compared to the CPU alternatives across multiple application domains, from Artificial Intelligence(AI) to High Performance Computing(HPC). NVIDIA libraries provide an easiest way to get started with GPU acceleration, whether for building a new application or accelerating an existing application.

MATH LIBRARIES:

GPU-accelerated math libraries lay the foundation for compute-intensive applications in areas such as molecular dynamics, computational fluid dynamics, computational chemistry, medical imaging, and seismic exploration. Some of the math libraries are **cuBLAS**(BLAS library), **cuFFT**(library for Fast Fourier Transforms), **CUDA Math Library** (standard mathematical functions), **cuRAND**(random number generation), **cuSOLVER**(dense and sparse direct solvers), **cuSPARSE**(BLAS for sparse matrices), **cuTENSOR** (tensor library algebra library), **AmgX**(linear solvers for simulations and implicit unstructured methods).

PARALLEL ALGORITHM LIBRARIES:

GPU-accelerated libraries of highly efficient parallel algorithms for several operations in C++ and for use with graphs when studying relationships in natural sciences, logistics, travel planning, and more. **Thrust** is a powerful library of parallel algorithms and data structures. Thrust provides a flexible, high-level interface for GPU programming that greatly enhances developer productivity.

IMAGE AND VIDEO LIBRARIES:

GPU-accelerated libraries for image and video decoding, encoding, and processing that leverage CUDA and specialized hardware components of GPUs. Some of the libraries are: **nvJPEG** (High performance library for JPEG decoding), **NVIDIA Performance Primitives** (image, video and signal processing functions), **NVIDIA Video Codec SDK** (hardware-accelerated video encode and decode on Windows and Linux), **NVIDIA Optical Flow SDK** (dedicated to computing the relative motion of pixels between images).

COMMUNICATION LIBRARIES:

Performance-optimized multi-GPU and multi-node communication primitives. Some of the NVIDIA provided libraries are: **NVSHMEM** (OpenSHMEM standard for GPU memory), **NCCL** (Open-source library for fast multi-GPU, multi-node communications that maximizes bandwidth while maintaining low latency).

DEEP LEARNING LIBRARIES:

GPU-accelerated libraries for Deep Learning applications that leverage CUDA and specialized hardware components of GPUs. Some of those libraries include: **NVIDIA cuDNN** (library of primitives for deep neural networks), **NVIDIA TensorRT**(high-performance deep learning inference optimizer and runtime for production deployment), **NVIDIA Jarvis** (platform for developing engaging and contextual AI-powered conversation apps), **NVIDIA DeepStream SDK** (real-time streaming analytics toolkit for AI-based video understanding and multi-sensor processing), **NVIDIA DALI**(portable, open-source library for decoding and augmenting images and videos to accelerate deep learning applications).

PARTNER LIBRARIES:

NVIDIA has some libraries that are not completely developed by NVIDIA but have contributed to those libraries. These include OpenCV, FFmpeg, ArrayFire, MAGMA, IMSL Fortran Numerical Library, Gunrock, CHOLMOD, Triton Ocean SDK, CUVilib.