TECHNICAL RESOURCES

Connecting our research communities to information,
tools, and resources across HPC, AI and Big Data.

Intro to GPU Programming
- Video: Introduction to NVIDIA GPU Computing
- Presentation: Present and Future of Accelerated Computing Programming Approaches
- Video: ACM Winter School 2019 on HPC IIT Kanpur

Libraries
- Documentation: Math and communication libraries.
- Video: How CUDA Math Libraries Can Help You Unleash the Power of the New NVIDIA A100 GPU.
- Video: Recent Developments in NVIDIA Math Library.
- Presentation: High-Performance Sparse Linear Algebra on NVIDIA GPUs with cuSPARSE.

ARM
- Presentation: Port, Profile, and Tune HPC Applications for Arm-based Supercomputers.
- Presentation: Port, Profile, and Tune HPC Applications for Arm-based Supercomputers.
- Web Page: NVIDIA Arm HPC Developer Kit.

Programming Models

OpenACC
- Web Page: OpenACC.org Resources
- Forum: OpenACC Slack Channel
- Presentation: Zero to GPU Hero with OpenACC
- Presentation: Directive-Based Programming with OpenACC
- Training Series: OpenACC Training Series
- Videos: Directive Based GPU Programming
- GitHub: OpenACC Training Materials
- Docker Container: OpenACC Training Materials
- Online Course: Fundamentals of Accelerated Computing with OpenACC (Fee-based)
- Video: OpenACC Multi-GPU

Standard Languages
- Presentation: Standard Language Parallelism
- Presentation: Developing HPC Applications with Standard C++, Fortran, and Python
- Presentation: The NVIDIA C++ Standard Library
- Presentation: Shifting through the Gears of GPU Programming: Understanding Performance and Portability Trade-offs

CUDA
- Training Series: CUDA Training Series
- Presentation: CUDA on NVIDIA Ampere GPU Architecture: Taking Your Algorithms to the Next Level of Performance
- Presentation: CUDA: New Features And Beyond
- Presentation: Developing CUDA Kernels to Push Tensor Cores to the Absolute Limit on NVIDIA A100
- Online Course: Fundamentals of Accelerated Computing with CUDA C/C++ (Fee-based)
- Presentation: How CUDA Programming Works (GTC Fall '22), How CUDA Programming Works (GTC Spring '22)
- Presentation: CUDA Programming Model for Hopper Architecture
- Presentation: Optimizing CUDA Applications for NVIDIA Hopper Architecture
- Presentation: Enabling Hopper-Specific Optimizations in CUDA Applications
- Presentation: Performance Optimization with Modern CUDA Programming Techniques

CUDA Fortran
- Documentation: Programming Guide

Python
- Presentation: CuPy Overview: NumPy Syntax Computation with Advanced CUDA Features
- Presentation: Accelerating Python with CUDA
- Online Course: Fundamentals of Accelerated Computing with CUDA Python (Fee-based)
- Tutorial: Python Profiling
- Presentation: GPU Acceleration in Python
- Presentation: Legate: Scaling the Python Ecosystem
- Presentation: Scale Python and NumPy Performance with Legate
- Presentation: GPU Acceleration in Python using CuPy and Numba
- Presentation: Accelerate Computing with CUDA Python

Kokkos
- GitHub: Kokkos Repository
- GitHub: Kokkos Tutorials
- Forum: Kokkos Slack Channel
- Video: Kokkos: C++ Performance Portability for Production

RAJA
- GitHub: RAJA Repository
- Video: A Tutorial Introduction to RAJA

Multi-GPU
- Presentation: Multi-GPU Programming with MPI (a Magnum IO Session)
- Presentation: Multi-GPU Programming Models
- Presentation: A Partitioned Global Address Space Library for Large GPU Clusters
- Presentation: NVSHMEM: CUDA-Integrated Communication for NVIDIA GPUs (a Magnum IO session)
- Presentation: NVSHMEM: GPU-Integrated Communication for NVIDIA GPU Clusters
- Presentation: NCCL: High-Speed Inter-GPU Communication for Large-Scale Training
- Github: Multi-GPU Programming Models

Tools
- Download: NVIDIA Nsight™ Systems, NVIDIA Nsight Compute, NVIDIA Nsight Graphics, NVIDIA Nsight Visual Studio Code Edition, NVIDIA Deep Learning Designer
- Documentation: Nsight Systems, Nsight Compute, Nsight Graphics, CUPTI, NVIDIA Tools Extension SDK (Nvtx), Compute Sanitizer, CUDA-memcheck, CUDA-GDB, Nsight Visual Studio Code Edition, Nsight Deep Learning Designer
- Video: NVIDIA Nsight Systems "Profiling GPU Applications with NVIDIA NSight Systems"
- Video: NVIDIA Nsight Compute "Introduction to NVIDIA Nsight Compute"
- Technical Session: CUDA is Evolving, and the Latest Developer Tools are Adapting to Keep Up
- Technical Session: It's Alive: CUDA in Visual Studio Code
- Technical Session: Optimizing CUDA Kernels in HPC Simulation and Visualization Codes Using NVIDIA NSight Compute
- Technical Session: What the Profiler is Telling You: How to Get the Most Performance out of Your Hardware
- Technical Session: Performance Tuning CUDA Applications with the Roofline Model
- Technical Session: Roofline and NVIDIA Ampere GPU Architecture Analysis
- GitHub: Nsight Developer Tools Training
- Technical Blog: Analysis-Driven Optimization: Analyzing and Improving Performance with NVIDIA Nsight Compute
- Technical Blog: Custom Application Profile Timelines with NVTX
- Demo: NVIDIA Nsight-Developer Tools
- Blog: Debugging CUDA More Efficiently with NVIDIA Compute Sanitizer
- Technical Session: How to Understand and Optimize Shared Memory Accesses using Nsight Compute
- Technical Session: Requests, Wavefronts, Sectors Metrics: Understanding and Optimizing Memory-Bound Kernels with Nsight Compute

Data Science
- Web Page: RAPIDS overview
- Web Page: RAPIDS-Getting Started (Hands-on labs and other materials)
- Forum: RAPIDS Community
- Presentations: Collection of RAPIDS Talks
- Online Course: Fundamentals of Accelerated Data Science with RAPIDS(Fee-Based)

AI/Deep Learning Libraries, Frameworks, SDKs

Introduction to AI
- Presentation: Dive into Deep Learning
- Presentation: Do-it-Yourself Automatic Speech Recognition with NVIDIA Technologies
- Online Course: Fundamentals of Deep Learning for Computer Vision (Fee-Based)
- GitHub: Deep Learning Examples(The latest deep learning example networks for training and Inference.)

cuNumeric
- Presentation: cuNumeric and Legate: How to Create a Distributed GPU Accelerated Library
- Presentation: Evaluating Your Options for Accelerated Numerical Computing in Pure Python

CuDNN
- Documentation: CuDNN Developer Guide
- Presentation: cuDNN v8 New Advances in Deep Learning Acceleration: APIs, Optimizations, and How to Tackle the Future Challenges in Hardware and Software
- Presentation: Deep Learning Training with cuDNN

NVIDIA CLARA
- Container: NVIDIA Clara Train SDK
- Jupyter Notebooks: Intro to Clara Train SDK
- Jupyter Notebook: Clara Federated Learning
- Presentation: Clara Developer Day: Federated Learning Using Clara Train SDK
- Presentation: Clara Developer Day: Clara Train SDK Performance Walkthrough and Deep Dive
- Presentation: Clara Developer Day: Scalable and Modular Deployment Powered by Clara Deploy SDK
- Presentation: Clara Developer Day: Getting Started with Clara Train for High Performance & Iterative Experimentation with AutoML
- Web Page: NVIDIA Clara™ Parabricks Overview

DeepStream
- Presentation: Developing IVA Software Using NVIDIA DeepStream SDK
- Online Course: Getting Started with DeepStream for Video Analytics on Jetson Nano(Fee-Based)
- Online Course: AI Workflows for Intelligent Video Analytics with DeepStream(Fee-Based)

JAX
- Presentation: Seamlessly Scale Out Complex Compute Workloads with Multi-Node JAX
- Presentation: What's New in JAX

TensorRT
- Presentation: PyTorch-TensorRT: Accelerating Inference in PyTorch with TensorRT
- Presentation: Inference with Tensorflow 2 Integrated with TensorRT
- Presentation: TensorRT inference with TensorFlow 2.0
- Online Course: Optimization and Deployment of TensorFlow Models with TensorRT(Fee-Based)

TensorFlow
- Presentation: What's New in TensorFlow-An Ecosystem Update
- Presentation: Extensions of TensorFlow-Based Computational Fluid Dynamics
- Online Course: Image Classification with TensorFlow: Radiomics - 1p19q Chromosome Status Classification(Fee-Based)
- Demo: Profiling TensorFlow using Tensorboard

Keras
- Presentation: Keras and TensorFlow: The Next Five Years
- Presentation: Porting Keras Lamba Layers to TensorFlow.js
- Online Course: Modeling Time Series Data with Recurrent Neural Networks in Keras(Fee-Based)

PyTorch
- Website: PyTorch
- Container: PyTorch
- Presentation: Latest Developments on PyTorch: Getting Started and Community Highlights
- Presentation: Dynamic Shapes First: Advanced GPU Fusion in PyTorch
- Presentation: Scaling Large Models Using PyTorch RPC
- Presentation: Named Tensors, Model Quantization, and the Latest PyTorch Features
- Demo: Profiling PyTorch with Nsight System

Containers
- Video: Making Containers Easier with HPC Container Maker.
- Presentation: Containers Democratize HPC.
- Online Course: High-Performance Computing with Containers(Fee-Based).
- Presentation: Simplifying and Accelerating HPC Workflows with NGC.

Advanced
- Presentation: Inside the NVIDIA Ampere Architecture.
- Presentation: Optimizing Applications for NVIDIA Ampere GPU Architecture.
- Blog: NVIDIA Hopper Architecture In-Depth.
- Presentation: Inside the NVIDIA Hopper Architecture.

Navigate
- Events
- Attendees
- Mentors
- Resources
- Log In

About Open Hackathons
- Blog
- News
- Join Our Community
- Contact Us
- Terms and Conditions
- About

Company
- About OpenACC Organization
- Code of Conduct
- Privacy Policy
- Contact Site Administrator

Stay in Touch