

SENAI CIMATEC

Tecnologia, Inovação
e Educação para a Indústria

Sistema FIEB

**SENAI
CIMATEC**

PELO FUTURO DA INOVAÇÃO

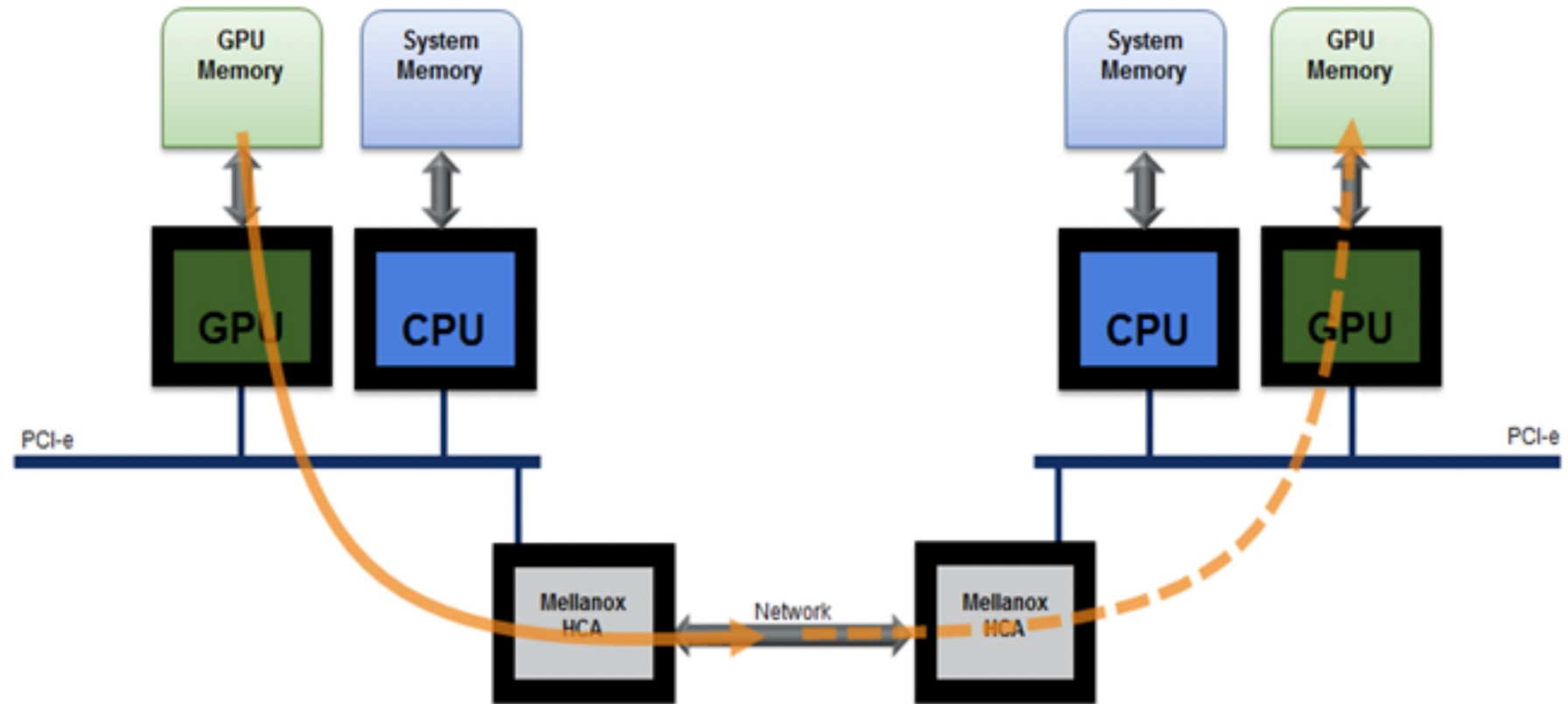


QMS Certification Services

GPU Programming Applied to Industrial Problems

Introduction to scaling multi-node applications

Murilo Boratto



Multi-GPU Environments

OPENACC

CUDA

OPENMP

NCCL

NVLINK

UCX

RDMA

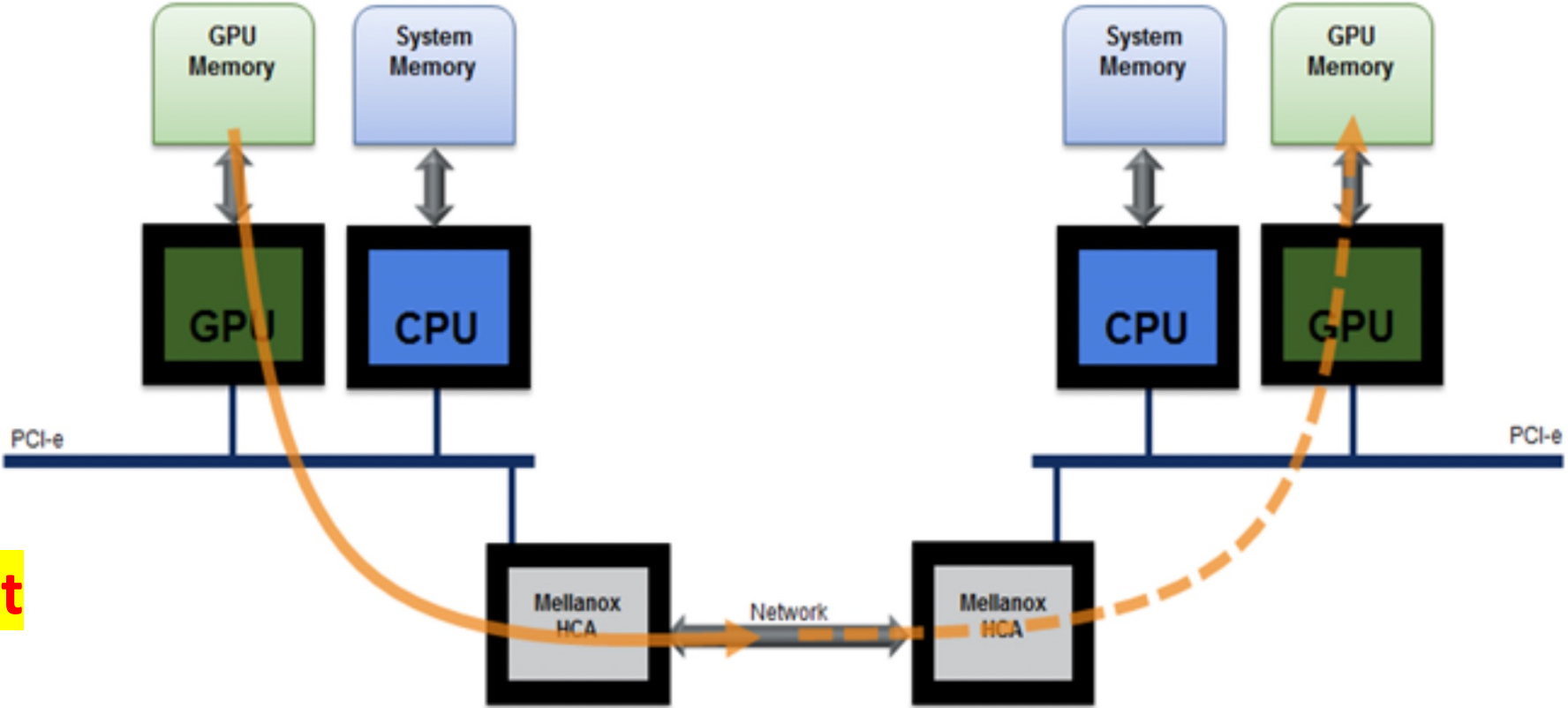
GPUDirect

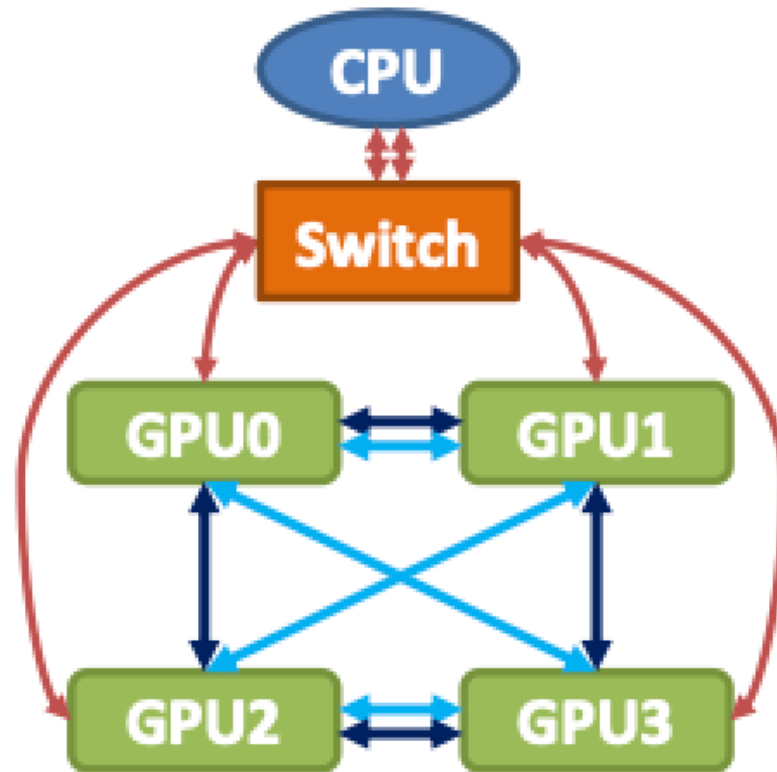
MPI+CUDA

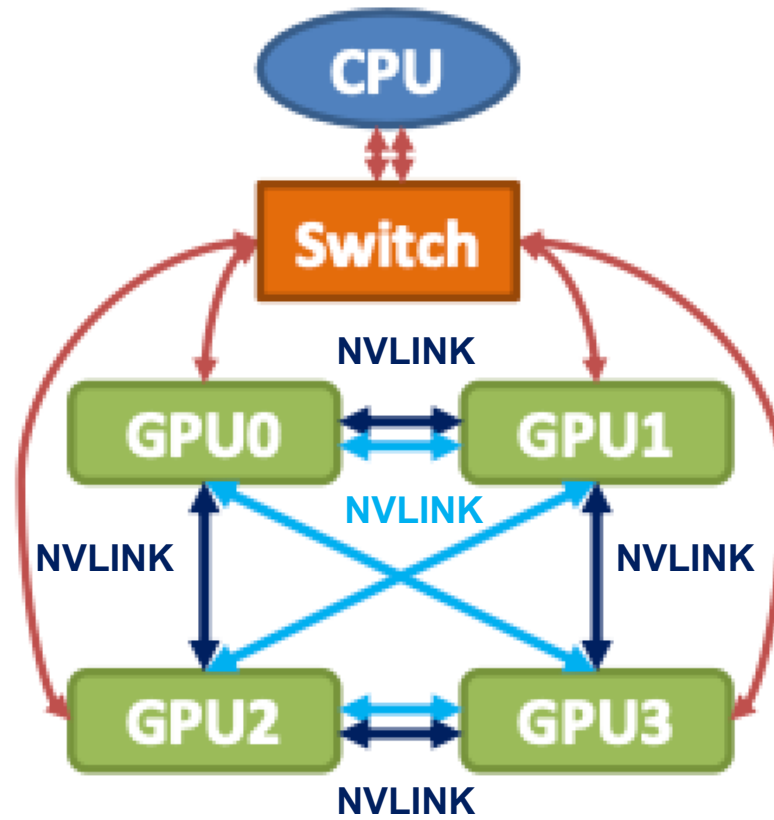
INFINIBAND (MELLANOX)

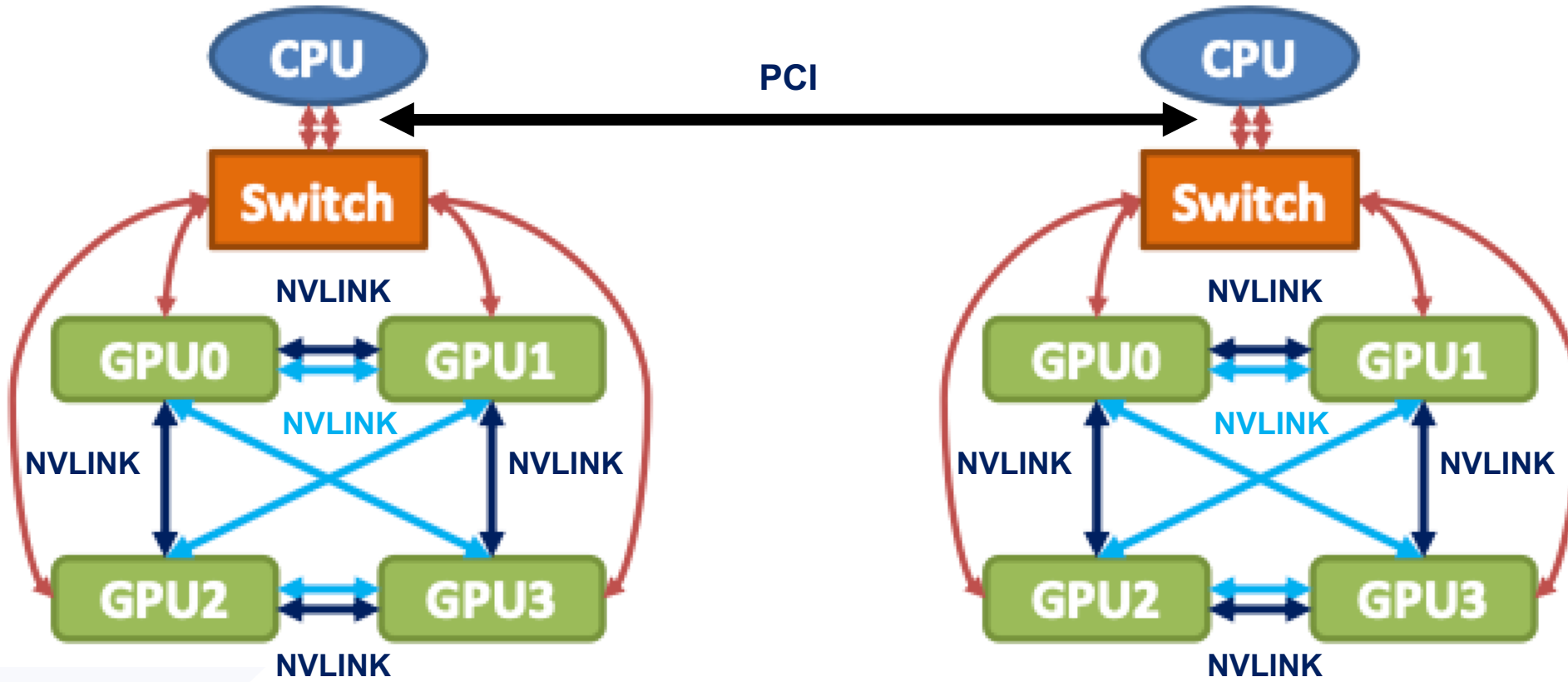
CUDAWARE

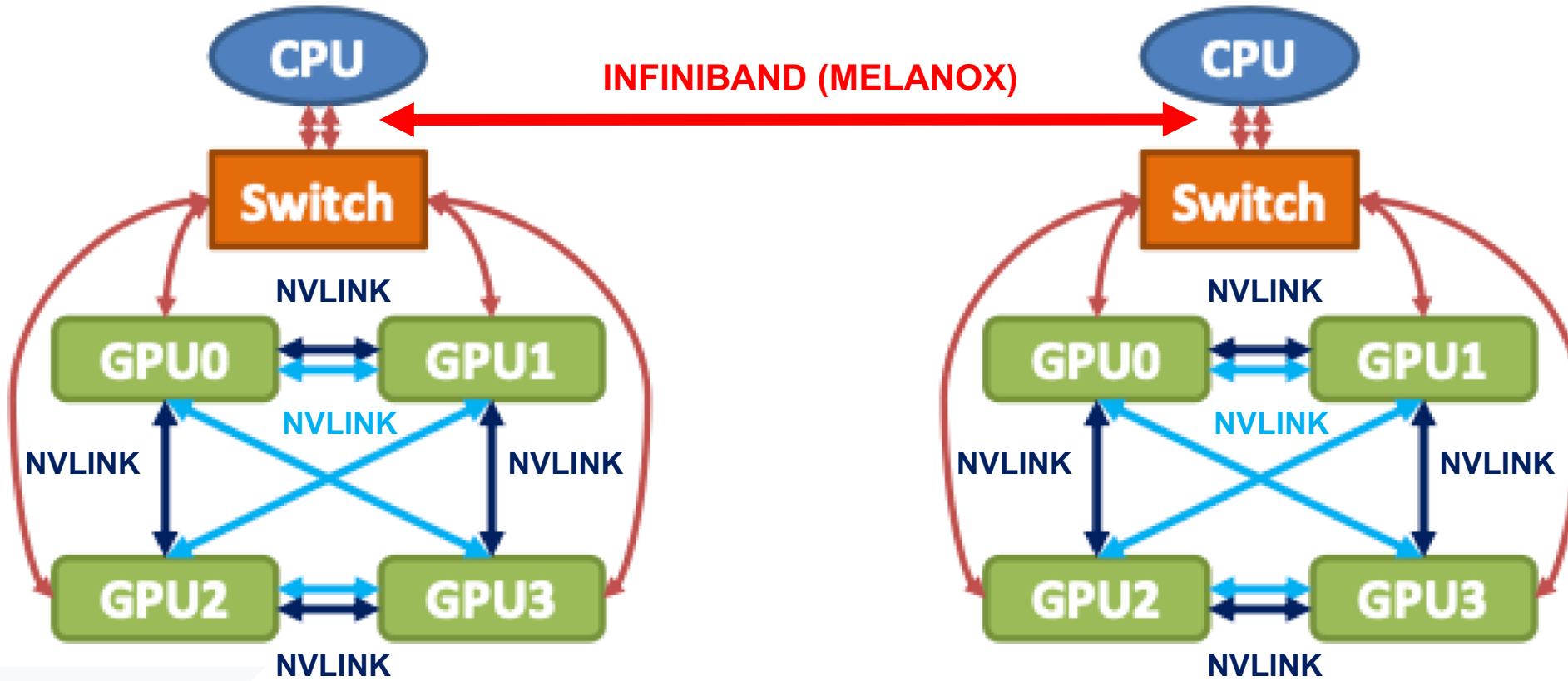
NVSHMEM



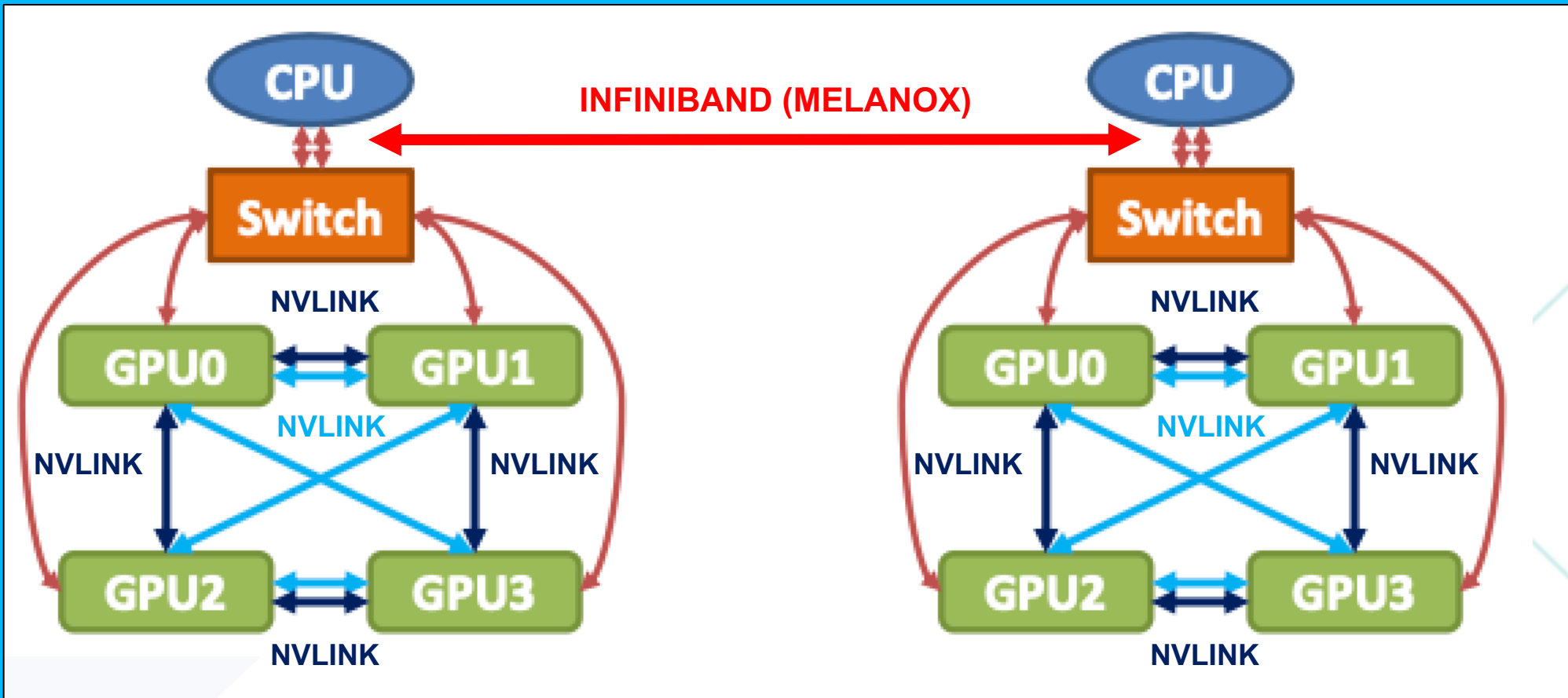








UCX (Unified Communication X Library)



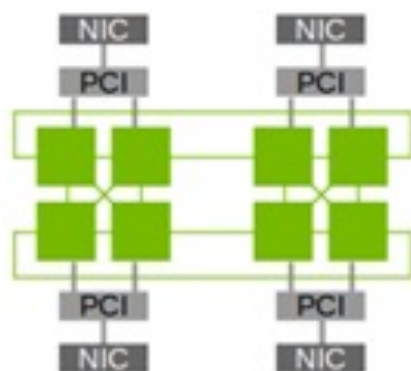
NCCL

**NVIDIA Collective
Communication Library**

Topology detection

Build graph with all GPUs, NICs, CPUs, PCI switches, NVLink, NVSwitch.

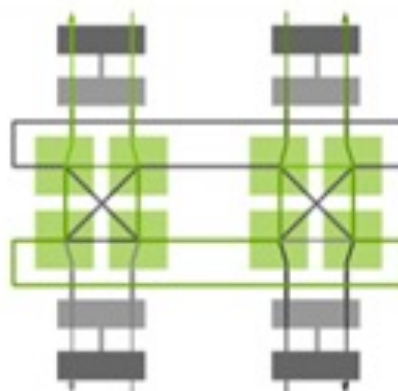
Topology injection for VMs.



Graph search

Extensive search to find optimal set of rings or trees.

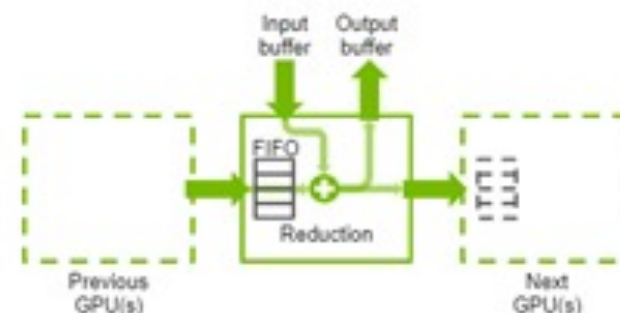
Performance prediction of each algorithm and auto-tuning.



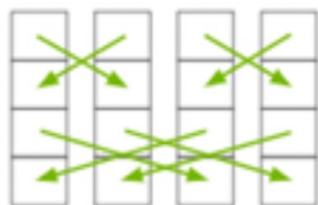
CUDA kernels

Optimized reductions and copies for a minimal SM usage.

CPU threads for network communication.



sendrecv



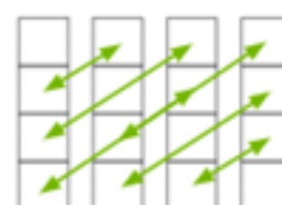
gather



scatter



alltoall

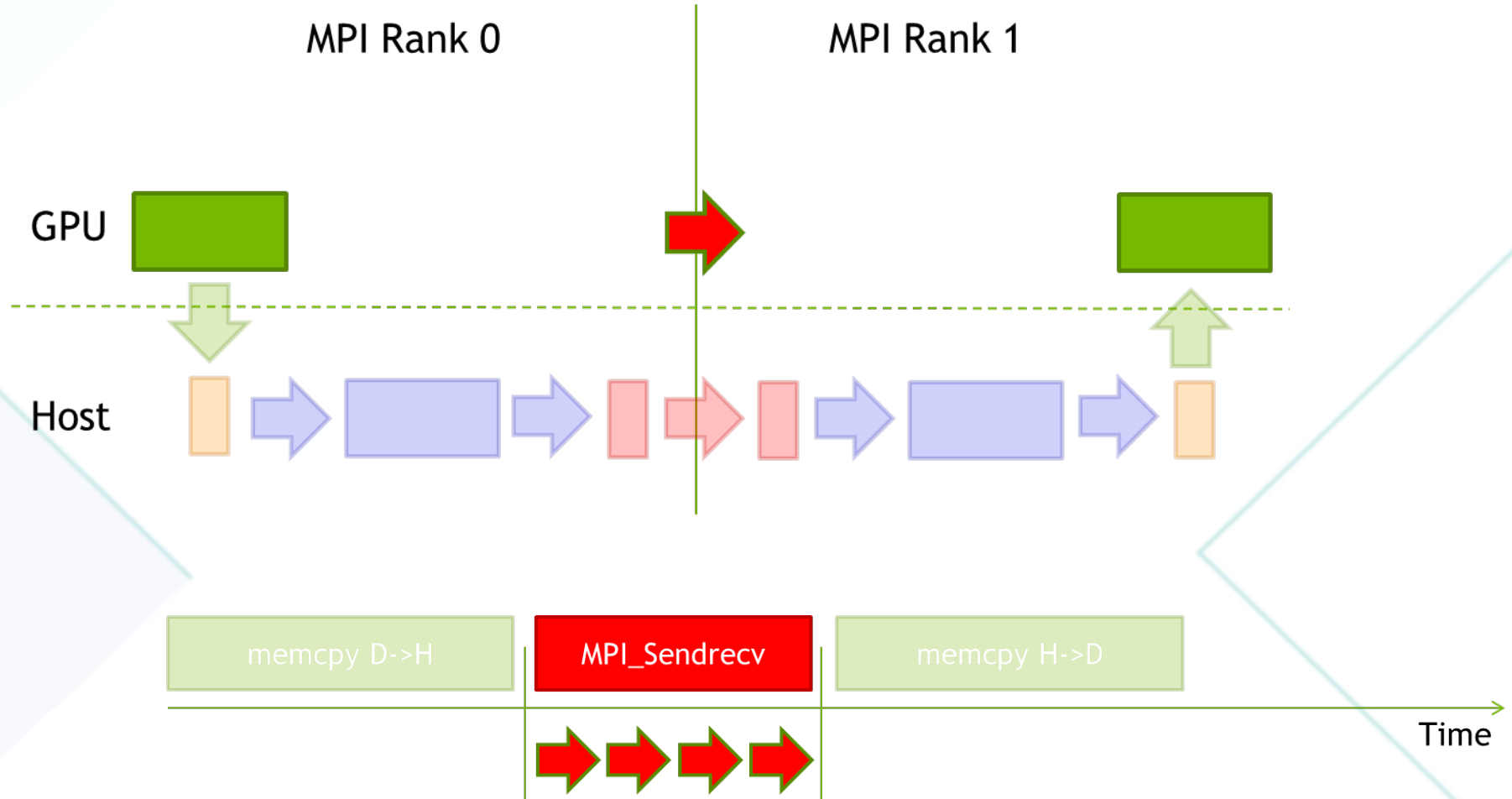


neighbor



CUDAWARE MPI

CUDA + MPI at GPUs



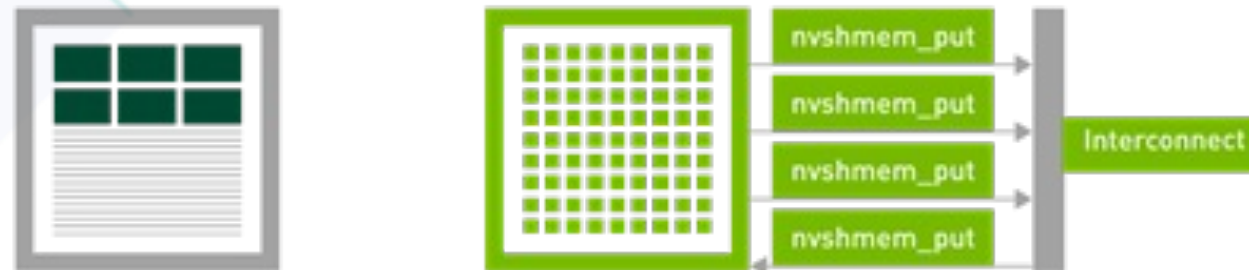
NVSHMEM

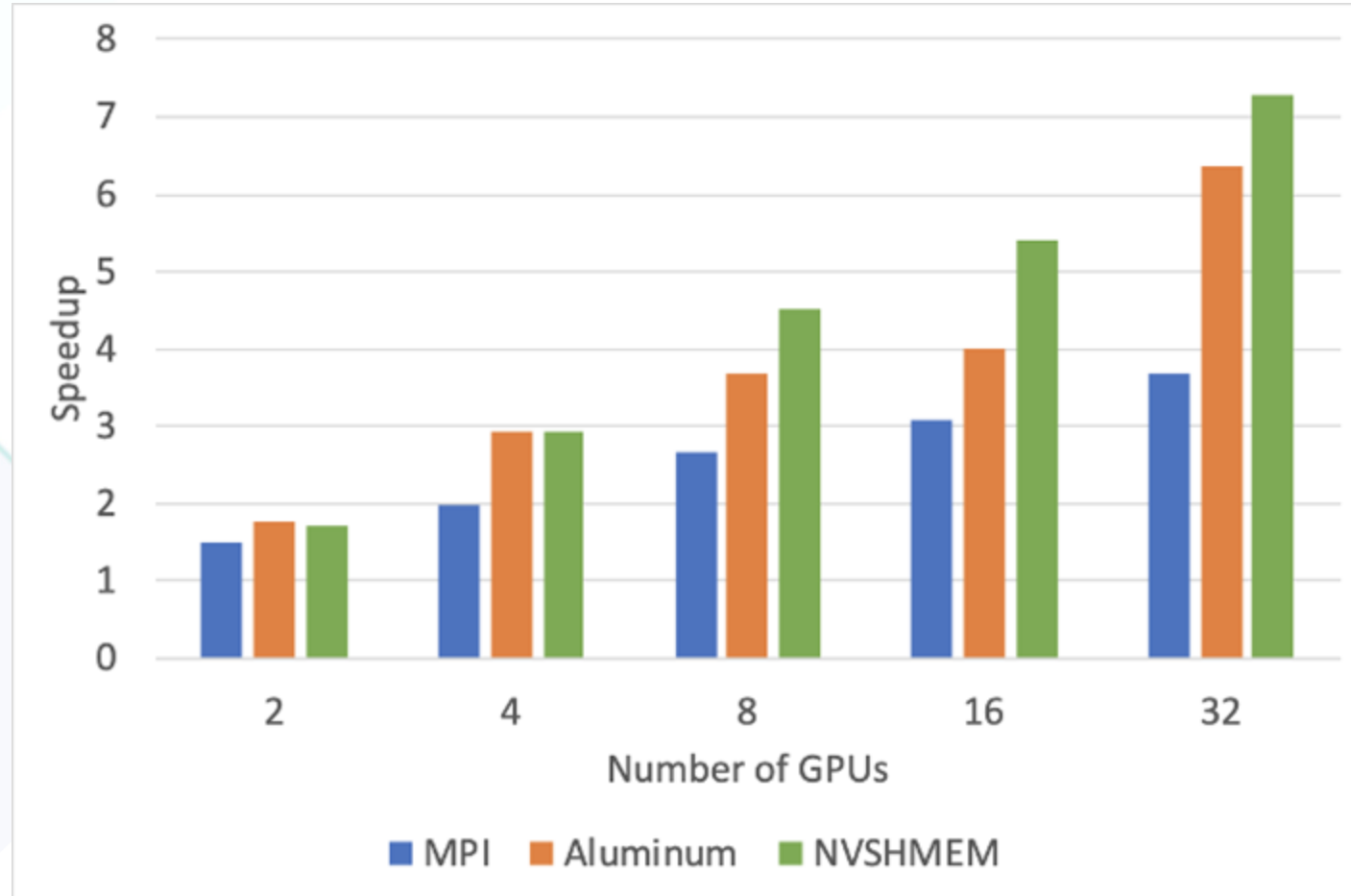
NVIDIA OpenSHMEM

MPI



NVSHMEM





<https://github.com/muriloboratto/parallel-computing-applied>

[https://github.com/muriloboratto/](https://github.com/muriloboratto/GPU-programming)GPU-programming

SENAI CIMATEC

Tecnologia, Inovação
e Educação para a Indústria

Murilo Boratto

murilo.boratto@fieb.org.br

Sistema FIEB

**SENAI
CIMATEC**

PELO FUTURO DA INOVAÇÃO



QMS Certification Services