

IWOCL 2024
2024-04-11

WELCOME FROM ARGONNE NATIONAL LABORATORY



BRICE VIDEAU
Computer Scientist



- Aerial view of Lemont campus
- Advanced Photon Source (bottom right)
- Chicago skyline (in the distance)

ARGONNE NATIONAL LABORATORY

Tackling Societal Challenges

- About 4000 employees, 2000 Researchers
- 6 user facilities serving about 8000 researchers around the world
- Tackling societal challenges
 - Environmental research
 - Energy research
 - Biology and medicine
 - Fundamental physics
 - AI for Science
 - Microelectronics

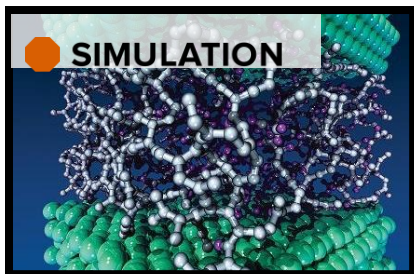


ARGONNE LEADERSHIP COMPUTING FACILITY



The Argonne Leadership Computing Facility provides world-class computing resources to the scientific community.

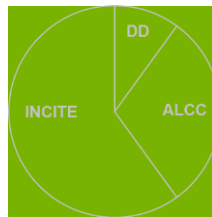
- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science



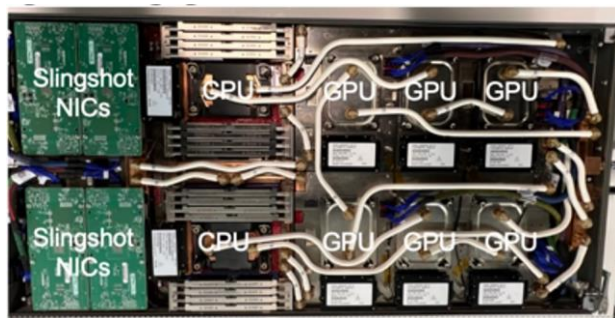
Architecture supports three types of computing

- Large-scale Simulation (PDEs, traditional HPC)
- Data Intensive Applications (scalable science pipelines)
- Deep Learning and Emerging Science AI (training and inferencing)

ALCF offers different pipelines for different applications



AURORA



Intel GPU
**Intel® Data Center GPU
Max Series**

Intel Xeon Processor
**4th Gen Intel XEON Max
Series CPU** with High
Bandwidth Memory

Platform
HPE Cray-Ex

Peak FP Performance
 ≥ 2 Exaflops DP

Memory
10.9PB of DDR @ 5.95 PB/s
1.36PB of CPU HBM @ 30.5 PB/s
8.16PB of GPU HBM @ 208.9 PB/s

Network
2.12 PB/s Peak Injection BW
0.69 PB/s Peak Bisection BW

Storage
230PB DAOS Capacity
31 TB/s DAOS Bandwidth

Racks - 166

Nodes - 10,624
CPUs - 21,248
GPUs - 63,744

Interconnect

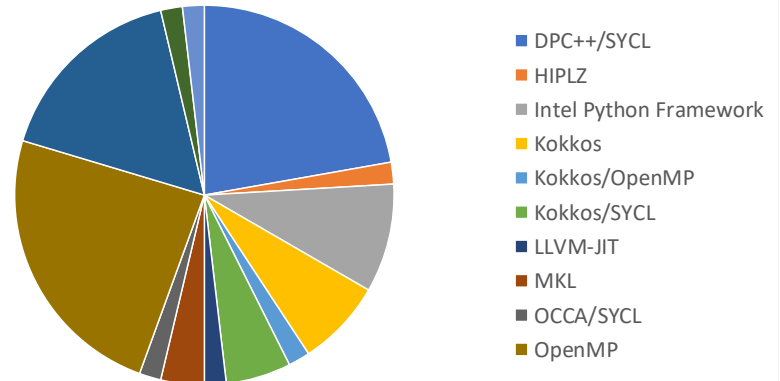
HPE Slingshot 11
Dragonfly topology with adaptive routing
Network Switch:
25.6 Tb/s per switch (64 200 Gb/s ports)
Links with 25 GB/s per direction

AVAILABLE AURORA PROGRAMMING MODELS

- **Aurora applications may use:**
 - DPC++/SYCL
 - OpenMP
 - Kokkos
 - Raja
 - OpenCL
- **Experimental**
 - HIP
- **Not available on Aurora:**
 - CUDA
 - OpenACC



Early Science Application Programming Model Distribution



RELIANCE ON OPEN STANDARDS

And Khronos' in Particular

- Most of applications running on Aurora are using Open Standards
- Around half leverage SYCL either directly or through a higher level programming model
 - Kokkos, OCCA, RAJA
- 100% of applications that use the GPU do so through SPIR-V
- OpenCL available directly or as an alternative backend for SYCL, OpenMP, or HIP
- ALCF is a Khronos member and participates in the SYCL and OpenCL working groups (and C++, OpenMP, MPI, ...)



HIP ON AURORA

Building HIP on SPIR-V and OpenCL (or Intel Level Zero)

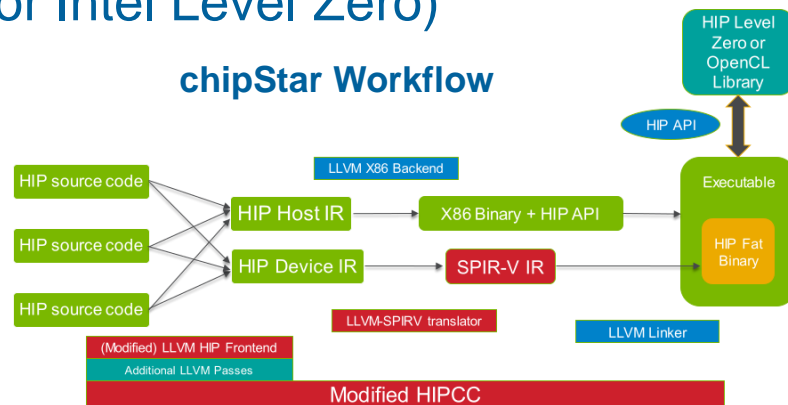
Overview

- Intel GPUs lack native support for the CUDA or HIP programming models
- A large fraction of GPU enabled HPC applications are written using CUDA or HIP
- Working to bring AMD's HIP and Nvidia CUDA to Aurora

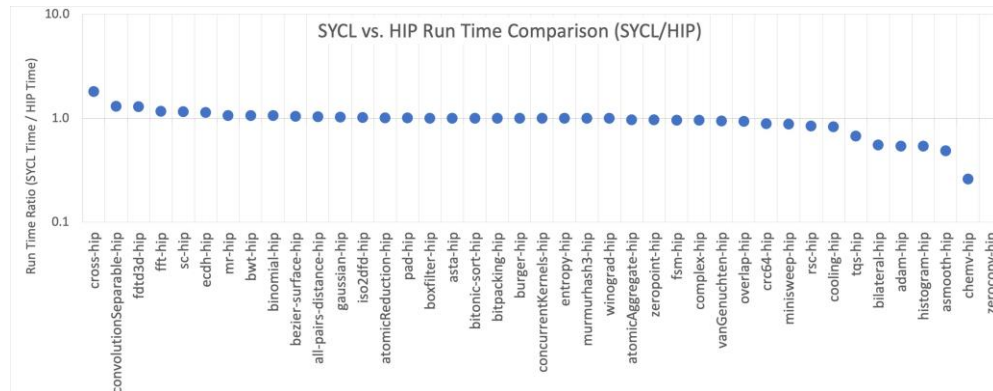
Progress

- Release 1.1 chipStar
 - Provides HIP support on LLVM 17, 16, and 15
 - <https://github.com/CHIP-SPV/chipStar>
- Supported applications: CP2K, GAMESS, libCEED
- Supported HIP libraries: HipBLAS, hipSOLVER, hipRTC
- Upcoming support for ExaBiome, hipFFT, and more...
- Experimental CUDA support
- Currently using extensions we are working on upstreaming into Khronos standards

chipStar Workflow



HecBench: HIP vs SYCL version of benchmarks



FINAL THOUGHTS

Importance of Open Standards ... and their Adoption

- Khronos standards, and the LLVM ecosystem, are
 - The basis of Aurora's software toolchain
 - Performant compilers and runtimes
 - Benefit broader community
 - Enabling development of layered implementations
- Wishlist
 - Make SYCL, OpenCL, and Vulkan "compatible"
 - Unify Vulkan and OpenCL compute and memory models, SPIR-V
 - Layered SYCL 2020 and OpenCL over Vulkan
 - Broader adoption of SYCL and OpenCL

BROADER OPENCL SUPPORT FOR LINUX DESKTOPS

The MESA Strategy

Keynote

Rusticl: Compute for the Linux Desktop?

Karol Herbst

Linux GPU Driver Software Engineer at Red Hat



Argonne
NATIONAL LABORATORY



**Argonne Leadership
Computing Facility**