Argonne
NATIONAL LABORATORY

Future ALCF Systems

Argonne Leadership Computing Facility

2021 Computational Performance Workshop, May 4-6

Scott Parker

www.anl.gov

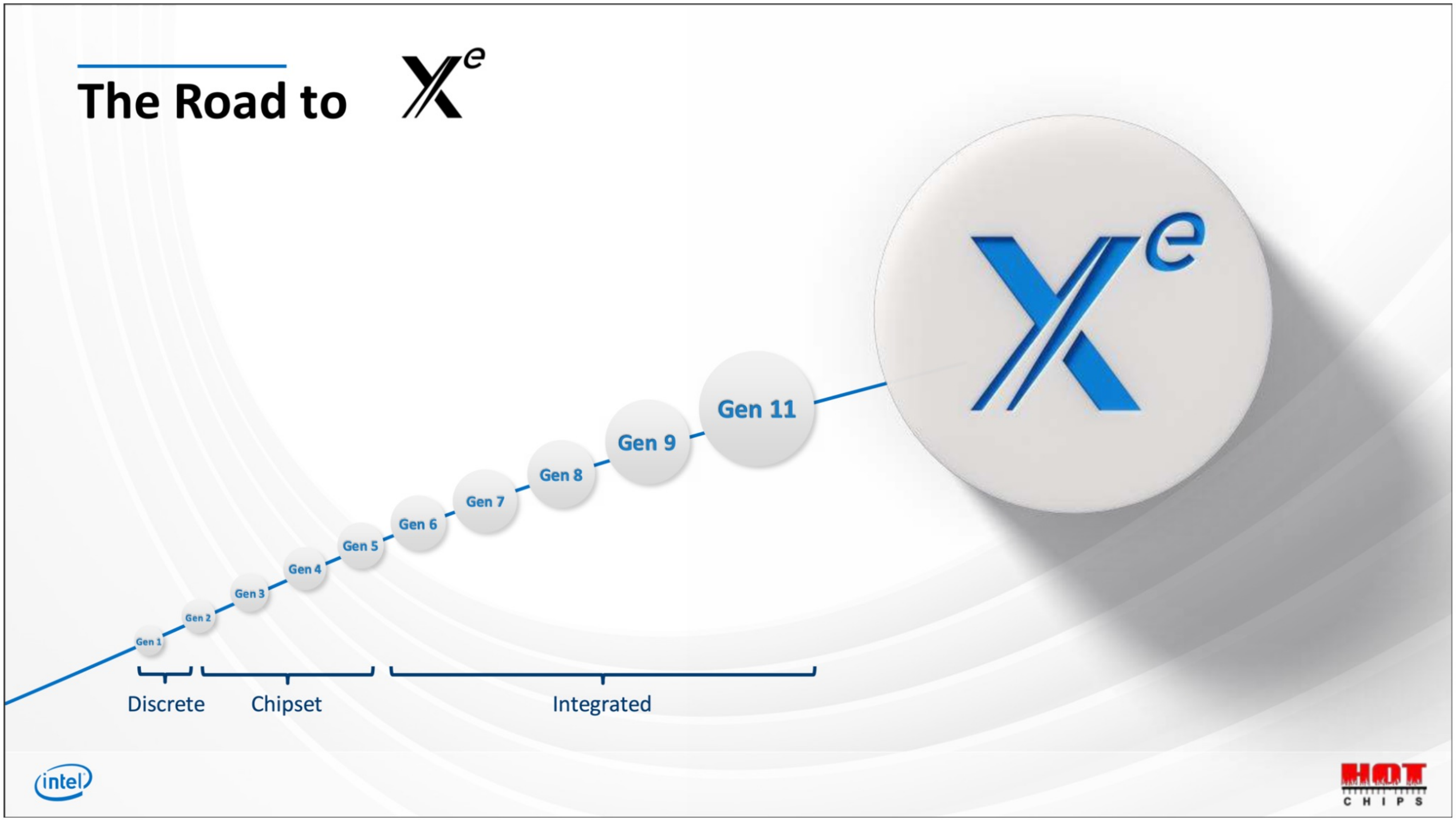
Aurora

Aurora: A High-level View

- Intel-HPE machine arriving at Argonne in 2022
  - Sustained Performance ≥ 1Exaflops DP
- Compute Nodes
  - 2 Intel Xeons (Sapphire Rapids)
  - 6 Intel X^e GPUs (Ponte Vecchio [PVC])
  - Node Performance > 130 TFlops
- System
  - HPE Cray XE Platform
  - Greater than 10 PB of total memory
  - HPE Slingshot network
- Fliesystem
  - Distributed Asynchronous Object Store (DAOS)
    - ≥ 230 PB of storage capacity; ≥ 25 TB/s
  - Lustre
    - 150PB of storage capacity; ~1 TB/s

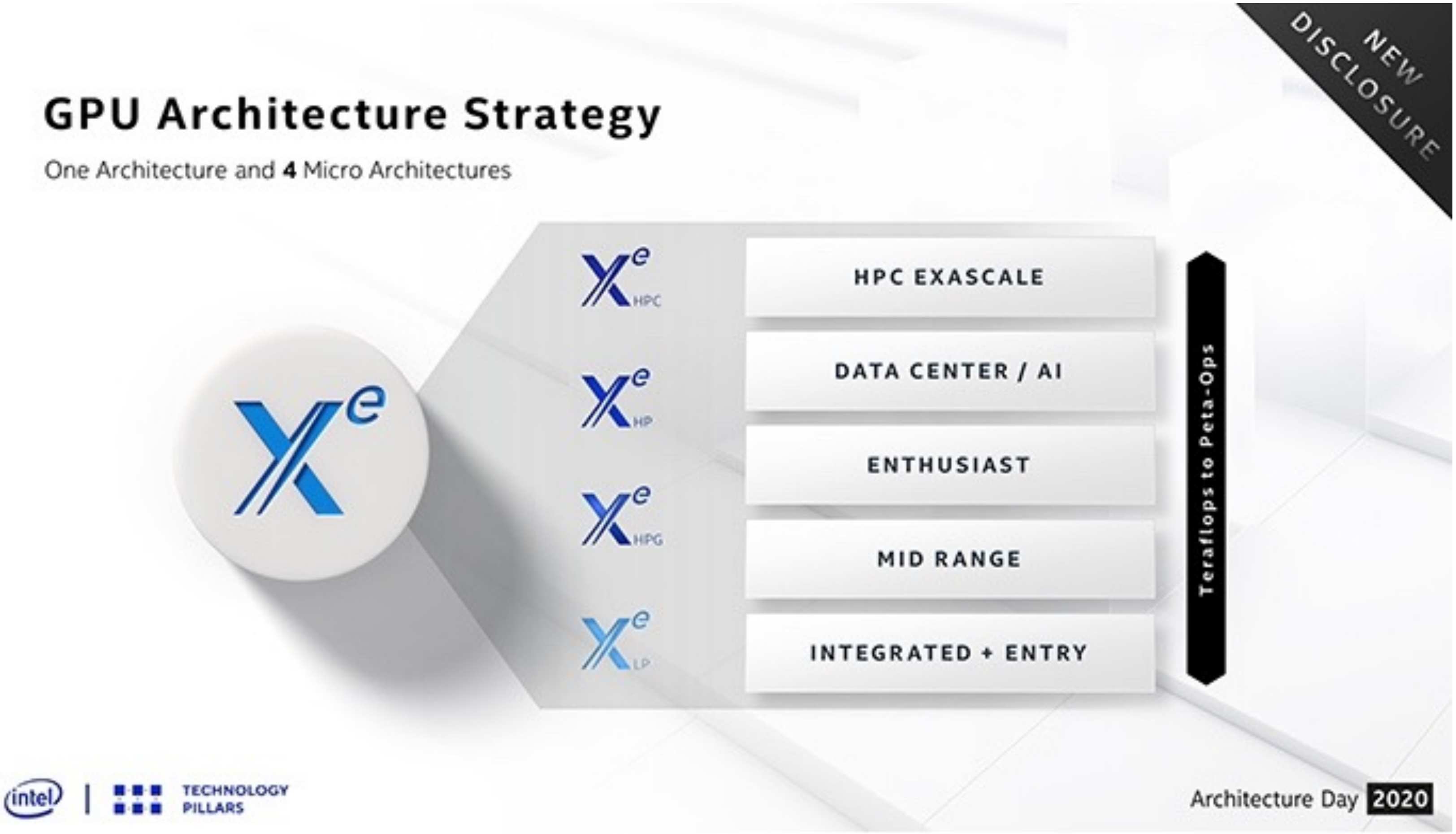
Argonne NATIONAL LABORATORY
U.S. DEPARTMENT OF ENERGY
intel
Hewlett Packard Enterprise
Aurora

The Evolution of Intel GPUs



The Road to X^e

The Evolution of Intel GPUs

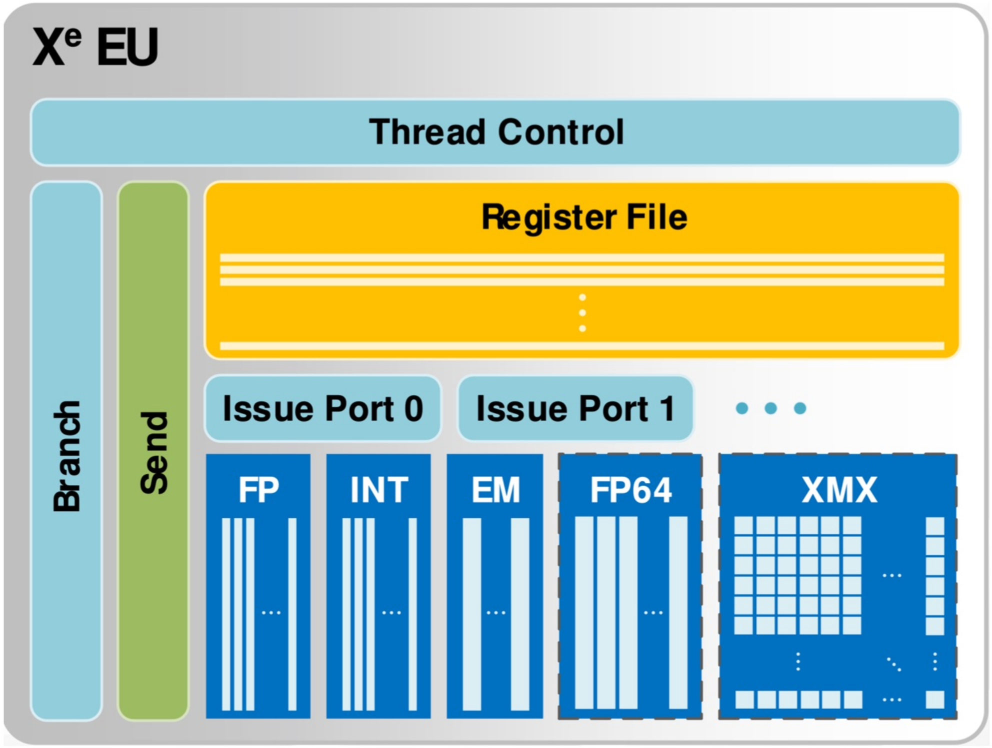


GPU Architecture Strategy

One Architecture and **4** Micro Architectures

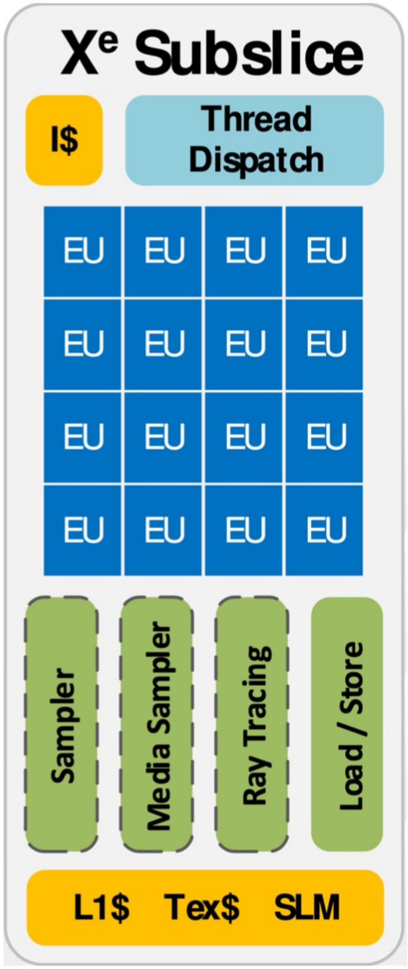
XE Execution Unit

- ❑ The EU executes instructions
  - ❑ Register file
  - ❑ Multiple issue ports
  - ❑ Vector pipelines
    - ❑ Float Point
    - ❑ Integer
    - ❑ Extended Math
    - ❑ FP 64 (optional)
    - ❑ Matrix Extension (XMX) (optional)
  - ❑ Thread control
  - ❑ Branch
  - ❑ Send (memory)



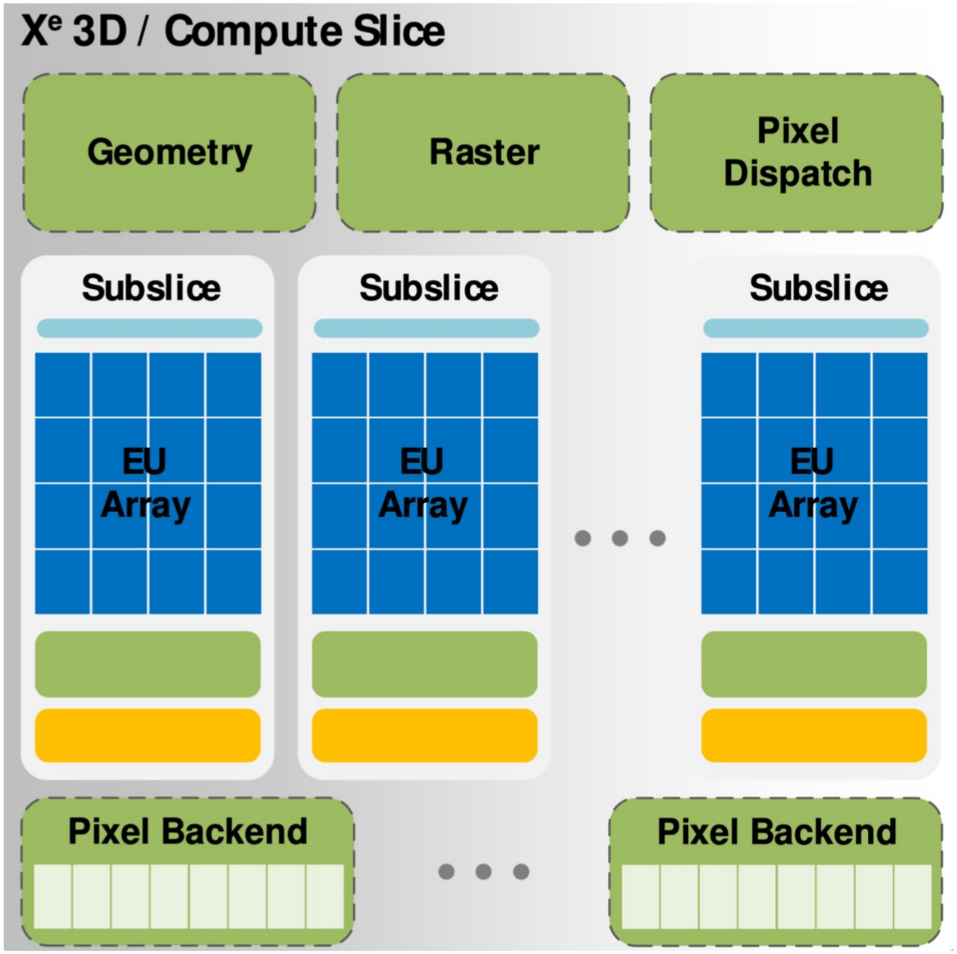
XE Subslice

- ❑A sub-slice contains:
  - ❑ 16 EUs
  - ❑ Thread dispatch
  - ❑ Instruction cache
  - ❑ L1, texture cache, and shared local memory
  - ❑ Load/Store
  - ❑ Fixed Function (optional)
    - ❑ 3D Sampler
    - ❑ Media Sampler
    - ❑ Ray Tracing



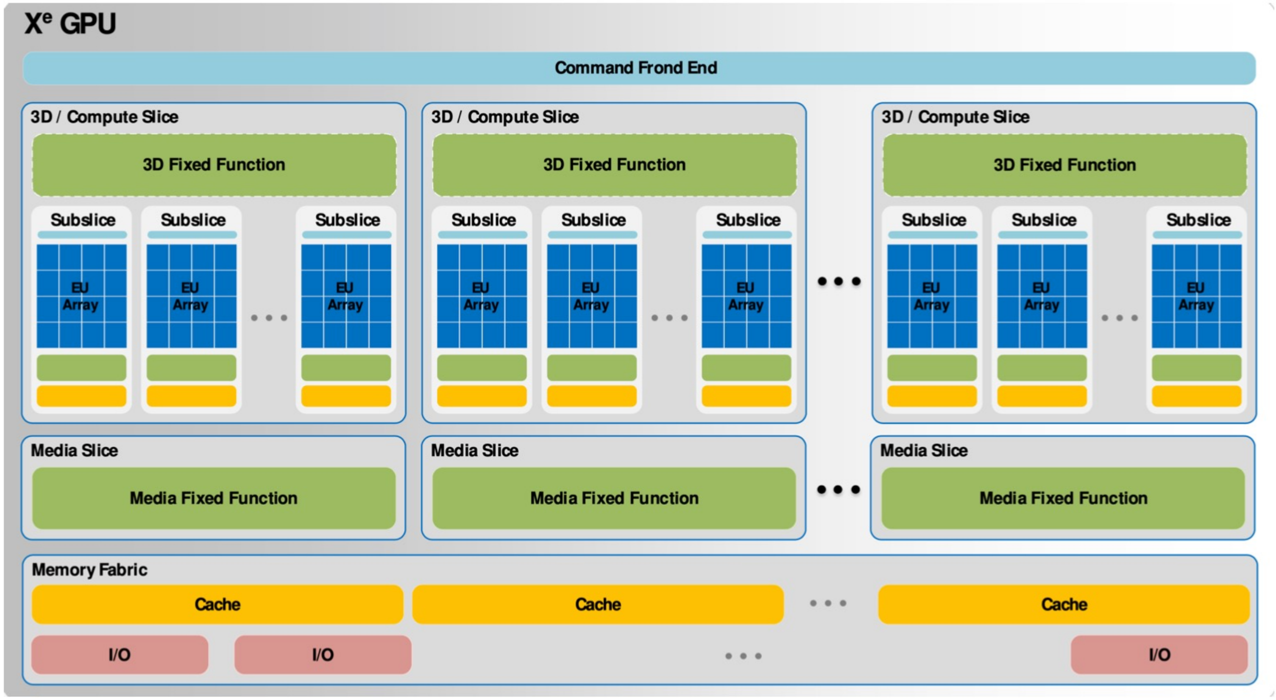
XE 3D/Compute Slice

- A slice contains
  - Variable number of subslices
  - 3D Fixed Function (optional)
    - Geometry
    - Raster



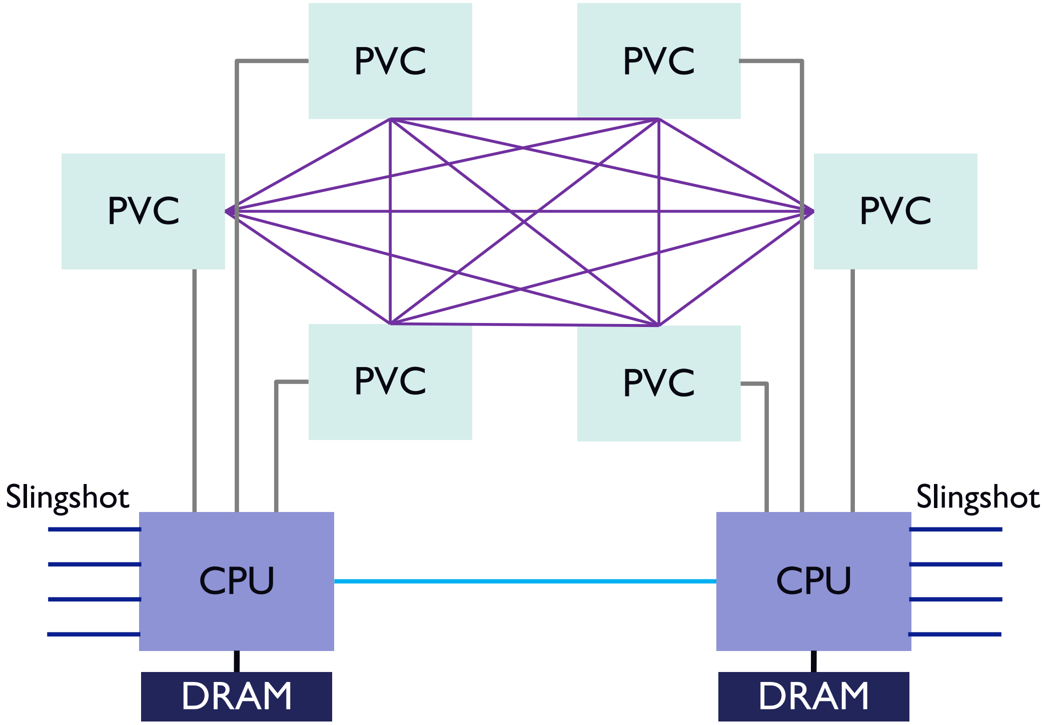
High Level Xe Architecture

- Xᵉ GPU is composed of
  - 3D/Compute Slice
  - Media Slice
  - Memory Fabric / Cache



Aurora Compute Node

- 6 X^e Architecture based GPUs (Ponte Vecchio)
  - All to all connection
  - Low latency and high bandwidth
- 2 Intel Xeon (Sapphire Rapids) processors
- Unified Memory Architecture across CPUs and GPUs
- 8 Slingshot Fabric endpoints



HPE Slingshot Network

- Slingshot is next generation scalable interconnect by HPE Cray
  - 8th major generation

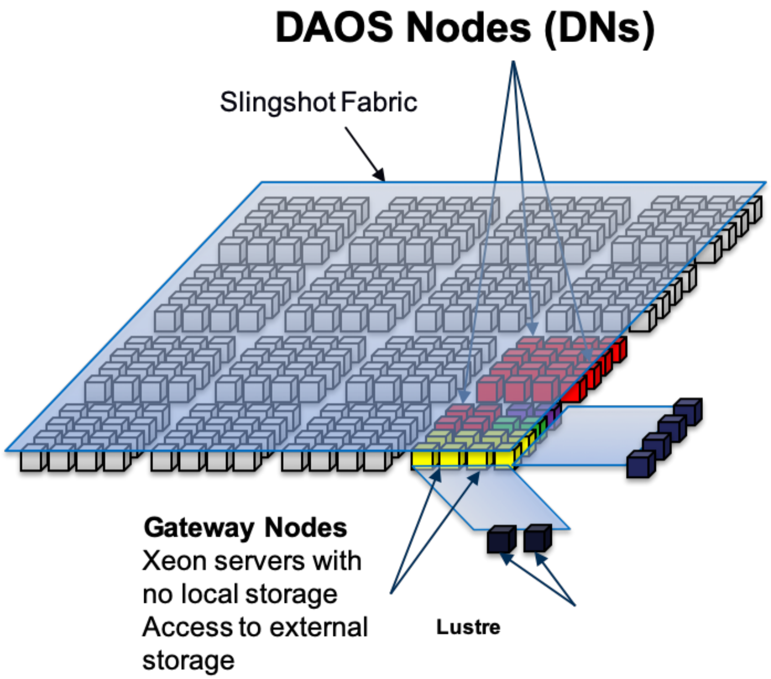
- Slingshot introduces:
  - Congestion management
  - Traffic classes
  - 3 hop dragonfly

https://www.hpe.com/us/en/compute/hpc/slingshot-interconnect.html

Argonne NATIONAL LABORATORY

Distributed Asynchronous Object Store (DAOS)

- ❑Primary storage system for Aurora
- ❑Offers high performance in bandwidth and IO operations
  - ❑ 230 PB capacity
  - ❑ ≥ 25 TB/s
- ❑Provides a flexible storage API that enables new I/O paradigms
- ❑Provides compatibility with existing I/O models such as POSIX, MPI-IO and HDF5
- ❑Open source storage solution



Pre-exascale and Exascale US Landscape

System	Delivery	CPU + Accelerator Vendor
Summit	2018	IBM + NVIDIA
Sierra	2018	IBM + NVIDIA
Perlmutter	2021	AMD + NVIDIA
Frontier	2021	AMD + AMD
Aurora	**2022**	**Intel + Intel**
El Capitan	2023	AMD + AMD

- Heterogenous Computing (CPU + Accelerator)
- Varying vendors

Heterogenous System Programming Models

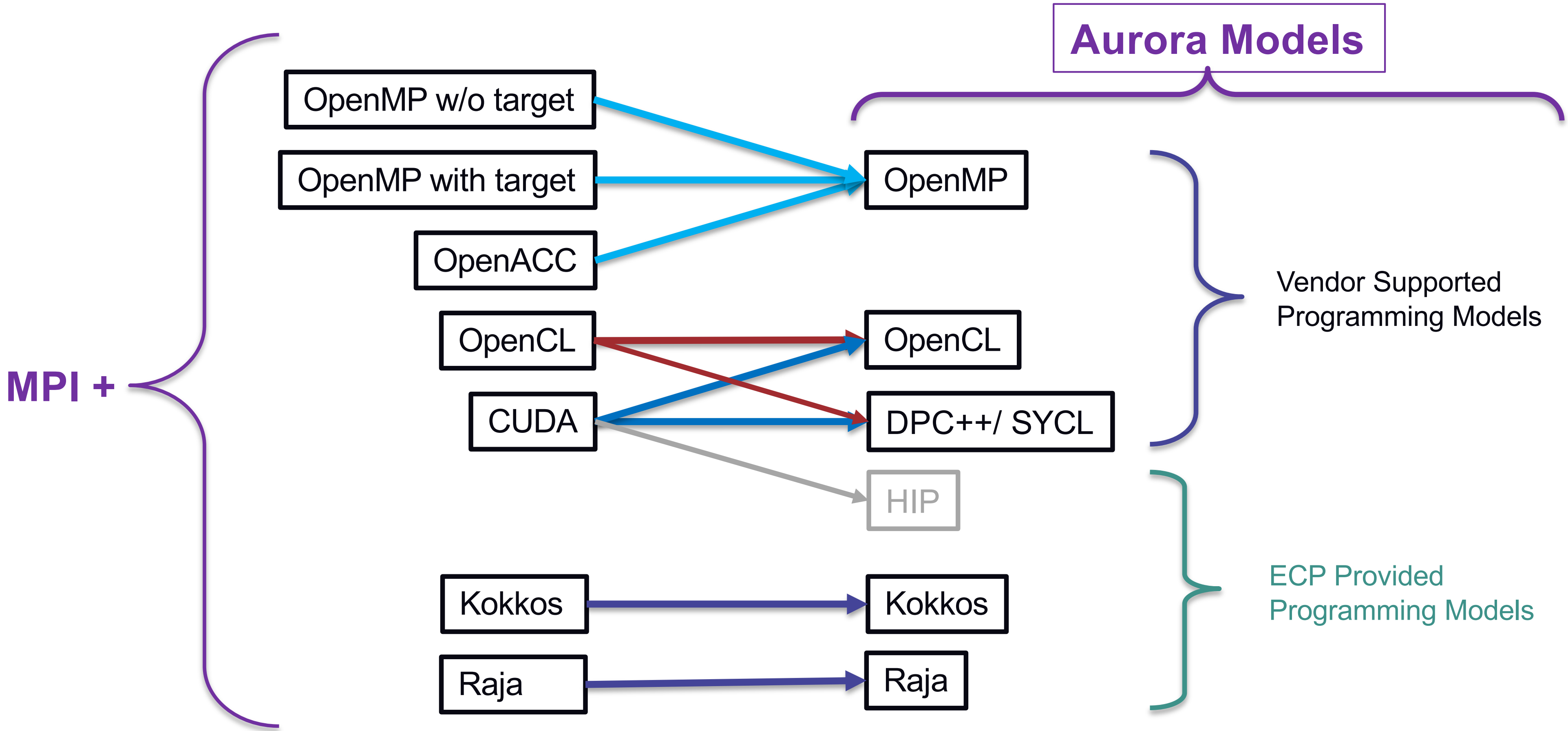
- Applications will be using a variety of programming models for Exascale:
  - CUDA
  - OpenCL
  - HIP
  - OpenACC
  - OpenMP
  - DPC++/SYCL
  - Kokkos
  - Raja

- Not all systems will support all models

Available Aurora Programming Models

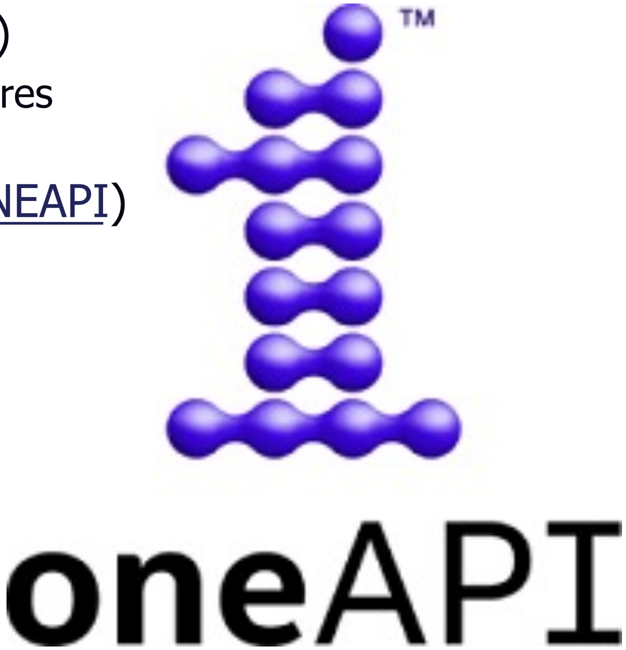
- □Aurora applications may use:
  - ~~□ CUDA~~
  - □ OpenCL
  - □ HIP
  - ~~□ OpenACC~~
  - □ OpenMP
  - □ DPC++/SYCL
  - □ Kokkos
  - □ Raja

Mapping of Existing Programming Models to Aurora



oneAPI

- Industry specification from Intel (https://www.oneapi.com/spec/)
  - Language and libraries to target programming across diverse architectures (DPC++, APIs, low level interface)
- Intel oneAPI products and toolkits (https://software.intel.com/ONEAPI)
  - Languages
    - Fortran (w/ OpenMP 5)
    - C/C++ (w/ OpenMP 5)
    - DPC++
    - Python
  - Libraries
    - oneAPI MKL (oneMKL)
    - oneAPI Deep Neural Network Library (oneDNN)
    - oneAPI Data Analytics Library (oneDAL)
    - MPI
  - Tools
    - Intel Advisor
    - Intel VTune
    - Intel Inspector

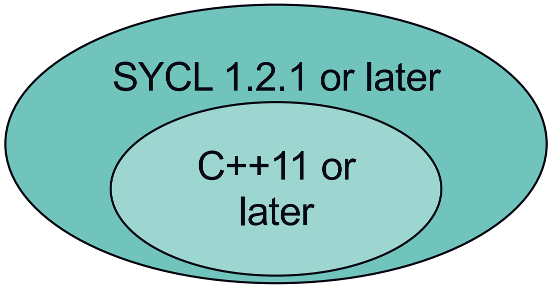


https://software.intel.com/oneapi

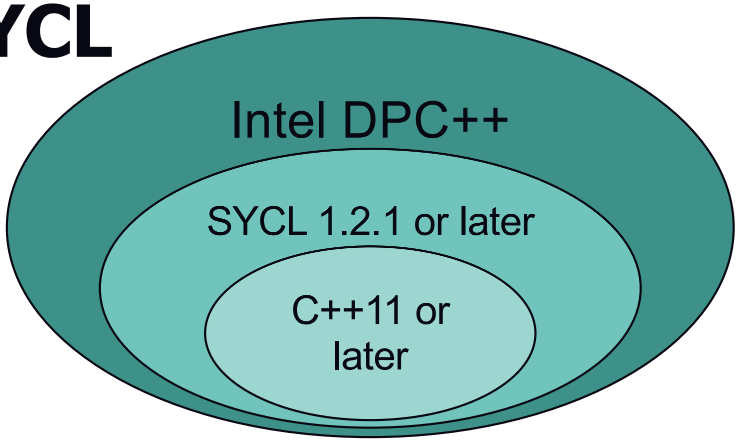
DPC++ (Data Parallel C++) and SYCL

- SYCL
    - Khronos standard specification
    - SYCL is a C++ based abstraction layer (standard C++11)
    - Builds on OpenCL **concepts** (but single-source)
    - *SYCL is designed to be as close to standard C++ as possible*

- Current Implementations of SYCL:
    - ComputeCPP™ (www.codeplay.com)
    - Intel SYCL (github.com/intel/llvm)
    - triSYCL (github.com/triSYCL/triSYCL)
    - hipSYCL (github.com/illuhad/hipSYCL)
  - **Runs on today's CPUs and nVidia, AMD, Intel GPUs**



DPC++ (Data Parallel C++) and SYCL



- SYCL
  - Khronos standard specification
  - SYCL is a C++ based abstraction layer (standard C++11)
  - Builds on OpenCL **concepts** (but single-source)
  - *SYCL is designed to be as close to standard C++ as possible*
- Current Implementations of SYCL:
  - ComputeCPP™ (www.codeplay.com)
  - Intel SYCL (github.com/intel/llvm)
  - triSYCL (github.com/triSYCL/triSYCL)
  - hipSYCL (github.com/illuhad/hipSYCL)
  - **Runs on today's CPUs and nVidia, AMD, Intel GPUs**
- DPC++
  - Part of Intel oneAPI specification
  - Intel extension of SYCL to support new innovative features
  - Incorporates SYCL 1.2.1 specification and Unified Shared Memory
  - Add language or runtime extensions as needed to meet user needs

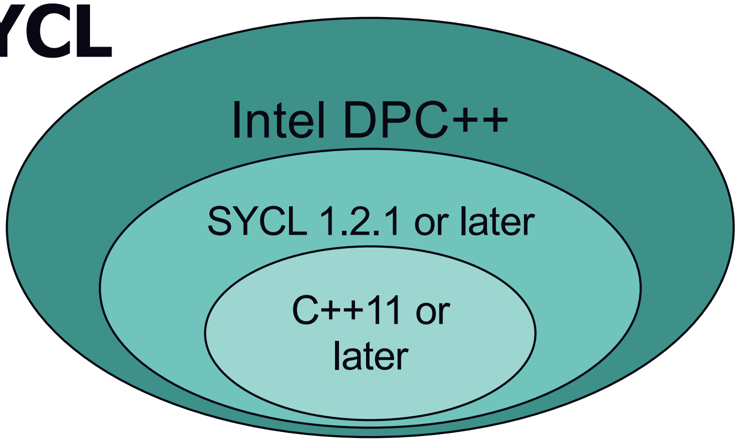
Extensions	**Description**
Unified Shared Memory (USM)	defines pointer-based memory accesses and management interfaces.
In-order queues	defines simple in-order semantics for queues, to simplify common coding patterns.
Reduction	provides reduction abstraction to the ND-range form of parallel_for.
Optional lambda name	removes requirement to manually name lambdas that define kernels.
Subgroups	defines a grouping of work-items within a work-group.
Data flow pipes	enables efficient First-In, First-Out (FIFO) communication (FPGA-only)

https://spec.oneapi.com/oneAPI/Elements/dpcpp/dpcpp_root.html#extensions-table

DPC++ (Data Parallel C++) and SYCL

- SYCL
  - Khronos standard specification
  - SYCL is a C++ based abstraction layer (standard C++11)
  - Builds on OpenCL **concepts** (but single-source)
  - *SYCL is designed to be as close to standard C++ as possible*
- Current Implementations of SYCL:
  - ComputeCPP™ (www.codeplay.com)
  - Intel SYCL (github.com/intel/llvm)
  - triSYCL (github.com/triSYCL/triSYCL)
  - hipSYCL (github.com/illuhad/hipSYCL)
  - **Runs on today's CPUs and nVidia, AMD, Intel GPUs**
- DPC++
  - Part of Intel oneAPI specification
  - Intel ex
  - Incorp Memo
  - Add language or runtime extensions as needed to meet user needs

Many DPC++ extensions included in SYCL 2020 specification released in February 2021.



Extensions	Description
Unified Shared Memory (USM)	defines pointer-based memory accesses and management interfaces.
In-order queues	defines simple in-order semantics for queues, to simplify common coding patterns.
Reduction	provides reduction abstraction to the ND-range form of parallel_for.
a	removes requirement to manually name lambdas that define kernels.
	defines a grouping of work-items within a work-group.
Data flow pipes	enables efficient First-In, First-Out (FIFO) communication (FPGA-only)

https://spec.oneapi.com/oneAPI/Elements/dpcpp/dpcpp_root.html#extensions-table

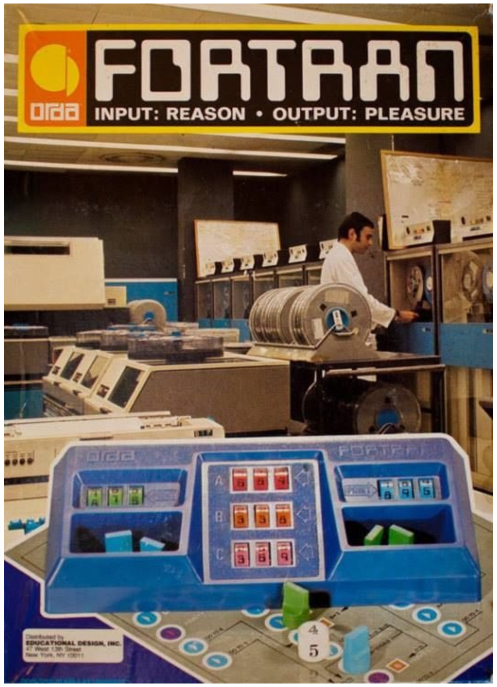
OpenMP 5

- OpenMP 5 constructs will provide directives based programming model for Intel GPUs
- Available for C, C++, and Fortran
- A portable model expected to be supported on a variety of platforms (Aurora, Frontier, Perlmutter, …)
- Optimized for Aurora
- For Aurora, OpenACC codes could be converted into OpenMP
  - ALCF staff will assist with conversion, training, and best practices
  - Automated translation possible through the clacc conversion tool (for C/C++)

https://www.openmp.org/

Intel Fortran for Aurora

- □ Fortran 2008
- □ OpenMP 5
- □ New compiler—LLVM backend
  - □ Strong Intel history of optimizing Fortran compilers
- □ Beta available today in OneAPI toolkits



https://software.intel.com/content/www/us/en/develop/tools/oneapi/components/fortran-compiler.html

Polaris

POLARIS

System Spec	Polaris
Planned Installation / Production	Q2CY2021 / Q3CY2021
System Peak (PF)	35-45
Peak Power (MW)	<2
Total System Memory (TB)	>250
System Memory Type	DDR, HBM
Node Performance (TF)	>70
Node Processors	1 CPU; 4 GPUs
System Size (nodes)	>500
Node-to-Node Interconnect	200Gb

Programming Models

- OpenMP 4.5/5
- SYCL
- Kokkos
- Raja
- HiP

Thank You