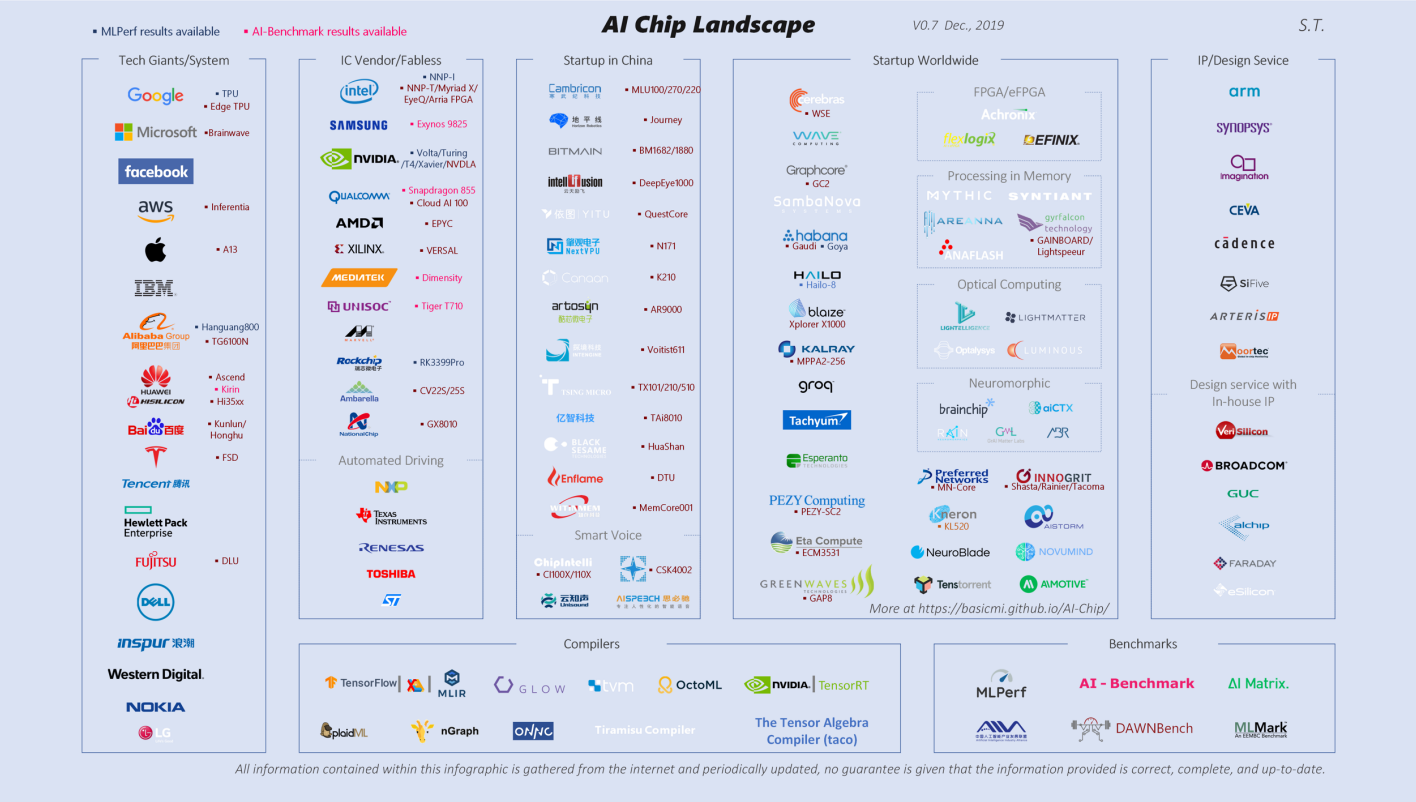
UXL Foundation

UXL FOUNDATION
Unified Acceleration

The computing landscape is changing

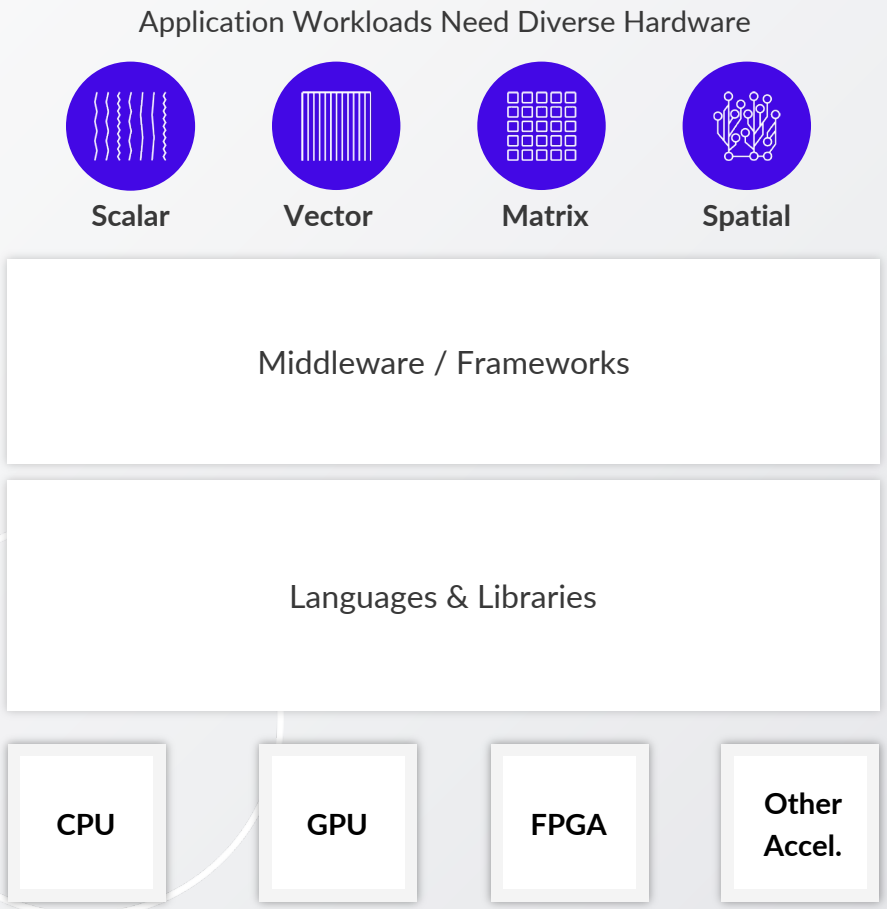
- Compute architectures are increasingly
  - Multi-vendor
  - Heterogeneous
  - Multi-Architecture



https://github.com/basicmi/AI-Chip

Software Challenges for Accelerator Computing

- Heterogeneous architectures are multi-vendor
- Significant investment to migrate software to new hardware
- Need an open standard way to develop software for accelerators



Unified Acceleration Foundation (UXL)

Mission

- Build a multi-architecture multi-vendor software ecosystem for all accelerators
- Unify the heterogeneous compute ecosystem around open standards
- Build on and expand open source projects for accelerated computing

Use case focus: AI, HPC, Edge AI and Edge Compute

Unified Acceleration Foundation (UXL)

Steering Members

arm

FUJITSU

Google Cloud

Imagination

intel

Qualcomm

SAMSUNG

vmware

Governance

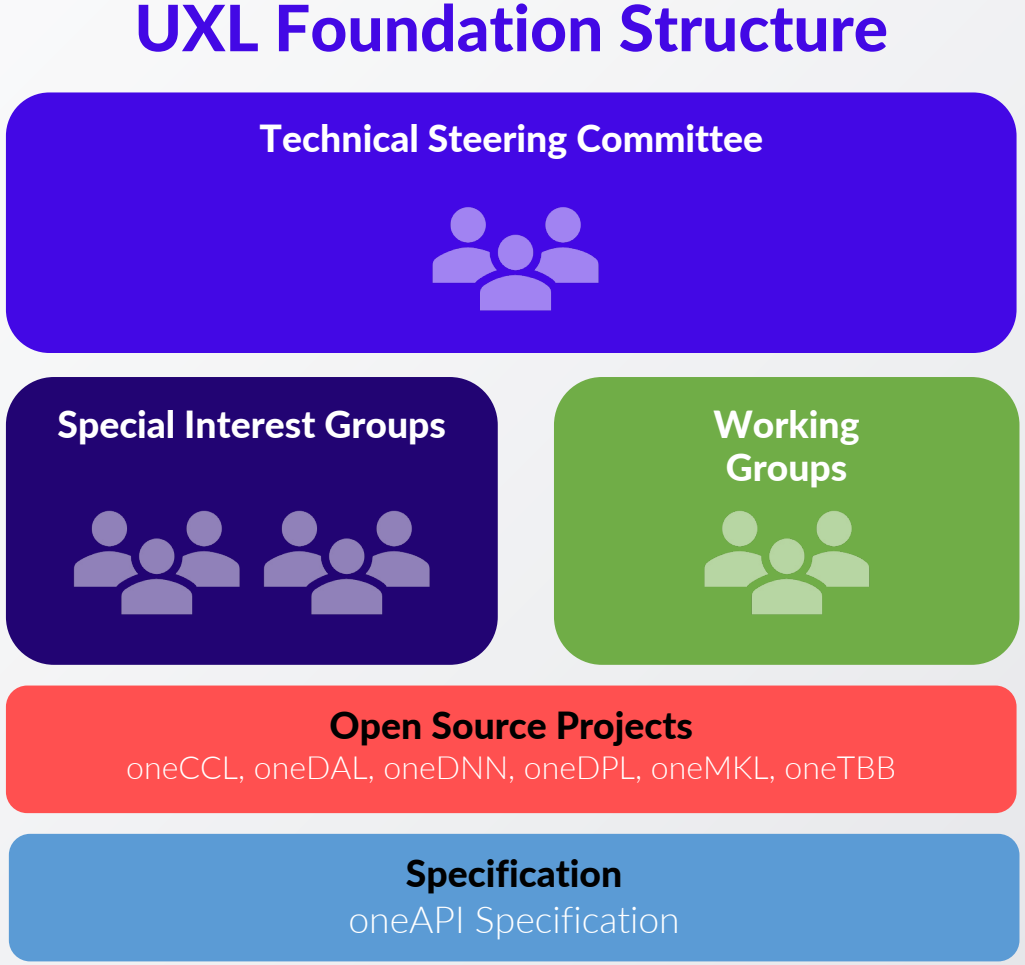
Joint Development Foundation governance

SIGs: AI, Hardware, Language, Math

Working Groups: Specification, Open Source

Join Us:

Participate in SIGs and Working Groups



Unified Acceleration Foundation (UXL)

oneAPI Specification

- Initial contribution:
  oneAPI Specification & Open Source

SYCL and C++

Dependencies on existing standards

oneDPL

Data Parallel C++ Library

oneDNN

Deep Neural Network Library

oneCCL

Collective Communications Library

Level Zero

Hardware Interface

oneDAL

Data Analytics Library

oneTBB

Threading Building Blocks

oneMKL

Math Kernel Library

APPROACH

The founding companies are seeding the project with highly valuable contributions to open source libraries

Working Groups

Specification – defining an open standard for accelerated libraries

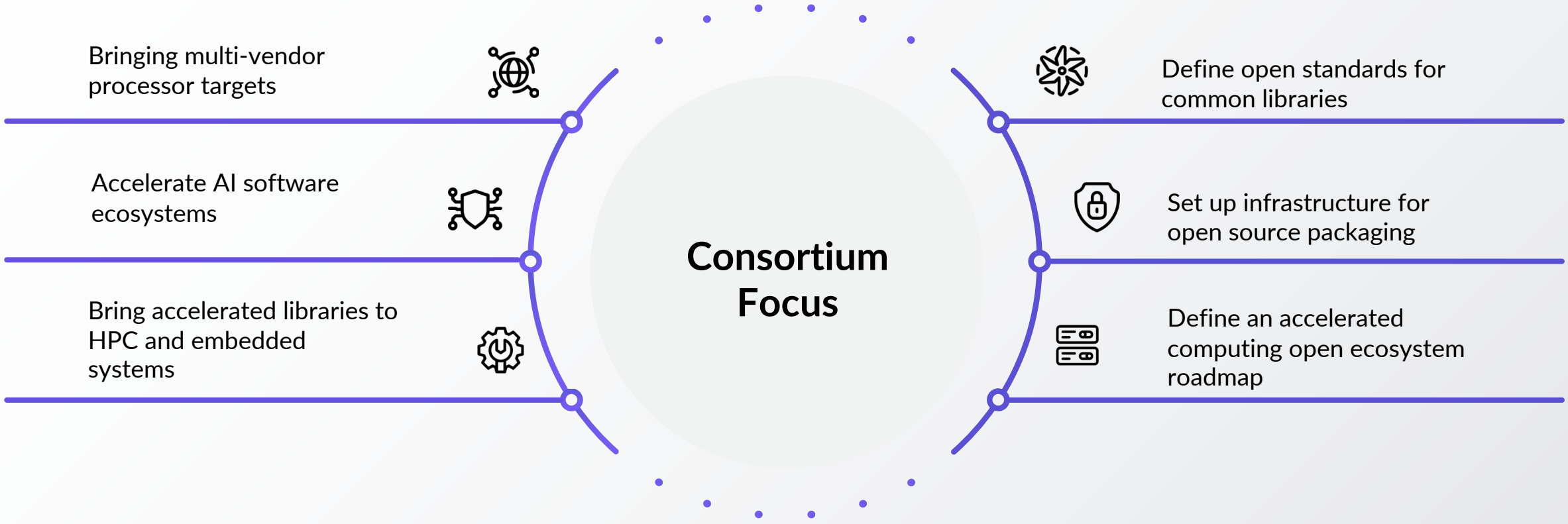
Open Source – coordinating community contributions and feedback

The group will work to drive the development of an open ecosystem for accelerated computing based on the fundamentals of open standards and open source

Project governed by the Joint Development Foundation (JDF), a part of the Linux Foundation

TECHNICAL GOALS

Open specifications, APIs, open source for AI and HPC, Edge Compute and Edge AI



Unified Acceleration Foundation (UXL)

Existing ongoing collaborations

- **Fujitsu:**
  oneAPI Deep Neural Network Library (oneDNN); oneAPI Data Analytics Library (oneDAL) optimizations for Arm processors
- **Google Cloud:**
  oneDNN optimizations for Intel processors
- **Argonne, Lawrence Berkeley & Oakridge:**
  DPC++ and oneDNN used on Intel, Nvidia and AMD GPUs
- **GROMACS:**
  SYCL and oneAPI used to target multi-vendor architectures

UXL Foundation SIGs

Language

Enabling languages for accelerated computing

Hardware

Bringing more target processors to the UXL projects

Math

Optimized operators for commonly used functions

AI

Accelerated libraries and frameworks for machine learning and AI algorithms

TIMELINE


Sept. 2023
Public announcement
Q4 2023
New members can join
Annual Specification Release
Q4 2023
Develop Roadmap
Plan 2024
2024
Members contributing to open source projects

Join The UXL Foundation

Steering Member
$20k*

- Seat on the Steering Committee
- Voting Rights
- Define the direction of the foundation

General Member
$5k*

- Working Group Voting Rights
- Influence Working Group direction
- Co-marketing

Contributor Member
$0

- Participate in Working Groups
- Contribute to the specification
- Contribute to the projects

https://uxlfoundation.org/membership

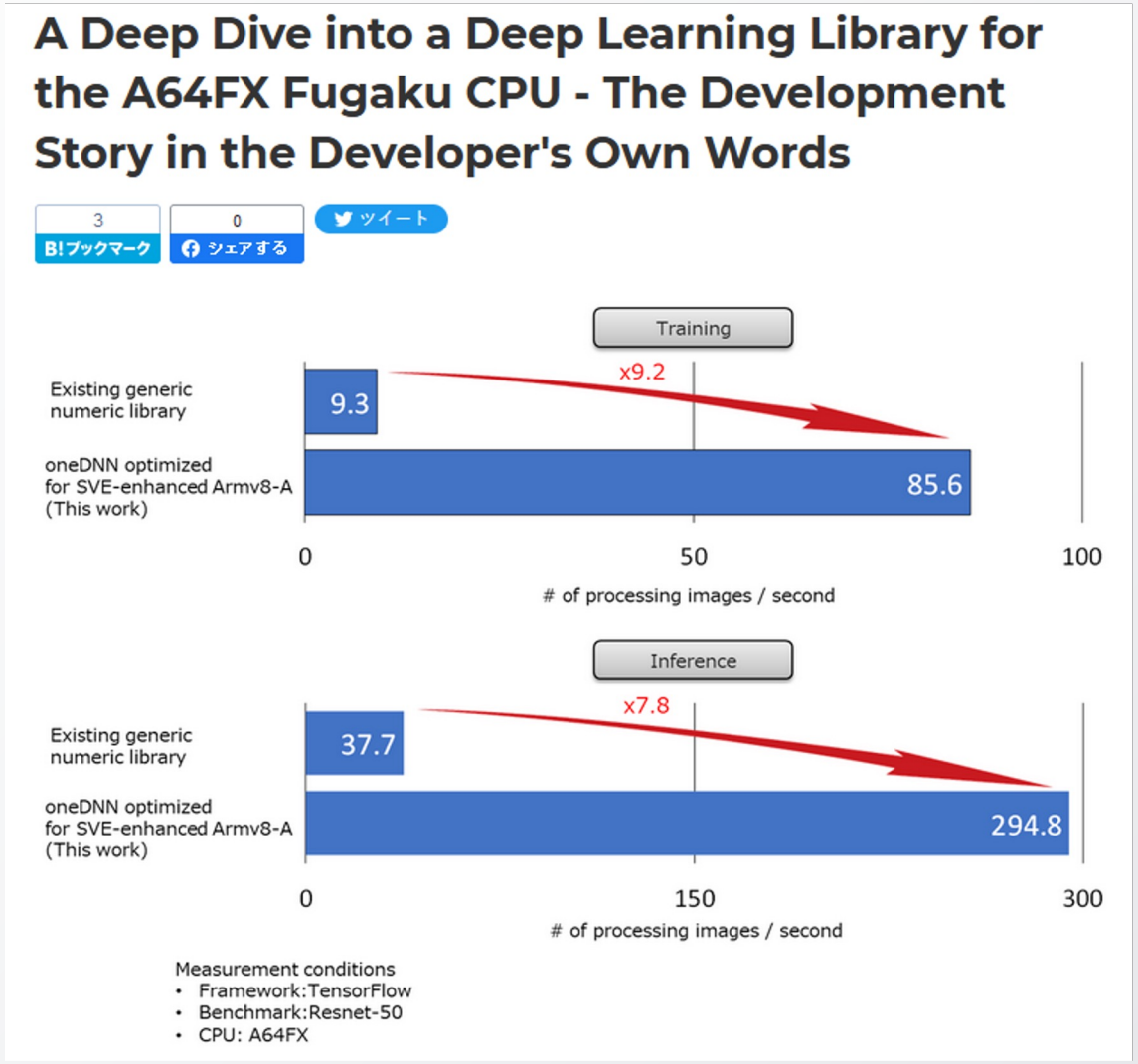
membership@uxlfoundation.org

* plus Linux Foundation membership

Case Studies

Fujitsu

- Adapted oneAPI Deep Neural Network library for Fugaku Arm CPU
- Achieved significant performance improvements using highly optimized implementations of deep learning building blocks
- Contributed to oneDNN open source project



https://blog.fltech.dev/entry/2020/11/19/fugaku-onednn-deep-dive-en

Google Cloud

- TensorFlow uses the oneAPI Deep Neural Network (oneDNN) library to accelerate models
- Significant improvements in performance were achieved using oneDNN

AMX · bfloat16 ·

Optimizing TensorFlow for 4th Gen Intel Xeon Processors

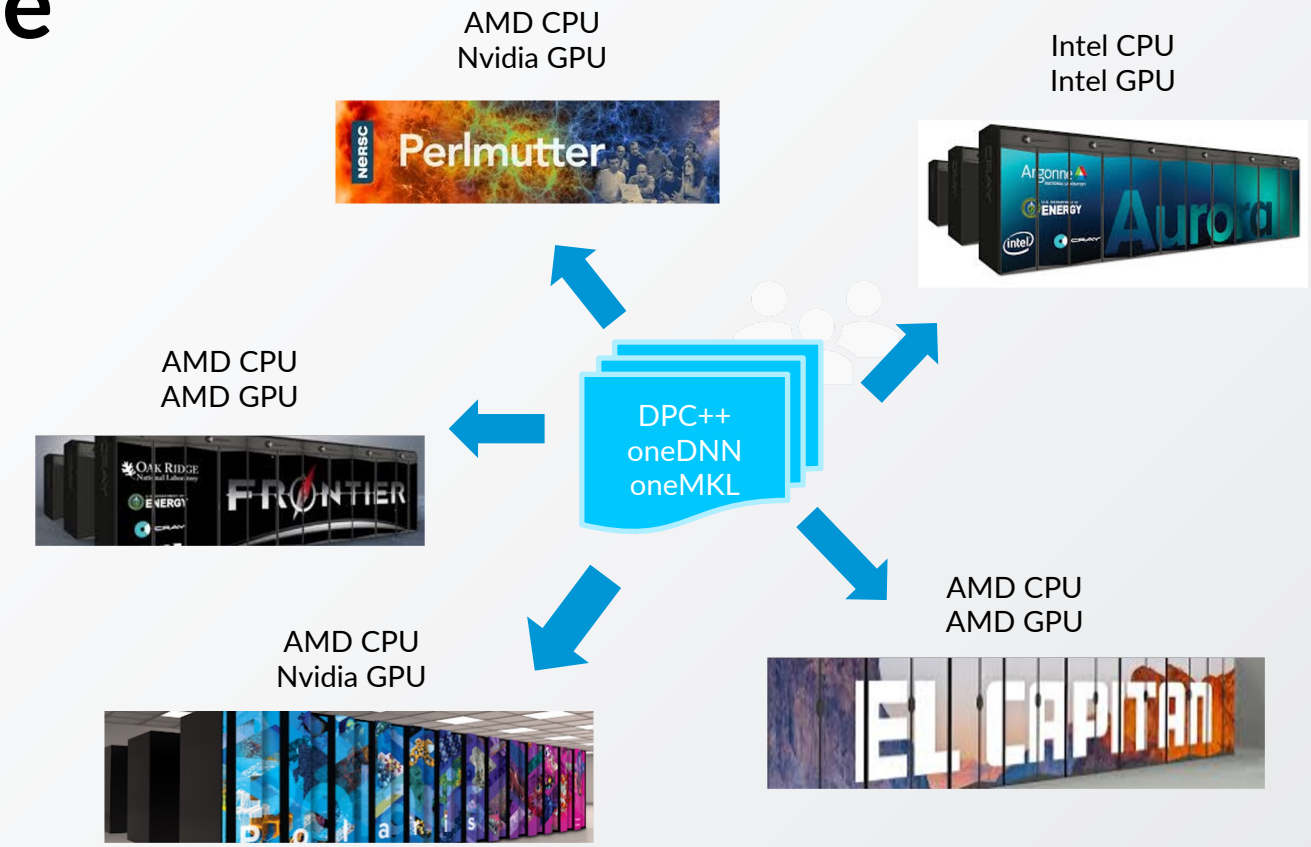
January 10, 2023

Posted by Ashraf Bhuiyan, AG Ramesh from Intel, Penporn Koanantakool from Google

https://blog.tensorflow.org/2023/01/optimizing-tensorflow-for-4th-gen-intel-xeon-processors.html

Argonne, Lawrence Berkeley and Oak Ridge

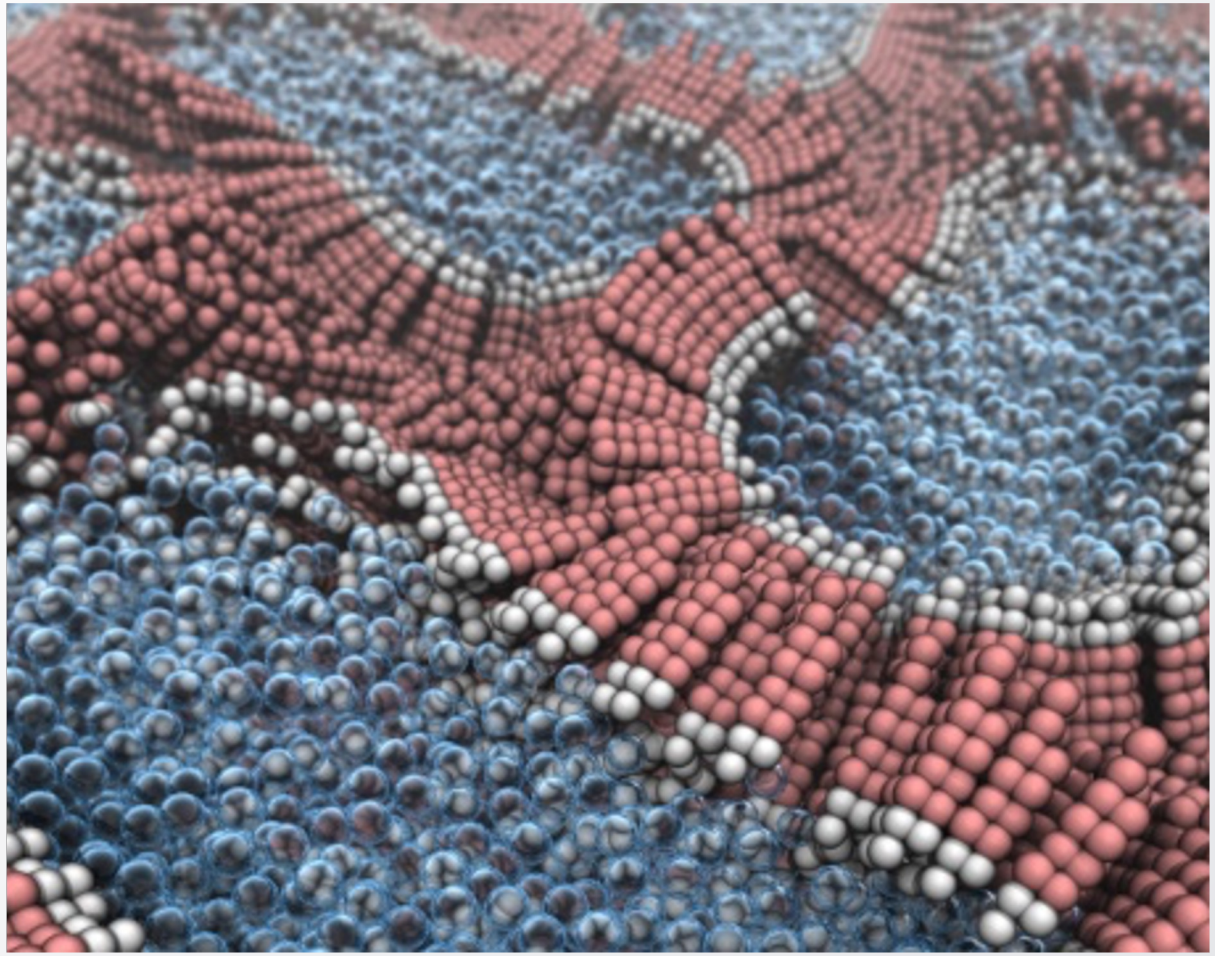
- US National Laboratories are ensuring researchers can target new supercomputers using a common programming model
- Partnerships enable the SYCL implementation DPC++ and oneDNN on Intel, Nvidia and AMD GPUs



GROMACS

FAST. FLEXIBLE. FREE.
GROMACS

- Adopted SYCL and oneAPI to target multi-vendor architectures
- GROMACS workload can be executed on AMD and Nvidia GPUs simultaneously, as well as Intel GPU and CPU from a single binary executable



Questions