



Unified Acceleration Foundation

Unified Acceleration Foundation (UXL)

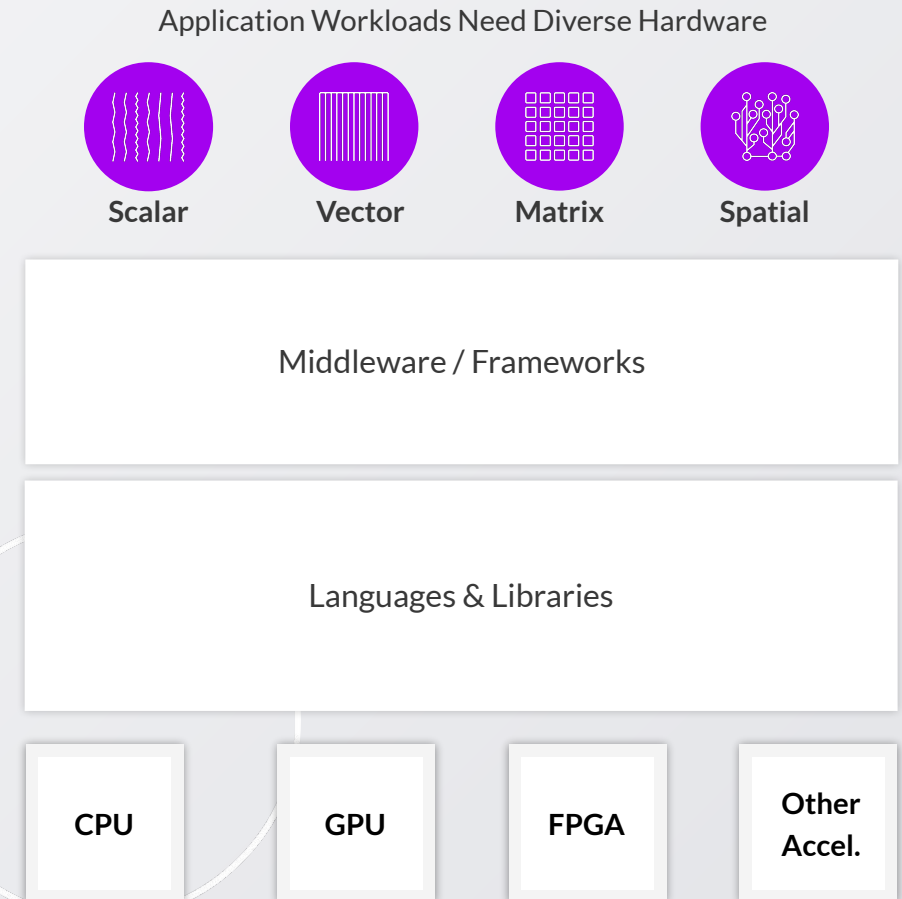
Mission

- 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

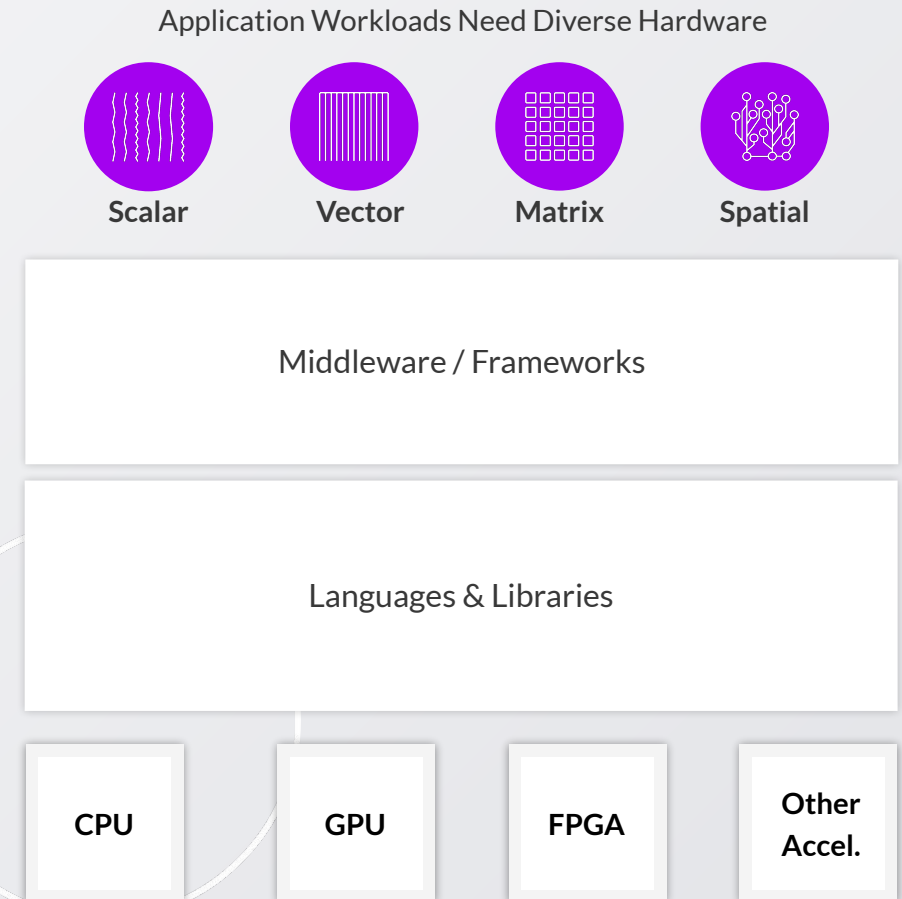
Software Challenges for Accelerator Computing

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



Hardware Challenges for Accelerator Computing

- Software developers demand a standard way to target processors
- Huge investment is required to deliver a software platform for new hardware architectures



Unified Acceleration Foundation

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

UXL Foundation Structure

Technical Steering Committee



Special Interest Groups



Working Groups



Open Source Projects

oneCCL, oneDAL, oneDNN, oneDPL, oneMKL, oneTBB

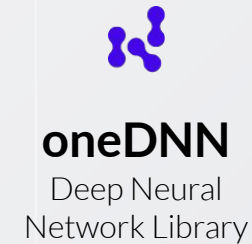
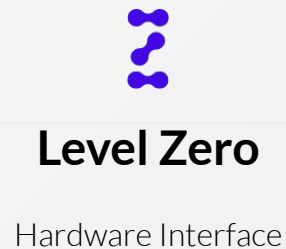
Specification

oneAPI Specification

Unified Acceleration Foundation

oneAPI Specification and Projects

- **Initial contribution:** oneAPI Specification & Open Source



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

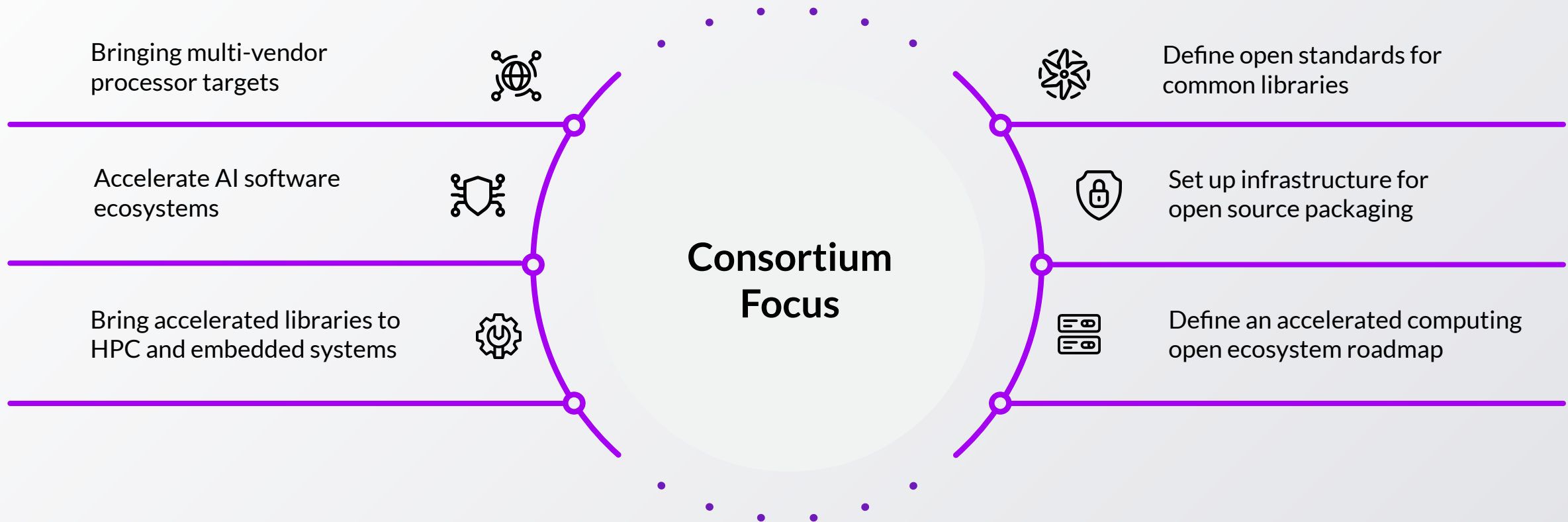


The group will work to drive the development of an open ecosystem for accelerated computing based on the fundamentals of open standards and open source

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

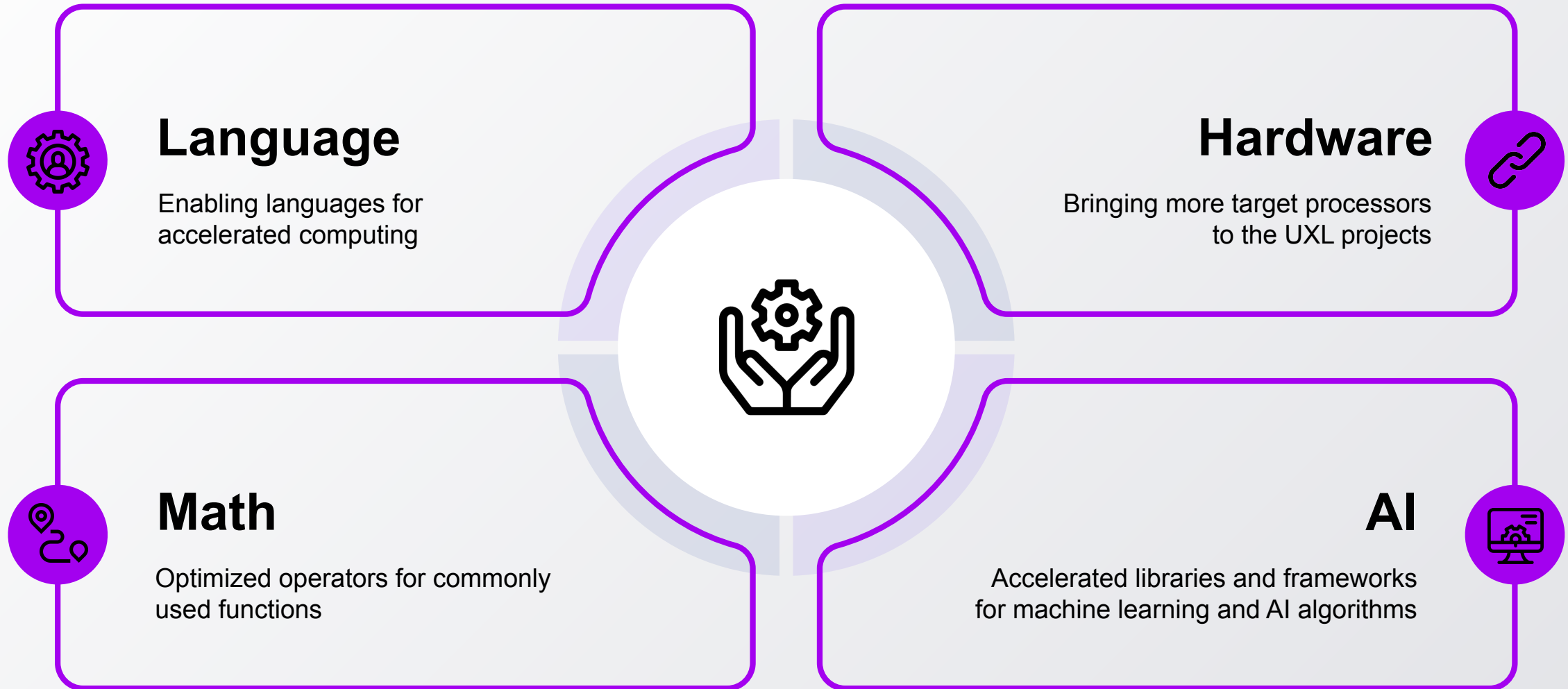


Unified Acceleration Foundation

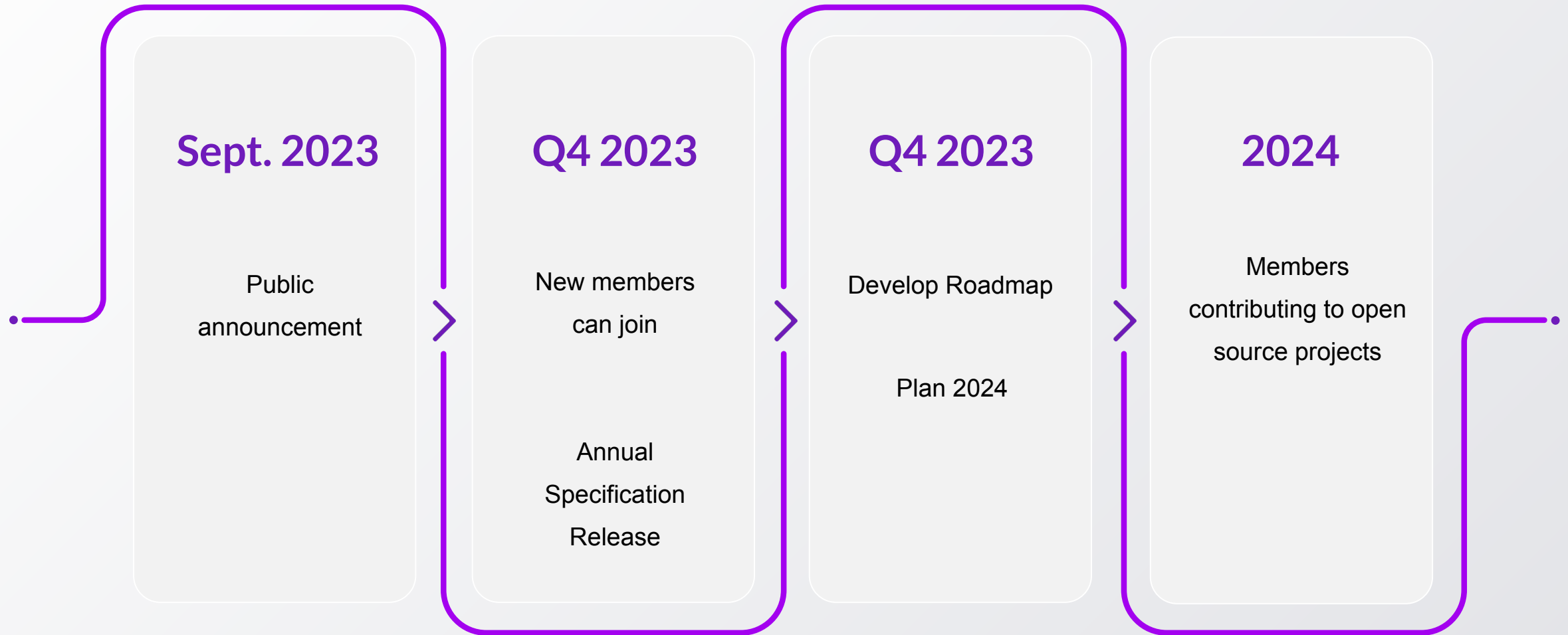
Existing Ongoing Collaborations

- **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++, oneMKL and oneDNN used on Intel, Nvidia and AMD GPUs
- **GROMACS:**
SYCL and oneAPI used to target multi-vendor architectures

UXL Foundation SIGs



Timeline



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

membership@uxlfoundation.org

Contributor Member \$0

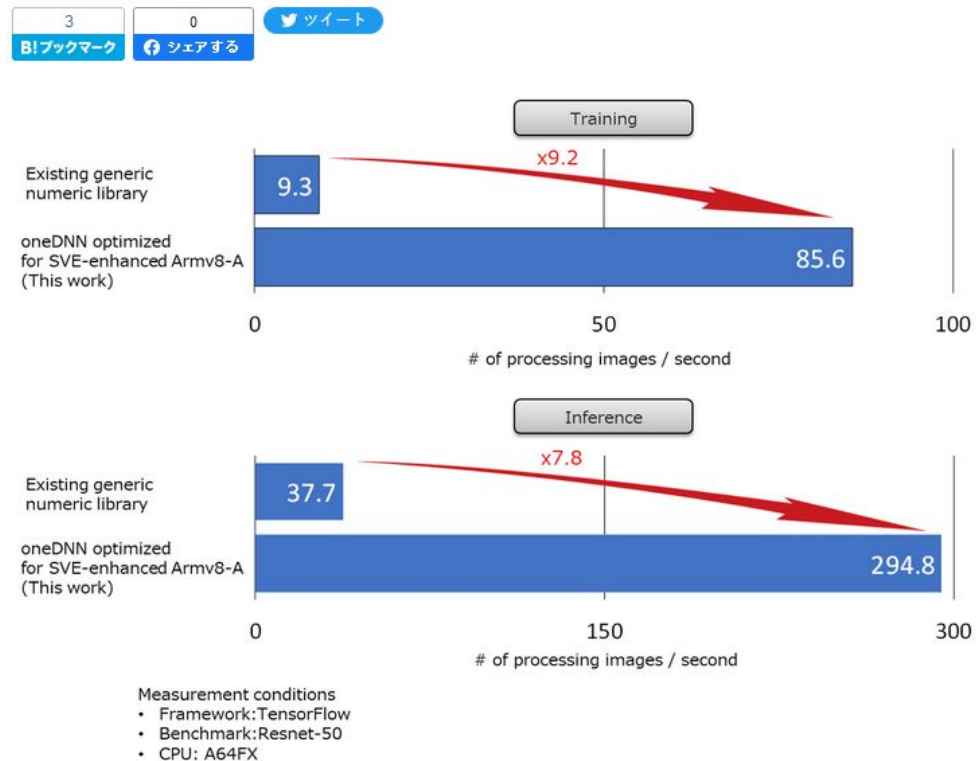
- Participate in Working Groups
- Contribute to the specification
- Contribute to the projects

* plus Linux Foundation membership

Case Studies

- 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

A Deep Dive into a Deep Learning Library for the A64FX Fugaku CPU - The Development Story in the Developer's Own Words



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

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

AMX · bfloat16 · G

Optimizing TensorFlow for 4th Gen Intel Xeon Processors

January 10, 2023



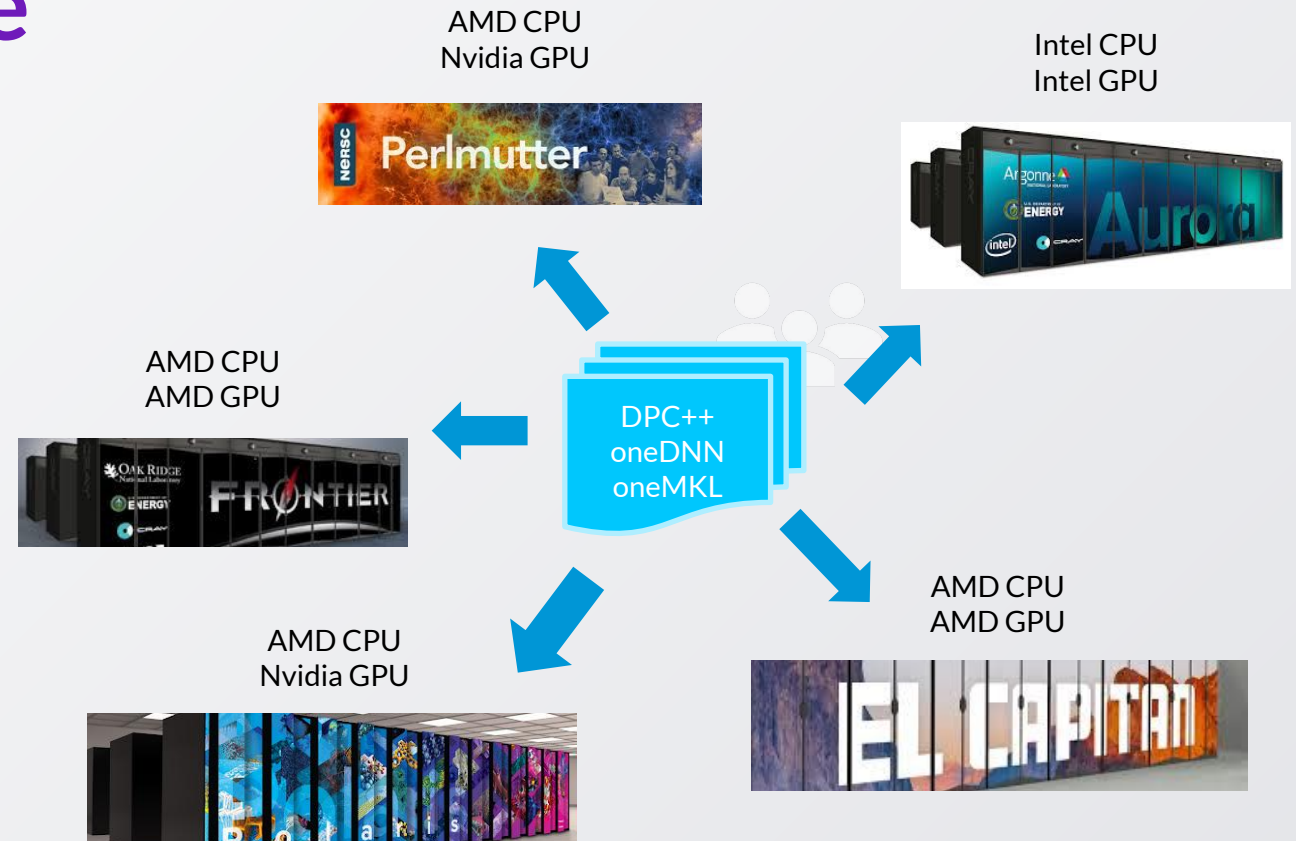
Posted by Ashraf Bhuiyan, AG Ramesh from Intel, Penporn Koanantakool from Google



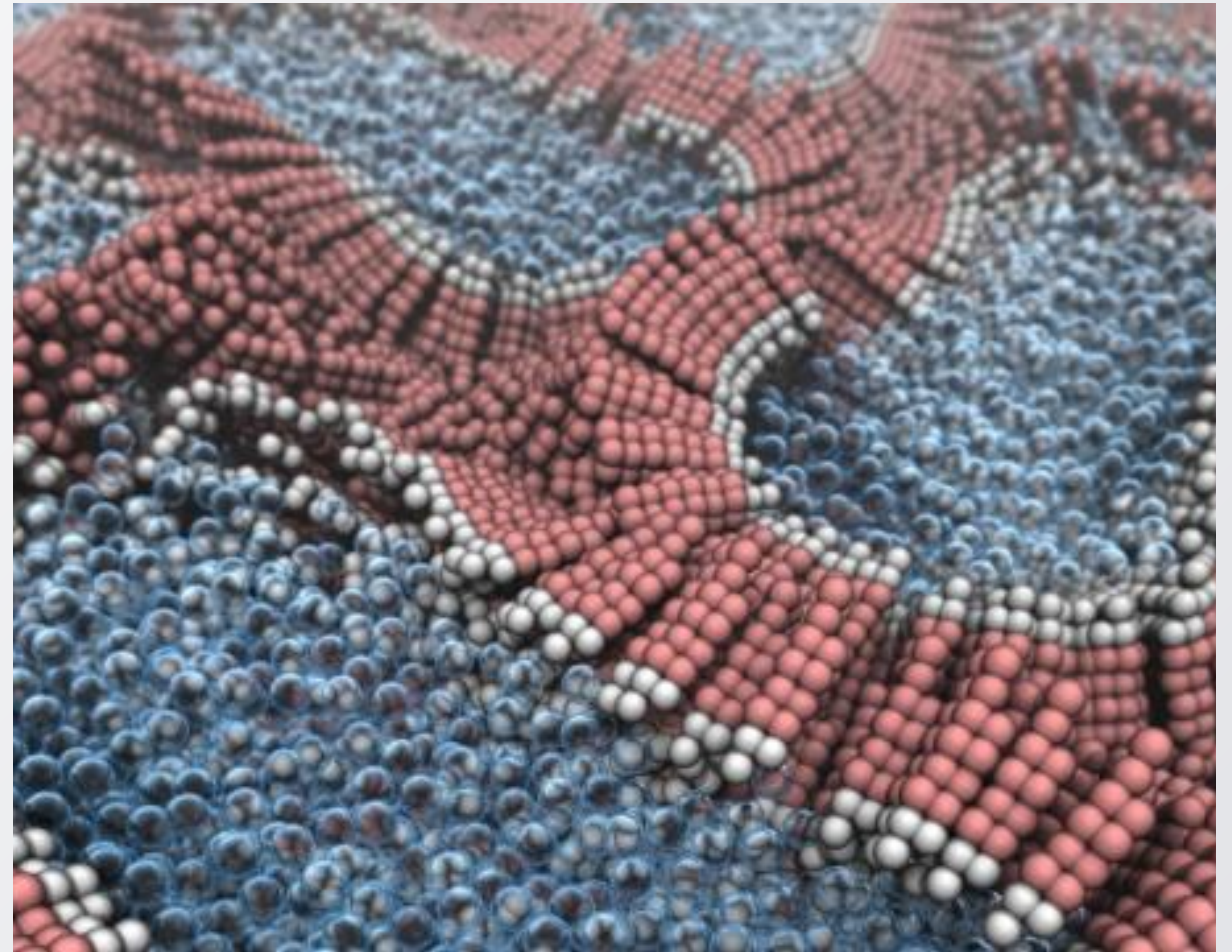
<https://blog.tensorflow.org/2023/01/optimizing-tensorflow-for-4th-gen-intel-xeon-processors.html>

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



- 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