

Intel® oneAPI, News and Advances 2024

Europar24-Madrid

CGS

August 26, 2024

- “Intel® oneAPI Programming Guide”,
<https://www.intel.com/content/www/us/en/develop/documentation/oneapi-programming-guide/top.html>



Outline

1 Introduction

2 Suite oneAPI

3 Intel Developer Cloud named Tiber

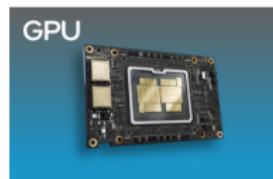
4 Other





Introduction

- High-Performance Computing (HPC) should be an exclusive concern of large-scale science
- ... but it is becoming a fundamental feature in other domains such as AI, data analysis, content creation, or graphics
- **Proliferation of architectures:** GPUs, FPGAs, ASICs...
 - Programmers have to adapt their developments to more and more complex programming on different type of accelerators



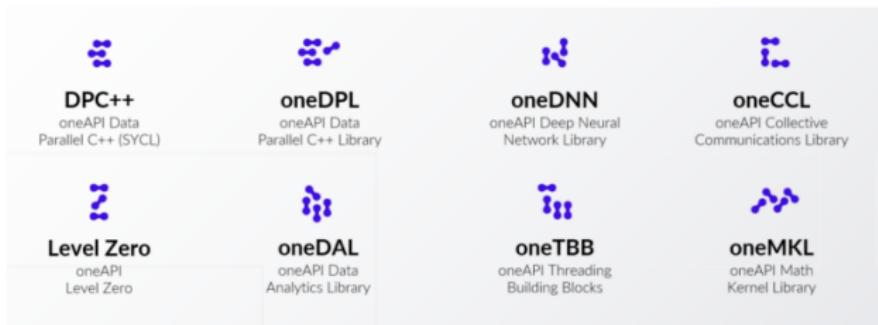
- Establishment of the [Unified Acceleration Foundation \(UXL\)](#) announced at Linux Foundation Open Source Summit Sept'23
- Executive members: Arm, Fujitsu, Google, Imagination Technologies, Intel, Qualcomm Technologies Inc, and Samsung
- **Objective:** an open and standards-based programming model for all accelerators, promoting support for multiple architectures and multiple vendors

UXL FOUNDATION

Unified Acceleration



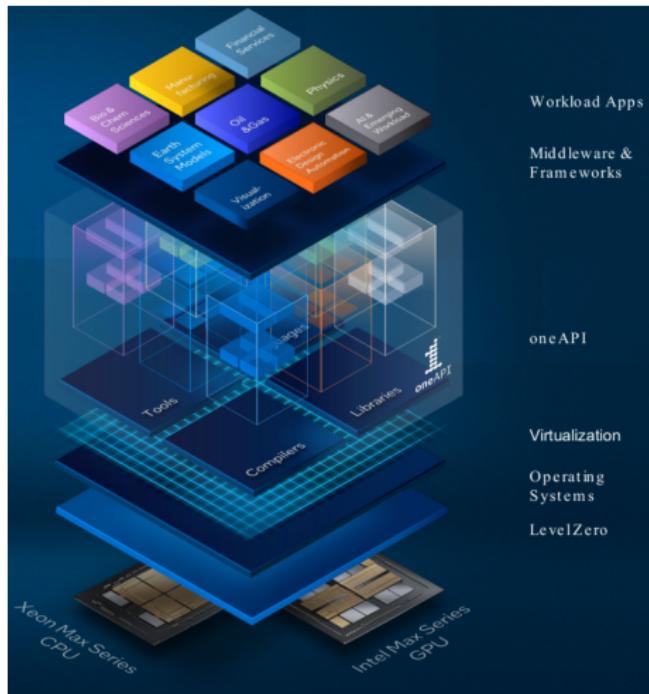
- **SYCL** as the open standard from Khronos and the **oneAPI** specification form the basis of UXL Foundation's efforts
- Collaboration with processor vendors and software developers, aligning with standards bodies such as Khronos and ISO C++



oneAPI elements Intel is donating to UXL

oneAPI Intro

- Unified programming model: diverse architectures
- Optimized language and libraries
- Performance equivalent to high-level native language
- Based on industry standards and open specifications
- Compatible with existing HPC programming models

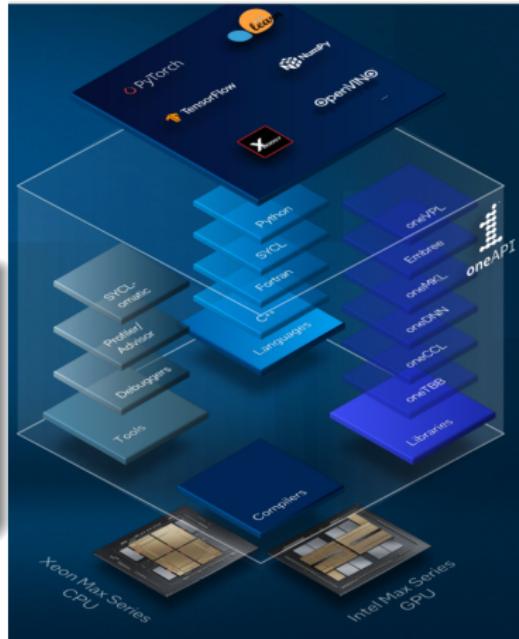


oneAPI Intro

- Standards-based language: C++ and SYCL
- Powerful APIs to accelerate domain-specific functions

Solution to single provider

- Open standard to promote community and industry support
- Allows code reuse across different architectures and providers





- A comprehensive set of development tools tested from CPU to XPU
- Available for installation in Intel® oneAPI Toolkits

Intel® oneAPI Base Toolkit	 The logo consists of the Intel 'i' icon followed by the number '1' and the text 'oneAPI' on a blue background, with 'BASE TOOLKIT' below it.	A core set of high-performance libraries and tools for building C++, SYCL, C/OpenMP, and Python applications
Add-on Domain-specific Toolkits	 The logo consists of the Intel 'i' icon followed by the number '1' and the text 'oneAPI' on a blue background, with 'HPC TOOLKIT' below it. For HPC developers	 The logo consists of the Intel 'i' icon followed by the number '1' and the text 'oneAPI' on a blue background, with 'IOT TOOLKIT' below it. For Edge & IoT developers
	 The logo consists of the Intel 'i' icon followed by the number '1' and the text 'oneAPI' on a blue background, with 'HPC TOOLKIT' below it. Intel® oneAPI Tools for HPC Deliver fast Fortran, OpenMP & MPI applications that scale	 The logo consists of the Intel 'i' icon followed by the number '1' and the text 'oneAPI' on a blue background, with 'IOT TOOLKIT' below it. Intel® oneAPI Tools for IoT Build efficient, reliable solutions that run at network's edge
<p>For visual creators, scientists, and engineers</p> <p>Intel® oneAPI Rendering Toolkit Create performant, high-fidelity visualization applications</p> <p>For AI developers and data scientists</p> <p>Intel® AI Analytics Toolkit Accelerate machine learning & data science pipelines end-to-end with optimized DL & ML frameworks & high-performing Python libraries</p> <p>OpenVINO™ For deep learning inference developers</p> <p>Intel® OpenVINO™ toolkit Deploy high performance inference & applications from edge to cloud</p> <p>Download at intel.com/oneAPI Or run tools in the Intel® Developer Cloud</p>		

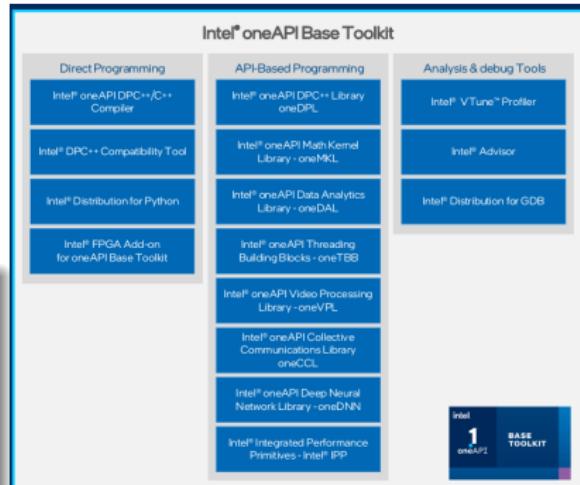


oneAPI Base Toolkit

- Basic set of high-performance tools and libraries
- C++ compiler with SYCL support (heterogeneous computing)

Features

- Data Parallel C++ Compiler
- Portability with SYCLomatic
- Python distribution (optimized libraries scikit-learn, NumPy)



Updates in DPC++

- Improves support for accelerators with more features from SYCL2020 and OpenMP 5.0 and v5.1
- Enhances performance in CPU-GPU applications
 - SYCL/DPC++ offers equivalent performance to OpenMP on CPU
 - Example: HPCBench achieves 106% performance compared to OpenMP
 - Example: HPCBench achieves 138% performance compared to CUDA@A100

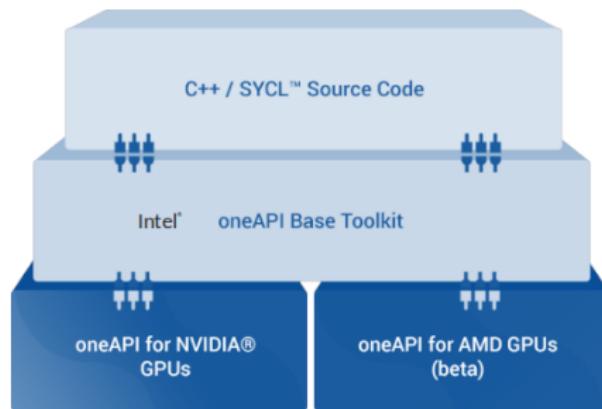
What is DPC++?

- Compiler with SYCL support
- Based on standards such as LLVM



DPC++/SYCL Portable

- DPC++ is an open project: [source code on GitHub](#) and allows the creation of a [Toolchain](#).
 - oneAPI for NVIDIA GPUs through a plugin developed by [Codeplay](#)
 - Facilitates creating SYCL code and running it on compatible NVIDIA GPUs (>sm50)
 - Support for AMD GPUs



[NVIDIA GPU plug-in](#) [AMD GPU plug-in](#)



Libraries

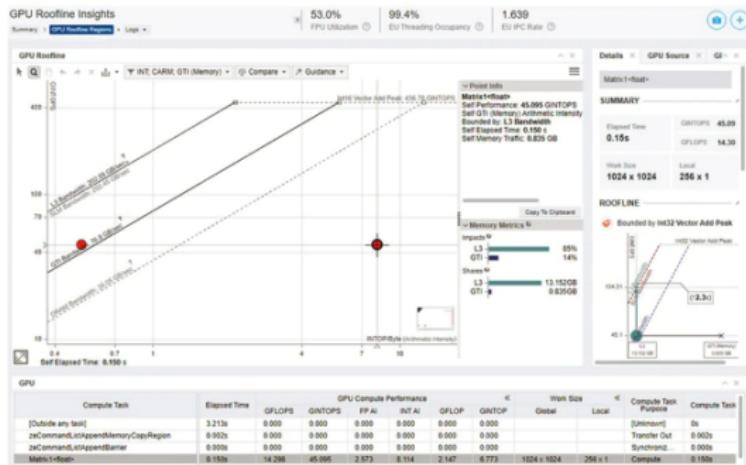
- Optimized libraries supported by compilers
 - oneDNN uses AMX, AVX-512, VNNI, and bfloat16 for acceleration in machine learning processes
- Optimized libraries
 - oneDNN supports AMX, AVX-512, VNNI, and bfloat16 to accelerate training in DL
 - oneVPL is compatible with AV1 hardware codec via Hardware (Intel® Data Center GPU Flex Series and Intel® Arc)
 - oneMKL improves portability and compatibility
 - OpenMP offload support and recent Intel® XMX repertoire support
 - Optimized operations in TF32, FP16, BF16, and INT8.
 - Interfaces for SYCL and C/Fortran OpenMP
 - Adds GPU synchronization support for BLAS



- Analysis tool
 - Supports C, C++, Fortran, SYCL, OpenMP, OpenCL, or Python code
- Facilitates efficient CPU code development
 - Monitors efficient threading, vectorization, and memory usage
- Possibility of efficient GPU offload suggestions
 - Identifies application parts that can be offloaded cost-effectively
 - Optimizes code for computation and memory.
- Flowchart design and analysis
 - Create, visualize, and analyze tasks and dependencies for heterogeneous algorithms



- Roofline model and performance on CPUs/GPUs
 - Practical advice for generating code on GPUs
 - Recommendations for CPU vs GPU usage (memory hierarchy)
 - User guide for GPUs



Intel® VTune™ Profiler

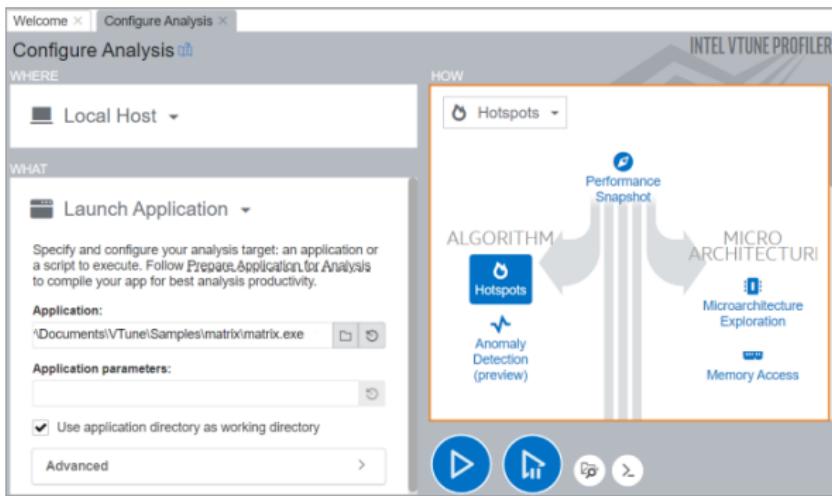
- Intel® VTune™ Profiler

- Tool for optimizing application performance, system performance, and system configuration for HPC, cloud, IoT, storage...
- CPU, GPU, and FPGA analysis
- Multi-language support: SYCL, C, C++, C#, Fortran, OpenCL, Python
- Tool to find out bottlenecks on several devices
- Several levels: system and/or application



Intel® VTune™ Profiler

- Various supported analyses
 - Performance Snapshot or overall performance
 - HPC characterization and Parallelism
 - Accelerators: GPUs and CPU/FPGAs interaction
 - Microarchitecture and Memory
 - I/O Monitoring

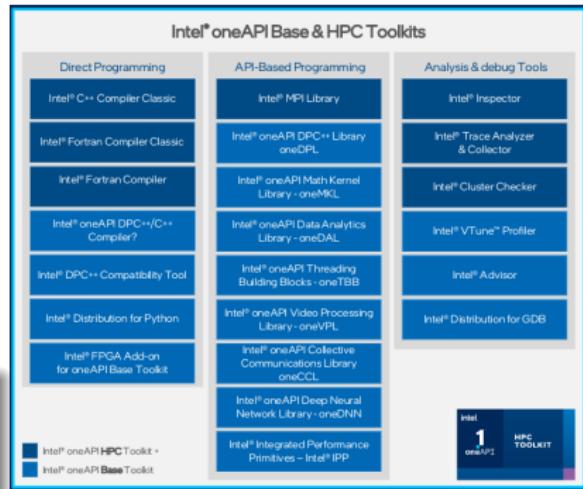


oneAPI HPC Toolkit

- Toolkit complementing the Intel® oneAPI Base Toolkit
- Scalable and high-performance parallel code in C++, Fortran, SYCL, OpenMP, and MPI (from enterprise to the cloud)

Features

- Acceleration and performance across the range of Intel Cores & Intel Accelerators
- Fast, scalable, and reliable parallel code with less effort based on industry standards



CGS

Intel® oneAPI, News and Advances 2024

Danysoft
Software Elite Reseller

August 26, 2024

19 / 32



Intel Developer
Cloud named
Tiber



New Tiber Developer Cloud

- Available [in the URL](#)



- Multiple configurations catering to various workloads
 - From AI training and inference
 - ... prototyping and evaluating the latest hardware using the environment that best suits your business needs...
- Learn with practical tutorials
 - Experiment with real-world code examples
 - Evaluate performance and acceleration with multiple hardware configurations.
 - Create heterogeneous applications



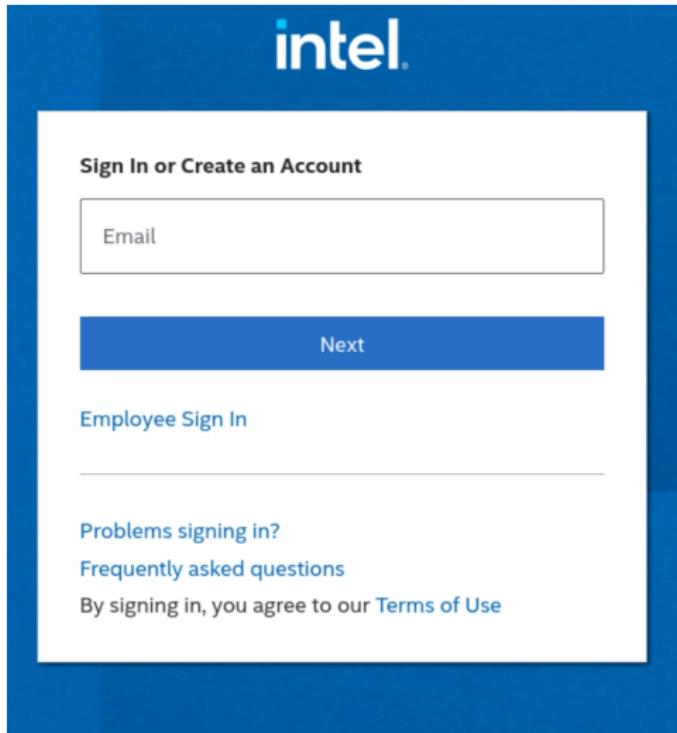
Available Hardware

- Test and evaluate a variety of virtual machines
 - Bare metal systems
 - Edge devices
 - Platforms for AI training
- Development environments
 - Containers
 - JupyterLabs
 - Direct SSH connection



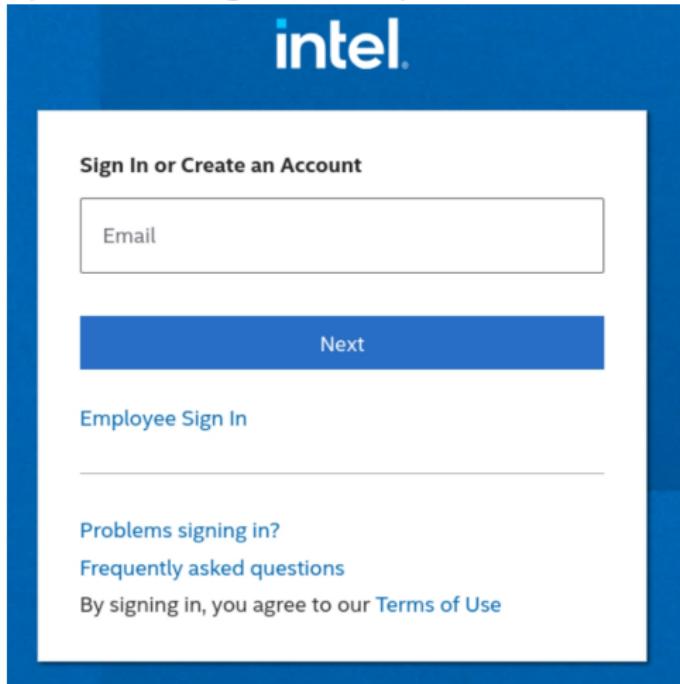
Access Instructions

- Documentation and updates available at <https://console.cloud.intel.com> with [Tutorials](#)



Register

- To have an account on Intel® Tiber Developer Cloud follow the link <http://cloud.intel.com>
- Follow the steps in the registration process:



Training

- Quickly launch a training node in the SLURM cluster.
 - **No SSH key** is required
- ① Visit the Intel® Tiber™ Developer Cloud [Console](#)
 - ② In the menu at left, click *Training*
 - ③ Click *Launch* to open a training notebook



Training C++ SYCL

- Essentials of SYCL (some examples will be shown in the workshop)
- Migrate from CUDA to C++ with SYCL (also shown in this workshop)
- Performance, Portability and Productivity
- Introduction to GPU



Example (I)

- Let's take an example SYCL code that returns the selected device, in this case, a GPU.

```
#include <sycl/sycl.hpp>
using namespace sycl;
int main() {
    // Create a device queue with device selector
    queue q(gpu_selector_v);
    // Print the device name
    std::cout << "Device: " << q.get_device().get_info<info::device::name>() << "\n";
    return 0;
}
```

- Compile it with the **icpx** compiler

```
uXXXX@idc-beta-batch-pvc-node-03:~$ icpx -o ex exampleSYCL.cpp -fsycl
```



Launch Instance

- More info in [Get Started](#)
 - Note: SSH key is required**
- Upload an SSH Key
 - Configure instance
 - Launch instance

The screenshot shows the intel liber: Developer Cloud interface. The top navigation bar includes links for Preview, Preview Catalog, Preview Instances, Preview Keys, and Documentation. A search bar is also present. The main content area is titled "Available hardware (14)" and features several categories:

- AIPC**: Shows a single item: "Intel® Core™ Ultra Processor Series I".
 - 1 socket, 16 cores, 64 GB memory, 512 GB disk
- CPU**: Shows four processor categories:
 - Intel® Xeon® processors, codenamed Sierra Forest (Scalable Processor)
 - 2 sockets, 192 cores, 1TB memory, 3TB disk
 - Intel® Xeon® processors, codenamed Siena Forest (Advanced Processor)
 - 2 sockets, 384 cores, 1.5 TB memory, 3 TB disk
 - 4th Generation Intel® Xeon® Scalable processors with TDX®
 - 2 sockets, 128 cores, 1TB memory, 1TB disk
 - 4th Gen Intel® Xeon® processors with high bandwidth memory (HBM)
 - 2 sockets, 192 cores, 1TB memory, 1TB disk

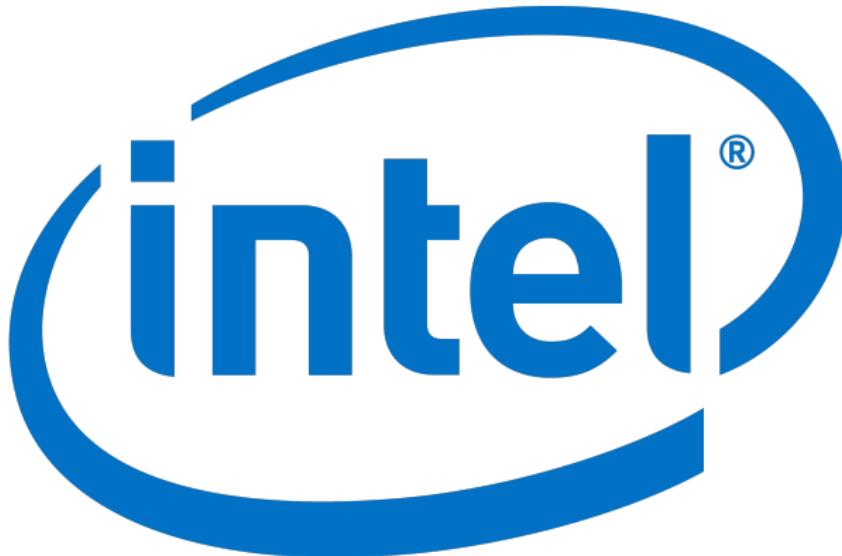




Available Resources

- Initiative [oneAPI](#)
- Intel oneAPI Base & HPC Toolkit
- Instructions for the Intel Tiber Developer Cloud access





Software



Thanks for your attention!!!



Avda. de la industria 4, edif. 1
28108 Alcobendas | Madrid | España



[+34] 91 663 8683



Email
info@danysoft.com



Website
www.danysoft.com/intel

