oneAPI

oneAPI

Multiarchitecture Programming for Accelerated Compute, Freedom of Choice for Hardware

oneAPI Industry Initiative & Intel® oneAPI Toolkits

Software & Advanced Technologies Group (SATG)
Software Products & Ecosystem
June 2023

Edmund Preiss ,
Intel Dev Tools - Business Dev Manager

intel.

All information provided in this deck is subject to change without notice.
Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Modern Applications Demand Increased Processing

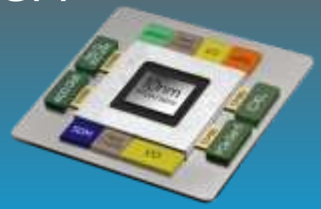
Diverse accelerators needed to meet today's performance requirements:

48% of developers target heterogeneous systems that use more than one kind of processor or core[1]

CPU

GPU

FPGA



Other Accelerators

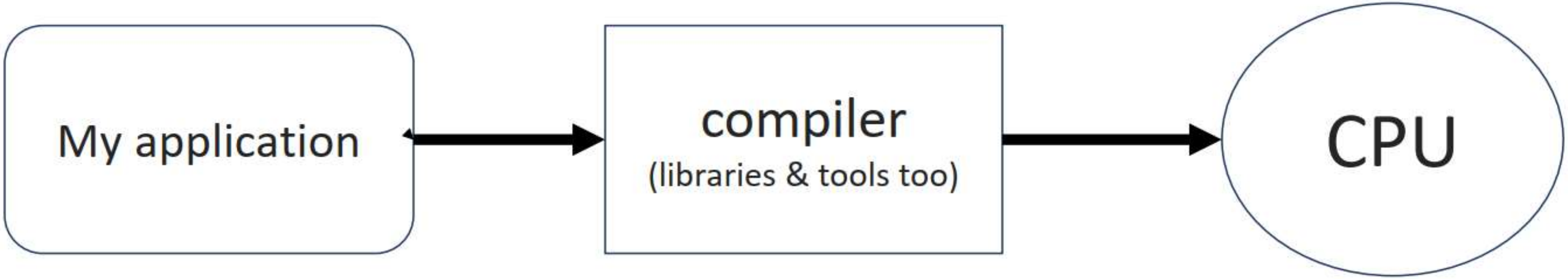
Developer Challenges: Multiple Architectures, Vendors, and Programming Models

oneAPI

Open, Standards-based, Multiarchitecture Programming

[1] *Evans Data Global Development Survey Report 22.1, June 2022*

Before heterogeneous systems

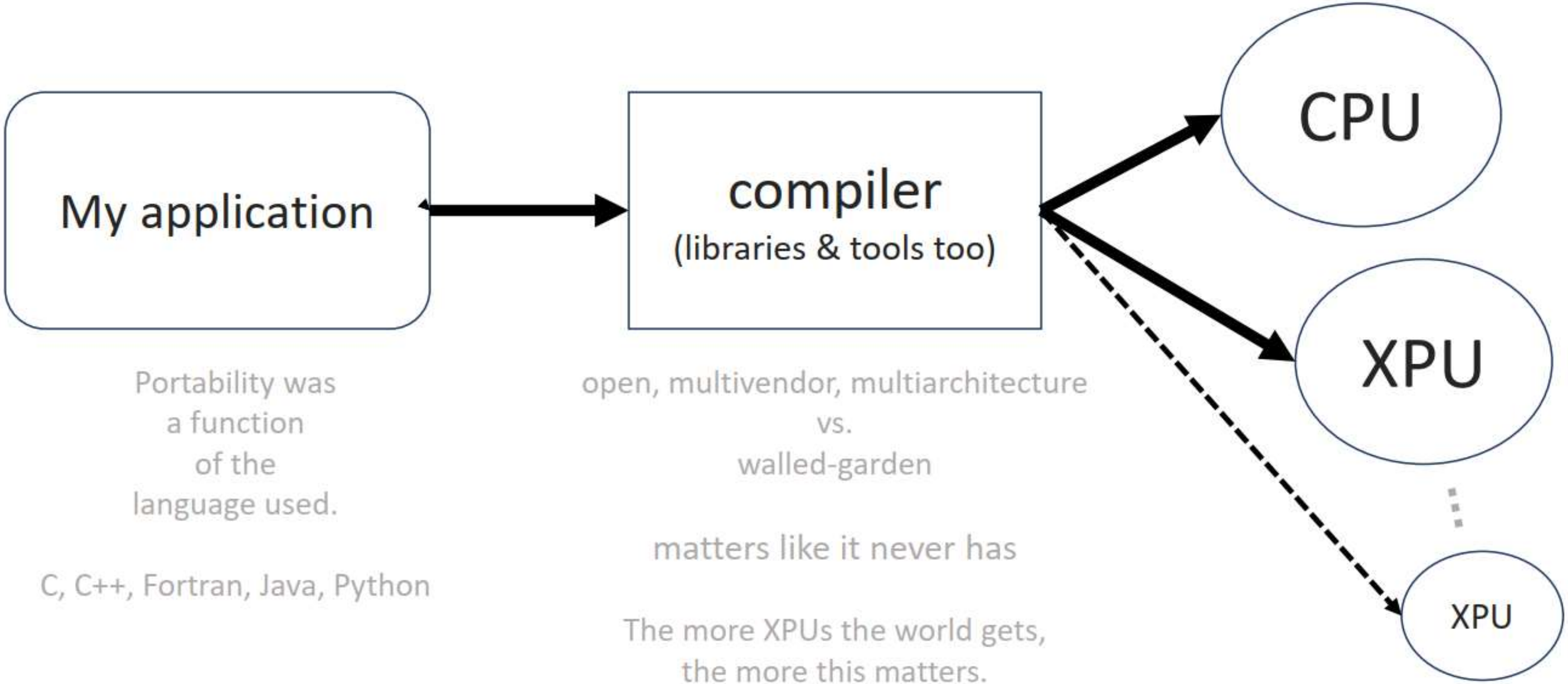


Portability was
a function
of the
language used.

C, C++, Fortran, Java, Python

I didn't care if
the compiler, etc.,
was proprietary or not –
since the target system was
single vendor, single architecture.

Now, with heterogeneous systems



My application

compiler

(libraries & tools too)

CPU

XPU

XPU

Portability was
a function
of the
language used.

C, C++, Fortran, Java, Python

open, multivendor, multiarchitecture
vs.
walled-garden

matters like it never has

The more XPUs the world gets,
the more this matters.

oneAPI Industry Initiative

Break the Chains of Proprietary Lock-in

Freedom to Make Your Best Choice

- C++ programming model for multiple architectures and vendors
- Cross-architecture code reuse for freedom from vendor lock-in

Realize all the Hardware Value

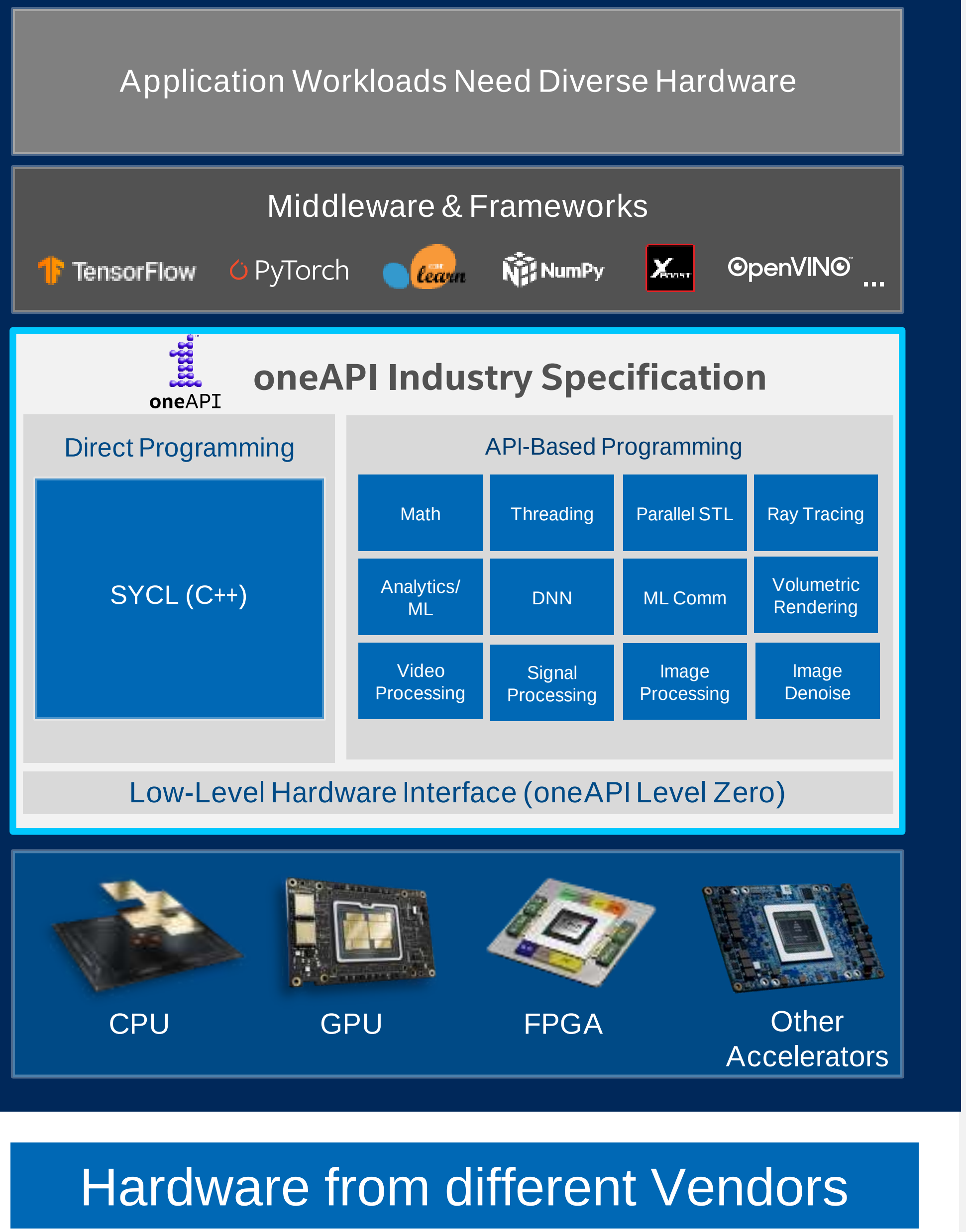
- Performance across CPU, GPUs, FPGAs, and other accelerators
- Expose and exploit cutting-edge features of the latest hardware

Develop & Deploy Software with Peace of Mind

- Open industry standards provide a safe, clear path to the future
- Interoperable with familiar languages and programming models including Fortran, Python, OpenMP, and MPI
- Powerful libraries for acceleration of domain-specific functions

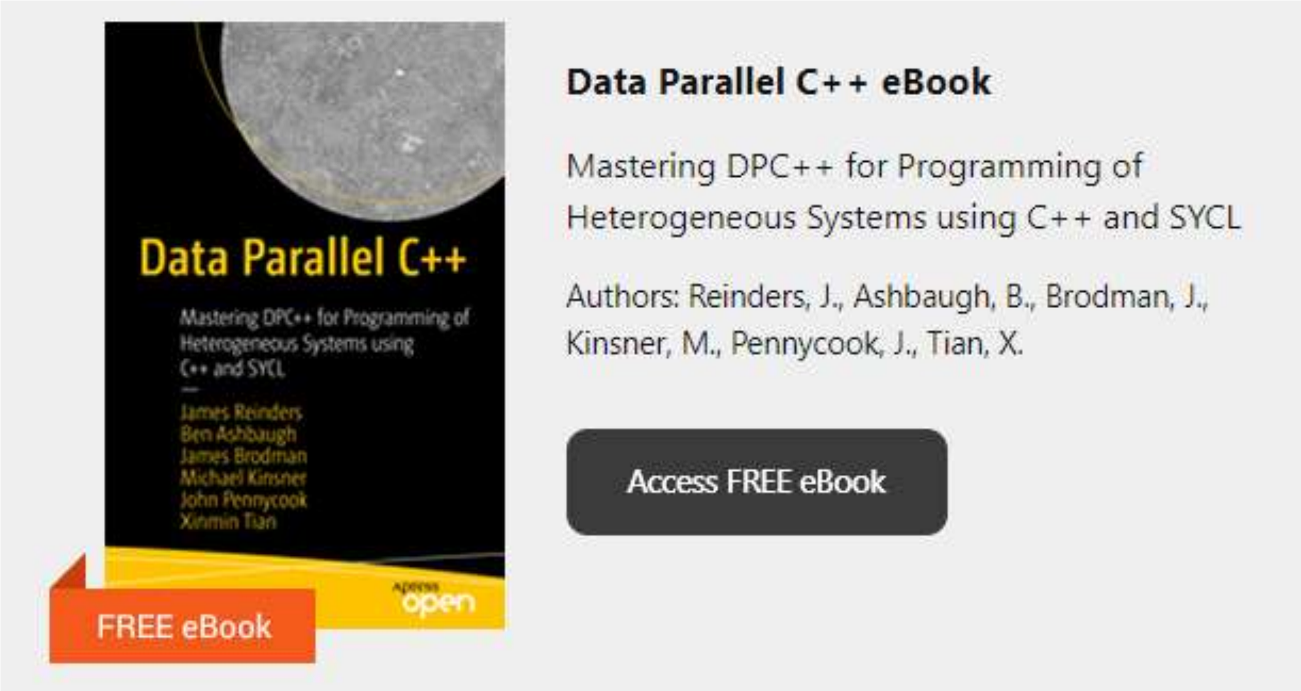
oneAPI

The productive, smart path to freedom for accelerated computing from the economic and technical burdens of proprietary programming models



Data Parallel C++

Standards-based, Most Comprehensive, Cross-architecture Implementation of SYCL



Data Parallel C++ eBook

Mastering DPC++ for Programming of Heterogeneous Systems using C++ and SYCL

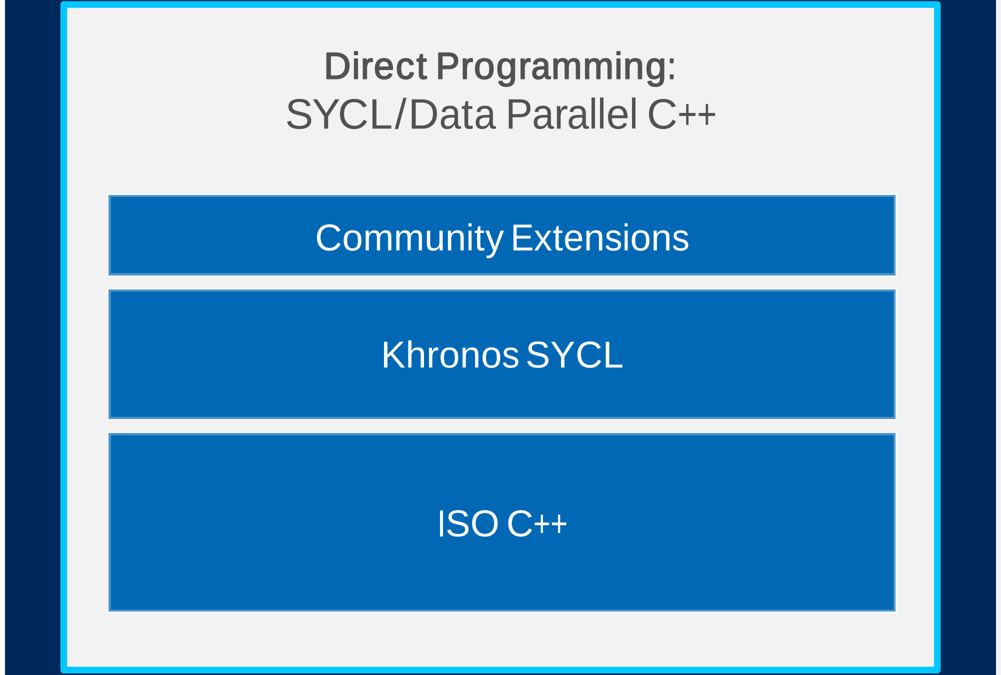
Authors: Reinders, J., Ashbaugh, B., Brodman, J., Kinsner, M., Pennycook, J., Tian, X.

Access FREE eBook

Or click here

ICX/DPC++ aims to be the best implementation of SYCL

oneAPI™



Direct Programming:
SYCL/Data Parallel C++

Community Extensions

Khronos SYCL

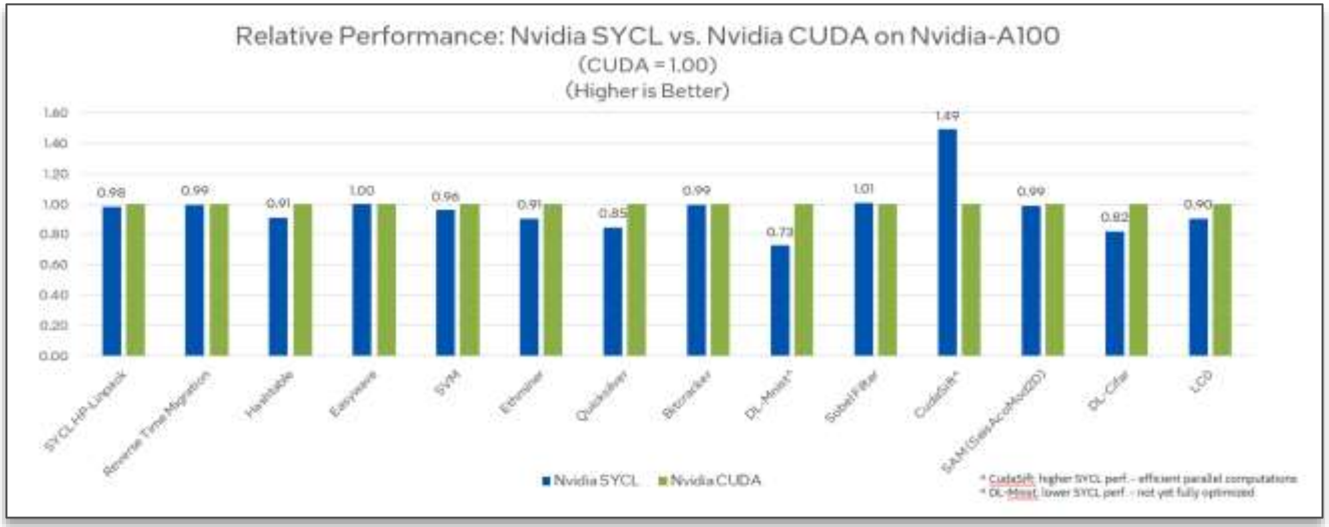
ISO C++

The open source and Intel DPC++/C++ compiler supports Intel CPUs, GPUs, and FPGAs.
Codeplay announced a DPC++ compiler that targets Nvidia GPUs.

Accelerating Choice with SYCL

Khronos Group Standard

- Open, standards-based
- Multiarchitecture performance
- Freedom from vendor lock-in
- Comparable performance to native CUDA on Nvidia GPUs
- Extension of widely used C++ language
- Speed code migration via open source SYCLomatic or Intel® DPC++ Compatibility Tool



Relative Performance: Nvidia SYCL vs. Nvidia CUDA on Nvidia-A100
(CUDA = 1.00)
(Higher is Better)

* CudaSift: higher SYCL perf. – efficient parallel computations
* DL-Mnist: lower SYCL perf. – not yet fully optimized

Architectures	Intel | Nvidia | AMD CPU/GPU | RISC-V | ARM Mali | PowerVR | Xilinx

Testing Date: Performance results are **based on testing by Intel as of April 15, 2023** and may not reflect all publicly available updates.

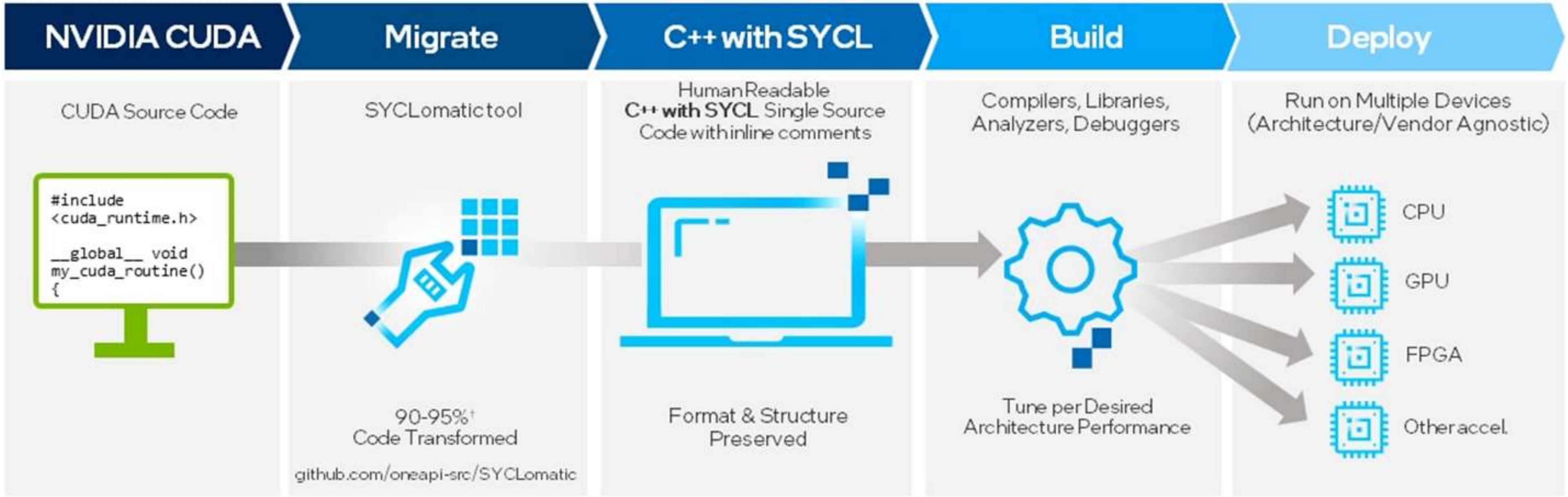
Configuration Details and Workload Setup: Intel® Xeon® Platinum 8360Y CPU @ 2.4GHz, 2 socket, Hyper Thread On, Turbo On, 256GB Hynix DDR4-3200, ucode 0xd000363. GPU: Nvidia A100 PCIe 80GB GPU memory. Software: SYCL open source/CLANG 17.0.0, CUDA SDK 12.0 with NVIDIA-NVCC 12.0.76, cuMath 12.0, cuDNN 12.0, Ubuntu 22.04.1. SYCL open source/CLANG compiler switches: -fscycl-targets=nvptx64-nvidia-cuda, NVIDIA NVCC compiler switches: -O3 –gencode arch=compute_80, code=sm_80. Represented workloads with Intel optimizations.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See configuration disclosure for details. No product or component can be absolutely secure.

Performance varies by use, configuration, and other factors. Learn more at www.Intel.com/PerformanceIndex. Your costs and results may vary.

SYCLomatic: CUDA* to SYCL* Migration Made Easy

Choose where to run your software, don't let the software choose for you.



Open source SYCLomatic tool assists developers migrating code written in CUDA to C++ with SYCL, generating **human readable** code wherever possible

~90-95% of code typically migrates automatically[1]

Inline comments are provided to help developers finish porting the application

Intel® DPC++ Compatibility Tool is Intel's implementation, available in the Intel® oneAPI Base Toolkit

github.com/oneapi-src/SYCLomatic

[1] Intel estimates as of March 2023. Based on measurements on a set of 85 HPC benchmarks and samples, with examples like Rodinia, SHOC, PENNANT. Results may vary.
*Other names and brands may be claimed as the property of others. SYCL is a trademark of the Khronos Group Inc.

Codeplay oneAPI Plug-ins for Nvidia* & AMD*

Support for Nvidia & AMD GPUs to Intel® oneAPI Base Toolkit

oneAPI for NVIDIA & AMD GPUs

- Free download of binary plugins to Intel® oneAPI DPC++/C++ Compiler:
- Nvidia GPU
- AMD beta GPU
- No need to build from source!
- Plug-ins updated quarterly in-sync with SYCL 2020 conformance & performance

Priority Support

- Available through Intel, Codeplay & our channel
- Requires Intel Priority Support for Intel® oneAPI DPC++/C++ Compiler
- Intel takes first call, Codeplay delivers backend support
- Codeplay provides access to older plug-in versions

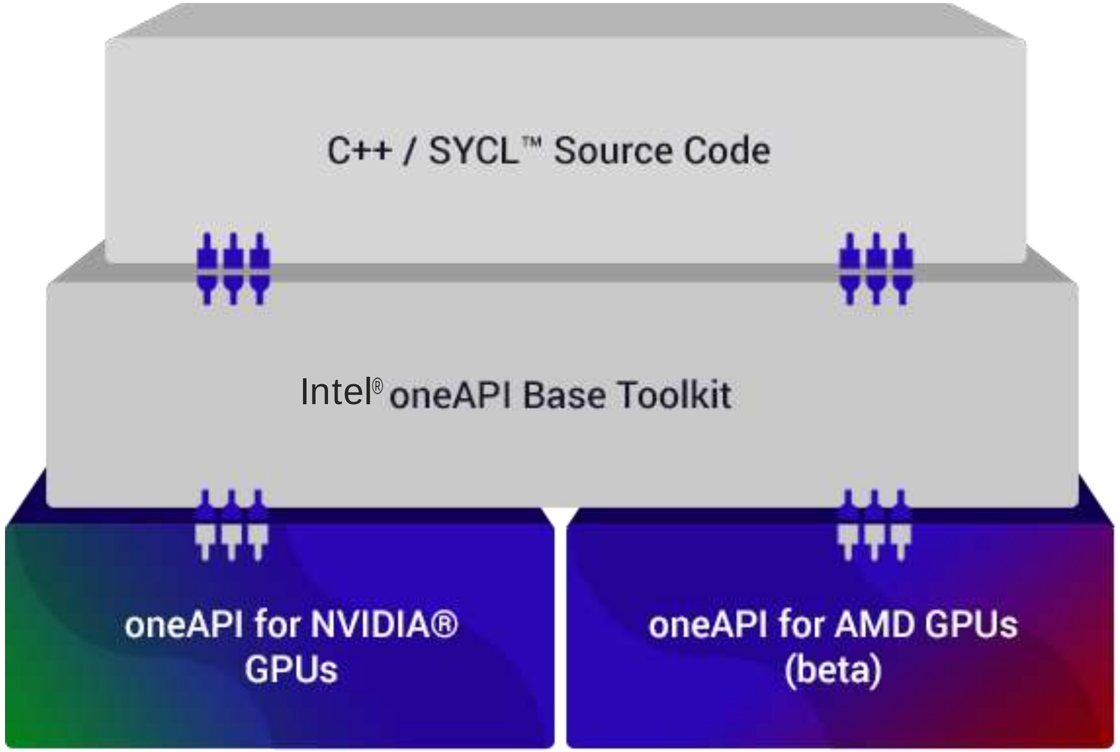


Image courtesy of Codeplay Software Ltd.

Nvidia GPU plug-in

AMD GPU plug-in

Codeplay blog

Codeplay press release

*Other names and brands may be claimed as the property of others.

oneAPI Industry Momentum

End Users

- Accrad Accelerated Radiology
- @rchanan
- FOUNDRY. MODO
- BioDataAnalysis GmbH
- GeoEast
- BENTLEY
- CINESITE
- LAIKA
- kt
- PHILIPS
- Brightskies
- AUTODESK ARNOLD
- 中国石油集团东方地球物理勘探有限责任公司 BGP INC. CHINA NATIONAL PETROLEUM CORPORATION
- iOmniscient
- EURECOM
- Verizon
- MediaKind
- 大势智慧 DASPATIAL
- WeBank
- SAMSUNG MEDISON
- TANGENT ANIMATION
- allegro.ai
- ILLUMINATION MACGUFF

National Labs

- Argonne National Laboratory
- UT-BATTELLE Oak Ridge National Laboratory
- CERN openlab
- CINECA
- Peraton Labs
- Laboratório Nacional de Computação Científica
- lrz Leibniz Supercomputing Centre
- SANKHYA SUTRA

ISVs & OSVs

- codeplay
- CANONICAL
- CERTI FACE
- ANACONDA
- Ansys
- MAX PLANCK COMPUTING & DATA FACILITY MPCDF
- AI SINGAPORE
- sas
- Red Hat OpenShift Data Science
- KATANA GRAPH
- CGG
- AIBLE
- CHAOSGROUP
- MAXON
- Ospirent
- FOUNDRY.
- mercenaries engineering Guerilla
- Hisense Medical 海信医疗
- SENAI CIMATEC FIEB
- SAP
- E4 COMPUTER ENGINEERING
- YUAN
- GIGASPACES innovate with confidence
- MEGH COMPUTING
- KFBIO
- Tech Mahindra
- AsiaInfo 亚信科技
- SUSE
- vmware
- UNITED IMAGING

OEMs & SIs

- BittWare a molex company
- Hewlett Packard Enterprise
- DELL Technologies
- Atos
- Lenovo
- HCL
- rENIAC
- MEGWARE

Universities & Research Institutes

- Lobachevsky University
- Universidad Complutense Madrid
- Technion Israel Institute of Technology
- University of Cambridge
- ICT 中国科学院计算技术研究所 Institute of Computing Technology, Chinese Academy of Sciences
- Ben-Gurion University of the Negev
- Berkeley University of California
- UC DAVIS University of California
- The University of Tennessee Knoxville
- Old Dominion University
- 北京大学软件与微电子学院 School of Software & Microelectronics
- University of Oregon
- Técnico Lisboa
- inesc id lisboa
- ZIB
- ILLINOIS
- The University of Utah
- Purdue University
- Faculty of Mathematics and Physics Charles University
- Indian Institutes of Technology Delhi / Kharagpur / Roorkee
- University College London
- Durham University
- University of Stuttgart Germany
- SDSC San Diego Supercomputer Center
- सी-डैक CDAC
- UKRI Science and Technology Facilities Council Scientific Computing
- TACC
- Indian Institute of Science Bangalore
- IISER PUNE Indian Institute of Science Education & Research Pune
- Universidad de Málaga
- URZ Heidelberg University Computing Centre
- NIU Northern Illinois University
- University of Bristol
- Stockholm University
- KTH

CSPs & Frameworks

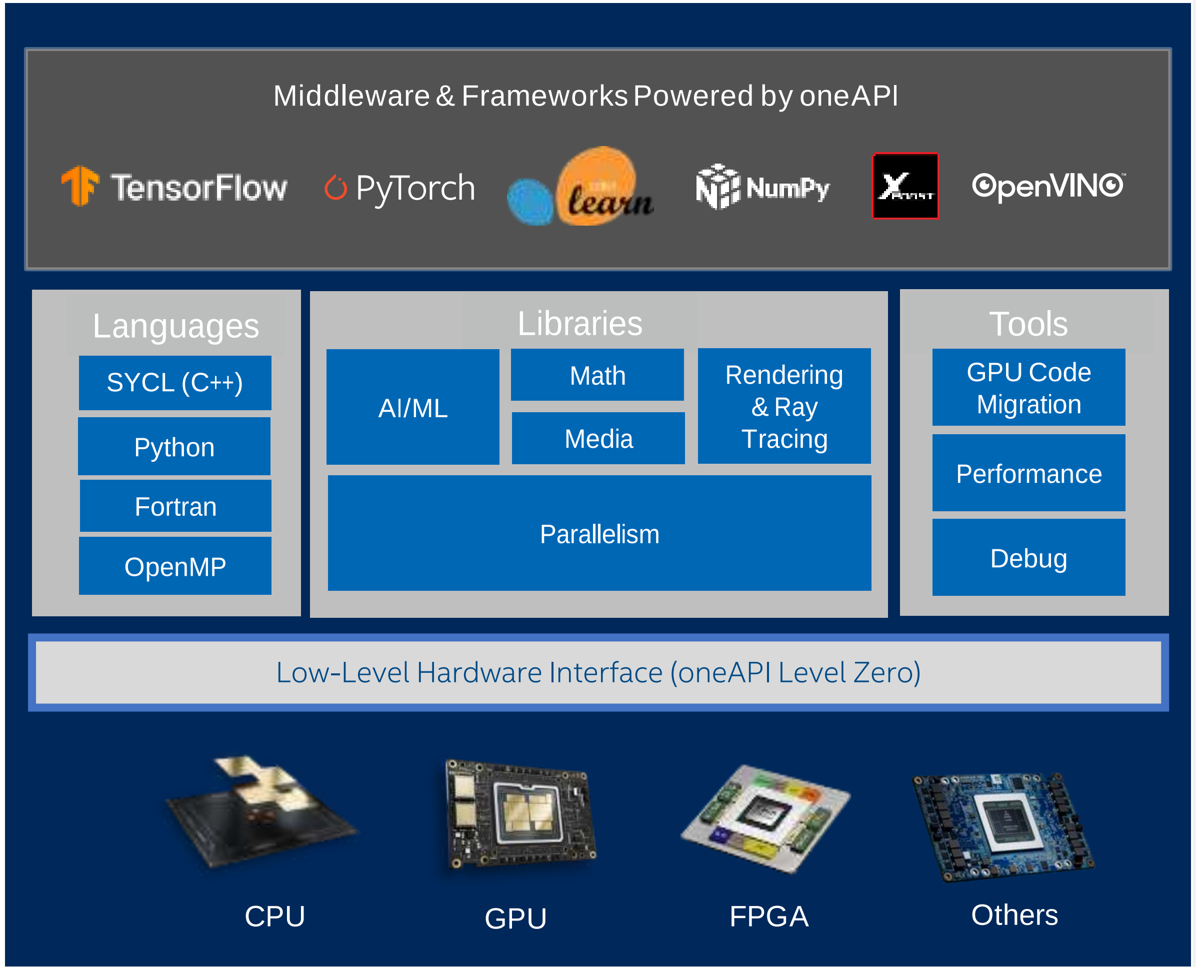
- Microsoft Azure
- Alibaba Cloud
- Google Cloud
- TensorFlow
- Taboola
- Tencent 腾讯
- DataRobot
- Bai百度
- 飞桨 PaddlePaddle
- NAVER CLOVA
- PyTorch

These organizations support the oneAPI initiative for a single, unified programming model for cross-architecture development.
It does not indicate any agreement to purchase or use of Intel's products. *Other names and brands may be claimed as the property of others.

Intel® Developer Tools Supporting oneAPI

A complete set of proven tools expanded from CPU to accelerators

- Advanced compilers, libraries, and analysis, debug, and porting tools
- Full support for C, C++ with SYCL, Python, Fortran, MPI, OpenMP
- Intel® Advisor determines device target mix before you write your code
- Intel's compilers optimize code to take full advantage of multiarchitecture workload distribution.
- Intel® VTune™ Profiler analyzes hotspots to optimize code performance
- Intel AI tools support acceleration of major deep learning and machine learning frameworks



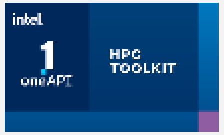
Middleware & Frameworks Powered by oneAPI

Intel® oneAPI Toolkits

Intel® oneAPI Base Toolkit

A core set of high-performance libraries and tools for building C++, SYCL, C/OpenMP, and Python applications

Add-on Domain-specific Toolkits



For HPC developers

Intel® oneAPI Tools for HPC
Deliver fast Fortran, OpenMP & MPI applications that scale

For visual creators, scientists & engineers

Intel® oneAPI Rendering Toolkit
Accelerate visual compute, deliver high-performance, high-fidelity visualization applications.

For edge & IoT developers

Intel® oneAPI Tools for IoT
Build efficient, reliable solutions that run at network's edge

Toolkits powered by oneAPI

For AI developers & data scientists

Intel® AI Analytics Toolkit
Accelerate machine learning & data science pipelines end-to-end with optimized DL & ML frameworks & high-performing Python libraries

OpenVINO™

For deep learning inference developers

Intel® OpenVINO™ toolkit
Deploy high performance inference & applications from edge to cloud

Download at intel.com/oneAPI
Or visit Intel® DevCloud for oneAPI

Latest version available 2023

Intel® oneAPI Tools for HPC

Intel® oneAPI HPC Toolkit

Deliver Fast Applications that Scale

What is it?

A toolkit that adds to the Intel® oneAPI Base Toolkit for building high-performance, scalable parallel code on C++, Fortran, SYCL, OpenMP & MPI from enterprise to cloud, and HPC to AI applications.

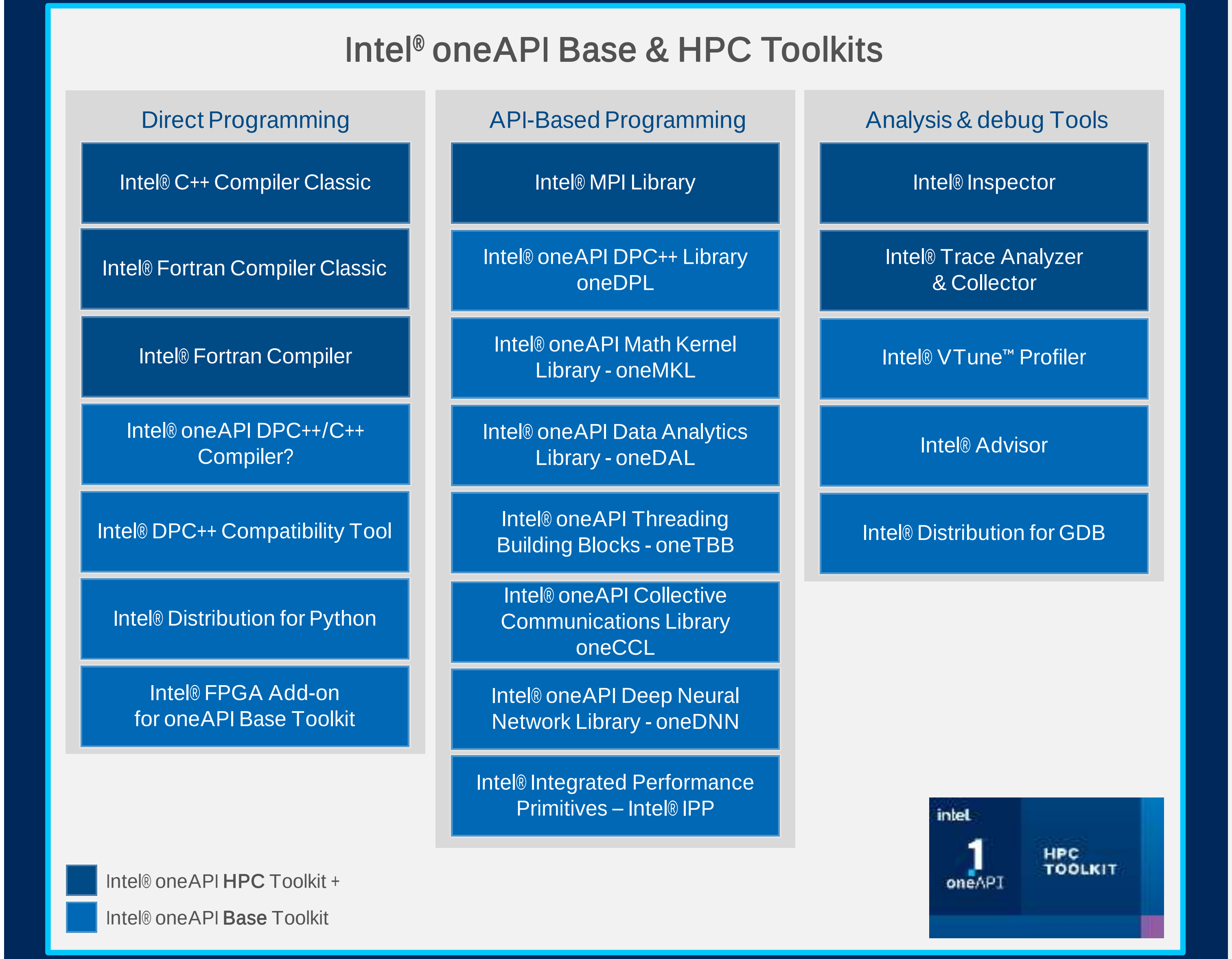
Who needs this product?

- OEMs/ISVs
- C++, Fortran, OpenMP, MPI Developers

Why is this important?

- Accelerate performance on Intel® Xeon® & Core™ processors & Intel accelerators
- Deliver fast, scalable, reliable parallel code with less effort built on industry standards

Learn More & Download



Intel® oneAPI Base & HPC Toolkits

Direct Programming

- Intel® C++ Compiler Classic
- Intel® Fortran Compiler Classic
- Intel® Fortran Compiler
- Intel® oneAPI DPC++/C++ Compiler?
- Intel® DPC++ Compatibility Tool
- Intel® Distribution for Python
- Intel® FPGA Add-on for oneAPI Base Toolkit

API-Based Programming

- Intel® MPI Library
- Intel® oneAPI DPC++ Library oneDPL
- Intel® oneAPI Math Kernel Library - oneMKL
- Intel® oneAPI Data Analytics Library - oneDAL
- Intel® oneAPI Threading Building Blocks - oneTBB
- Intel® oneAPI Collective Communications Library oneCCL
- Intel® oneAPI Deep Neural Network Library - oneDNN
- Intel® Integrated Performance Primitives – Intel® IPP

Analysis & debug Tools

- Intel® Inspector
- Intel® Trace Analyzer & Collector
- Intel® VTune™ Profiler
- Intel® Advisor
- Intel® Distribution for GDB

■ Intel® oneAPI **HPC** Toolkit +

■ Intel® oneAPI **Base** Toolkit

Back to Domain-specific Toolkits for Specialized Workloads

Intel® Developer Tools 2023.1 – Highlights

Optimized, Standards-based Support for Powerful New Architectures

Optimized support for Intel's portfolio of CPU and GPU architectures

4th Gen Intel® Xeon® Scalable Processors with Intel® Advanced Matrix Extensions, Quick assist Technology, Intel® AVX-512, bfloat16, and more built-in accelerators

Intel® Xeon® Max Series CPUs with high-bandwidth memory

Intel® Data Center GPUs, including Flex Series with hardware AV1 encode and Max Series with datatype flexibility, Intel® X^e Matrix Extensions, vector engine, XE-Link, and other features

Compilers & SYCL Support

Intel® oneAPI DPC++/C++ Compiler

Bfloat16 is now a full feature (vs. experimental) to accelerate machine learning algorithms (especially deep learning training on the latest platforms.

Adds auto cpu_dispatch, kernel properties for SYCL*, more SYCL 2020 features to improve developer productivity

The compiler enhances developer productivity with auto cpu_dispatch, kernel properties for SYCL*, more SYCL 2020 features, and better compiler error handling for SYCL and OpenMP code.

The Intel® oneAPI DPC++ Library has improved performance of the sort, scan, and reduce algorithms.

Intel® DPC++ Compatibility Tool (based on open source SYCLomatic) supports the latest release of CUDA headers and adds more CUDA APIs to the equivalent SYCL language and oneAPI library functions including runtime, math, and neural network domains.

Intel® Fortran Compiler enhances OpenMP 5.0, 5.1 compliance and improves performance.

Performance Libraries

Intel® oneAPI Math Kernel Library improves Intel® Data Center GPU Max Series performance via new real FFTs plus 1D and 2D optimizations, new random number generators, as well as additional Sparse BLAS and LAPACK inverse optimizations for Cholesky, triangle matrix, and batch LU routines.

Intel® MPI Library improves performance for collectives using GPU buffers and through default process pinning on CPUs with E-cores and P-cores

Intel® Integrated Performance Primitives - Cryptography:

Expands offerings with CCM/GCM modes, enabling Crypto Multi-Buffer to provide greater performance benefit when compared to scalar implementation

Eliminates need for redundant buffer, increases efficiency, ease of use for API and adoptability; with new bug fixes for CBC/CFB modes for SM4 algorithm

Adds support for asymmetric cryptographic algorithm SM2, for key exchange protocol and encryption/decryption APIs. Other bug fixes and security enhancements.

oneAPI Commercial Support Available

Priority Support for Intel® oneAPI Toolkits

Every paid version of Intel® oneAPI Developer Toolkits includes Priority Support for that toolkit (Intel oneAPI Base, HPC, IOT, & Rendering Toolkits)

- Direct, private interaction with Intel software support engineers
- Accelerated response time
- Access to—and support for—previous Intel products such as Fortran compiler versions, previous toolkit versions, and more
- Intel Technical Consulting Engineers for on-site or online training and consultation at a reduced cost

oneAPI Resources

software.intel.com/oneapi

Get Started

- software.intel.com/oneapi
- Documentation + dev guides
- Code Samples
- Intel® Developer Cloud

Industry Initiative

- oneAPI.io
- oneAPI open Industry Specification
- Open-source Implementations

Learn

- Training: Webinars & courses
- Intel® DevMesh Innovator Projects
- Summits & Workshops: Live & on-demand virtual workshops, community-led sessions
- Training by certified oneAPI experts worldwide for HPC & AI

Ecosystem

- Community Forums
- Intel® DevMesh Innovator Projects
- Academic Programs: oneAPI Centers of Excellence: research, enabling code, curriculum, teaching

Useful resources to further your learning

TechDecoded – Technical Articles and Tutorials
https://www.intel.com/content/www/us/en/developer/tools/oneapi/tech-articles-how-to/overview.html#gs.zur2dq

Featured Workflows https://www.intel.com/content/www/us/en/developer/tools/oneapi/training/overview.html

Training Catalogue
https://www.intel.com/content/www/us/en/developer/tools/oneapi/training/catalog.html?f:@stm_10184_en=%5BIntel%C2%AE%20oneAPI%20HPC%20Toolkit%5D

The Parallel Universe Magazine

https://www.intel.com/content/www/us/en/developer/community/parallel-universe-magazine/overview.html?wapkw=Parallel%20Universe%20Magazine#gs.zurmaq

intel software

THE PARALLEL UNIVERSE

Supply Chain Optimization at Enterprise Scale

Optimizing Transformer Model Inference on Intel® Processors

Optimize Utility Maintenance Prediction for Better Service

Issue 52 2023

Maximize Your Performance

With Intel Developer Tools & Hardware Platforms

oneAPI

HPC & Data Center

intel oneAPI BASE TOOLKIT

intel oneAPI HPC TOOLKIT

intel XEON

intel DATA CENTER GPU FLEX SERIES

intel DATA CENTER GPU MAX SERIES

intel AGILEX

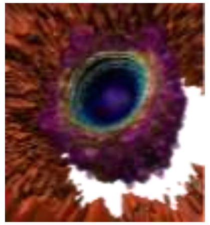
Performance

- Optimize compute performance on the latest Intel CPUs, GPUs and FPGAs
- Maximize built-in accelerators
- Accelerate across AI frameworks

AI & Visualization

intel AI ANALYTICS TOOLKIT

intel oneAPI RENDERING TOOLKIT



intel XEON

intel DATA CENTER GPU FLEX SERIES

intel DATA CENTER GPU MAX SERIES

intel AGILEX

Productivity

- Familiar languages and standards
- Easily integrate w/ legacy code
- Easily migrate CUDA to SYCL
- Minimize code re-writes

Embedded Systems & IoT

intel oneAPI IoT TOOLKIT

intel XEON

intel ATOM

intel CORE i3

intel CYCLONE

Freedom

- Open alternative to proprietary lock-in
- Enables easy architecture retargeting
- Code longevity for future hardware

Notices & Disclaimers

Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex. Results may vary.

Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details.
No product or component can be absolutely secure.

Texas Advanced Computing Center (TACC) Frontera references
Article: *HPCWire: Visualization & Filesystem Use Cases Show Value of Large Memory Fat Notes on Frontera*.
www.intel.com/content/dam/support/us/en/documents/memory-and-storage/data-center-persistent-mem/Intel-Optane-DC-Persistent-Memory-Quick-Start-Guide.pdf
software.intel.com/content/www/us/en/develop/articles/introduction-to-programming-with-persistent-memory-from-intel.html
wreda.github.io/papers/assise-osdi20.pdf

KFBIO
KFBIO m. tuberculosis screening detectron2 model throughput performance on 2nd Intel® Xeon® Gold 6252 processor: NEW: Test 1 (single instance with PyTorch 1.6: Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel® Xeon® Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated Test 2 (24 instances with PyTorch 1.6: Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel Xeon Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated BASELINE: (single instance with PyTorch 1.4): Tested by Intel as of 5/22/2020. 2-socket 2nd Gen Intel Xeon Gold 6252 Processor, 24 cores, HT On, Turbo ON, Total Memory 192 GB (12 slots/16 GB/2666 MHz), BIOS: SSE5C620.86B.02.01.0008.031920191559 (ucode: 0x500002c), Ubuntu 18.04.4 LTS, kernel 5.3.0-51-generic, mitigated.

Tangent Studios
Configurations for Render Times with Intel® Embree, testing conducted by Tangent Animation Labs. Render farm: 8x Intel® Core™ processors +hyperthread*2 + 128gig. In-office workstations: Intel® Xeon® processors HP blade c7000 chassis, with HP460 gen8 blades - 2x Intel Xeon E5-2650 V2, Eight Core 2.6GHz-128GB. Software: Blender 2.78 with custom build using Intel® Embree. For more information on Tangent's work with Embree, watch this video:
www.youtube.com/watch?time_continue=251&v=_2Ia4h8q3xs&feature=emb_logo
Recreation of the performance numbers can be recreated using Agent327, Blender and Embree.

Chaos Group - Up to 90% Memory Reduction for Displacement
Testing conducted by Chaos Group with Intel® Embree 2020. Software Corona Renderer 5 with Intel Embree. Up to 90% memory reduction calculated using Corona Renderer 5 with regular displacement grids per triangle of 154 bytes versus Corona Renderer 5 with Intel Embree, which has a displacement capability grid of 12 bytes per grid triangle. (12/154 = 7.8% usage or >90% memory reduction.) Recreation of the performance numbers can be accomplished using Corona Renderer 5 and Embree. For more information, visit the Corona Renderer Blog: blog.corona-renderer.com/corona-renderer-5-for-3ds-max-released/

The Addams Family 2 - **Gained a 10% to 20%—and sometimes 25%—efficiency in rendering, saving thousands of hours in rendering production time**.
Testing Date: Results are based on data conducted by Cinesite 2020-21. 10% to up to 25% rendering efficiency/thousands of hours saved in rendering production time/15 hrs per frame per shot to 12-13 hrs.
Cinesite Configuration: 18-core Intel® Xeon® Scalable processors (W-2295) used in render farm, 2nd gen Intel Xeon processor-based workstations (W-2135 and -2195) used. Rendering tools: Gaffer, Arnold, along with optimizations by Intel® Open Image Denoise.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

intel®