OpenCL Working Group Update IWOCL 2024

Kevin Petit, Arm

IWOCL 2024
APRIL 8-11 | CHICAGO, USA
12th International Workshop on OpenCL and SYCL
OpenCL
SYCL

Khronos Connects Software to Silicon

Close to 200 members worldwide
Any organization is welcome to join

KHRONOS GROUP
AMD
arm
EPIC GAMES
Google
HUAWEI
IKEA
Imagination
intel
NVIDIA
Qualcomm
SAMSUNG
SONY
VALVE
VeriSilicon
3dMD
7INVENSUN
acer
adidas
Adobe
AMOTIVE
ALL3D
Alibaba.com
amazon.com
Ambarella
AGI
Argonne
AUTODESK
avataar
BADVISOR
BASEMARK
BINOMIAL
BLIZZARD
blender
BOLT GRAPHICS
BOSCH Invented for life
BRENWILL
BROADCOM
ByteDance
Canon
CESIUM
codeweavers
Collins Aerospace
Continental
CoreAVI
DGG
DASSAULT SYSTEMES
DENSO Crafting the Core
DMP
dopple
DoubleMe
EA
emersya
ERICSSON
ETHEREAL ENGINE
FORGE
Fraunhofer
GLENFLY
HITACHI Inspire the Next
HOLOCHIP
htc
id
igalia
Imperial College London
INNOSILICON
Inseye
JUICE
KDAB
KITTYCAD
KNU
Leia Inc.
Linaro
LW
Loft3Di
London Dynamics
Los Alamos
logitech
LUNARG
Magic Leap
matrox
MEDIATEK
Mercedes-Benz
Meta
METAVRSE
METAX
Microsoft
MIFLY
mobica
MOORE THREADS
moz://a
National Institute of Standards and Technology
NaturalPoint
NETFLIX
NIHON UNIVERSITY
Nintendo
NOKIA
NXP
oppo
OSU Oregon State
oloy
OVR
OXIDE
Penumbra
PHASMATIC
PLAY FOR DREAM
pluto
ptc
RASTERGRID
RAZER
Red Hat
RENESAS
RIVE
Rokid
SEOUL NATIONAL UNIVERSITY
SIDEQUEST
socionext
STREAM
StretchSense
SuperDNA 3D Lab
Synaptics
SYNOPSYS
TAKUMI
TAMPERE UNIVERSITY OF TECHNOLOGY
Tencent 腾讯
TEXAS INSTRUMENTS
Think Silicon
TRAVERSE RESEARCH
THREEDY instant 3Dhub
ThreeKit
tobii
ultraleap
Unity
University of BRISTOL
ILLINOIS
universität innsbruck
UNIVERSITÄT DES SAARLANDES
TOHOKU
UX3D
VARJO
VIRTUAL DIMENSION CENTER
vivo
VNTANA
Wearable Devices
XE3D
XIANGDIXIAN
XREAL
ZUSE INSTITUTE BERLIN

Liaisons: Cooperation with industry associations and organizations

AOUSD
AREA
AUTOSAR
emva
ETSI
GTI
HAPTICS INDUSTRY FORUM
ISO JTC1 IEC
Open Geospatial Consortium
Smithsonian
TIFCA
VESA
W3C
VRM CONSORTIUM
web3D CONSORTIUM

Founded in 2000
~ 200 Members | ~ 40% US, 30% Europe, 30% Asia

KHRONOS GROUP

Open, royalty-free interoperability standards to harness the power of GPU, XR and multiprocessor hardware

3D graphics, augmented and virtual reality, parallel programming, inferencing and vision acceleration

Non-profit, member-driven standards organization, open to any company

Proven multi-company governance and Intellectual Property Framework

Khronos Compute Acceleration Standards

Higher-level Languages and APIs

Streamlined development and performance portability

SYCL SC / **SYCL**

Single source C++ programming with compute acceleration

NNEF →

Neural Network Exchange Format Trained Networks

OpenVX

Graph-based vision and inferencing acceleration

gstreamer **OpenCV** **FFmpeg**

Third party vision, streaming and inferencing libraries

Applications, libraries, and higher-level languages and APIs can use lower-level Khronos standards to access hardware acceleration

Lower-level Languages and APIs

Explicit hardware control

Multiple programming abstractions to meet the needs of diverse software stack architectures

Vulkan / **Vulkan SC**

GPU rendering + compute acceleration

↓

GPU

← Shaders

SPIR

Intermediate Representation (IR) language compiler target supporting parallel execution and graphics

Kernels →

OpenCL

Heterogeneous compute acceleration

↓

CPU | GPU

FPGA | DSP

AI/Tensor HW

Custom Hardware

OpenCL Complements Vulkan

Not just GPU acceleration
Simpler programming model
Relatively lightweight run-time
More language flexibility, e.g., pointers
Rigorously defined numeric precision
Framework for connecting custom processors

KHRONOS GROUP

Apps, Libraries and Engines using OpenCL

The industry’s most pervasive, cross-vendor, open standard for low-level heterogeneous parallel programming
https://en.wikipedia.org/wiki/List_of_OpenCL_applications

Desktop Creative Apps

F. Adobe otoy RED DIGITAL CINEMA REDCINE-X PRO acdsee DASSAULT SYSTEMES Modo SONY SideFX darktable AUTODESK CHAOSGROUP Vegas Pro Capture One SILHOUETTE CyberLink ptc GIMP ArcSoft RADEON PRORENDER Blackmagicdesign REALFLOW LuxCoreRender OPEN SOURCE PHYSICALLY BASED RENDERER

Parallel Languages

SYCL OpenACC DIRECTIVES FOR ACCELERATORS aparapi PyOpenCL

Machine Learning Libraries and Frameworks

Intel OPENVINO™ Intel clDNN Xiaomi MACE Mobile AI Compute Engine portDNN Caffe NNAPI Acuity ML/AI Software Foundation VeriSilicon Qualcomm Neural Processing SDK for AI Synopsis MetaWare EV TI DL Library (TIDL) Arm Compute Library 飞桨 PaddlePaddle MNN Mobile Neural Network

Molecular Modelling Libraries

siremol.org FOLDING @HOME ForceBalance OpenMM CHARMM Chemistry at HARvard Macromolecular Mechanics GROMACS FAST. FLEXIBLE. FREE.

Machine Learning Compilers

GLOW plaidML tvm.ai

Vision, Imaging and Video Libraries

FAST Framework for Heterogeneous Medical Image Computing and Visualization OpenCV ImageMagick Halide FFmpeg Metashape OpenVX™

Math and Physics Libraries

BULLET PHYSICS LIBRARY GNU Octave Wolfram Mathematica ArrayFire C, C++, Fortran Matlab

Linear Algebra Libraries

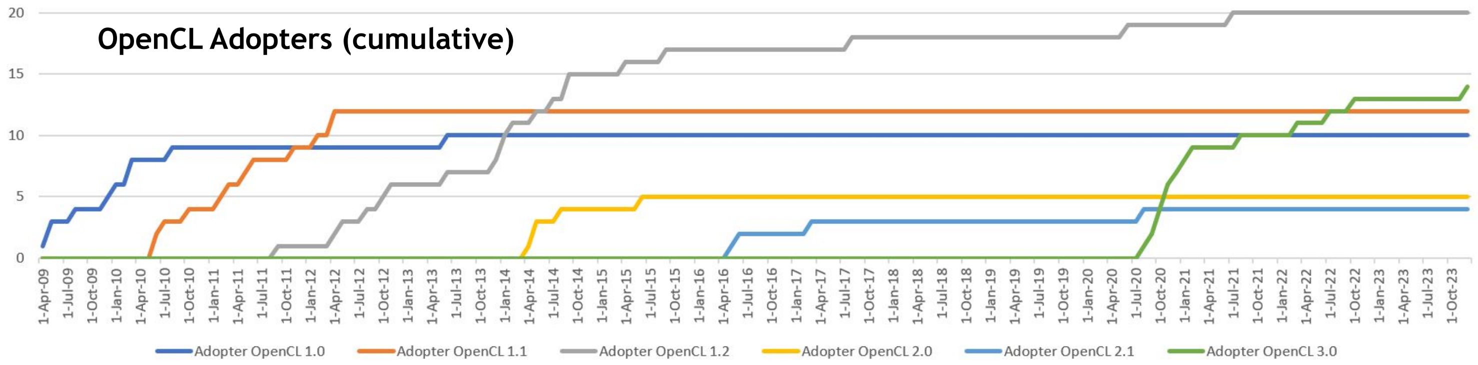
portBLAS ViennaCL CLBlast

OpenCL State-of-the Union

- **OpenCL 3.0 adoption is strong and growing**
  - 14 OpenCL 3.0 Adopters, second only to OpenCL 1.2 (Vulkan 1.3 has 13 Adopters)
- **Significant open-source activity**
  - Mesa Rusticl for Linux
  - clang/LLVM compilation front-ends
  - Layered implementations clspv and Ancle over Vulkan, OpenCLon12 over DX12
- **OpenCL is a popular substrate layer for higher-level models, especially SYCL**
  - The second most common offload path, after CUDA, but before SYCL, Vulkan, HIP
- **Emerging acceptance of OpenCL as compute layer over Vulkan**
  - Especially for ML, simpler programming model, more language flexibility, e.g., pointers
  - First conformant layered OpenCL 3.0 implementation
- **Regular (roughly) quarterly Releases with new unified specification format!**
  - 3.0.16 is released for IWOCL 2024 with External Memory and Semaphores finalized
- **Active extension pipeline - driven by mobile, embedded and desktop markets**
  - Recordable Command Buffers, Cooperative Matrix, Unified Shared Memory, YUV Images, Tiling Controls…

IWOCL 2024
APRIL 8-11 | CHICAGO, USA
12th International Workshop on OpenCL and SYCL
OpenCL
SYCL

OpenCL 3.0 Adoption



Currently 14 OpenCL 3.0 Adopters, 9 already submitted conformant products - ***second only to OpenCL 1.2***

https://www.khronos.org/conformance/adopters/conformant-products/opencl

arm ✓ codeplay® Google ✓ HUAWEI Imagination ✓ intel ✓ MESA ✓ Microsoft

NVIDIA ✓ QNX Qualcomm ✓ SAMSUNG ✓ VeriSilicon ✓ Tampere University

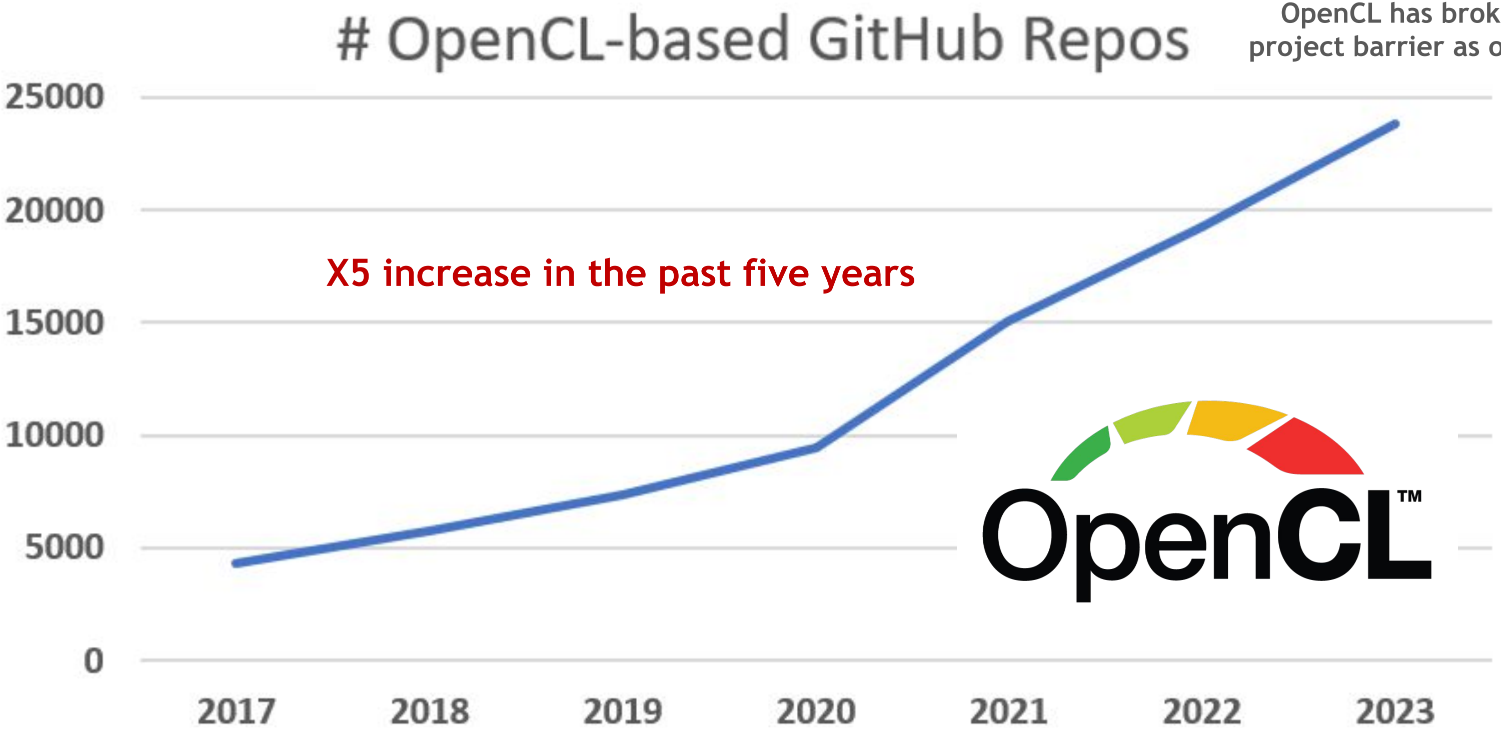
✓ Shipping OpenCL 3.0 Conformant Implementations

AMD KALRAY MARVELL MEDIATEK ST life.augmented Texas Instruments

Adopters of previous OpenCL Versions

KHRONOS® GROUP

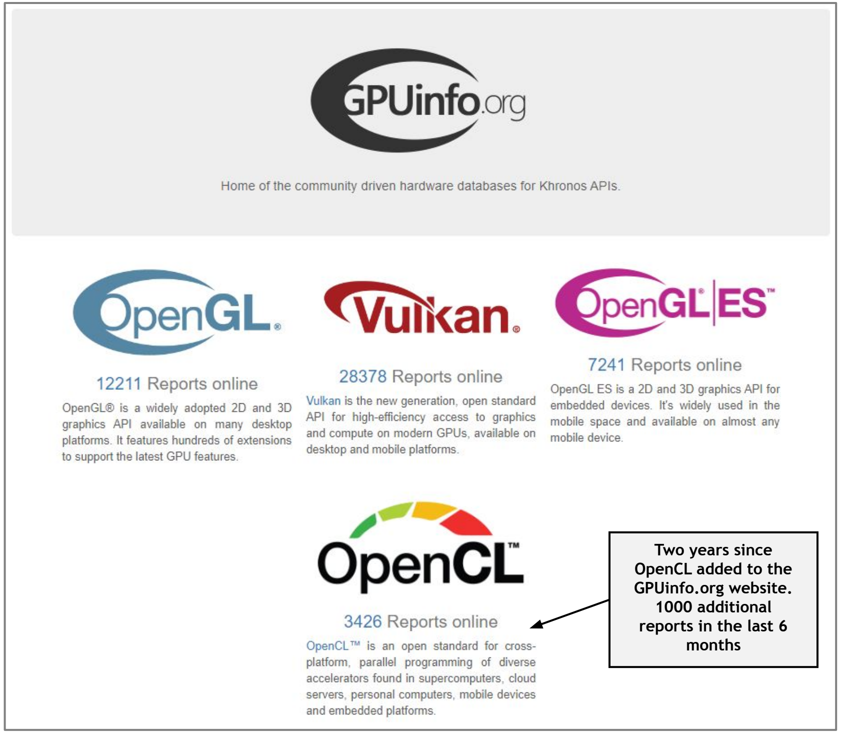
OpenCL Open-Source Project Momentum



X5 increase in the past five years

OpenCL has broken the 25K project barrier as of March 2024

OpenCL on GPUInfo.org



Home of the community driven hardware databases for Khronos APIs.

12211 Reports online

OpenGL® is a widely adopted 2D and 3D graphics API available on many desktop platforms. It features hundreds of extensions to support the latest GPU features.

28378 Reports online

Vulkan is the new generation, open standard API for high-efficiency access to graphics and compute on modern GPUs, available on desktop and mobile platforms.

7241 Reports online

OpenGL ES is a 2D and 3D graphics API for embedded devices. It's widely used in the mobile space and available on almost any mobile device.

3426 Reports online

OpenCL™ is an open standard for cross-platform, parallel programming of diverse accelerators found in supercomputers, cloud servers, personal computers, mobile devices and embedded platforms.

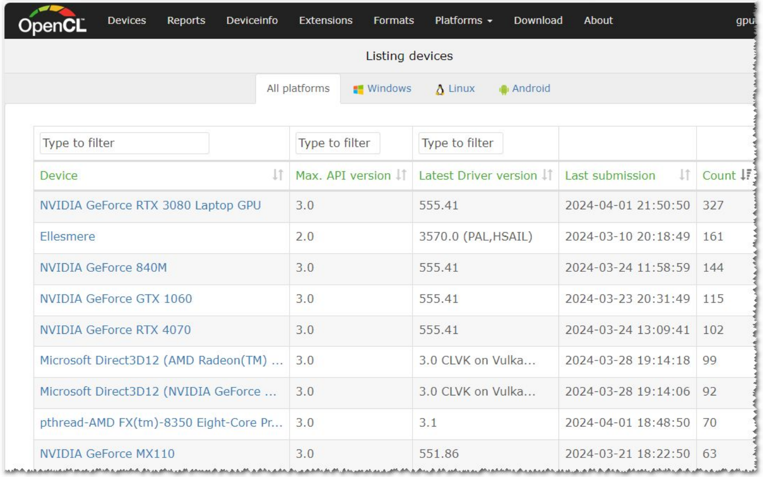
Two years since OpenCL added to the GPUinfo.org website. 1000 additional reports in the last 6 months

The online GPUinfo.org database is populated using the OpenCL Hardware Capability Viewer application

Available for Windows, Linux and Android

Reads and displays OpenCL information and uploads to the database

Please download and run to help populate the database!

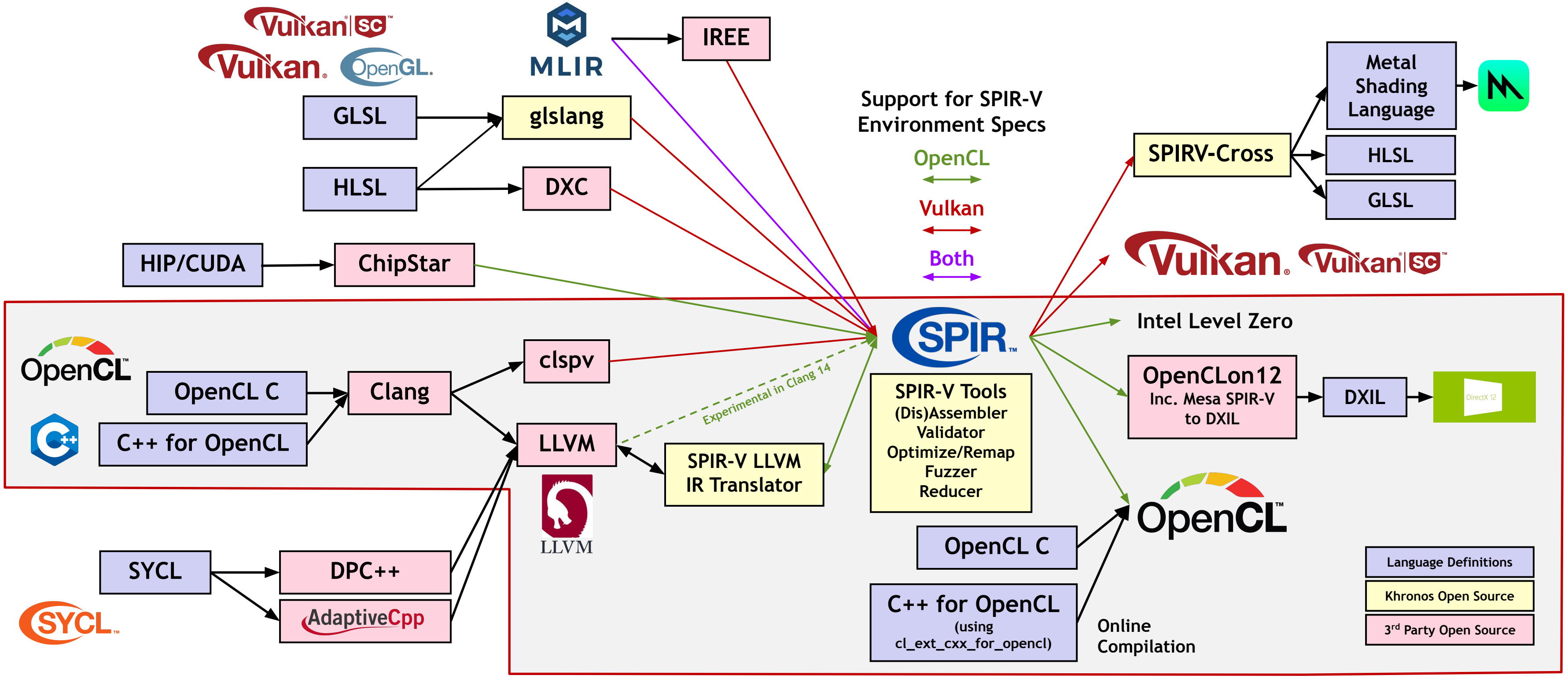


Listing devices

All platforms | Windows | Linux | Android

Device	Max. API version	Latest Driver version	Last submission	Count
Type to filter	Type to filter	Type to filter		
NVIDIA GeForce RTX 3080 Laptop GPU	3.0	555.41	2024-04-01 21:50:50	327
Ellesmere	2.0	3570.0 (PAL,HSAIL)	2024-03-10 20:18:49	161
NVIDIA GeForce 840M	3.0	555.41	2024-03-24 11:58:59	144
NVIDIA GeForce GTX 1060	3.0	555.41	2024-03-23 20:31:49	115
NVIDIA GeForce RTX 4070	3.0	555.41	2024-03-24 13:09:41	102
Microsoft Direct3D12 (AMD Radeon(TM) ...	3.0	3.0 CLVK on Vulka...	2024-03-28 19:14:18	99
Microsoft Direct3D12 (NVIDIA GeForce ...	3.0	3.0 CLVK on Vulka...	2024-03-28 19:14:06	92
pthread-AMD FX(tm)-8350 Eight-Core Pr...	3.0	3.1	2024-04-01 18:48:50	70
NVIDIA GeForce MX110	3.0	551.86	2024-03-21 18:22:50	63

OpenCL Deployment Flexibility



Vulkan SC

Vulkan

OpenGL

MLIR

IREE

GLSL

glslang

HLSL

DXC

HIP/CUDA

ChipStar

Support for SPIR-V Environment Specs

OpenCL ↔

Vulkan ↔

Both ↔

SPIRV-Cross

Metal Shading Language

HLSL

GLSL

Vulkan

Vulkan SC

OpenCL

OpenCL C

C++ for OpenCL

Clang

clspv

LLVM

Experimental in Clang 14

SPIR-V LLVM IR Translator

SPIR

SPIR-V Tools
(Dis)Assembler
Validator
Optimize/Remap
Fuzzer
Reducer

Intel Level Zero

OpenCLon12
Inc. Mesa SPIR-V to DXIL

DXIL

DirectX 12

LLVM

SYCL

DPC++

AdaptiveCpp

SYCL

OpenCL

OpenCL C

C++ for OpenCL
(using cl_ext_cxx_for_opencl)

Online Compilation

Language Definitions

Khronos Open Source

3rd Party Open Source

API Layering

Enabled by growing robustness of open-source compiler ecosystem using SPIR-V

Layers Over	Vulkan	OpenGL	OpenCL	**OpenGL ES**	**DX12**	**DX9-11**
Vulkan		Zink	clspv + clvk Ancle RustiCL/Zink	GLOVE Angle	vkd3d-Proton vkd3d	DXVK WineD3D
OpenGL	gfx-rs Ashes			Angle		WineD3D
DX12	Dozen gfx-rs	Microsoft ‘GLOn12’	Microsoft ‘CLOn12’			Microsoft D3D11On12
DX9-11	gfx-rs Ashes			Angle		
Metal	MoltenVK gfx-rs			MoltenGL Angle		

ROWS Benefit Platforms by adding APIs

COLUMNS Benefit ISVs by making an API available everywhere

Layered OpenCL Implementations

clspv + clvk

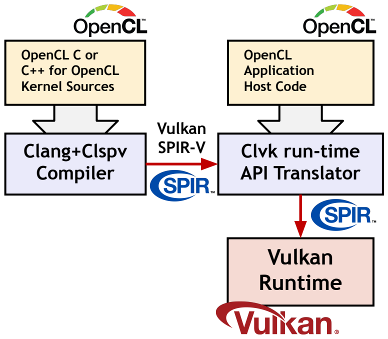
OpenCL over Vulkan
Google

clspv open-source OpenCL kernel to Vulkan SPIR-V compiler - tracks top-of-tree LLVM and Clang - not a fork

clvk - prototype open-source OpenCL to Vulkan run-time API translator

Used by shipping apps and engines on Android e.g., Adobe Premiere Rush video editor - 200K lines of OpenCL C kernel code

Android chrome Google


OpenCL C or C++ for OpenCL Kernel Sources
Clang+Clspv Compiler
Vulkan SPIR-V
OpenCL Application Host Code
Clvk run-time API Translator
Vulkan Runtime

clspv + Ancle

OpenCL over Vulkan
Samsung

Integrates clspv and OpenCL runtime into Angle code base

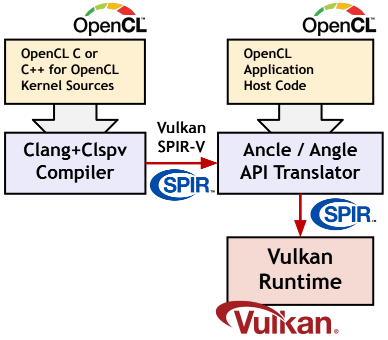
Samsung Motivation

“OpenCL is widely used and deployed and is making a comeback thanks to ML”

“OpenCL is a favored high-level (front-end) compute language! Easier to write than Vulkan”

Ancle makes OpenCL a first-class citizen in Android by relying on Vulkan as its Native Driver”

Android SAMSUNG

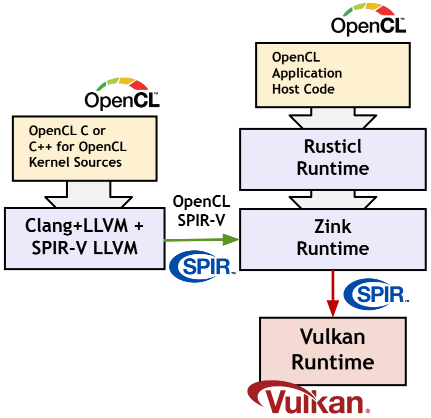

OpenCL C or C++ for OpenCL Kernel Sources
Clang+Clspv Compiler
Vulkan SPIR-V
OpenCL Application Host Code
Ancle / Angle API Translator
Vulkan Runtime

Rusticl over Zink

OpenCL over Vulkan
Mesa

The Zink Gallium driver emits Vulkan API calls and now supports OpenCL Kernels

MESA

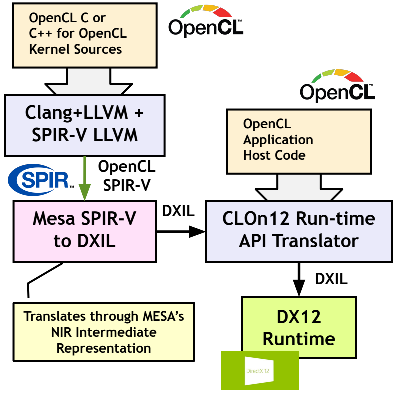


OpenCLOn12

OpenCL over DX12
Microsoft

GPU-accelerated OpenCL on any DX12 PC and Cloud instance (x86 or Arm)

Microsoft COLLABORA


OpenCL C or C++ for OpenCL Kernel Sources
Clang+LLVM + SPIR-V LLVM
OpenCL SPIR-V
Mesa SPIR-V to DXIL
DXIL
Translates through MESA's NIR Intermediate Representation
OpenCL Application Host Code
CLOn12 Run-time API Translator
DXIL
DX12 Runtime

OpenCL Acceleration in Many ML Stacks

Machine Learning Compilers	tvm.ai	plaidML	GLOW	
	amazon	intel	f	Google
Import Formats	Caffe, Keras, MXNet, ONNX	TensorFlow Graph, MXNet, PaddlePaddle, Keras, ONNX	PyTorch, ONNX	TensorFlow Graph, PyTorch, ONNX
Front-end / IR	NNVM / Relay IR	nGraph / Stripe IR	Glow Core / Glow IR	XLA HLO MLIR
Output	OpenCL OpenCL, LLVM, CUDA, Metal	OpenCL OpenCL, LLVM, CUDA	OpenCL OpenCL LLVM	OpenCL LLVM, TPU IR, XLA IR TensorFlow Lite / NNAPI (inc. HW accel)

SPIR Vulkan

Common Steps

1. Import Trained Network Description

2. Graph-level optimizations e.g., node fusion, node lowering and memory tiling

3. Decompose to primitive instructions and emit programs for accelerated run-times

Additional Machine Learning Compilers and Frameworks using OpenCL Acceleration

Inferencing Libraries and Frameworks

Alibaba MNN
Arm Compute Library
Baidu PaddlePaddle/Paddle-Lite
Berkeley Caffe
Intel clDNN and OpenVINO
Google TensorFlow and NNAPI
portDNN
Synopsis MetaWare EV
Texas Instruments DL Library (TIDL)
VeriSilicon Acuity
Xiaomi Mace

Embedded NN Compilers

CEVA Deep Neural Network (CDNN)
Cadence Xtensa
Neural Network Compiler (XNNC)

Acuity ML/AI Software Foundation · Caffe · MACE Mobile AI Compute Engine · MNN Mobile Neural Network · OPENVINO™ · 飞桨 PaddlePaddle

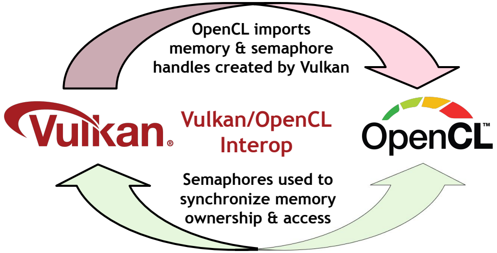
OpenCL Specification Releases and Roadmap

OpenCL 3.0.16 shipped on April 4th, 2024

Continues the regular release cadence for new functionality and bug fixes

External memory objects and semaphores for external sharing and Interop finalized

Kernel Clock extension provisional release


OpenCL imports memory & semaphore handles created by Vulkan
Vulkan
Vulkan/OpenCL Interop
OpenCL
Semaphores used to synchronize memory ownership & access

OpenCL Extension Pipeline

Provisional, EXT and Vendor extensions - candidates for final ratification

We are listening to your input!

Support C++ for OpenCL (EXT)
Command Buffer Record/Replay (provisional)
Unified Shared Memory
Floating Point Atomics (EXT)
Required Subgroup Size
Generalized Image from buffer
Image Tiling Controls

YUV Multi-planar Images
Cross-workgroup Barriers
Cooperative Matrices
Timeline Semaphores
32 and 64-length vectors
Indirect Dispatch
ML Operations

OpenCL SDK Upgrades

Open-source OpenCL SDK includes all components to develop OpenCL applications

OpenCL Headers (include/api)
OpenCL C++ bindings (include/cpp)
OpenCL Utility Libraries (include/utils)
Build system and CI

Documentation and Sample Code

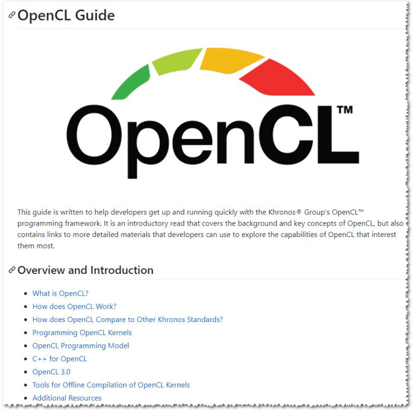
OpenCL Guide
Code samples (samples/)
Documentation (docs/)

Loader and Layers

SDK and Layers Tutorial

Khronos funds SDK upgrades

Community contributions also welcome!



OpenCL Guide

This guide is written to help developers get up and running quickly with the Khronos® Group's OpenCL™ programming framework. It is an introductory read that covers the background and key concepts of OpenCL, but also contains links to more detailed materials that developers can use to explore the capabilities of OpenCL that interest them most.

Overview and Introduction

- What is OpenCL?
- How does OpenCL Work?
- How does OpenCL Compare to Other Khronos Standards?
- Programming OpenCL Kernels
- OpenCL Programming Model
- C++ for OpenCL
- OpenCL 3.0
- Tools for Offline Compilation of OpenCL Kernels
- Additional Resources

Spring 2022 SDK Updates

More details in the SDK Blog

Enhanced Cmake-based build system
Subprojects and components

Binary releases
Tagged SDK versions

Enhanced SDK documentation
In OpenCL Guide

OpenCL 3.0 Samples
C, C++, Python and Ruby

Utility Libraries
For loading kernel source and binary files

Coming Soon!

Upstream to Kitware's FindOpenCL.cmake
Enhances OpenCL:: namespace

Packaging and Distribution Support
Build packages from the SDK
Package newer versions of OpenCL
Ease cross-platform installation, including PPAs

Enhanced SDK Validation Layers
Object lifetime, Input parameters, SPIR-V

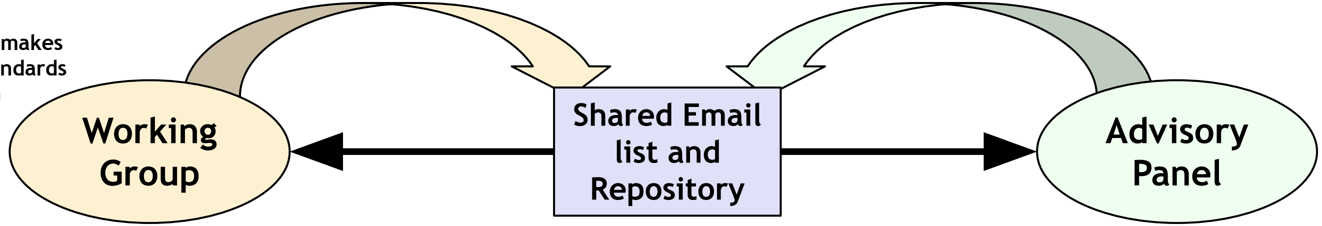
Discussion Topics

- **How can we reduce desktop fragmentation**
  - Need of universal SPIR-V ingest
  - Promote the idea of SPIR-V ingestion front-end to LLVM?
  - Leverage Microsoft's SPIR-V in LLVM?
  - Layered implementations may help?
- **Provide more support and encouragement for layered OpenCL implementations?**
  - Clspv/Ancle, Microsoft OpenCLon12, Rusticl/Zink
  - Does Rusticl over Zink on MoltenVK work for OpenCL on Apple?
  - OpenCL on Pi – maybe through Rusticl over Zink/Vulkan?
- **How encourage Tensorflow and PyTorch direct support for OpenCL (not just TensorFlow Lite)**
  - Increased investment in TVM as an open source path to other stacks?
  - Strengthen operations for ML: coop matrix, Subgroup requirements for wavefront/warp size, Built-in Kernels?
- **How increase effectiveness as target layer e.g., for SYCL and OpenMP**
  - Approach OpenMP for backend cooperation once we have SPIR-V backend in LLVM?
- **Market demand for OpenCL Safety Critical Profile?**
  - OpenCL IS already being deployed in SC markets
  - Backend for SYCL SC?

Your input and feedback is welcome!

OpenCL Advisory Panel

Working Group makes decisions on standards evolution



Working Group

Shared Email list and Repository

Advisory Panel

Hosted by Khronos.
Under Khronos NDA

Khronos Members

Any company can join.
Membership Fee.
Covered by NDA *and*
IP Framework

Working groups can share draft specifications and accept detailed design contributions as Panel Members are covered by IP Framework

Panel Members

Invited industry experts.
$0 Cost.
Covered by NDA *and*
IP Framework

Regular meetings to give feedback on roadmap and draft specifications

Please reach out to opencl-chair@lists.khronos.org if you wish to apply

Developers - Please Give Us Feedback!

- **Give us your feedback on the OpenCL spec GitHub**
  - What could be added to the OpenCL ecosystem to make you more productive?
  - What API and Language features do you most need?
  - https://github.com/KhronosGroup/OpenCL-Docs
- **Please download and run the GPUinfo OpenCL Hardware Capability Viewer**
  - https://opencl.gpuinfo.org/download.php
- **Consider applying to join the OpenCL Advisory Panel!**
  - Email opencl-chair@lists.khronos.org

OpenCL Resources

- OpenCL Home Page
  - https://www.khronos.org/opencl/
- OpenCL Registry for OpenCL core and extension specifications
  - https://www.khronos.org/registry/OpenCL/
- C++ for OpenCL Documentation
  - https://github.com/KhronosGroup/Khronosdotorg/blob/master/api/opencl/assets/CXX_for_OpenCL.pdf
- OpenCL SDK
  - https://github.com/KhronosGroup/OpenCL-SDK
- OpenCL Guide
  - https://github.com/KhronosGroup/OpenCL-Guide
- OpenCL Specification Source
  - https://github.com/KhronosGroup/OpenCL-Docs
- OpenCL Conformant Products
  - https://www.khronos.org/conformance/adopters/conformant-products/opencl
- GPUinfo.org Hardware Database
  - https://www.gpuinfo.org/
- Layered OpenCL implementations - clspv/clvk and OpenCLon12
  - https://github.com/google/clspv
  - https://github.com/kpet/clvk
  - https://github.com/microsoft/OpenCLOn12