K H R S N O S G R O U P

State of the Union OpenCL Working Group

Neil Trevett

Khronos President OpenCL Working Group Chair NVIDIA VP Developer Ecosystems <u>ntrevett@nvidia.com</u> | <u>@neilt3d</u>





© Khronos[®] Group 2021

Primary Developments Since IWOCL 2020

OpenCL 3.0 Finalized and Released Including subgroups and embedded processor extensions

Multiple OpenCL 3.0 Implementations Shipping With many more Adopters in the pipeline

Conformance Testing Improvements >250 commits to the OpenCL test suite since IWOCL 2020

Regular Releases Including OpenCL 3.0.7 here at IWOCL with new extensions!

OpenCL Guide Released and SDK Enhanced! Tutorial on SDK Layers here at IWOCL

C++ for OpenCL gaining momentum Interaction with LLVM community deepening

Increased activity in layered implementations Microsoft's OpenCLon12 in addition to Google's clspv

> Roadmap Discussions Underway Building Advisory Panel Interactions



This work is licensed under a Creative Commons Attribution 4.0 International License

2

Η×

OpenCL Open-Source Project Momentum

OpenCL-based GitHib Repos



This work is licensed under a Creative Commons Attribution 4.0 International License

© The Khronos[®] Group Inc. 2021 - Page 3

K H R O N O S

OpenCL 3.0

Increased Ecosystem Flexibility

All functionality beyond OpenCL 1.2 queryable Macros for optional OpenCL C language features Widely adopted extensions to be integrated into core

OpenCL C++ for OpenCL

Open-source <u>C++ for OpenCL</u> front end compiler combines OpenCL C and C++17 replacing OpenCL C++ language spec

Unified Specification

All versions of OpenCL in one specification for easier maintenance, evolution and accessibility <u>Source</u> on Khronos GitHub for community feedback, functionality requests and bug fixes

New Functionality

Subgroups with SPIR-V 1.3 in core (optional) Asynchronous DMA extension for embedded processors

Easy OpenCL 3.0 migration for applications

OpenCL 1.2 applications - no change OpenCL 2.X applications - no code changes if all used functionality present Queries recommended for future portability



OpenCL is Widely Deployed and Used



This work is licensed under a Creative Commons Attribution 4.0 International License

ົ

O°

Z

2

т

 $\mathbf{\Sigma}$

https://en.wikipedia.org/wiki/List_of_OpenCL_applications

© The Khronos® Group Inc. 2021 - Page 5





This work is licensed under a Creative Commons Attribution 4.0 International License

S O Z V

HR

 $\mathbf{\mathbf{x}}$

C++ for OpenCL

- Open-Source Compiler Front-end
 - Replaces the OpenCL C++ kernel language spec
 - Official releases published in OpenCL-Docs repo
- Enables full OpenCL C and most C++17 capabilities
 - OpenCL C code is valid and fully compatible
 - Enables gradual transition to C++ for existing apps
 - Language documentation
- Supported in Clang since release 9.0
 - Generates SPIR-V 1.0 plus SPIR-V 1.2 where necessary
 - Full details are provided in OpenCL-Guide
- Online compilation via <u>cl_ext_cxx_for_opencl</u>







Asynchronous DMA Extensions

OpenCL embraces a new class of Embedded Processors

Many DSP-like devices have Direct Memory Access hardware

Transfer data between global and local memories via DMA transactions

Transactions run asynchronously in parallel to device compute enabling wait for transactions to complete Multiple transactions can be queued to run concurrently or in order via fences

OpenCL abstracts DMA capabilities via extended asynchronous workgroup copy built-ins

(New!) 2- and 3-dimensional async workgroup copy extensions support complex memory transfers (New!) async workgroup fence built-in controls execution order of dependent transactions New extensions complement the existing 1-dimensional async workgroup copy built-ins



S O N N

2

Т

 $\mathbf{\Sigma}$

Async Fence controls order of dependent transactions



All transactions prior to async_fence must complete before any new transaction starts, without a synchronous wait

The first of significant upcoming advances in OpenCL to enhance support for embedded processors

This work is licensed under a Creative Commons Attribution 4.0 International License

© The Khronos[®] Group Inc. 2021 - Page 8

OpenCL 3.0.7 Release at IWOCL

Second Maintenance release since OpenCL 3.0 in September 2020

Clarifications, formatting, bug fixes Adds optional extensions

cl_khr_spirv_extended_debug_info

Enables SPIR-V modules to use the OpenCL.DebugInfo.100 extended instruction set

cl_khr_pci_bus_info

Query PCI domain, bus, device, and function information for an OpenCL device

cl_khr_extended_bit_ops

Adds OpenCL C built-in functions to insert, extract, and reverse bits in a bitfield

cl_khr_suggested_local_work_size

Adds a query for a suggested local work group size for a kernel running on an OpenCL device

cl_khr_spirv_linkonce_odr

Enables LinkOnceODR SPIR-V link type to separately compile and link C++ programs

This work is licensed under a Creative Commons Attribution 4.0 International License

Specification available on the <u>OpenCL Registry</u>

OpenCL SDK - In Development

- Bringing together all the components needed to develop OpenCL applications
 - OpenCL Headers (include/api)
 - OpenCL C++ bindings (include/cpp)
 - OpenCL Utility Libraries (include/utils)
 - Build system and CI
- Other resources useful to OpenCL developers
 - OpenCL Guide
 - Code samples (samples/)
 - Documentation (docs/)
- Loader and Layers
 - Initial layers implemented
 - SDK and Layers Tutorial here at IWOCL
- Watch GitHub Repo for updates
 - Community contributions welcome!

	()nen(;
	•
	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 contains links to more detailed materials that developers can use to explore the capabilities of OpenCL that intere- them most.
0	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



API Layering

Enabled by growing robustness of open-source compiler ecosystem

			\frown			
Layers Over	Vulkan	OpenGL	OpenCL	OpenGL ES	DX12	DX9-11
Vulkan		Zink	clspv clvk	GLOVE Angle	vkd3d-Proton vkd3d	DXVK WineD3D
OpenGL	gfx-rs Ashes			Angle		WineD3D
DX12	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 Enable content without additional kernel level drivers

COLUMNS Benefit ISVs by making an API available everywhere

Application deployment flexibility by fighting platform fragmentation Making an API available across multiple platforms even if no native drivers available

SPIR-V Language Ecosystem



Layered OpenCL Implementations

clspv + clvk

clspv - Google's 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

OpenCLOn12

Microsoft and COLLABORA GPU-accelerated OpenCL on any DX12 PC and Cloud instance (x86 or Arm) Leverages Clang/LLVM AND MESA OpenGLOn12 - OpenGL 3.3 over DX12 is already conformant



OpenCL Roadmap

Extensions in Ratification Expected Public Release 2Q 2021

cl_khr_integer_dot_product Adds support for SPIR-V instructions and OpenCL C built-in functions to compute the dot product of vectors of integers

External Sharing Extensions (Provisional)

cl_khr_external_memory Create OpenCL memory objects from OS-specific memory handles (similar to VK_KHR_external_memory)

cl_khr_semaphore

Semaphore synchronization object that can be signaled and reset multiple times and signaled from outside OpenCL

The cl_khr_external_semaphore and cl_khr_external_semaphore_sync_fd Create OpenCL semaphore objects from OS-specific semaphore handles

cl_khr_vk_sharing extension Associate an OpenCL context with a Vulkan physical device

The External Sharing Extensions are Provisional to enable developer feedback before finalization



External Sharing Extensions

Generic extensions to import external memory and semaphores exported by other APIs API-specific interop extensions e.g., Vulkan More flexible than previous interop APIs using implicit resources

Longer Term Roadmap Discussions

Command Buffer Recording and Replay Unified Shared Memory **Floating-point Atomics Global Barriers** YUV Multi-planar Images Generalized Image from Buffer Indirect Dispatch **Collective Programming Expect and Assume Optimization Hints Required Subgroup Size** Machine Learning Operations **Extended Async Copies** 2D and 3D Prefetch Built In Functions

Developer Feedback Welcome!

What is your highest priority? What is missing? Requirements and use cases

See 'Extensions Feedback' issue on GitHub https://github.com/KhronosGroup/OpenCL-Docs/issues/604

New functionality is proven as extensions before being added to core

OpenCL Advisory Panel



Chaired by Máté Ferenc Nagy-Egri at StreamHPC

OpenCL Advisory Panel meeting here at IWOCL 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!

- How is your transition to using OpenCL 3.0?
 - Are you encountering any issues?
- Which optional features do you expect to use in your application or library?
 - Usage data drives which optional features should be made mandatory in future
- What new features do you most need?
 - What roadmap extensions would you prioritize, and are there any gaps?
 - https://github.com/KhronosGroup/OpenCL-Docs/issues/604
- Consider applying to join the OpenCL Advisory Panel!
 - Email opencl-chair@lists.khronos.org

ູ

0° Z°

Ŕ

×H

More OpenCL information!

https://www.khronos.org/opencl/

Feedback Welcome!

https://github.com/KhronosGroup/OpenCL-Docs

