Vision and Inferencing Acceleration

OpenVX Performance, Portability and Memory Footprint Advantages

14th October 2020

Welcome!

This slide deck will be posted at www.khronos.org today

The webinar will be recorded and made freely available

Please take a moment to answer the polls at the end of the webinar

Ask any questions via the Q&A Panel at any time

We will answer questions at the end of the webinar

Today's Speakers

Introduction to Khronos and OpenVX

Neil Trevett
Khronos President

AMD and OpenVX

Mike Schmit
AMD
Director of Software System Design ML Computer Vision

Cadence and OpenVX

Frank Brill
Cadence Systems
Design Engineering Director

Using OpenVX

Kiriti Nagesh Gowda
OpenVX Group Chair

Khronos Connects Software to Silicon

Open interoperability standards to enable software to effectively harness the power of 3D and multiprocessor acceleration

Over 150 members worldwide
Any company is welcome to join

Founded in 2000
>150 Members ~ 40% US, 30% Europe, 30% Asia

3D graphics, XR, parallel programming, vision acceleration and machine learning

Non-profit, member-driven standards-defining industry consortium

Open to any interested company

All Khronos standards are royalty-free

Well-defined IP Framework protects participant's intellectual property

Khronos Active Initiatives

3D Graphics
Desktop, Mobile, Web
Embedded and Safety Critical

Vulkan, ANARI, OpenGL ES, OpenGL, WebGL, EGL, OpenGL SC, Vulkan SC

3D Assets
Authoring
and Delivery

glTF, COLLADA, 3DCommerce

Portable XR
Augmented and
Virtual Reality

OpenXR

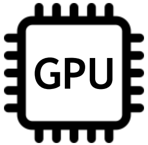
Parallel Computation
Vision, Inferencing,
Machine Learning

OpenVX, OpenCL, NNEF, SYCL, SPIR

The Origin of OpenVX

Engines and Applications

3D Graphics API Driver



Driver Model

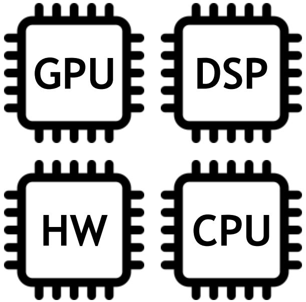
An open API standard enables multiple silicon vendors to ship drivers with their silicon

Silicon vendors can aggressively optimize drivers for their own silicon architecture

OpenVX is the industry’s only API standard enabling portable access to vendor-optimized vision drivers

Engines and Applications

Vision API Driver

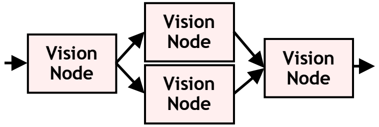


High-level Abstraction

3D graphics is always accelerated by a GPU - so a low-level GPU-centric API still provides cross-vendor portability

Vision processing can be accelerated by a wide variety of hardware architectures

OpenVX needs a higher-level graph abstraction to enable optimized cross-vendor drivers

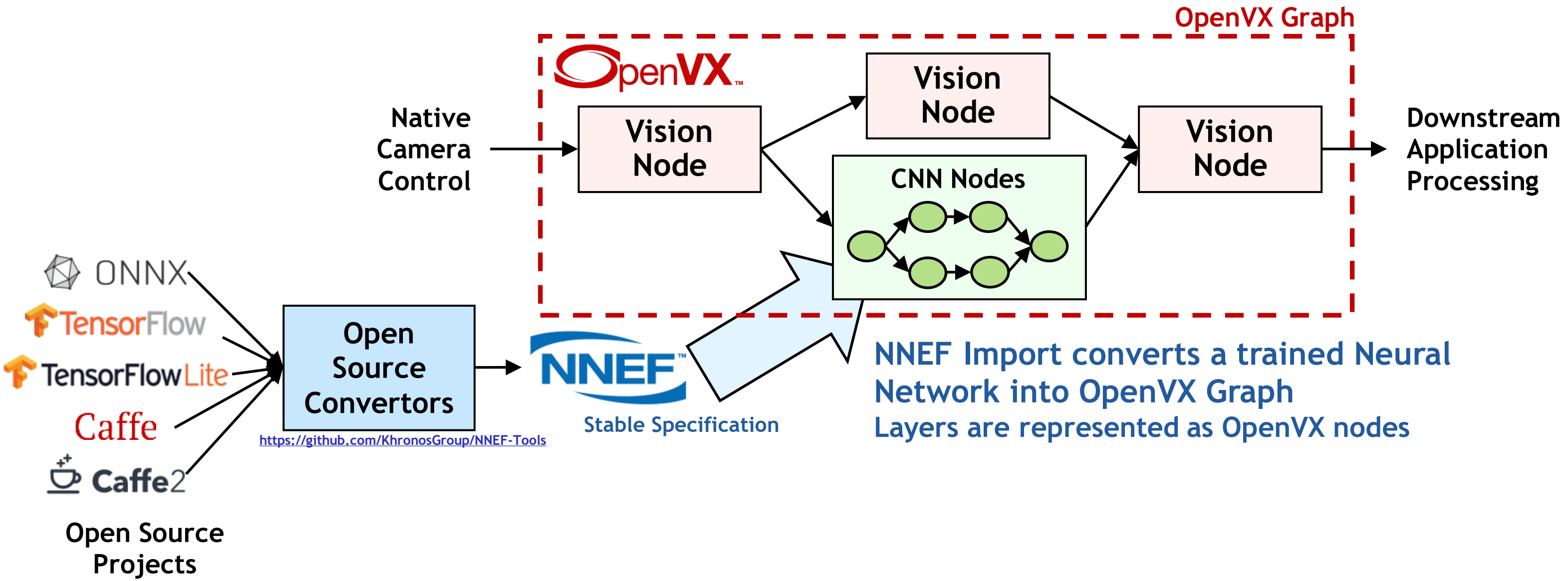


Vision Processing Graph

OpenVX Cross-Vendor Vision and Inferencing

High-level graph-based abstraction for portable, efficient vision processing

Optimized OpenVX drivers created, optimized and shipped by processor vendors
Implementable on almost any hardware or processor with performance portability
Graph can contain vision processing and NN nodes for global optimization
Run-time graph execution need very little host CPU interaction



OpenVX Efficiency through Graphs..

Graph Scheduling

Split graph execution across the whole system: CPU / GPU / dedicated HW

Faster execution or lower power consumption

Memory Management

Reuse pre-allocated memory for multiple intermediate data

Less allocation overhead, more memory for other applications

Kernel Fusion

Replace a sub-graph with a single faster node

Better memory locality, less kernel launch overhead

Data Tiling

Execute a sub-graph at tile granularity instead of image granularity

Better use of data cache and local memory

Performance comparable to hand-optimized, non-portable code

Real, complex applications on real-world hardware

Much lower development effort and higher portability than hand-optimized code

OpenVX and OpenCV are Complementary

	OpenCV	OpenVX
Implementation	Community driven open source library	API drivers implemented, optimized and shipped by hardware vendors
Conformance	Extensive OpenCV Test Suite but no formal Conformance program	Implementations pass Khronos Conformance Testing to use trademark and ensure cross vendor consistency
Scope	100s of imaging and vision functions Multiple camera APIs/interfaces	Tight focus on dozens of core hardware accelerated functions plus extensions and accelerated custom nodes. Uses external camera drivers
Inferencing	Deep Neural Network module to construct networks from layers for forward pass computations only. Import from ONNX, TensorFlow, Torch, Caffe	Neural Network layers and operations represented directly in the OpenVX Graph. NNEF direct import with multiple convertors
Acceleration	OpenCV 3.0 Transparent API (or T-API) enables function offload to OpenCL devices	Implementation free to use any underlying API such as OpenCL. OpenCL for accelerated Custom Nodes
Efficiency	OpenCV 4.0 G-API graph model for some filters, arithmetic/binary operations, and well-defined geometrical transformations	Graph-based execution of all Nodes. Optimizable computation and data transfer
IP Protection	None. Source code licensed under BSD. Some modules require royalties/licensing	Protected under Khronos IP Framework - Khronos members agree not to assert patents against API when used in Conformant implementations

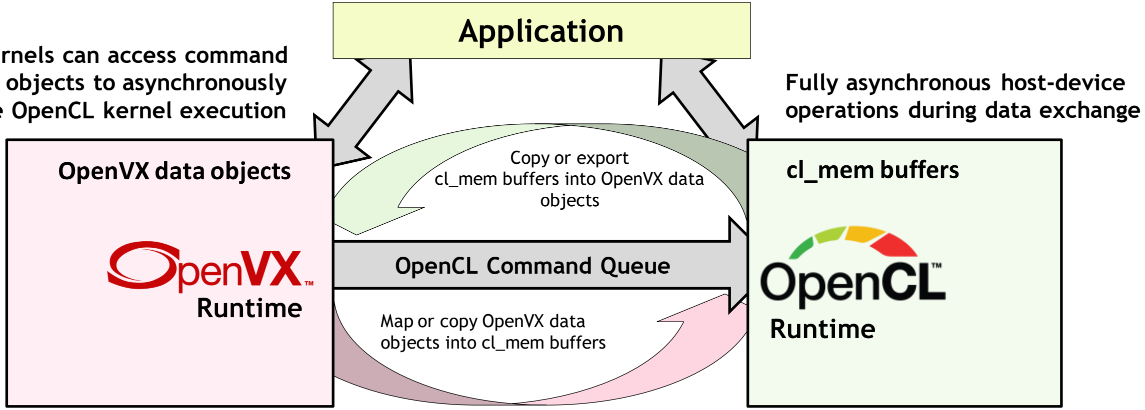
OpenVX Extensibility

OpenVX core specification defines market-targeted feature sets

Baseline Graph Infrastructure (enables other Feature Sets)
Default Vision Functions
Enhanced Vision Functions
Neural Network Inferencing (including tensor objects)
NNEF Kernel import (including tensor objects)
Binary Images
Safety Critical (reduced features and graph import for easier safety certification)

OpenVX is Extensible

Fully accelerated custom nodes can be integrated into the OpenVX graph with OpenCL interop



OpenVX/OpenCL Interop

OpenVX Industry Deployment

KHRONOS GROUP

OpenVX Working Group

- OpenVX Specifications
- Open source Conformance Test Suite
- Adopters Program (fee)
- Open source sample implementation

Porting and optimization

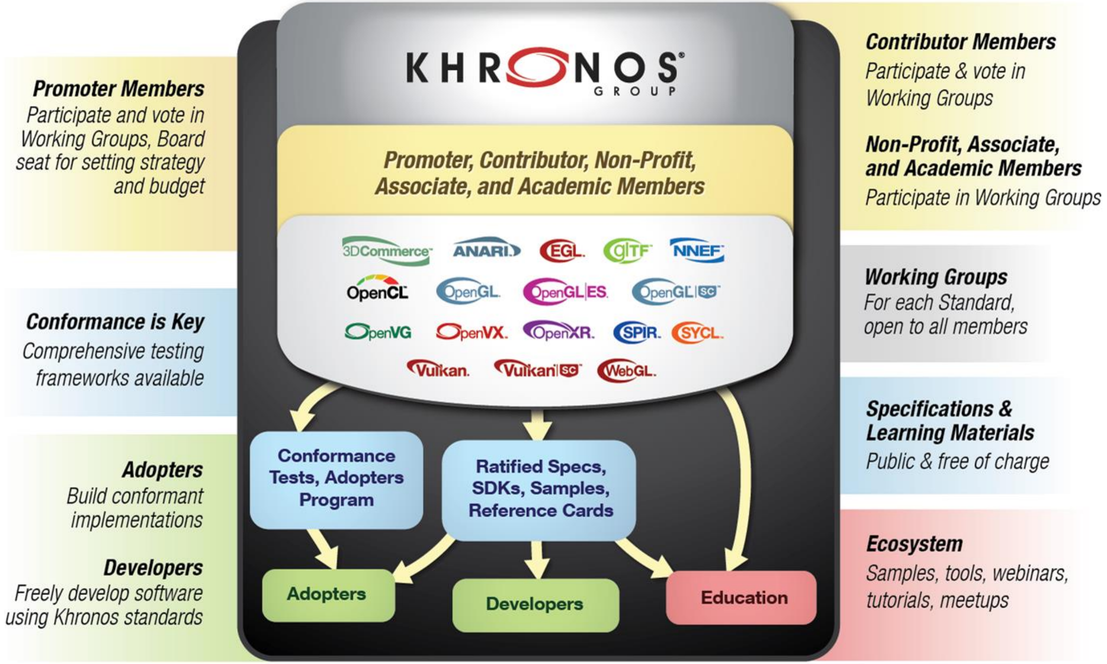
Vendors optimize and ship drivers for their platform

Full list of conformant products here:
https://www.khronos.org/conformance/adopters/conformant-products/openvx

Conformant Implementations can use OpenVX logo

Developers can freely use any OpenVX Implementation

Khronos for Global Industry Collaboration



Khronos membership is open to any company

Influence the design and direction of key open standards that will drive your business

Accelerate time-to-market with early access to specification drafts

Provide industry thought leadership and gain insights into industry trends and directions

Benefit from Adopter discounts

www.khronos.org/members/