2023 embedded VISION SUMMIT

Open Standards Unleash Hardware Acceleration for Embedded Vision

Neil Trevett

President

Khronos Group

KHRONOS GROUP

Topics

- The need for parallel offload of vision and inferencing processing
- The need for open standard APIs for embedded vision
- Introduction to Khronos and its family of open API standards
- Khronos and safety-critical APIs
- Introducing the Kamaros embedded camera API in development
- How to get involved!

Khronos Connects Software to Silicon

KHRONOS GROUP

Close to 200 members worldwide
Any organization is welcome to join

AMD
arm
EPIC GAMES
Google
HUAWEI
IKEA
Imagination
intel
NVIDIA
Qualcomm
SAMSUNG
SONY
Tencent腾讯
VALVE
VeriSilicon
3dMDY
acer
adidas
Adobe
AIMOTIVE
ALL3D
Alibaba.com
Almalence
amazon.com
Ambarella
AGI
Argonne
ASHLEY
AUTODESK
avataar
BADVISOR
BASEMARK
BINOMIAL
blaize
BLIZZARD
blender
BOLT GRAPHICS
BOSCH
botspot
BRENWILL
BROADCOM
ByteDance
CableLabs
CAPPASITY
CESIUM
codeweavers
Collins Aerospace
Continental
CoreAVI
DGG
DASSAULT SYSTEMES
DMP
DIMENCO
DisplayLink
dopple
DoubleMe
EA
emersya
ERICSSON
FIXSTARS
FORGE
Fraunhofer
GLENFLY
GO2 VIRTUAL
hp
Hexa
HITACHI Inspire the Next
HOLOCHIP
HONOR
htc
id
igalia
Imperial College London
INNOSILICON
MTA SZTAKI
Inseye
JUICE
KDAB
KITTYHAWD
KNU
Linaro
LIV
Loft3Di
Los Alamos
logitech
LUNARG
Magic Leap
MASTERPIECE STUDIO
matrix
matrox
MEDIATEK
Mercedes-Benz
Meta
METAVERSE
METAX
Microsoft
mobica
moz://a
National Institute of Standards and Technology
NaturalPoint
The New York Times
NETFLIX
NIHON UNIVERSITY
Nintendo
NOKIA
nreal
NSI-TEXE
NXP
oppo
OSU
otoy
OTSL
OVR
OXIDE
Penumbra
pluto
ptc
Red Hat
RENESAS
RIVE
Rokid
shopify
shutterstock
SIRU
SIDEQUEST
SILICONARTS
SKYWORTH VR
socionext
STREAM
SuperDNA 3D Lab
Synaptics
SYNOPSYS
TAKUMI
TAMPERE UNIVERSITY OF TECHNOLOGY
TELEDYNE IMAGING
TEXAS INSTRUMENTS
Think Silicon
TRAVERSE RESEARCH
THREEDY instant 3Dhub
ThreeKit
tobii
ultraleap
Unity
University of BRISTOL
ILLINOIS
universität innsbruck
UNIVERSITÄT DES SAARLANDES
UNIVERSITY OF TORONTO
TOHOKU
University of Windsor
UX3D
VARJO
vivo
vmware
VNTANA
YVR
wayfair
XE3D
XIANGDIXIAN
ZUSE INSTITUTE BERLIN
AREA
AUTOSAR
emva
ETSI
GTI
HAPTICS INDUSTRY FORUM
ISO JTC1 IEC
Open Geospatial Consortium
Smithsonian
TIFCA
VESA
W3C
web3D CONSORTIUM

Liaisons: Cooperation with industry associations and organizations

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

KHRONOS GROUP

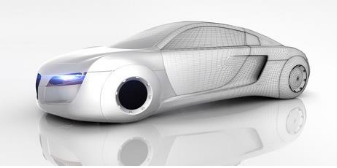
Open, royalty-free interoperability standards to harness the power of GPU, XR and multiprocessor hardware

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

Increasing Embedded Vision Processing Demands

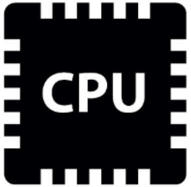


Increasing number and diversity of sensors per system
Including camera arrays and depth sensors such as Lidar

Increasing sophistication of sensor and vision processing
Including inferencing

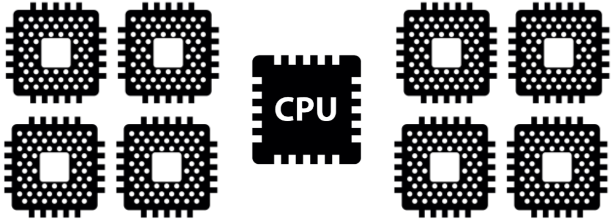
Rapidly increasing demand for higher performance vision processing

The Need for Parallel Processing



Single Processor

Simple to program
but
May not provide enough performance
especially
As Moore’s Law frequency/power
scaling is slowing



Multi-Processor

Additional processors
can process expanded workloads
but
Adds complexity to system design
and programming, as there is the need to:

(i) Distribute workload across processors

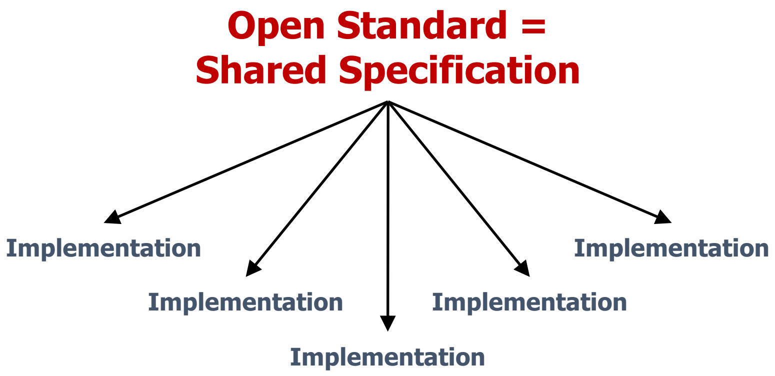
(ii) Synchronize use of compute and memory resources

(iii) Communicate intermediate data and results

Can open standards help solve this complexity?

Open Standards Make Technology Pervasive

INTEROPERABILITY standards define precise COMMUNICATION
E.g., software to hardware, client to server



Open Standard =
Shared Specification

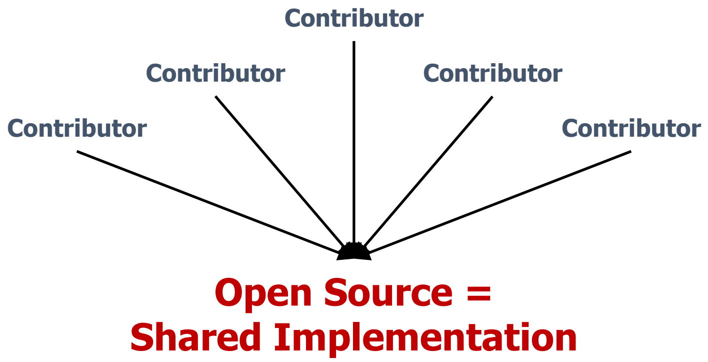
Implementation

Implementation

Implementation

Implementation

Implementation



Contributor

Contributor

Contributor

Contributor

Contributor

Open Source =
Shared Implementation

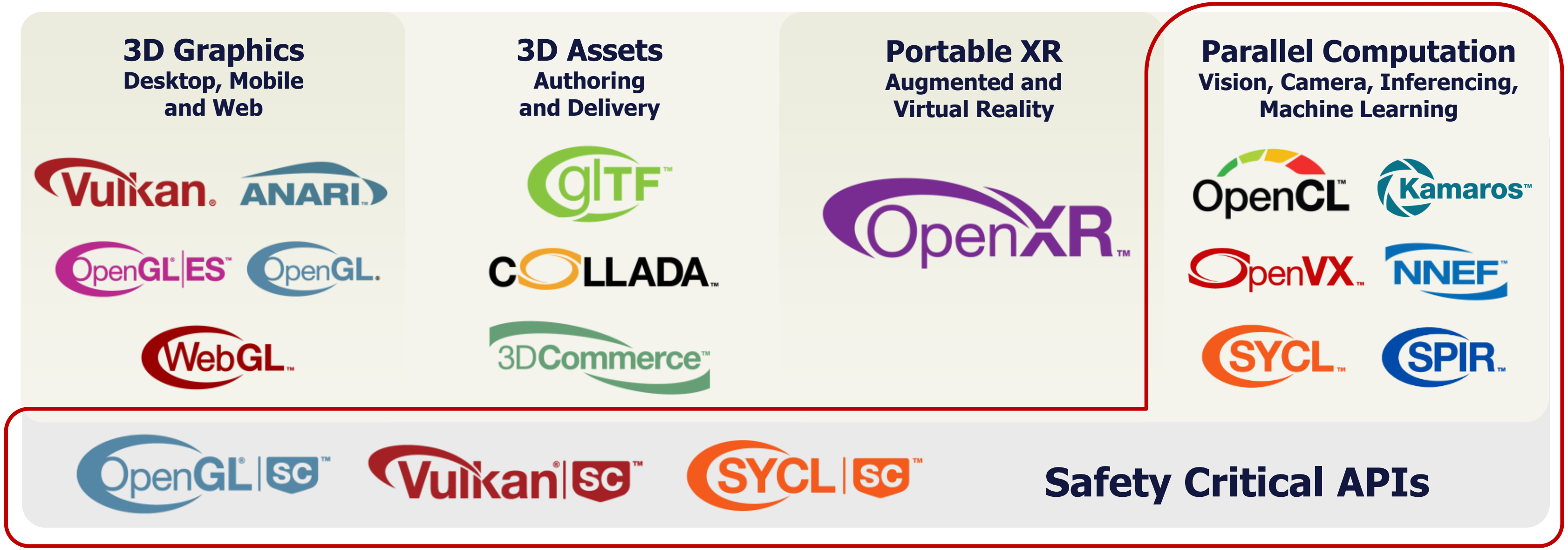
Open standards with rigorous conformance testing enable consistency across multiple implementations that can meet the needs of diverse markets, price points, and use cases

Open standards often use open source to spread the implementation effort for sample implementations, tools, samples, conformance tests, validators, etc.

Open Standard Why’s, When’s and How’s

<table>
<tr><th colspan="2">Why</th></tr>
<tr><td><b>Grow Markets</b>
By reducing consumer confusion and increasing capabilities and usability</td><td><b>Reduce Costs</b>
By sharing ecosystem development between many companies and driving volume</td></tr>
<tr><td><b>Speed Time to Market</b>
With well-proven functionality, testing and interoperability</td><td><b>Enable Innovation</b>
Free companies to compete on value differentiators: quality, performance, power etc.</td></tr>
<tr><th colspan="2">When</th></tr>
<tr><td><b>Proven Technologies</b>
Don’t do R&D by standards committee</td><td><b>Consensus Need</b>
The downside of not having a standard is clear to all</td></tr>
<tr><th colspan="2">How</th></tr>
<tr><td><b>Multi-company governance</b>
Avoid single-company control or dependency</td><td><b>Well-defined IPR Policy</b>
Royalty-free standards have clearer path to wide adoption</td></tr>
<tr><td><b>With Extensibility</b>
Enable a responsive innovation pipeline to meet customer and market needs</td><td><b>Thoughtful Abstraction</b>
Focus on interoperability and avoid over specifying implementation which stifles innovation</td></tr>
</table>

Khronos Active Standards



3D Graphics
Desktop, Mobile and Web

3D Assets
Authoring and Delivery

Portable XR
Augmented and Virtual Reality

Parallel Computation
Vision, Camera, Inferencing, Machine Learning

Safety Critical APIs

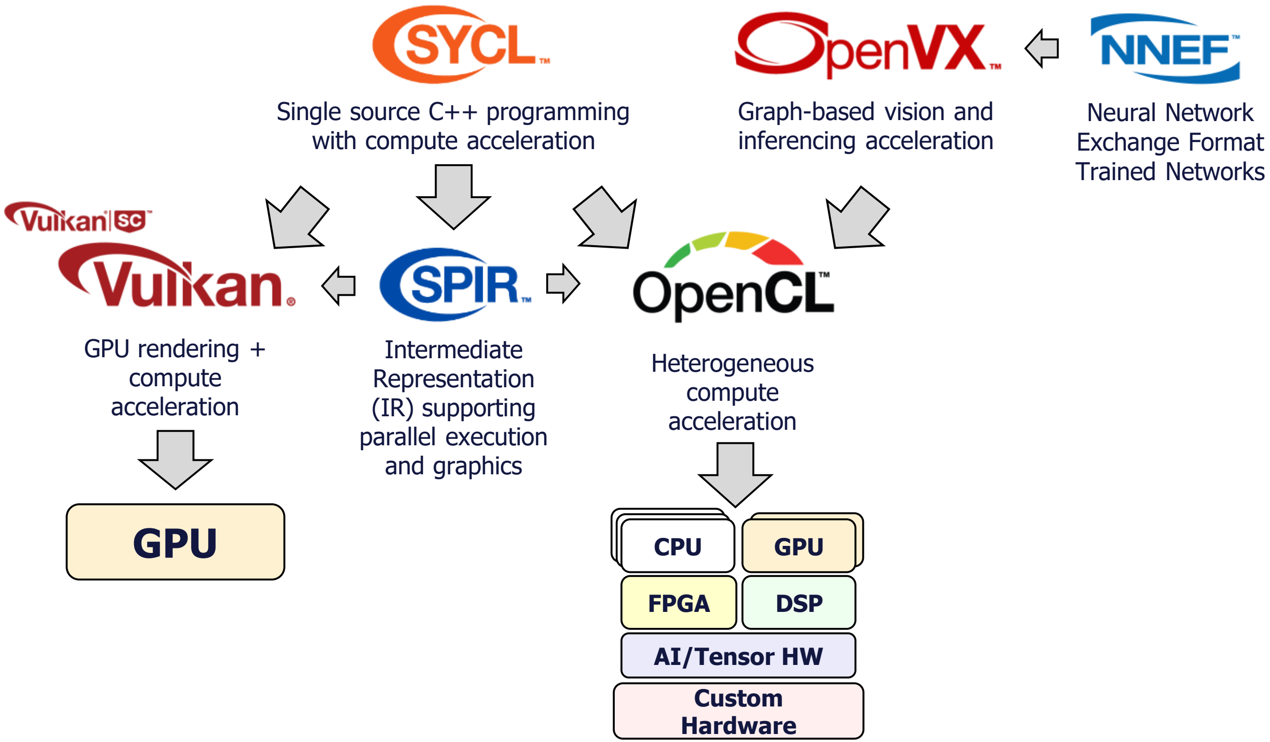
Open Standard APIs in Embedded Markets

- Enable cross-platform software reusability
- Decouple software and hardware for easier development and integration of new components
- Provide cross-generation reusability
- Facilitate field upgradability

Khronos Compute Acceleration Standards

Higher-level Languages and APIs
Streamlined development and performance portability

Lower-level Languages and APIs
Direct hardware control and optimization



Single source C++ programming with compute acceleration

Graph-based vision and inferencing acceleration

Neural Network Exchange Format Trained Networks

GPU rendering + compute acceleration

GPU

Intermediate Representation (IR) supporting parallel execution and graphics

Heterogeneous compute acceleration

CPU

GPU

FPGA

DSP

AI/Tensor HW

Custom Hardware

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

OpenCL – Low-level Parallel Programing

Programming and Runtime Framework for Application Acceleration

Offload compute-intensive kernels onto parallel heterogeneous processors
CPUs, GPUs, DSPs, FPGAs, Tensor Processors
OpenCL C or C++ kernel languages

Platform Layer API

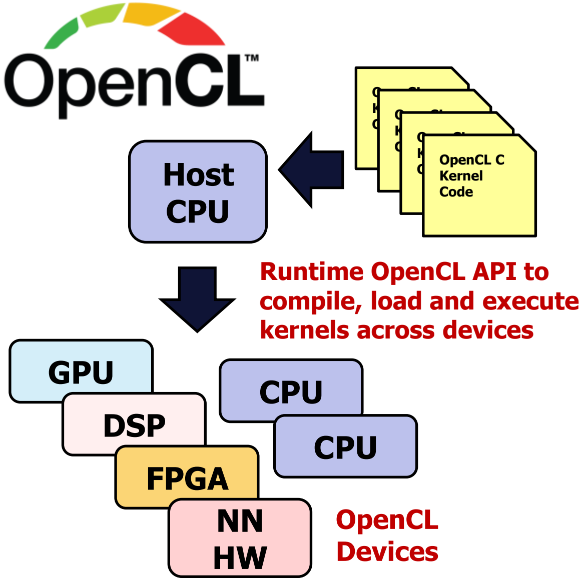
Query, select and initialize compute devices

Runtime API

Build and execute kernels programs on multiple devices

Explicit Application Control

Which programs execute on what device
Where data is stored in memories in the system
When programs are run, and what operations are dependent on earlier operations



Runtime OpenCL API to compile, load and execute kernels across devices

OpenCL Devices

Complements GPU-only APIs

Simpler programming model
Relatively lightweight run-time
More language flexibility, e.g., pointers
Rigorously defined numeric precision

Executing OpenCL Programs

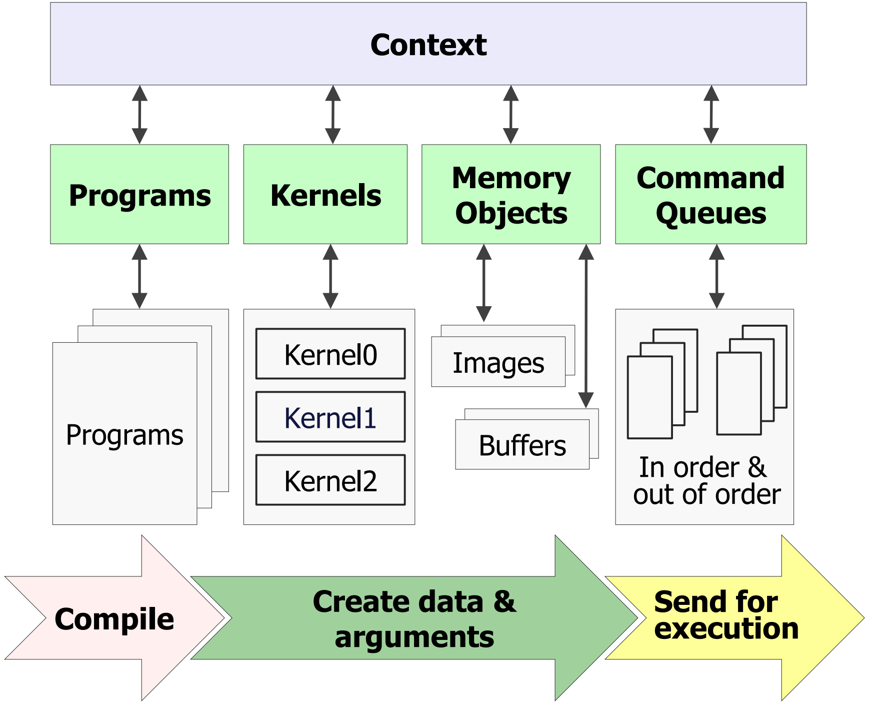
A ***kernel*** program is the basic unit of executable code (similar to a C function)

An OpenCL ***program*** is a collection of kernels and functions

An OpenCL ***command queue*** is used by the host application to send kernels and data transfer functions to a device for execution

By *enqueueing* commands into a command queue, kernels and data transfer functions may execute asynchronously and in parallel with application host code

As an open standard, OpenCL is a well proven design, available from many silicon vendors with an extensive ecosystem of available tools, compilers, libraries and educational materials



Growing Need for APIs for Functional Safety

Growing demand for accelerated graphics and compute growing where functional safety is paramount

Autonomous vehicles and devices, avionics, medical, industrial, and energy

1990s
Avionics

2010s
Automotive

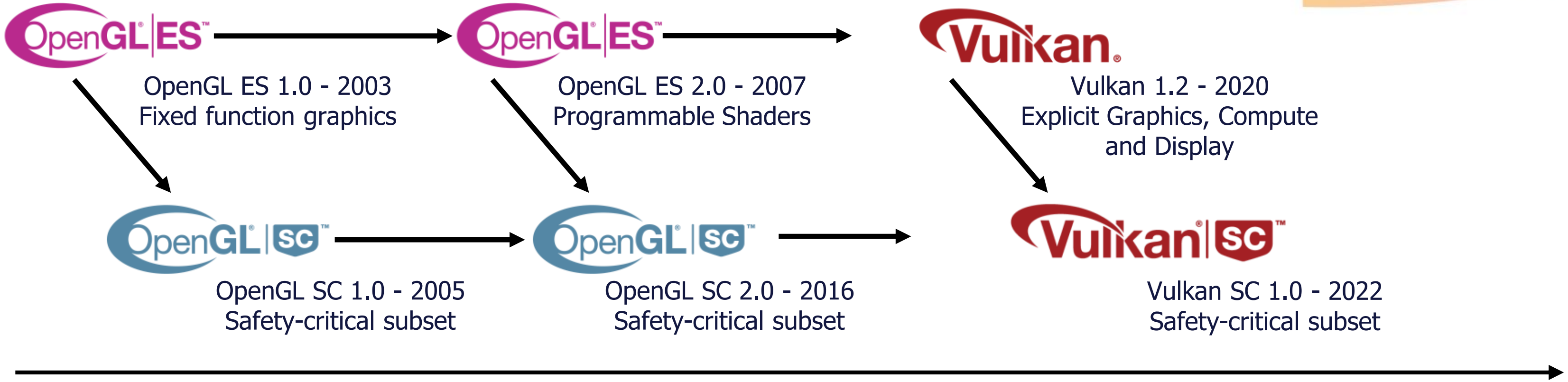
2020s…
Everywhere

Safety-critical APIs are designed to reduce system-level certification effort and costs

1) Streamlined to reduce documentation and testing surface area
2) Deterministic behavior to simplify system design and testing
3) Unambiguous and comprehensive fault handling

Industry safety-critical standards include
RTCA DO-178C (avionics) | ISO 26262 (automotive)
IEC 61508 (industrial) | IEC 62304 (medical)

Khronos Safety Critical Standards Evolution



OpenGL ES 1.0 - 2003
Fixed function graphics

OpenGL ES 2.0 - 2007
Programmable Shaders

Vulkan 1.2 - 2020
Explicit Graphics, Compute and Display

OpenGL SC 1.0 - 2005
Safety-critical subset

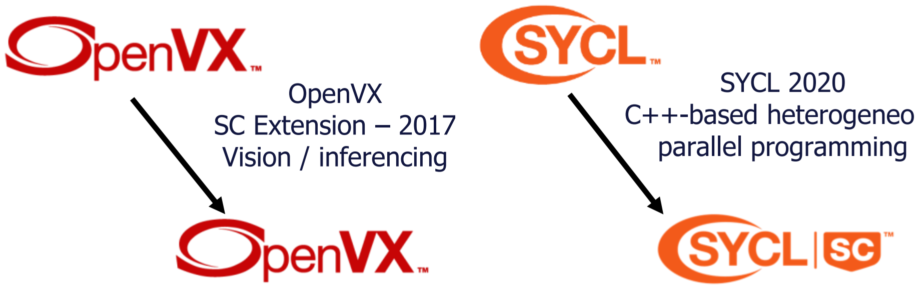
OpenGL SC 2.0 - 2016
Safety-critical subset

Vulkan SC 1.0 - 2022
Safety-critical subset

Khronos has 20 years experience in standards for safety-critical markets

Leveraging proven mainstream standards with shipping implementations and developer tooling and familiarity

A choice of abstraction levels to suit different markets and developer needs



OpenVX
SC Extension – 2017
Vision / inferencing

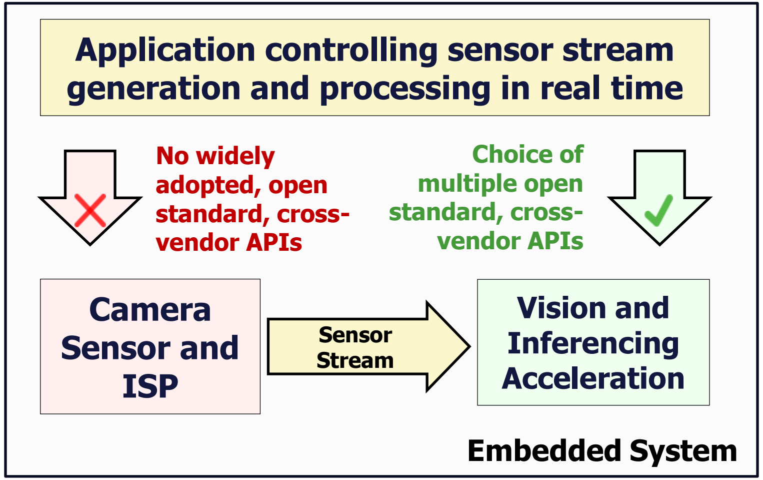
SYCL 2020
C++-based heterogeneous parallel programming

OpenVX 1.3 – 2019
SC Extension integrated into core OpenVX

March 2022
SYCL SC Working Group announced

Need for Embedded Camera API Standard

An effective open, cross-vendor open standard for camera, sensor and ISP control is an ecosystem gap compared to processing APIs


Application controlling sensor stream generation and processing in real time
No widely adopted, open standard, cross-vendor APIs
Choice of multiple open standard, cross-vendor APIs
Camera Sensor and ISP
Sensor Stream
Vision and Inferencing Acceleration
Embedded System

The cost and time to integrate and utilize sensors in embedded systems using proprietary APIs has become a major constraint on innovation and efficiency in the embedded vision market

Benefits of a Cross-vendor Embedded Camera API

Portability of camera/sensor code for easier system integration of new sensors

Preservation of application code across multiple generations of cameras and sensors

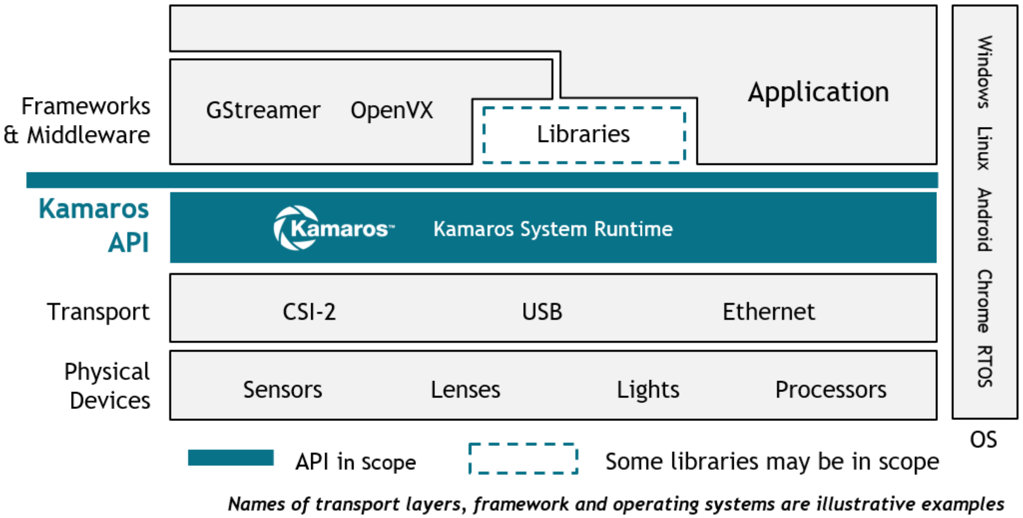
Sophisticated control over sensor stream generation increases effectiveness of downstream accelerated processing

Kamaros Embedded Camera API

Camera Programming Model based on Pipeline Templates

Defines available streams through sensor, ISP and other processing elements
Exposes static and per-frame controls - ranging from simple to more complex
Balances 'simplicity and portability' and 'detailed HW control'
Tested pipeline templates can be vendor-supplied

In development – first draft spec expected in 2023



Names of transport layers, framework and operating systems are illustrative examples

Application-facing API for cross-vendor, cross-platform portability

Multiple Language Bindings

IDL based API design with automatic generation of multiple language bindings to C, C++, Python, and other programming languages

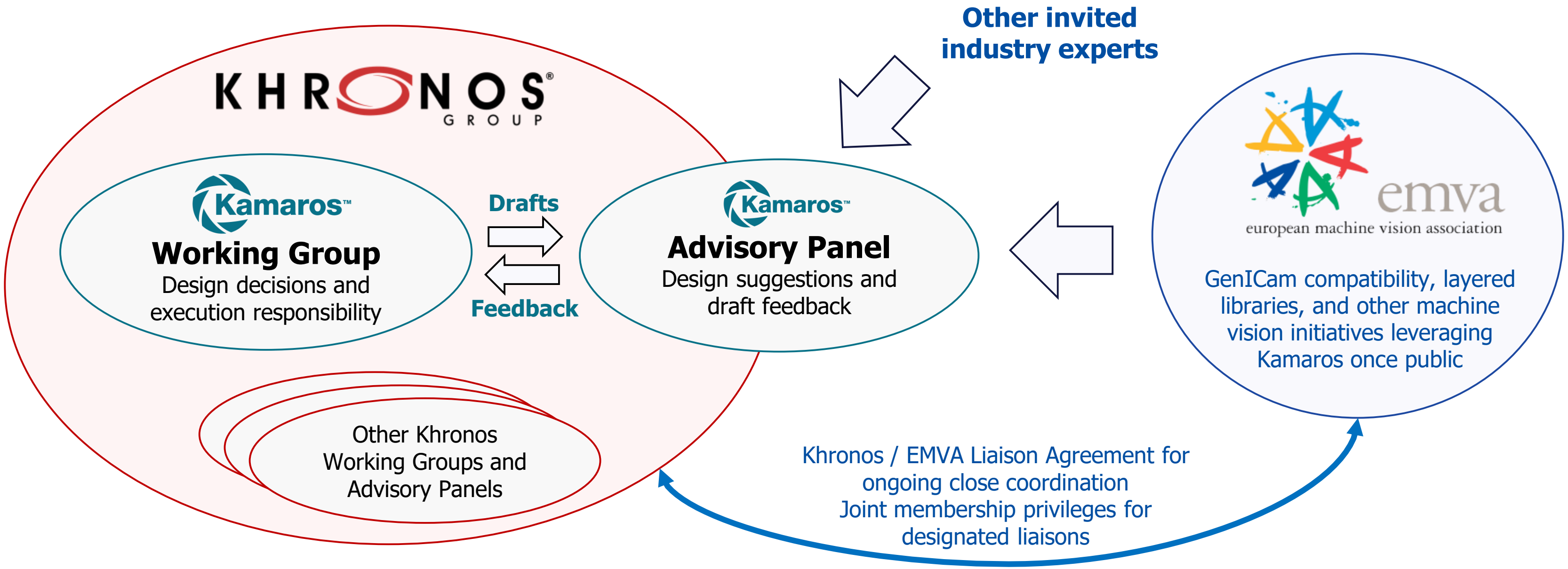
Loadable Layers

Command dispatch system to enable developers to use installable layers for validation, profiling, and debugging, etc.

Resource Management and Synchronization

Vulkan-style command queues well-suited to camera request processing

Kamaros Working Group Organization


KHRONOS GROUP
Kamaros™
Working Group
Design decisions and execution responsibility
Drafts
Feedback
Kamaros™
Advisory Panel
Design suggestions and draft feedback
Other invited industry experts
Other Khronos Working Groups and Advisory Panels
emva
european machine vision association
GenICam compatibility, layered libraries, and other machine vision initiatives leveraging Kamaros once public
Khronos / EMVA Liaison Agreement for ongoing close coordination
Joint membership privileges for designated liaisons

Any company is welcome to join Khronos for access to any Khronos Working Group

Get Involved!

Khronos is developing a growing family of open, royalty-free API standards for embedded and safety-critical markets

Any company is welcome to join Khronos to influence standards development
https://www.khronos.org/members/ or email memberservices@khronosgroup.org

More information on any Khronos APIs
https://www.khronos.org/

Khronos members can participate in the Kamaros Camera Working Group
https://www.khronos.org/kamaros