This webinar will begin shortly at
18:00 CET | 12:00 EST | 09:00 PDT

Accelerating Machine Learning with Vulkan

May 5, 2022
How to Participate

Speaker Questions
Please submit speaker questions at any time using the Zoom Q&A button (not Zoom chat). During the panel, we will put as many questions as possible to the speakers.

Help with Meeting Logistics
Please use Zoom Chat for logistical questions or if you are having issues with Zoom.

Recording and Slides
We are recording this webinar and will publicly post the slides and video at the Forum Home Page at https://www.khronos.org/machine-learning.

Survey
To help us design future Khronos Machine Learning Forum events, we appreciate you completing the short survey form that we will distribute after the session.
Accelerating Machine Learning with Vulkan

Khronos ML Forum
Neil Trevett, Khronos

ML Primitives Extension
Jeff Leger, Qualcomm

Cooperative Matrix Extension
Pierre Boudier

IREE Compiler Targeting Vulkan
Lei Zhang, Google

Q&A Panel Moderator
Kevin Petit, Arm
Introduction to the Khronos Machine Learning Forum

Neil Trevett
Khronos President
Khronos Open Standards Mission

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

Founded in 2000

~200 Members ~ 40% US, 30% Europe, 30% Asia
Khronos Compute Acceleration Standards

Increasing industry interest in parallel compute acceleration to combat the ‘End of Moore’s Law’

Higher-level Languages and APIs
- Streamlined development and performance portability
  - Single source C++ programming with compute acceleration
  - Intermediate Representation (IR) supporting parallel execution and graphics
  - Heterogeneous compute acceleration

Lower-level Languages and APIs
- Direct Hardware Control
  - Trained Neural Networks
  - Intermediate Representation (IR) from framework formats

Available APIs:
- Vulkan
- SPIR
- OpenCL
- SYCL
- OpenVX
- NNEF

Supporting technologies:
- GPU rendering + compute acceleration
- Intermediate Representation (IR) supporting parallel execution and graphics
- Heterogeneous compute acceleration

Platforms:
- CPU
- GPU
- FPGA
- DSP
- AI/Tensor HW
- Custom Hardware
Khronos Machine Learning Forum

Productive communication and cooperation on ML Acceleration
... between Machine Learning hardware and software communities

Forum is free to join, no NDA or IP commitments
Dedicated meetings, email and slack channels for group communication

Use cases, requirements, updates and presentations

Machine Learning Community

API guidelines, updates and roadmaps

Khronos Hardware and Software Members
Machine Learning Forum Meeting Series

Public Meetings

- **Khronos ML Summit**
  - October 2021
  - Session #1 Video
  - Session #2 Video

- **ML Summit Response Session**
  - January 2022

- **Vulkan ML Webinar**
  - You are here!

- **OpenCL ML Webinar**
  - 11th May 2022

- **SYCL and OpenVX Webinars..**
  - Being scheduled

Forum Member Meetings will start in July 2022
Input and requests for specific topics welcome!

https://www.khronos.org/machine-learning has all the information you need to join!
Khronos Machine Learning Resources

- Khronos Machine Learning Forum
  - https://www.khronos.org/machine-learning

- Khronos homepage for all Khronos Standards
  - https://www.khronos.org/

- OpenCL Resources and C++ for OpenCL documentation
  - https://github.com/KhronosGroup/OpenCL-Guide
  - https://www.khronos.org/opencl/assets/CXX_for_OpenCL.html

- OpenVX Tutorial, Samples and Sample Implementation
  - https://github.com/rgiduthuri/openvx_tutorial
  - https://github.com/KhronosGroup/openvx-samples
  - https://github.com/KhronosGroup/OpenVX-sample-impl/tree/openvx_1.3

- NNEF Tools
  - https://github.com/KhronosGroup/NNEF-Tools

- SYCL Resources
  - http://sycl.tech

- SPIR-V User Guide
  - https://github.com/KhronosGroup/SPIRV-Guide
Accelerating Machine Learning with Vulkan

Khronos ML Forum
Neil Trevett, Khronos

ML Primitives Extension
Jeff Leger, Qualcomm

Cooperative Matrix Extension
Pierre Boudier

IREE Compiler Targeting Vulkan
Lei Zhang, Google

Q&A Panel Moderator
Kevin Petit, Arm
Accelerating Machine Learning with Vulkan
5 May 2022

Ask the Experts Panel
Use Zoom Q&A to ask your questions at any time!
Thank you!

https://www.khronos.org/events/accelerating-machine-learning-with-vulkan