IWOCL 2026

清華大學
Tsinghua University

OpenCL—Building the Foundation for Computing in China

Hu He | Tsinghua University

May 6-8, 2026 | Heilbronn, Germany

清華大學 Tsinghua University

CONTENTS

01 The Landscape

02 The Solution

03 The New Era

04 The Contribution

THE LANDSCAPE

01. Huge Computational Demand

- Technology Imperative
  - Model parameter inflation
  - Inference overtaking training
- Market Dynamics
  - World's largest internet user base
  - Enterprise AI adoption
- Infrastructure Demand

￥140.19 billion (2024)

￥1.3 trillion (2029, projected)

China AI Chip Market Size

Source :
https://static.sse.com.cn/stock/disclosure/announcement/c/202509/002098_20250905_EBZW.pdf

THE LANDSCAPE

02. The De Facto Standard Ecosystem

NVIDIA CUDA

NVIDIA CUDA has become the de facto standard ecosystem for global compute research and development.

"As of GTC 2025 (March 2025), NVIDIA reports over **6 million** developers using CUDA globally .
The platform has recorded more than **53 million** cumulative downloads according to NVIDIA's official documentation"

Ecosystem Dependence

- Deep binding between technology stack and developer tools
- Talent with monolithic skill sets
- The software adaptation challenge for in-house developed chips

Diversity Decline

- Dominant ecosystem drives architectural convergence
- The de facto standard weakens the evolutionary momentum of open standards
- Demands outside the mainstream market are overlooked

Systemic Risk

- Supply chain and business continuity risk
- Imbalanced cost structure and bargaining power
- Technology homogeneity reduces long-term resilience

THE LANDSCAPE

03. Prosperous Hardware and Fragmented Ecosystem

Huawei Ascend

Optimized for CANN Stack

Hygon DCU

ROCm-compatible Architecture

Moore Threads

Native MUSA Architecture

Cambricon

MLU for Cloud & Edge AI
Neuware / BANG

Biren

BIRENSUPA

... and more architectures

Challenge : How do we enable developers to write once and deploy seamlessly across this entire hardware spectrum?

THE SOLUTION

01. OpenCL as The Unified Software Abstraction Layer

Open Standard & Vendor-Neutral

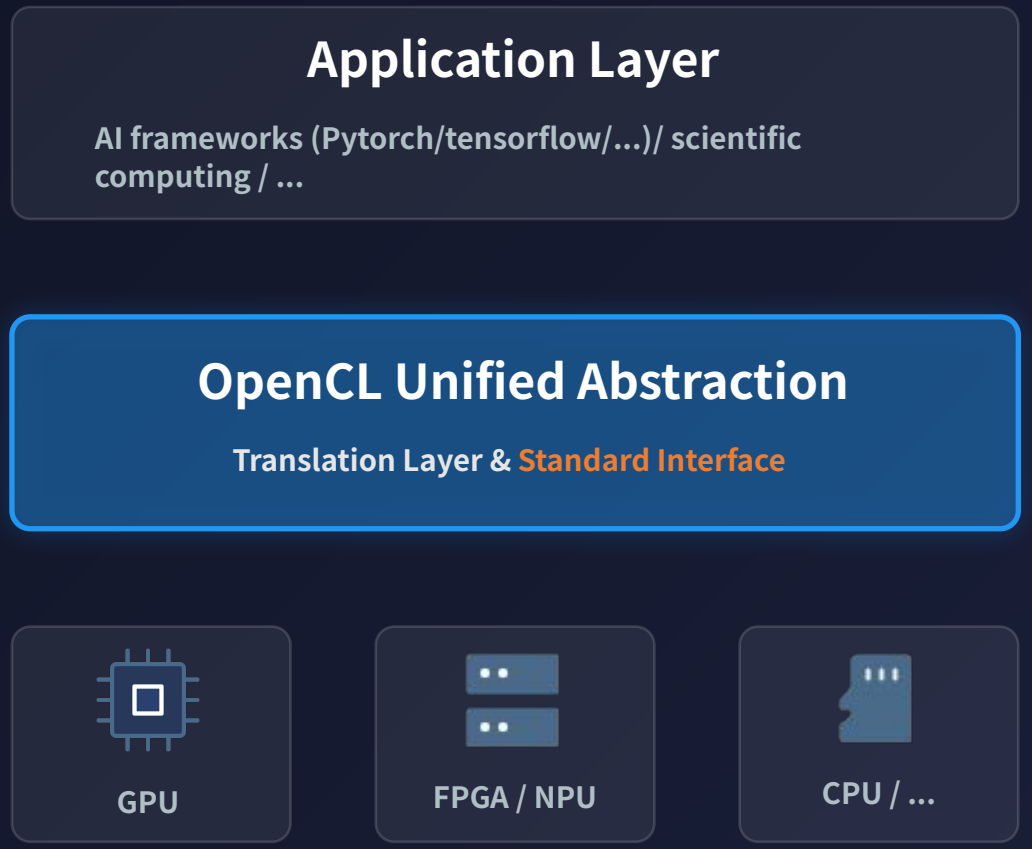
OpenCL is an open, royalty-free standard that is not limited by any proprietary technology and provides the possibility for collaborative technological innovation.

Native Heterogeneous Support & CTS

OpenCL natively supports almost all mainstream hardware types, including NVIDIA/AMD/Intel GPUs, CPUs, FPGAs, DSPs, RISC-V chips, etc., unifying various computing resources.

Reduce cross-platform migration costs

AI frameworks and scientific computing libraries only need to be adapted to OpenCL to support all hardware.



THE SOLUTION

02. Application Cases

Chinese chip companies are using OpenCL as a key breakthrough to break down ecosystem barriers, and are accelerating their deployment in areas such as edge computing, graphics, and computing performance.

Vendor	OpenCL Version	Profile	Native	Core Software Stack	Typical Products
MetaX	1.2	Full	Y	MXMACA	Xiyun C-series, Xisi N-series
VeriSilicon	1.1 / 1.2 / 3.0	Full/Embedded	Y	Verisilicon OpenCL SDK + Drivers	Vivante GC-series GPU IP, VIP9000-series NPU IP
Lisuan Tech	3.0	Embedded	Y	TrueGPU native OpenCL driver	G100 (7G100)
Moore Threads	3.0	Full	Y	Proprietary GPU driver + OpenCL ICD	S70/S80/S2000/S3000
Jingjia Micro	1.2 (JM7) / 3.0 (JM9)	Embedded	Y	Proprietary GPU Driver	JM9100, JM7200
Loongson	1.1 (Integrated GPU) / 3.0 (Discrete GPU)	Full/Embedded	Y	Proprietary GPU driver	2K3000/3B6000M (Integrated), 9A1000 (Discrete)

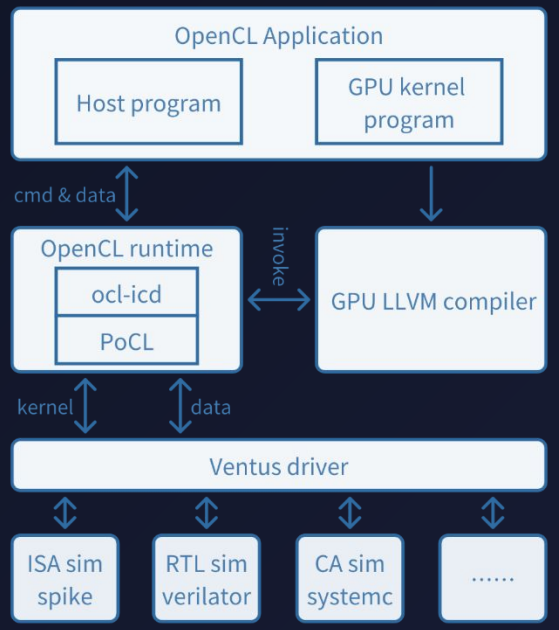
The New Era

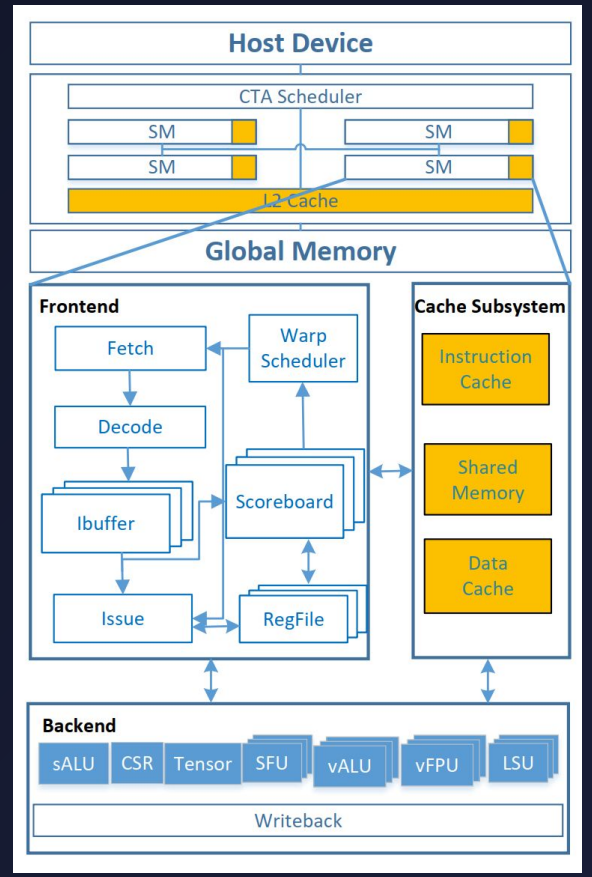
清華大學 Tsinghua University

01. Ventus OpenGPGPU Program ： OpenCL + RISC-V

Ventus GPGPU is China's first and only open-source GPGPU project, ranking among the top-performing in its category.

- Ventus GPGPU, based on RISC-V Vector Extensions (RVV), defines and open-sources the complete GPGPU instruction set architecture (ISA).
- Ventus GPGPU provides a full-stack open-source solution including instruction set architecture, microarchitecture design, and the OpenCL 2.0 software toolchain.
- The OpenCL software stack includes the OpenCL compiler, the OpenCL drivers, and the OpenCL libraries. The driver portion is implemented based on PoCL.
- The overall pass rate for OpenCL-CTS exceeds 85%.





https://github.com/THU-DSP-LAB

J. Li et al., "RISC-V-Based GPGPU With Vector Capabilities for High-Performance Computing," in IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 33, no. 8, pp. 2239-2251, Aug. 2025

The New Era

02. Ventus-PyTorch ： OpenCL + RISC-V GPGPU + AI

- **Core Architecture**

Layer	Component	Function
AI Layer	PyTorch Frontend	Native PyTorch code
Backend Layer	ventus-pytorch	PyTorch ops → OpenCL kernel translation
Compiler Layer	Ventus OpenCL Compiler (LLVM)	OpenCL C → Ventus GPGPU ISA, supports FP16/BF16/TF32
Runtime Layer	PoCL + Device Driver	OpenCL API scheduling & hardware dispatch
Hardware Layer	Ventus GPGPU (RISC-V + RVV)	SIMT execution, multi-precision tensor cores

- **Verified Results**

  - GPT-2 (TF32) — Foundational text generation validation
  - Pythia-410M (FP16) — Mid-scale model inference
  - Qwen2.5-0.5B (BF16) — End-to-end dialogue model execution

Significance : Validates the full-stack feasibility of OpenCL + RISC-V GPGPU + PyTorch !

https://github.com/THU-DSP-LAB/ventus-pytorch

The Contribution

01. CKAP (China Khronos Advisory Panel)

Current structure

China Khronos Advisory Panel (CKAP)

- Compute
- Graphics

- Professor He Hu (Tsinghua University) is the convenor/chair for CKAP
- Dennis Fu provides backend support

- Email reflector has been tested and working
- Meeting minutes and other docs have been uploaded to the official CKAP area
- Online meetings and F2F meetings held

OpenCL TSG chaired by Tsinghua University	SYCL TSG chaired by BOSC	glTF/Vulkan TSG chair hasn't been officially selected

Communicate with WG in Causeway CKAP area, email reflector, etc.

OpenCL SYCL SPIR-V Vulkan glTF

From "Standards Follower" to "Standards Co-builder"

The Contribution

02. OpenCL Certification Program

- Jointly held by Khronos and Tsinghua University
  - A local Tsinghua University institution (the licensee) is authorized to conduct certification.
  - Tsinghua Shenzhen International Graduate School organizes training
  - Certificates are provided to students who pass the certification exam
- Program will cover multi-level multi-segments
  - basic graduate level course
  - Engineers with working experience
  - financial sector Quantitative Trading
  - Management level people
- Pilot project to test out the pipeline
  - If successful, glTF, Vulkan, OpenXR could be next

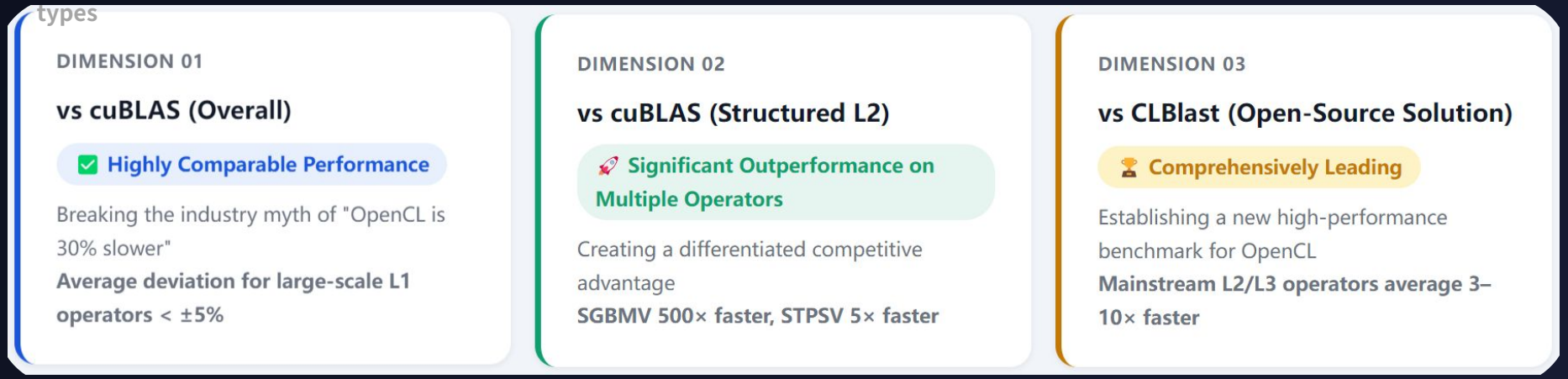
Significance: To cultivate heterogeneous talent, promote operator library migration, and contribute to the global OpenCL ecosystem.

The Contribution

03. Self-Developed OpenCL BLAS Library

- **Core Performance Results (on NVIDIA RTX, OpenCL 3.0)**

  Full BLAS coverage (L1/L2/L3), single/double precision, real & complex types

  Level-1 : Highly on par with cuBLAS, confirming the implementation quality of basic operators
  Level-2 : Achieving orders-of-magnitude performance leadership in banded/packed storage scenarios, the biggest differentiation highlight
  Level-3 : Gap with cuBLAS continues to narrow; already established clear leadership over open-source CLBlast

- **Strategic Value**
  - Proves OpenCL can match cuBLAS performance — shatters the "open = slower" myth
  - Vendor-agnostic math backbone for AI frameworks, scientific computing, and domestic chips (e.g., RISC-V GPGPU)
  - Lowers migration cost for new hardware platforms

IWOCL 2026

清華大學 Tsinghua University

Thanks

Grounded in open standards, empowered by collaborative ecosystems.