

VCK5000 Versal Development Card





Product Description

The AMD VCK5000 Versal development card is built on the AMD 7nm Versal[™] adaptive SoC architecture and is designed for (AI) Engine development with Vitis endto-end flow and AI Inference development with partner solutions. For full Vivado[™] flow and device customization, please contact sales.

Product Overview

The AMD VCK5000 Versal[™] development card is built on the AMD 7nm Versal adaptive SoC architecture and is designed to optimize 5G, data center compute, AI, signal processing, radar, and many other applications. Fully supported by Vitis[™], Vitis AI, and partner solutions like Mipsology Zebra® and Aupera® VMSS, the VCK5000 domain-specific architecture brings strong horsepower per watt while keeping ease-of-use in mind with C/C++, software programmability.

Delivering the near 100% compute efficiency per watt in standard AI benchmarks and 2x TCO compared to the flagship nVidia GPUs, the VCK5000 development platform is ideal for CNN, RNN, and NLP acceleration for your cloud and edge applications.



Al Inference Development

If you are an AI developer, bring your TensorFlow and PyTorch trained models to directly infer on Versal using Mipsology Zebra and build, configure, and deploy computer vision applications on FPGA platforms with Aupera Video Machine Learning Streaming Server solution.

Key Features

Explore partner solutions and articles, and learn about the key features for AI Inference Development with the VCK5000

2x TCO Reduction vs Mainstream nVidia GPUs

- 2x perf/w and perf/\$ compared to Nvidia Ampere with standard MLPerf Models
- Achieves 90% compute efficiency
- Consume less than 100W at card level

2x End-to-End Video Analytics Throughput vs nVidia GPUs

- Full pipeline from H.264 decode to computer vision to up to 10 AI models
- Video decode and CV run on x86 CPU or discrete U30 Alveo card
- Plug-in based pipeline composition from FFmpeg / Gstreamer
- ML Heavy: H.264 Decode + Yolov3 + 3x ResNet-18
- Video Heavy: H.264 Decode + tinyYolov3 + 3x ResNet-50



Easy to Use with Familiar Frameworks

- Easy-to-use software flow for any CPU & GPU users, no hardware programming required
- Run inference from Tensorflow framework directly on board
- State-of-the-art model supported with mainstream frameworks PyTorch, TensorFlow, TensoFlow 2 and Caffe

Partner Solutions

Mipsology Zebra AI Inference Solutions & Aupera Video Machine Learning Streaming Server

Mipsology Zebra Al Inference Solution

Mipsology Zebra Software

Zebra combines high performance with unprecedented ease of use and is AMD's preferred AI accelerator for computing neural networks for image-based AI applications.

Zebra seamlessly replaces GPU/CPU to compute any image based neural network faster and with lower power consumption. And with Zebra, there is no need to retrain or make any changes to your network or application. Zebra deploys with a simple Linnux command, so you don't need FPGA knowledge to use Zebra.



Aupera Intelligent Video Analytics Solution

Aupera Video Machine Learning Streaming Server Solution

Aupera VMSS is a software framework for Video AI inferencing applications. VMSS2.0 allows users to rapidly build, configure, and deploy Computer Vision pipelines using a graphical user interface (GUI); with no coding. Custom pipelines can be easily built using Aupera's node toolkit, decoding, pre-processing, post-processing, etc.; or by creating custom nodes which can be uploaded, built, tested, and used all from the GUI.

Aupera's commercial video AI applications can also be configured, launched, and visualized through this framework. Users have a choice of checking the results of their custom pipeline through video overlays or by sending text output.

VCK5000 in Action

Data centers are increasingly turning to artificial intelligence to manage various tasks from monitoring equipment to server optimization. At the heart of the data center, FPGA-based adaptive computing is proving itself to be, in many cases, the most efficient and cost-effective solution for running complex AI workloads.

Here are the best uses of the VCK5000 developer card combined with Vitis AI from our 2021 Adaptive Computing Challenge



AI Engine Development

If you are looking to implement algorithm acceleration with AI engine and programmable logic, we provide AI engine C/C++ high-level abstraction APIs and Vitis Accelerated Libraries. The Vitis end-to-end flow is developed using C/C++ to run on X86 or embedded processors and manage runtime interactions with the accelerator with XRT. The hardware component, or kernel, can be developed using C/C++, or RTL target on PL or AI Engines.

Key Features

Power and Performance

Up to 10x performance improvements compared to previous generation AMD UltraScale+[™] with less power in diverse applications

Industry-leading compute power: up 145 TOPS (int8); 37 TOPS (int16); 12T FLOPs (fp32)

Software Familiarity

A familiar software development flow with Vitis unified software platform

Accelerate your applications faster with AI engine C/C++ high-level abstraction APIs and Vitis Accelerated Libraries

Mixed Kernel Development

Customize your own data pipeline with mixed kernels

Develop AIE kernels in C/C++, PL kernels in RTL or HLS, and let Vitis stitch together the full system



Get Started with AI Engine Development

Step 1: Purchase

Purchase VCK5000 production silicon-based card

Step 2: Access Secure Site

Request access to the VCK5000 Versal Development Card Secure Site

Step 3: Get Started

Follow the getting started steps and installation guide in the VCK5000 Versal Development Card Secure Site

Articles for AI Engine Development

BOARD SPECIFICATIONS

Card Specifications	VCK5000		
Device	VC1902		
Compute	Active	Passive*	
INT8 TOPs (peak)	145	145	



Dimensions			
Height	Full	Full	
Length	Full	3/4	
Width	Dual Slot	Dual Slot	
Memory			
Off-chip Memory Capacity	16 GB	16 GB	
Off-chip Total Bandwidth	102.4 GB/s	102.4 GB/s	
Internal SRAM Capacity	23.9 MB	23.9 MB	
Internal SRAM Total Bandwidth	23.5 TB/s	23.5 TB/s	
Interfaces			
PCI Express	Gen3 x 16 / Gen4 x 8	Gen3 x 16 / Gen4 x 8	
Network Interfaces	2x QSFP28 (100GbE)	2x QSFP28 (100GbE)	



Logic Resources				
Look-up Tables (LUTs)	899,840	899,840		
Power and Thermal				
Maximum Total Power	225W	225W		
Thermal Cooling	Active	Passive		

* We will ship the Active board only. If you remove the fans from the VCK5000, following the Hardware Installation Guide, it becomes Passive.