Arm Ethos-N Processor Series

KEY FEATURES & BENEFITS

+ Scalable Performance

Delivering up to 10 to 1 TOP/s of single core performance with multicore scalability, supporting up to eight NPUs in a cluster and up to 64 NPUs in mesh systems.

+ Highly Efficient

Achieving up to 5 TOPs/W with low DRAM traffic (MB/Infr) through compression, clustering, and operator cascading.

Optimized Design

Driving up to 225% convolution performance uplift using Winograd on 3x3 kernels, delivering up to 90%MAC utilization.

Unified Software and Tools

Develop, deploy and debug with the Arm AI platform using online or offline compilation and Arm Development Studio 5 Streamline.

arm Ethos<u>-N NPUs</u>



Arm Ethos-N NPUs address the ML inference requirements of multiple markets with a unified software platform.

Powering AI Inference from Cloud to Edge to Endpoint

Highlights

+ Arm Al Platform

Arm AI platform optimally supports all popular frameworks and operators with the ability to add new networks/operators on Cortex CPU, Mali GPU, or through Ethos NPU driver updates.

+ Inference Deployment Flexibility

Efficiently execute your models using ahead of time compilation with TVM or online interpreted with Arm NN or through Android Neural Networks API (NNAPI).

+ Target Multiple Market Segments

Scalable single core from 1 to 10 TOP/s single core and up to 640 TOP/s through multi-core and mesh technologies for mobile, auto and infrastructure devices.

+ Longer Battery Life

Up to 40% lower DRAM traffic (MB/Infr) through improved weight and feature map compression, clustering, and cascading.

+ Advanced Operator Cascading

Optimized NPU Cascading to improve battery life by automatically scheduling graphs into sub-graphs of connected operators to optimally utilize on device SRAM avoiding power intensive DRAM access.

+ Early Network Performance Feedback

Ethos-N NPU Static Performance Analyzer (SPA) allows early performance feedback before hardware availability.

+ Security

Supports comprehensive security solution in conjunction with Arm SMMU & CryptoCell IP.

System Integration (SMMU)

Allows for support and protection of memory and easy handling of multiple users via tight system integration through the ACE-Lite master port and optional SMMU integration.

KEY USE CASES FOR THE ETHOS PROCESSOR SERIES

- ÷ Object classification
- ÷ Object detection
- ÷ Face detection/identification
- ÷ Human pose detection/ hand-gesture recognition
- 4 Image segmentation
- ٠ Image beautification
- Super resolution ÷
- 4 Framerate adjustment (super slow-mo)
- ÷ Speech recognition
- 4 Sound recognition
- Noise cancellation ÷
- ÷ Speech synthesis
- ÷ Language translation

MARKET SEGMENTS





Smart Camera





 \triangleright

Smart Home

쁥 Consumer















IVI/ADAS





Infrastructure

Specifications

				r	r	
Key Features		Ethos-N78	Ethos-N77	Ethos-N57	Ethos-N37	
	Performance	10, 5, 2, 1 TOP/s	4 TOP/s	2 TOP/s	1 TOP/s	
	MAC/Cycle (8x8)	4096, 2048, 1024, 512	2048	1024	512	
	Efficient convolution	Winograd support delivers 2.25x peak performance over baseline				
	Configurability	90+ Design Options	Singl	ngle Product Offering		
	Network support		CNN and RNN Int-8 and Int-16			
	Data types					
	Secure mode		TEE or SEE			
	Multicore capability		8 NPUs in a cluster 64 NPUs in a mesh			
Memory System	Embedded SRAM	384KB - 4MB	1-4 MB	512 KB	512 KB	
	Bandwidth reduction	Enhanced Compression	Extended compression technology, layer/operator fusion, clustering, and workload tilling			
	Main interface	1xAXI4 (128-bit), ACE-5 Lite				
Development Platform	Neural frameworks	TensorFlow, TensorFlow Lite, Caffe2, PyTorch, MXNet, ONNX				
	Inference deployment	Ahead of time compiled with TVM Online interpreted with Arm NN Android Neural Networks API (NNAPI)				
	Software components	Arm NN, Arm NPU software (compiler and support library, driver) Heterogeneous layer-by-layer visibility in Development Studio 5 Streamline Ethos-N Static Performance Analyzer (SPA), Arm Juno FPGA systems, Cycle Models				
	Debug and profile					
	Evaluation and early prototyping					

To find out more about the Ethos processor series, visit developer.arm.com/ethos



All brand names or product names are the property of their respective holders. Neither the whole nor any part of the information contained in, or the product described in, this document may be adapted or reproduced in any material form except with the prior written permission of the copyright holder. The product described in this document is subject to continuous developments and improvements. All particulars of the product and its use contained in this document are given in good faith. All warranties implied or expressed, including but not limited to implied warranties of satisfactory quality or fitness for purpose are excluded. This document is intended only to provide information to the reader about the product. To the extent permitted by local laws Arm shall not be liable for any loss or damage arising from the use of any information in this document or any error or omission in such information.