Arm Ethos-N57
NPU

AT A GLANCE
Optimized for the most cost- and power-sensitive designs, Ethos-N57 delivers premium AI experiences in mainstream phones and DTVs. With the highest performance, open-source software framework and the largest AI ecosystem, the Arm AI platform makes it easy to develop on Arm.

Balanced Efficiency and Performance
ML Inference Processor

What’s New?

- **Network Support**
  Flexible design supports a variety of popular neural networks, including CNNs and RNNs, for classification, object detection, image enhancements, speech recognition and natural language understanding.

- **Futureproof Operator Coverage**
  The MAC engine flexibly decomposes arbitrarily sized kernels with stride and dilation modes including convolution, deconvolution, depthwise separable, and vector product. Programable Layer Engines execute layers not supported by the MAC engine, supporting various primitives, activation functions and future operators.

- **Mixed Precision**
  Supports both Int-8 and Int-16; lower precision Int-8 for classification and detection tasks; high-precision Int-16 for HDR image enhancements and audio tasks.

- **Compression and Winograd Convolution**
  MAC engines provide decompression, activation, Winograd transformation and scaling. Winograd accelerates common filters by 225% compared to other NPUs, allowing actual performance to far exceed architectural performance.

- **Multicore**
  Supports up to eight processors in a tightly coupled cluster, with the ability to process multiple networks in parallel or a single, large network split across cores. Larger configurations of up to 64 cores are supported through Arm CoreLink mesh technology.

- **Weight and Feature Map Compression**
  Minimizes system memory bandwidth by 1.5-3x, reducing off-chip memory accesses by 90% through extended compression technologies, targeting both weight and activations.

- **Security**
  Supports TrustZone system security with configurable secure queues for multiple users and flexible processing in the TEE or SEE, providing layered security to protect both ML models and input data.

- **System Integration (SMMU)**
  ACE-Lite master port and optional SMMU (System Memory Management Unit) integration allows for support and protection of memory and easy handling of multiple users.
### Key Use Cases for Ethos-N57
- Object classification
- Object detection
- Face detection/identification
- Human pose detection/hand-gesture recognition
- Image segmentation
- Image beautification
- Super resolution
- Framerate adjustment (super slow-mo)
- Speech recognition
- Sound recognition
- Noise cancellation
- Speech synthesis
- Language translation

### Specifications

<table>
<thead>
<tr>
<th>Key Features</th>
<th>Performance (at 1GHz)</th>
<th>Up to 1 TOP/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>MACs (8x8)</td>
<td>512</td>
<td></td>
</tr>
<tr>
<td>Data Types</td>
<td>Int-8 and Int-16</td>
<td></td>
</tr>
<tr>
<td>Network Support</td>
<td>CNN and RNN</td>
<td></td>
</tr>
<tr>
<td>Efficient Convolution</td>
<td>Winograd support</td>
<td></td>
</tr>
<tr>
<td>Sparsity</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Secure Mode</td>
<td>TEE or SEE</td>
<td></td>
</tr>
<tr>
<td>Multicore Capability</td>
<td>8 NPUs in a cluster</td>
<td>64 NPUs in a mesh</td>
</tr>
</tbody>
</table>

### Memory System
- Embedded SRAM: 512KB
- Bandwidth Reduction: Extended compression technology, layer/operator fusion
- Main Interface: 1xAXI4 (128-bit), ACE-5 Lite

### Development Platform

- **Neural Frameworks**: TensorFlow, TensorFlow Lite, Caffe2, PyTorch, MXNet, ONNX
- **Neural Operator API**: Arm NN, AndroidNN
- **Software Components**: Arm NN, neural compiler, driver and support library
- **Debug and Profile**: Layer-by-layer visibility
- **Evaluation and Early Prototyping**: Arm Juno FPGA systems and cycle models

### Market Segments

- **Mainstream Mobile**
- **STB/DTV**
- **Smart camera**
- **Robotics and Drones**
- **Rich Embedded/IoT Endpoints**
- **Consumer Devices**
- **Automotive**: Cockpit, IVI, Entry

To find out more about the Ethos-N57 processor, visit developer.arm.com/ethos-n57

---

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.