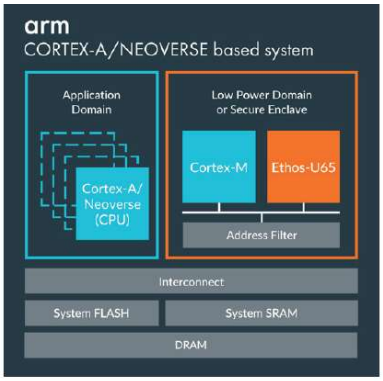
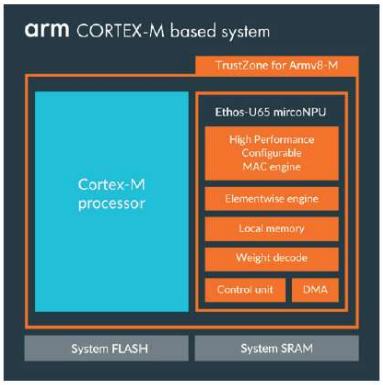
在 AI 裝置的新世界中帶動創新

利用 Arm Ethos-U 微型神經網路處理器(NPU)家族的最新新力軍，在各種嵌入式裝置打造低成本、高效率的 AI 解決方案。Ethos-U65 保有 Arm Ethos-U55 的功耗效率，並將應用性延伸到 Arm Cortex-A 與 Arm Neoverse 架構系統，且達成兩倍的終端機器學習(ML)效能。





Ethos-U65 可以結合高性能的 Cortex-A、Neoverse 系統或 Cortex-M 低功耗裝置進行許多不同的應用。

重點:

- 全新使用場景

能應用於要求嚴苛的 AI 使用場景，如物件檢測與分段，效能(推論/秒)提升 150%，並支援從 DRAM 進行讀/寫。

- 支援複雜的模型

在具更寬的 AXI 介面(128 位元)與支援 DRAM 的 Cortex-A 系統中，在豐富的作業系統下能處理複雜的作業負載，能為流行的網路帶來平均 150%的效能提升(推論/秒)。

- 整合的直接記憶體存取(DMA)

透過 AXI5 主介面連接到系統記憶體的 DMA，提前提取權重與啟動。

- 節能

相較前世代 Cortex-M，對於如 ASR 等 ML 作業負載，最高可減少 90%能源。

市場區隔:

- 面向未來的運算子覆蓋範圍

負責繁重運算的運算子直接在微型 NPU 上運行，例如卷積(convolution)、LSTM、RNN、共用、啟動函數與原始元素級別函數。其它核心則在緊密耦合的 Cortex-M 上，使用 CMSIS-NN 自動運行。

- 離線優化

對神經網路、執行的運算子、層融合以及層重定序進行離線編譯與優化，以提升效能且最高可以降低 90%的系統記憶體需求。與非優化的定序相比，可以達成效能提升並降低功耗。

- 元素級別引擎

針對常用的元素級別運算進行優化，如經常使用的 scaling、LSTM、GRU 運算的加、乘與減。可以促成由這些類似的原始運算組成的未來運算子。

Ethos-U65 的關鍵特色與優點:

- 延展效能與效率

在最小的面積解鎖全新視覺與語音使用場景，效能提升兩倍(與 Ethos-U55 相比)，在 0.6mm^2(16 奈米製程)面積上達成 1 TOP/秒。

- 彈性的整合

利用豐富的作業系統與 DRAM 支援性，在 Cortex-A 與 Neoverse 系統打造低成本、高效率系統；以及利用極為成功的 Ethos-U 架構，在 BareMetal

或 RTOS SRAM/FLASH 系統上，以 Cortex-M 打造系統。

- 統一的軟體與工具

利用橫跨 Arm Cortex、Neoverse 與 Ethos-U 處理器的共用工具，以 Arm Endpoint AI 解決方案打造、部署與除錯 AI 應用。

- 強化的設計

支援具延伸運算子支援性的流行網路，提供更寬的 AXI 介面，並把 ECC 加入內部 RAM 以提升可靠性。

- 混合精度

支援 Int-8 與 Int-16：供分類與檢測任務使用的較低精度；供音訊與有限 HDR 影像強化使用的高精度 Int-16。

- 無損耗的壓縮

先進、無損的模型壓縮，最多可以讓模型的尺寸縮小 75%，提升系統推論效能並降低功耗。

Ethos-U65 關鍵使用場景:

- 物件分類
- 物件檢測
- 人臉檢測/識別
- 人類姿勢檢測
- 手勢辨識
- 影像分段
- 影像美化
- 超解析
- 語音辨識
- 聲音辨識
- 降噪

Ethos-U65 規格:

Key Features	Performance (At 1GHz)	512 GOPS/s to 1 TOP/s
	MACs (8x8)	256, 512
	Utilization on popular networks	Up to 85%
	Data Types	Int-8 and Int-16
	Network Support	CNN and RNN/LSTM
	Winograd Support	No
	Sparsity	Yes
Memory System	Internal SRAM	55 to 104 KB
	System Interfaces	Two 128-bit AXI
	External On Chip SRAM	KB to Multi-MB
	Compression	Weights only
	Memory Optimizations	Extended compression, layer/operator fusion
Development Platform	Neural Frameworks	TensorFlow Lite Micro
	Operating Systems	Bare-metal, RTOS, Linux
	Software Components	TensorFlow Lite Micro Runtime, CMSIS-NN, Optimizer, Driver
	Debug and Profile	Layer-by-layer visibility with PMUs
	Evaluation and Early Prototyping	Performance Model, Cycle Accurate Model, or FPGA Evaluations

想了解 Ethos-U65 處理器，請點此 https://developer.arm.com/ethos-u65