



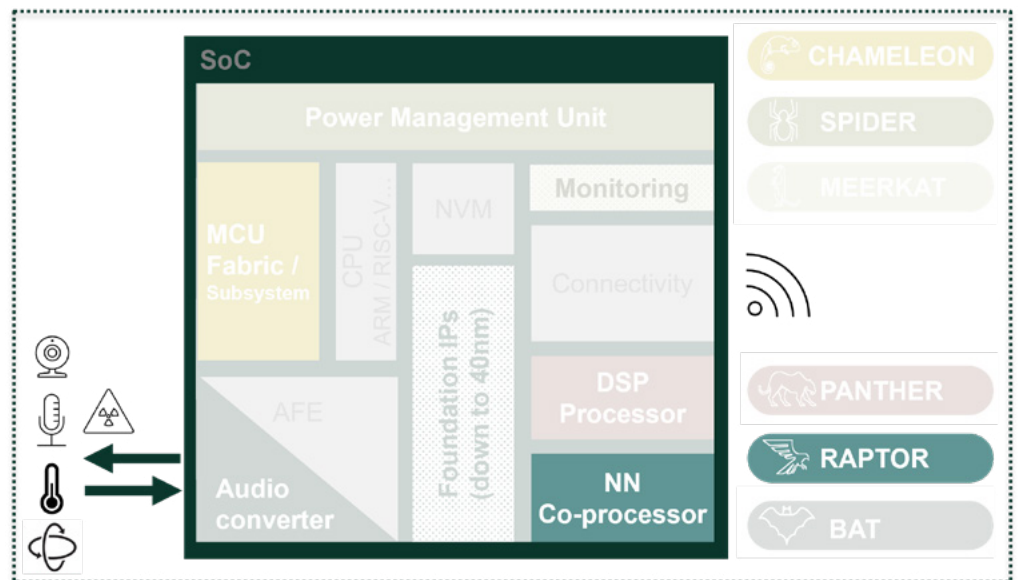
Speed up your algorithms at a fraction of the energy spent by your CPU

APPLICATIONS

- Home appliance: door lock, smart speaker, voice controlled system
- Image classification: General Purpose and Ultra Low Power MCU
- Voice controlled equipment: Audio, health, industrial, metering, metrology
- Connectivity and low power IoT end-nodes: Smartwatches, TWS, Smart Shoes

Tiny RAPTOR is a Power efficient Neural Processing IP Platform specialized in sound and vision. Tiny RAPTOR's near-memory computing architecture is composed of a DMA, local memory and up to 128 processing elements in its most powerful version.

Tiny RAPTOR is a fully programmable accelerator designed to execute deep neural networks (DNN) in an energy-efficient way. It reduces the inference time needed to run Machine Learning (ML) Neural Networks (NN). Tiny RAPTOR fits particularly well within any MCU subsystem, in particular Dolphin Design's solution (CHAMELEON).



KEY FIGURES

- 3x higher power efficiency compared to state-of-the-art NPU (KWS-TinyML)
- 30mW power @ mobilenet v2 image classification (224x224-60FPS-500MHz)
- Up to 256GOPS peak at 1GHz
- 2.2TOPS/W computing efficiency
- Small footprint (0.045mm² in 22nm for 32 GOPS)

TECHNOLOGY FOR BETTER FUTURE

The solution to deal with data deluge, while preventing the increase of power consumed by data centers, is known as Edge AI. This solution consists in transferring most of the processing intelligence from the cloud to the sensor. It translates into an unprecedented need to increase performances of «smart devices» by a factor of 1,000 at constant energy consumption.

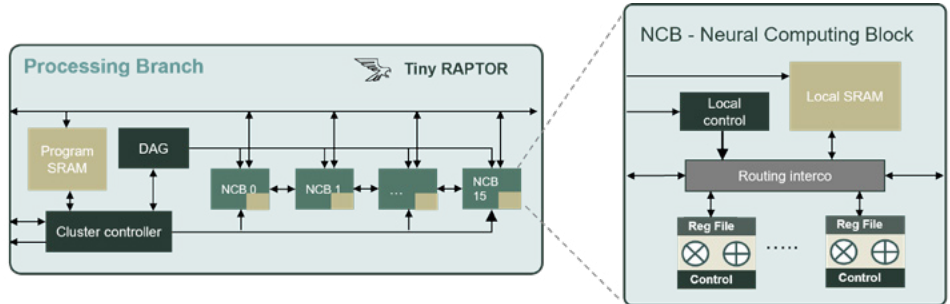
With its SPEED IP platform, Dolphin Design is positioned as THE provider of solutions for Edge AI System-on-Chip designers. We enable our customers to do much more with less energy resulting in major benefits on environment.





KEY BENEFITS

- Specialized computing to achieve ultimate energy efficiency
- Plugin accelerator to enhance the performance of your MCU
- Hardware flexibility to cover various NN structures
- Real time image and audio processing
- Reduced memory footprint with compression
- Native compatibility with standard AI frameworks



	MACs	Embedded memory	DMA	Peak processing power	Area (mm ²) (@ 22nm)	Market
Tiny RAPTOR V1	32	128kB	No	32 GOPs	0.045	
Tiny RAPTOR V2	32-128	512kB	Yes	128 GOPs	0.18	

KEY FEATURES

- Pre-verified and highly Flexible TCDM for near memory processing
- Complete SDK (AI Toolchain, ISS, Drivers)
- Scalable from 32 to 128 MAC/ cycle
- Signed and unsigned 8-bit architecture with SIMD support
- Lossless weight compression
- Embedded DMA for blind weight and activation transfer
- Events manager

TURNKEY NPU HARDWARE

Tiny RAPTOR is a specialized neural processor that implements up to 128 processing elements and the necessary control and arithmetic logic to execute Machine Learning Algorithms of predictive models.

FLEXIBLE NETWORK SUPPORT

Tiny Raptor’s scalable architecture is composed of a single branch to optimize the memory transfer within this branch embedding up to 16 Neuro Computing Blocks. Each of these block contains 8 processing elements, the embedded data memory and the routing interconnect.

COMPLETE SDK

Tiny RAPTOR comes with a complete toolchain running the most popular frameworks of machine learning models AI environments (SDK with Dolphin Design’s specific HAL drivers and documentation package) which complements CHAMELEON deliverables.

NEW USE CASE

Tiny RAPTOR specific architecture will accelerate the performance of common machine learning tasks and inference sequences for image classification, object detection, keyword spotting and other predictive models.

