IMG Series4 NNA

BEYOND NEXT-GEN AI

What is the IMG Series4 NNA?

IMG Series4 is a ground-breaking neural network accelerator (NNA) for the automotive industry to enable ADAS and autonomous driving.

Current systems deployed for testing and developing autonomous driving are physically large and very power-hungry, yet they are also underpowered in terms of performance. As such, they are not suitable for large scale commercial deployment, where power consumption is a significant issue, particularly in electric cars. What is required is an "edge" solution that offers:

- Very high performance
- Low latency
- Power efficiency

Meeting, and indeed exceeding these demands, Series 4 is set to become the edge platform of choice for the automotive industry. The key metric for neural network accelerators is the number of tera operations per second (TOPS) that can be performed. Current shipping solutions top out at 72 TOPS. The Series4 is a multi-core solution that can scale beyond 600 TOPS: it is a beyond next-gen solution.

IMG Series4 NNA features

Multi-Core Scalability and Flexibility

Multi-core allows for flexible allocation and synchronisation of workloads across the cores. Imagination's software, which provides fine-grained control and increases flexibility through batching, splitting and scheduling of multiple workloads, can now be exploited across any number of cores.

Phenomenal bandwidth efficiency

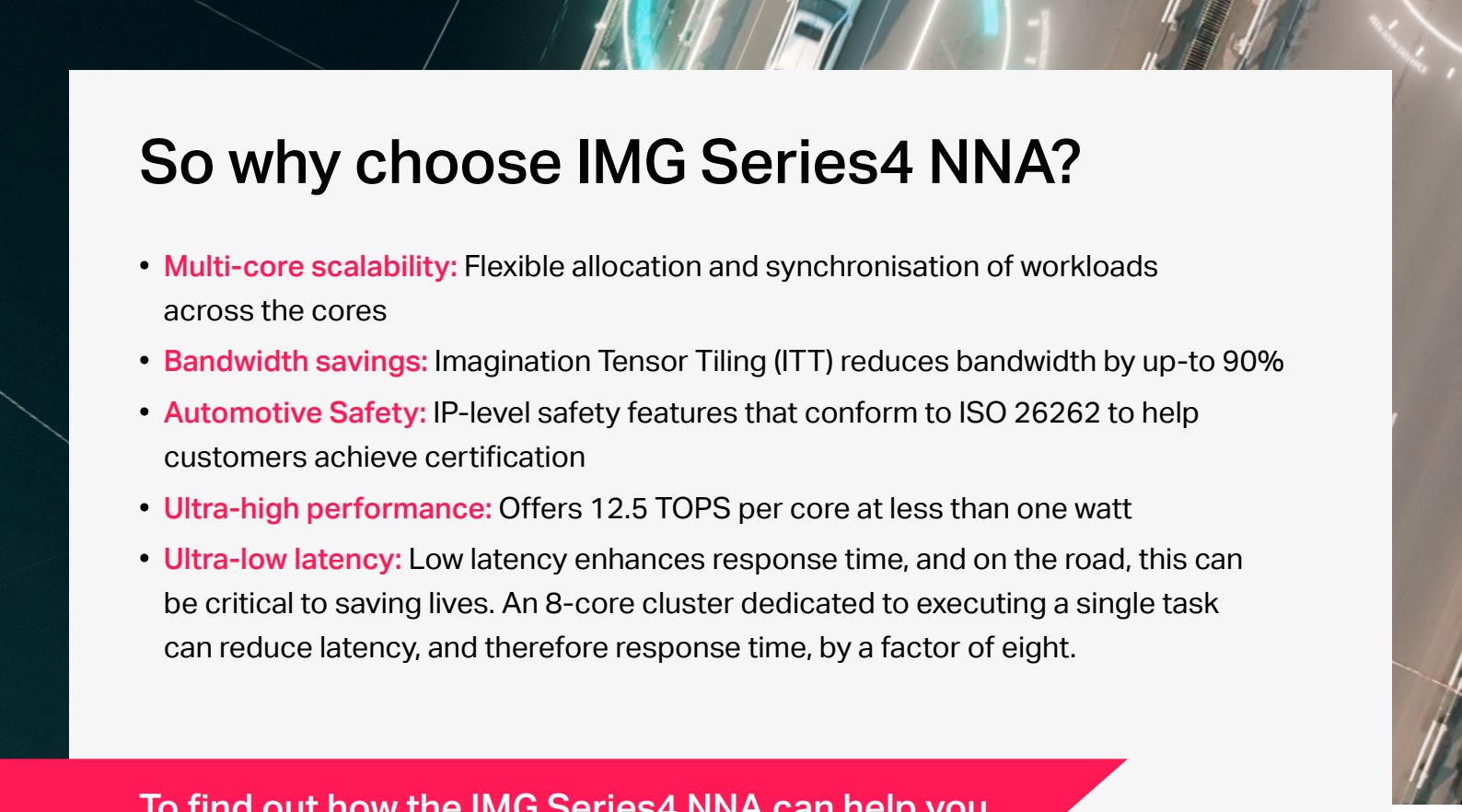
Processing tensors, the complex 3D maths performed inside a neural network, requires going out to memory and back across the bus. This takes time, consumes memory bandwidth, and absorbs power. Imagination Tensor Tiling (ITT), new to IMG Series4, is a patent-pending technology that solves this problem. It efficiently packages up tensors into tiles, which are then processed in groups for which all the intermediate data is stored in local on-chip memory, greatly minimising data transfers between layers of the network. This reduces bandwidth consumption by up to an incredible 90%.

Ultra-low latency

By combining all the cores into a 2, 4, 6 or 8-core cluster, they can be dedicated to executing a single task, reducing latency, and therefore response time, by a corresponding factor.

Built-in safety

IMG Series4 includes IP-level safety features that are built using a design process that conforms to ISO 26262, the standard for safety in automotive electronics, thus helping our customers to achieve certification. Via hardware safety mechanisms that protect the compiled network, the execution of the network and the data-processing pipeline, IMG Series4 enables functionally safe neural networks inferencing, without impacting performance.



So why choose IMG Series4 NNA?

- **Multi-core scalability:** Flexible allocation and synchronisation of workloads across the cores
- **Bandwidth savings:** Imagination Tensor Tiling (ITT) reduces bandwidth by up-to 90%
- **Automotive Safety:** IP-level safety features that conform to ISO 26262 to help customers achieve certification
- **Ultra-high performance:** Offers 12.5 TOPS per core at less than one watt
- **Ultra-low latency:** Low latency enhances response time, and on the road, this can be critical to saving lives. An 8-core cluster dedicated to executing a single task can reduce latency, and therefore response time, by a factor of eight.

To find out how the IMG Series4 NNA can help you visit imaginationtech.com/IMGSeries4NNA