

Accelerator for the Automotive Stream

The objective of the EPI automotive stream is to support Autonomous Driving Systems (ADS) based on a mainstream automotive safety/security MCU and two or more 'number crunchers' derived from the General Purpose Processor (GPP) (Figure 1). Each 'number cruncher' will implement the perception and path planning functions. As the GPP is based on ARM cores, it provides high performances for 64-bit and 32-bit floating-point computations. These are applicable to the path planning functions of ADS. The GPP also embeds compute units dedicated to acceleration, in particular the MPPA tiles for vehicle perception. Key vehicle perception functions include: sensor data segmentation and fusion; object detection and tracking; coupling with the localization functions. These functions require an integrated mix of CPU, GPU, and NPU (neural processing unit) capabilities, while meeting the soft real-time constraints of the AUTOSAR Adaptive standard.

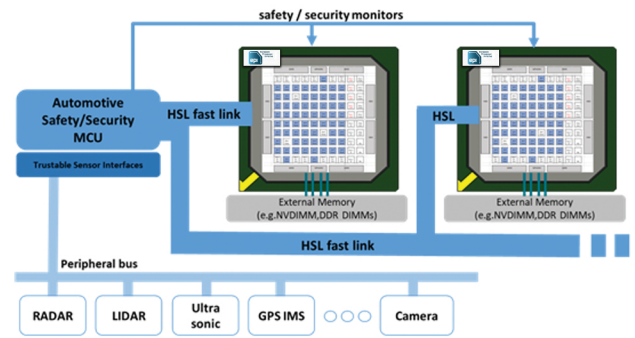


Figure 1 Automotive Stream embedded HPC platform

The MPPA accelerator tile for EPI features:

- Four Processing Engine (PE) cores.
- Resource Management (RM) core.
- Full crossbar local interconnect.
- 2MB local memory (SPM or L2\$).
- Micro-programmable DMA engine.
- Debug/trace Support Unit (DSU).
- Connection to the GPP AMBA 5 NoC.

Each PE or RM core features:

- 6-issue 64-bit VLIW architecture.
- 256-bit wide Load/Store Unit (LSU).
- 128-bit wide high-performance FPU.
- Tightly-coupled tensor coprocessor.

The Kalray software development framework is based on the C/C++/OpenMP/POSIX standards. It provides optimized application libraries (OpenCV, CNN, FFT, BLAS) and implements offloading API supported in GCC or LLVM. Offloading from ARM CPUs is based on three key developments:

- Retargeting of the LLVM compiler to the Kalray VLIW core architecture, in order to compile the OpenCL-C dialect.
- Extending the POCL open-source project with the ability to launch GCC C/C++/OpenMP 'native kernels' for execution on MPPA clusters under the lightweight POSIX ClusterOS. POCL provides the OpenCL 1.2 Embedded Profile host API.
- Implementing the OpenAMP open standard from the Multicore association, which allows standard Linux running on the ARM cores to manage the life cycle and the system call remoting requests from the lightweight POSIX ClusterOS.

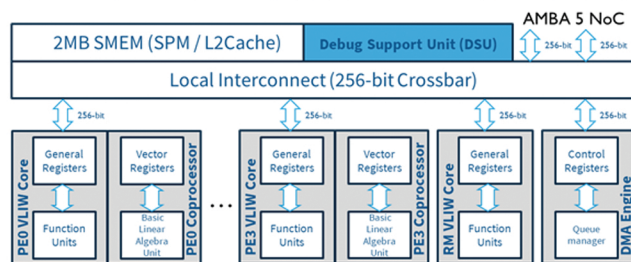


Figure 2 Overview of the MPPA accelerator tile for EPI

The MPPA accelerator tile (Figure 2) is developed for vehicle perception in the EPI Automotive Stream. This tile is based on the compute unit ('compute cluster') of the Kalray MPPA3 processor. Indeed, the MPPA processors already excel at vehicle perception, thanks to their architecture designed for high-performance signal processing, image processing, bit-level processing, and deep learning inference. They are fully programmable in C/C++/OpenMP under GCC or LLVM, and are able to host RTOS, SMP POSIX and Linux operating systems. Other tools include a deep learning compiler that produces highly optimized code from trained neural networks provided under the standard Caffe, TensorFlow and ONNX formats. In the EPI Automotive Stream, the MPPA3 processor will be used in the Intermediate Reference Platform (IRP) to support software development and performance analysis of ADS functions until the first iteration of the GPP is available.