MPPA® for
High-Performance
Embedded Computing

Kalray’s MPPA® processors will find their way into cars, drones and aerospace

Benoît Dupont de Dinechin, CTO, Kalray
May 11, 2017

www.kalrayinc.com
KALRAY COMPANY OVERVIEW

ACTIVITY

**Launched in 2008**
Spin-off of the “CEA” (French Department of Energy)

**Pioneer in manycore processors**
- Focus on low latency, low-power, compute intensive processing
- Core market: Data Centers and Critical Embedded Applications

**Comprehensive product offer**
- MPPA® processors, acceleration cards, associated software
- Strategic partnership with TSMC

TEAM

- **21** NUMBER OF GRANTED PATENTS HELD BY KALRAY FOR ITS MANYCORE ARCHITECTURE
- **TOP 20** “MOST PROMISING HPC SOLUTION PROVIDERS” – CIO REVIEW - 2016
- **1st** COMMERCIAL PROCESSOR TO BREAK THE “100 COMPUTING CORE WALL” - 2013

Two offices
Grenoble, France & Los Altos, CA, USA

Experienced technical team
World-leading multicore/manycore talent, with experience in both hardware & software
AUTONOMOUS VEHICLE CHALLENGES

... leading to processing challenges:

- High performance computing
  - Computer vision, machine learning
- Low-latency I/O from sensors
  - Camera, lidar, radar: Ethernet and LVDS
- Functional safety
  - Freedom from interference (ISO 26262)
- Cyber-security
  - Root of Trust, software IP protection

Failing to address these challenges will put passengers’ lives in danger

Autonomous cars will have to do a lot...
KALRAY’S MPPA® OFFERS A UNIQUE SOLUTION

MPPA® (Multi-Purpose Processing Array) Platform

**Hardware**
- Manycore CPU architecture
- Compute clusters of 16 high-performance CPU cores with local memory
- DSP-like timing predictability
- ‘Fully timing compositional’ cores for accurate static timing analysis
- Service guarantees of local memory system and network-on-chip
- FPGA-like I/O capabilities
  - 8x 10Gbps Ethernet, 16x PCIe Gen3, 2x DDR3 controllers

**Software**
- CPU programming
  - Standard C/C++/OpenCL, OpenVx
  - Model-based (SCADE Suite®, Simulink®)
THE MPPA® VALUE PROPOSITION

1. ENERGY EFFICIENT
   The MPPA®’s architecture is highly energy-efficient on integer, simple and double precision floating-point computations.

2. FUNCTION CONSOLIDATION
   The MPPA®’s clustered architecture enables consolidation of independent functions onto the same chip.

3. MIXED-CRITICALITY COMPUTING
   Depending on usage domains, MPPA® clusters are configured to operate in hard real-time or soft real-time.

4. SCALABLE PERFORMANCES
   MPPA® processors can be tiled through network-on-chip extensions to offer more compute and I/O capabilities.

All on the same chip! >>>
MPPA® PROCESSOR EVOLUTION

**Today (2016 – 2017)**

**MPPA® BOSTAN**
- TSMC 28HP
- 256 cores at 600 MHz
- Up to 700 GFLOPS 32-bit and 16-bit FMA
- 32 MB of local memory
- Core accelerators for AES including GCM and secure hashing

**Tomorrow (2018)**

**MPPA® COOLIDGE**
- TSMC 16FFC
- 80 cores at 1200 MHz & 80 co-processors
- Up to 3 TFLOPS 16-bit FMA in core accelerators for computer vision and machine learning
- 20 MB of local memory
- Hardware root of trust
- ISO 26262 ASIL-B

**Autonomous Cars Hit the Roads with Kalray Processing Solutions**

*Prototypes with MPPA® currently in development*
DEEP LEARNING SOLUTION

THE OFFER:

Kalray Neural Network (KaNN) Evaluation Kit & Development Kit Code generator for deep learning (Berkeley Caffe & Google TensorFlow)

MPPA® Processor Hardware & AccessCore SDK

Performance on GoogleNet = 65 fps @500Mhz
Un-modified Berkeley Caffe

360 fps @1.2Ghz
MPPA® Coolidge

Optimized architecture for deep learning

Supports both floating and integer high performance computing

Dedicated co-processor for vision and learning

Fully programmable solution, enabling custom neural networks and vision algorithms