Using Manycore Processors in Complex Embedded Systems

with Kalray MPPA® customer cases

Teratec Forum - June 25-26, 2013

Visit us on booth #10
Presentation Outline

- Kalray: the company, products and markets
- The rising of Manycore architectures
  - Features and benefits
- Using Manycore processors
  - Concrete customer cases
- Q&A
Kalray at A Glance

- Founded in 2008 – located in Paris, Grenoble (France) & Tokyo (Japan)
- **55+ people**
- Independent and unique technology: **MPPA® MANYCORE processor** (Multi-Purpose Processing Array) and software programming environment
- Targeting the *industrial, embedded and computing intensive markets*
- Large patent portfolio
- Several awards over the past years
  - Kalray ranked in “EETimes’ silicon 60 : Hot start up to watch” in 2012
  - “Best technology award” from “Les Trophées de l’Embarqué” in 2012
  - “Startup of the year” by ElectroniqueS magazine in 2013
First MPPA®-256 Chips with CMOS 28nm TSMC

- High processing performance
  700 GOPS – 230 GFLOPS
- Low power consumption - 5W
- High execution predictability
- Software programmable

Released November 2012
KALRAY, a global solution

Powerful, Low Power and Programmable Processors

C/C++ based Software Development Kit (SDK) for massively parallel programing

Development platform
Reference Design Board

Reference Design board
Application specific boards
Multi-MPPA or Single-MPPA boards
Target Application Areas

**INTENSIVE COMPUTING**
- Finance
- Numerical Simulation
- Geophysics
- Life sciences

**IMAGE & VIDEO**
- Broadcast
- Medical Imaging
- Digital Cinema
- Augmented reality
- Vision

**EMBEDDED SYSTEMS**
- Signal Processing
- Aerospace/Defence
- Transport
- Industrial Automation
- Video Protection

**TELECOM / NETWORKING**
- Packet Switching
- Network Optimisation
- Security Services
- Software Defined Radio
- Software Defined Network
MPPA MANYCORE Roadmap

Architecture scalability for high performances and low power

MPPA®-256 V1

1st core generation
50 GFLOPS/W

MPPA®-256 V2

2nd core generation
80 GFLOPS/W

MPPA®-64

Low Power
5 W

Q4 2012

MPPA®-1024

Low Power
7 W

Very Low Power
75 mW - 1,8 W

Q2 2014

MPPA®-1024

Low Power
5 W

Q2 2015

1st core generation
50 GFLOPS/W

2nd core generation
80 GFLOPS/W

3rd core generation
100 GFLOPS/W
Kalray Software Development Kit
MPPA ACCESSCORE – MPPA ACCESSLIB

Today

Q4 2013

Standard C/C++ Programming Environment

Dataflow Programming FPGA Style

Simulators & Profilers, Debuggers & System Trace

POSIX-Level Programming DSP Style

Operating Systems & Device Drivers

Streaming Programming GPU Style
The rising of Manycore architectures
Single-Threaded Integer Performance

- +52% increase yoy from 1996 to 2004
- +21% from 2007 to 2011
- Current figures suggest a +10% increase yoy since 2011 for most CPUs

No more progress in time scale, so enlarge in space with more cores
Multicore CPUs vs GPUs vs Manycore

**CPU**
- CPU are optimized for sequential code performance
- Sophisticated Control Logic to execute several instruction at the same time
- Very large cache to reduce access time to instruction and data of complex applications
- Clock frequency limits reached long ago
- 50 to 150 Watt

**CONstrained By**
- Power consumption, Complexity, Scalability

**GPU**
- Originates from the video game industry
- Very simple control units and huge number of floating point units working in parallel
- SIMD processing model
- Smaller cache than CPUs
- 20 to 300 Watt

**CONstrained By**
- Power consumption, Programming model, Communication overhead

**Manycore**
- Rich CPU-like Control units + FPU
- Cluster of processors share a local memory
- Clusters communicate through a high speed low latency network on chip
- MIMD paradigm (Multiple instructions Multiple Data) rather than MIMD
- 5 to 10 Watt

**CONstrained By**
- Disruptive technology adoption curve
MPPA® Technology Compared to GPU & CPU

MPPA-256
20pJ/Instruction
Optimized for GFlops/Watt and $
Distributed Memory model
Low Latency

230 GFlops
5 W
~50

GPU
200pJ/Instruction
Optimized for Throughput
Explicit Management of On-chip Memory

3000 GFlops
300 W
~10

CPU
2000pJ/Instruction
Optimized for Latency
Caches

700 GFlops
130 W
~5

Source: Bill Dally, “To ExaScale and Beyond” - NVidia
MPPA®-256 Processor Hierarchical Architecture

- **VLIW Core**
  - Instruction Level Parallelism

- **Compute Cluster**
  - Thread Level Parallelism
  - Shared Memory
  - Process Level Parallelism

- **Manycore Processor**
Kalray’s MPPA customer cases
Intensive Computing Driving Factors

- **Computing efficiency (GFlops/Watt)**
  - Energy consumption becomes an absolute barrier whether in high-end embedded systems or data-centers

- **Hardware efficiency (GFlops/$)**
  - Remove unnecessary hardware overhead

- **Bandwidth (MB/s)**
  - Bring data in and out fast and avoid bus bottleneck
Signal Processing Example

- Radar applications: STAP, ...
- Beam forming: Sonar, Echography
- Software Defined Radio (SDR)
- Dedicated libraries (FFT, FTFR, ...)

Well suited for massively parallel architectures
MPPA as an alternative to DSP+FPGA or CPU+GPU platforms
Customer case #1

- Healthcare industry sector
- Electronic device performing real time, very intensive signal and image processing tasks
- Current status
  - Uses conventional CPU+GPU
- Problems identified
  - Excessive energy footprint: won’t fit conventional electric setup at customer site, too bulky, too noisy.
  - GPUs hard to program and almost complete lack of tools for debugging/profiling/optimization work.
  - Totally inadequate for a forthcoming portable version of the device
Kalray’s 3-step Approach

1. Analyze
   - Profile legacy code using MPPA ACCESSCORE simulation tools

2. Parallelize
   - Use MPPA Dataflow to express parallelism at a high level

3. Optimize
   - Automatically map application on the 256 cores of MPPA processor
   - Run the code, trace and fine tune
Customer Benefits

### Hardware benefits
- **GFlops/W** is 4x better
- No need for host CPU (better GFlops/$)
- Extensible processing array without PCI bus
- MPPA v2 triples the GFlops/W

### Software benefits
- Porting effort < one month of work
- Powerful and unique optimization tools
- Unified toolset  Host + MPPA
- No need to change the code when extending the processing array
Q&A

www.kalray.eu - info@kalray.eu

Visit us at Teratec Booth #10