CMA-CUBE: A SCALABLE RECONFIGURABLE ACCELERATOR WITH 3-D WIRELESS INDUCTIVE COUPLING INTERCONNECT

Yusuke Koizumi[†], Eiichi Sasaki[†], Hideharu Amano[†], Hiroki Matsutani[†], Yasuhiro Take[†], Tadahiro Kuroda[†],
Ryuichi Sakamoto[††], Mitaro Namiki[††], Kimiyoshi Usami[†††], Masaaki Kondo[††††], Hiroshi Nakamura[†††††]

† Keio Univ. †† Tokyo Univ. of Agri. And Tech. ††† Shibaura Institute of Technology
†††† University of Electro-Cmmunications ††††† University of Tokyo

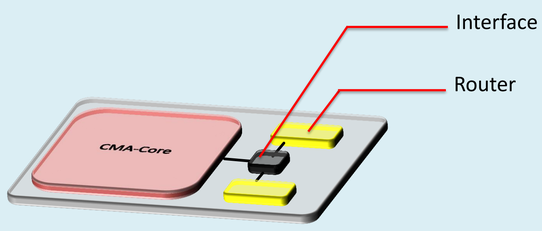
Introduction

- Recent battery driven IT devices require versatile functions and high performance with low energy consumption
- The target applications require wide range of performance
- It is difficult to develop an SoC (System-on-a Chip) for each product because of the grown initial cost of LSI

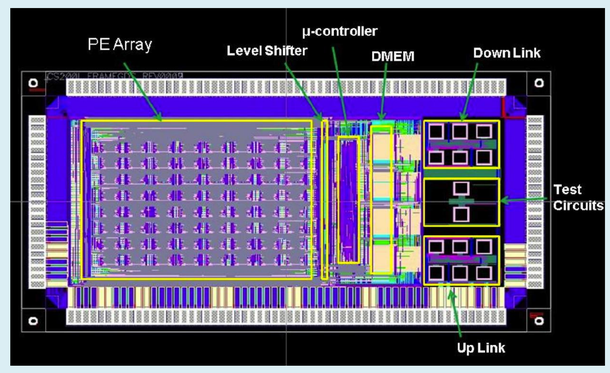
A scalable reconfigurable accelerator CMA-Cube has been proposed

CMA-Cube

- CMA-Cube consists of the following components
  - ✓ Reconfigurable accelerator core called CMA (Cool Mega Array)
  - ✓ Inductive Coupling interconnect
  - ✓ 2 Routers with bubble flow controll
  - ✓ Interface between a core and routers



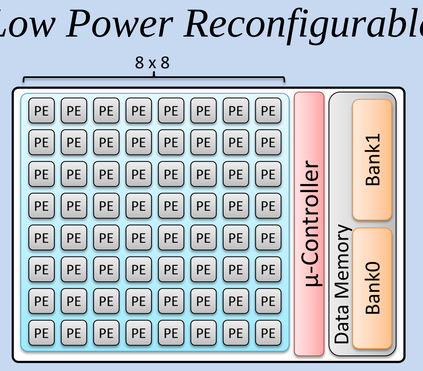
scheme of CMA-Cube



Layout of CMA-Cube

CMA Core

-Low Power Reconfigurable Accelerator-



PE(Processing Element) Array

- 8 x 8 PEs only with combinational circuit
- Eliminating the energy consumption for storing the intermediate data in registers and their clock distribution
- Supply voltage ranging from 0.5V to 1.2V

μ-Controller

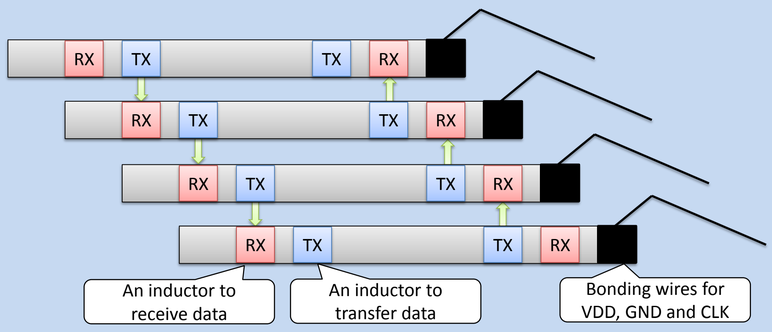
- Flexible data management between PE Array and Data Memory

Data Memory

- 25bits x 1024 entry
- 2 memory bank
- Data translation from external and operation on PE Array can be overlapped

Inductive Coupling

-Wireless Interconnection Technology-



- An inductor is implemented as a square coil with metal in common CMOS layout
- Chips are stacked with a certain length of gap so that the TX channel is located exactly on the same place of the RX channel of the next chip

Specifications of Inductilve Coupling	
Clock Freq.	4GHz for both edges
Bandwidth	36bits/5ns
Latency	2clocks/chip

Network on Chip

Router

- Standard 3-stage pipeline structure with 3-input/output
- Bubble flow control for dead lock free
- The following simple rule is applied

(1) A packet on a ring can move to the next hop along the ring when the input buffer of the next hop has an empty space of at least one packet.
(2) The network interface can inject a packet to a ring when the vertical input buffer of the ingress router has an empty space of at least two packets.
(3) A packet can exit from a ring only when the horizontal output buffer of the egress router has an empty space of at least one packet.

Interface

- Decode packets from a router
- Create packets to a router
- A packet consists of flits

Flit format

Types of packet	
single flit packet	Command
2-flit packet (header flit, data flit)	Single data transfer
5-flit packet (header flit, data flit x 4)	Block data transfer

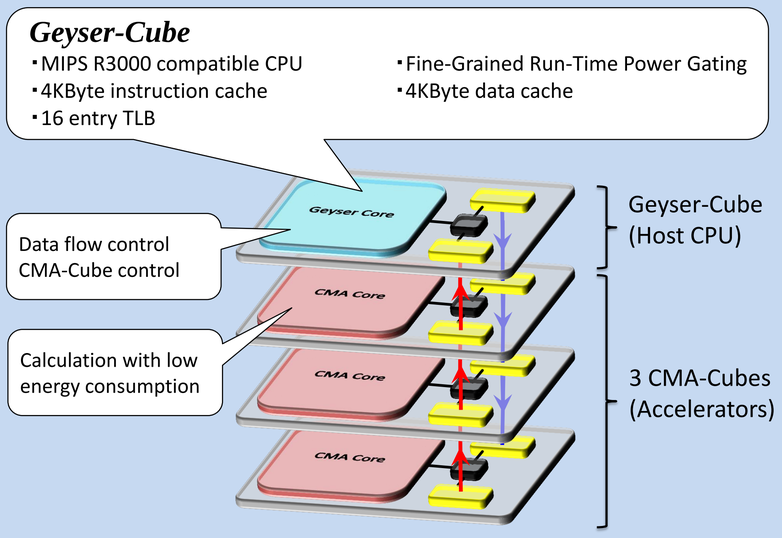
Evaluation

Target System

- Cube-1 Quad-Core

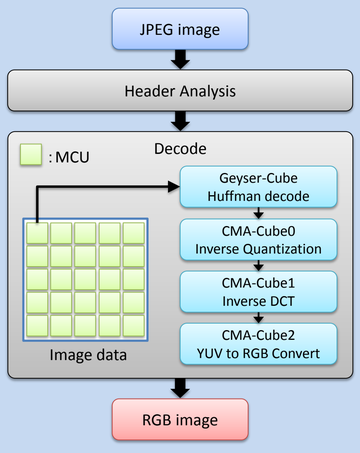
Geyser-Cube

- MIPS R3000 compatible CPU
- 4KByte instruction cache
- 16 entry TLB
- Fine-Grained Run-Time Power Gating
- 4KByte data cache



Benchmark Application

- JPEG decoder



Result

- Cube-1 Quad-Core vs No Acceleration

Simulation tool	Cadence NC-verilog 8.1
Frequency	Geyser-Cube:200MHz
	CMA-Cube:200MHz
	Router:200MHz
	Data bus:100MHz

- Cube-1 Quad-Core vs No Acceleration