Demonstration of Low Power Stream Processing Using a Variable Pipelined CGRA

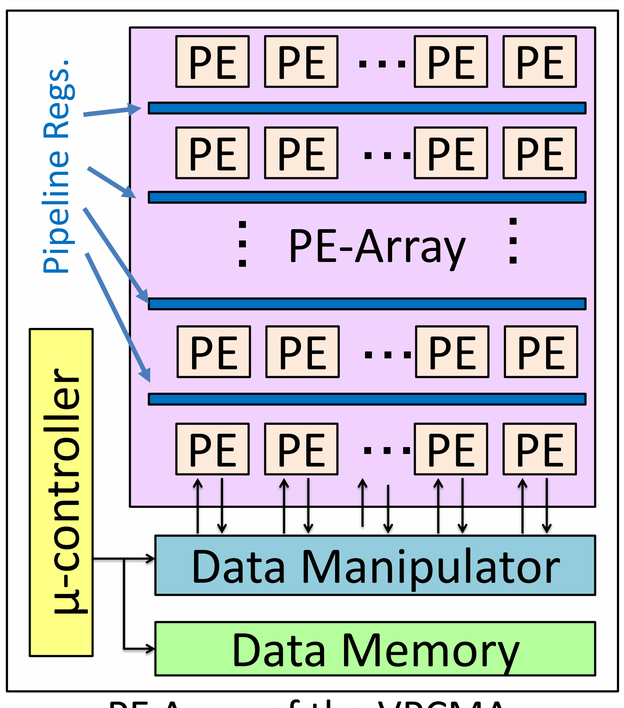
Takuya Kojima, Naoki Ando, Yusuke Matsushita, and Hideharu Amano Keio University

Introduction

CGRAs (Coarse-Grained Reconfigurable Architectures) are expected to be used for IoT devices and edge computing due to their high energy efficiency. VPCMA (Variable Pipelined Cool Mega Array) is a low power CGRA which we previously proposed in [1]. CC-SOTB2 is a real chip implementation of the VPCMA using Renesas 65-nm SOTB technology [2]. In this demonstration, we will show the low power consumption of the CC-SOTB2 while performing a real image processing with a tiny solar cell battery.

Architecture Overview

VPCMA (Variable Pipelined Cool Mega Array)[1]



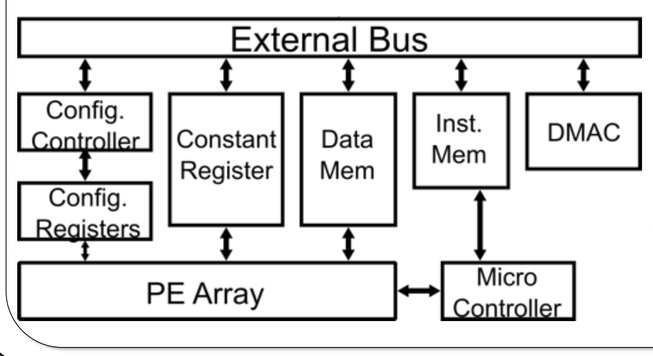
PE Array of the VPCMA

➢ PE(Processing Element)
- Composed of
  1. Simple ALU
  2. Switching Element
- No Register file

→ No need of clock signal

➢ PE Array
- 12 cols x 8 rows PEs
- 7 configurable pipeline regs. (Latch mode/Bypass mode)

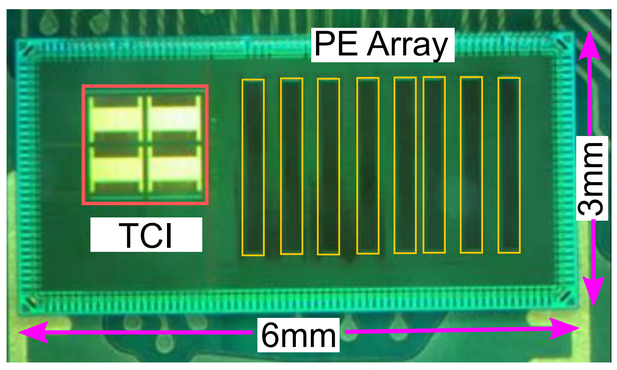
➢ Variable Pipeline
- Trade-off b/w performance & power consumption



➢ Micro-controller
- Controls data transfer b/w data memory & PE array

➢ External host processor
- Uses common data bus for data transfer, reconfiguration, and other controls

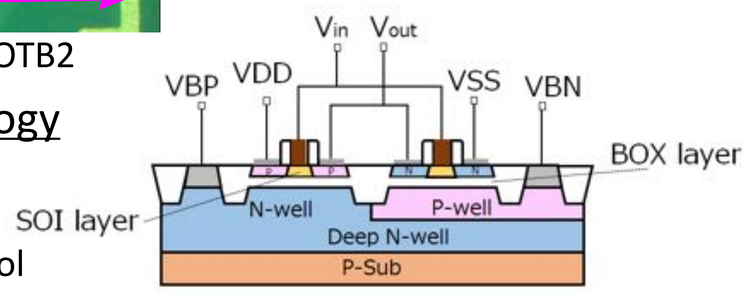
A Real Chip Implementation: *CC-SOTB2*



Chip Photograph of CCSOTB2

➢ A prototype chip of VPCMA: CC-SOTB2
- Fabricated with Renesas SOTB technology
- 3mm x 6mm die

➢ About SOTB technology
- 65 nm process
- FD-SOI
- Good for body bias control
  - Trade-off b/w leak power & transistor performance



Transistor of SOTB technology

➢ Five body bias domains in CC-SOTB2
- PE array's domains vs. micro-controller's domain

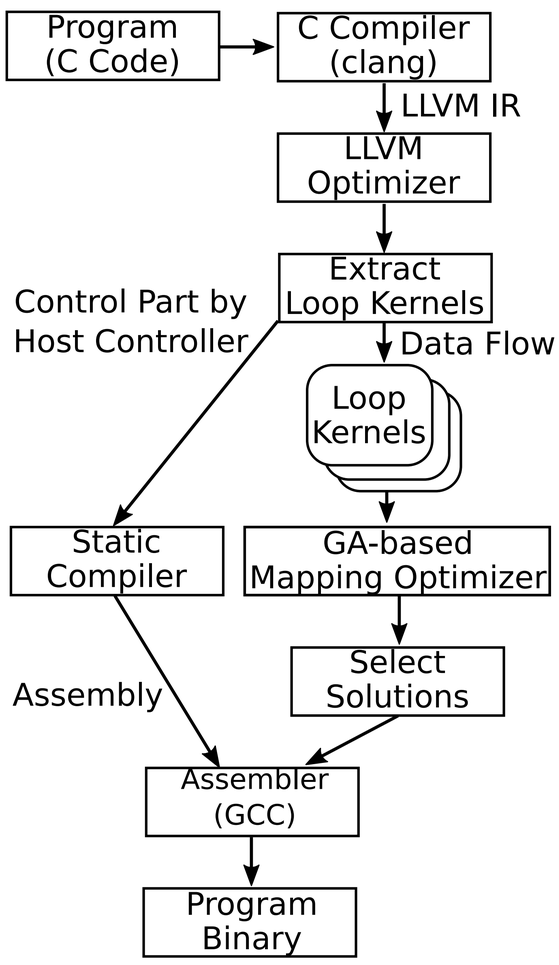
→ Adjust the balance of performance b/w PE array & micro-controller
- Four divided PE array's domains

→ Boost only bottleneck PEs & slow down other PEs

System Overview

Programming and Computing with a Host CPU

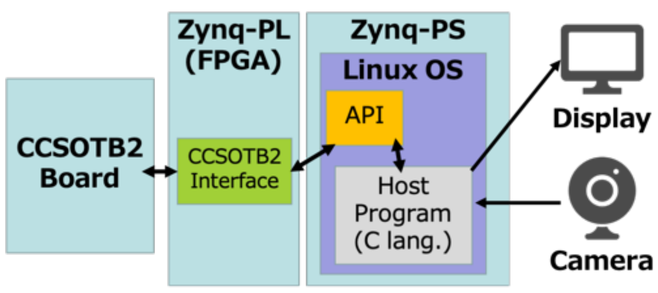
➢ Application Development Flow



- Program Codes
  - Written in C language
  - Offloading compute-intensive parts to CC-SOTB2
  - Leaving control parts to a host processor

  Full automated flow is under development
- Mapping Optimization[2]
  - Mapping Data-Flow-Graph in the loop to PE Array
  - Genetic-Algorithm-based Optimization tool
  - Optimizing followings considering target freq.
    1. Pipeline Structure
    2. Body bias voltages
    3. Place and Route
- Host Processor
  - Zynq-7000 FPGA
  - Linux OS working
- Interface of CC-SOTB2
  - Implemented on Zynq-PL
  - API for the Linux OS

➢ Computation with Zynq FPGA



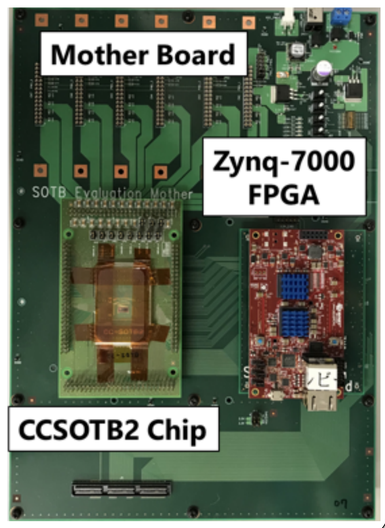
Demonstration Environment

➢ Motherboard for experiment & demo
- Connects CC-SOTB2 with Zynq FPGA
- Power supply boards are available

→ Voltage control by Zynq

➢ In this demonstration, a tiny solar cell battery
- Large internal resistance

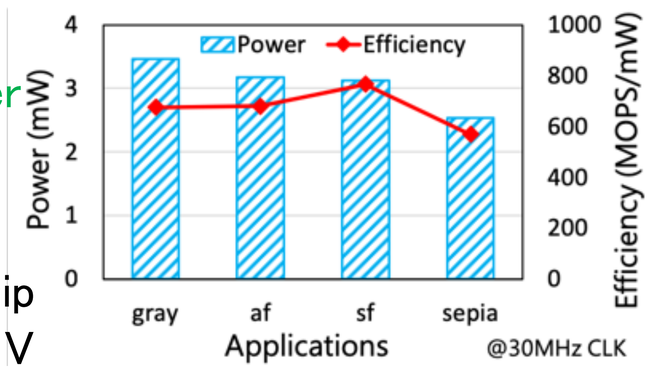
→ Voltage is kept only for extremely low power systems



Results of Image Processing

➢ In the best case (sf),
- About 3 mW peak power
- 80 PEs utilized

→ 2.4 GOPS / 3 mW



➢ Up to 30MHz, the real chip can work stably with 0.55 V

➢ Gray Scaling of RGB image

Input Image | At 20MHz, 0.55V | At 20MHz 0.40 V

Reference: [1] Ando, Naoki, et al. "Variable pipeline structure for coarse grained reconfigurable array CMA." FPT 2016, p. 217-220.
[2] Kojima, Takuya, et al. "Real Chip Evaluation of a Low Power CGRA with Optimized Application Mapping.", HEART 2018.