香港中文大學
The Chinese University of Hong Kong

Introduction to Gemmini

Shixin CHEN
Department of Computer Science & Engineering
Chinese University of Hong Kong
shixinchen@smail.nju.edu.hk

March 26, 2022

Outline

1. ISA: RISC-V
2. Core: Rocket & BOOM
3. Co-Processor: Gemmini

ISA: RISC-V

What is ISA?

Instrcution Set Architecture (ISA) a list of all the commands that a processor can execute. It provides a rule to guide the design of Microarchitecture.

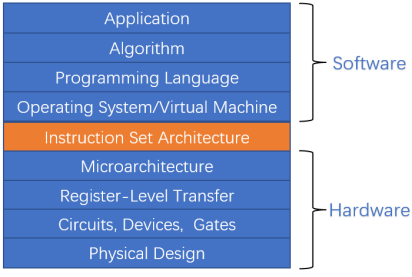

Figure 1: The location of ISA in Computer System

- Computational operation: such as `add` `mul`, `and`, `or` and `not`
- Load/Store: such as `move`, `load` or `store`
- Control flow: such as `goto`, `return`
- Information about General Registers, Memory Mode, Interrupts and Exceptions table...

RISC & CISC

ISAs can be classified as two kinds, Complex Instruction Set Computer (**CISC**) and Reduced Instruction Set Computer (**RISC**)

- **CISC**
  - using fewer instructions specialized for complex tasks
  - only 20% of instructions are commonly used
  - making it complicated for CPU design
- **RISC**
  - including some basic and simple instructions
  - using more instructions to complete complex tasks
  - most popular in modern CPU, e.g., ARM, MIPS, Power, Alpha, **RISC-V**

RISC-V

A free and open ISA from Berkeley. It is simplicity-oriented, only about 40 instructions. Moreover, it avoids many shortcomings in the classical ISAs.

Composable extensions Including:

- Consists of a base ISA – RV32I (32 bit), RV64I (64 bit)
- 'M': Math extension. Multiply and divide
- 'F', 'D': Floating point extensions, single and double precision
- 'V': Vector operators
- Designer can choose to implement combinations: e.g., RV64IMFT

Figure 2: The logo of RISC-V

The Aim of RISC-V

- A fully open-source ISA that can be used freely by any academic institutions or commercial organizations
- A standard ISA that is truly suitable and stable for hardware implementation

Figure 3: the Industry Backers of RISC-V

CPU: ROCKET & BOOM

Rocket Core

- **Rocket Core** is an open-source RISC-V processor core from Berkeley, which can be generated from the **Rocket-Chip Generator** written in Chisel.
- The configurations are specified through **Chisel** parameters most of which can be freely changed.

Category	ARM Cortex-A5	RISC-V Rocket
ISA	32-bit ARM v7	64-bit RISC-V v2
Architecture	Single-Issue In-Order	Single-Issue In-Order 5-stage
Performance	1.57 DMIPS/MHz	1.72 DMIPS/MHz
Process	TSMC 40GPLUS	TSMC 40GPLUS
Area w/o Caches	0.27 mm^2	0.14 mm^2
Area with 16K Caches	0.53 mm^2	0.39 mm^2
Area Efficiency	2.96 DMIPS/MHz/mm^2	4.41 DMIPS/MHz/mm^2
Frequency	>1GHz	>1GHz
Dynamic Power	<0.08 mW/MHz	0.034 mW/MHz

Figure 4: ARM vs Rocket Core[1]

[1] Krste Asanovic, Rimas Avizienis, et al. (2016). "The rocket chip generator". In: *EECS Department, University of California, Berkeley, Tech. Rep. UCB/EECS-2016-17* 4.

Rocket-Chip Generator

The Generator can generate different configurations of a SoC (System-on-Chip)

Configurable Parameters include:

- number of cores
- number of floating-point units, vector units
- cache sizes, number of TLB entries, memory sizes
- number of floating-point pipeline stages
- width of I/O ports
- ...

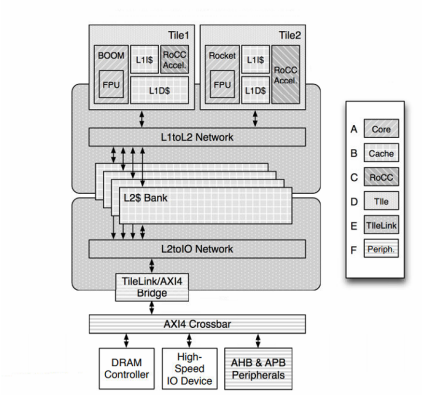

Figure 5: A Typical Generated Rocket SoC

BOOM Core

Berkeley Out-of-Order Machine (**BOOM**[2]) is another core that can be generated from Rocket-Chip.

- Comparing to Rocket core, it aims at higher performance, synthesizable, and parameterizable core for architecture research.

[2]Krste Asanovic, David A Patterson, and Christopher Celio (2015). *The berkeley out-of-order machine (boom): An industry-competitive, synthesizable, parameterized risc-v processor*. Tech. rep. University of California at Berkeley Berkeley United States.

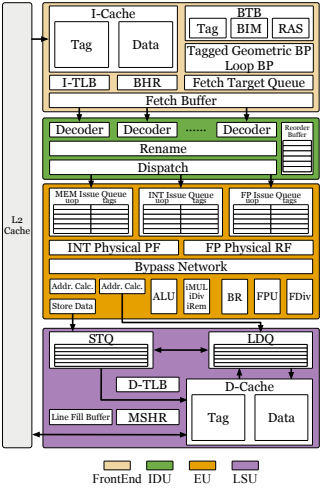

Figure 6: BOOM Architecture[3]

[3]Chen Bai et al. (2021). "BOOM-Explorer: RISC-V BOOM Microarchitecture Design Space Exploration Framework". In: *2021 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*. IEEE, pp. 1–9.

CO-PROCESSOR: GEMMINI

Motivation

Deep Neural Networks (DNNs) are exploding popularity, and DNN accelerators have attracted lots of attention.

Accelerator solutions:

- **Systolic Array** architecture
- **Fast Flourier Transform** (FFT) architecture
- **Winograd Algorithm** architecture

It is hard to design specialized hardware for emerging new DNN architectures. **DNN Accelerator Generator** can provide a flexible and efficient co-processor to boost the performance of DNNs and meet the demand of different DNNs.

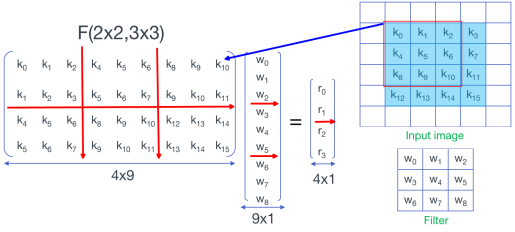

Figure 7: Image to Column Operation

$$C = A * B + D \tag{1}$$

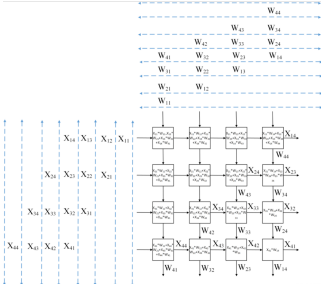

Figure 8: Systolic Opeartion

Gemmini

Gemmini is a hardware accelerator generator written in Chisel. It provides a full-system, full-stack DNN hardware exploration and evaluation platform.

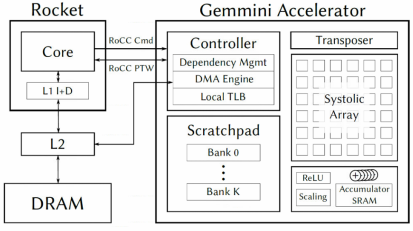

Figure 9: Gemmini Architecture

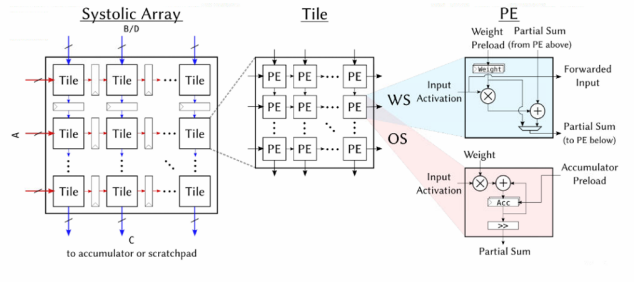

Figure 10: Systolic Architecture in Gemmini

Gemmini Parameters

Gemmini is a co-processor of Rocket or BOOM. Many DNN operations such as `matmul`, `relu`, `max-pooling` can be compiled as customed RISC-V instruction and sent form core to Gemmini.

Configurable Parameters:

- Dimensions of array tiles
- Capacity of memory, including scratchpads SRAM and accumulator SRAM
- Dataflow model: Output Stationary or Weight Stationary
- ...

Full-stack Evaluation

Gemmini is a full-stack platform, especially with other toolchains in Chipyard[4], which reports the metrics (power, performance, and area) of accelerator.

- **SPIKE:** Functionally accurate, fast to run and build
- **Verilator:** Cycle accurate, slow to run and build
- **Firesim:** Cycle accurate with realistic DRAM, very slow to build, very fast to run

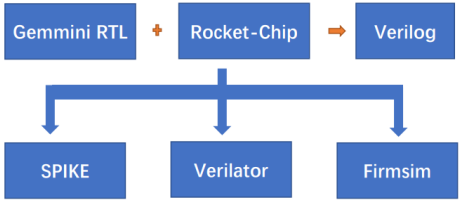

Figure 11: Output of Gemmini

[4] Alon Amid et al. (2020). "Chipyard: Integrated Design, Simulation, and Implementation Framework for Custom SoCs". In: *IEEE Micro* 40.4, pp. 10–21. ISSN: 1937-4143. DOI: 10.1109/MM.2020.2996616.

Recent Progress

- using different configures to generated SoC and report the performance of Gemmini
- exploring configures of Gemmini with higher efficiency through DSE methods

THANK YOU!