Out of Order Execution RISC-IV Superscalar Pipelined Processor in SystemVerilog

Abbas Bakhshandeh

March 27, 2024

Abstract

Computer processors play a crucial role in the coordination of various components in computer systems, execution of instructions, and efficient completion of computing tasks. In an effort to improve processor performance, a number of microarchitectural optimizations such as OoO (Out of Order) execution, pipelining, and superscalar execution can be implemented. The goal for this capstone project is to create an OoO RISC-IV superscalar (2 issue) pipelined processor in SystemVerilog. The RISC-IV ISA was chosen due to its reduced complexity, open-source nature, and overall efficiency. In order to successfully enable OoO, architectural structures such as a reservation station, ROB (Reorder Buffer), free pool, and a physical register file were incorporated. Doing so improved processor throughput and efficiency by allowing execution of instructions out-of-order while preventing data hazards from occurring. Pipelining further increased throughput by allowing different modules to work on different instructions concurrently. Furthermore, the superscalar aspect of the processor allowed for two instructions to be processed in parallel throughout the pipeline, which increased throughput and efficiency.

1 Introduction

Computer processor design is a complex task that requires the combination of various microarchitectures in a manner that meets power, throughput, and resource utilization metrics. A simple processor can execute instructions one at a time, with separate modules each performing one aspect of the process. The general processing workflow involves fetching the instruction, decoding it, executing it, and reflecting the results in the architectural state or memory. In an effort to increase throughput, most processors use the concept of pipeling. Pipelining allows each module to process one part of a separate instruction, so that no modules - in most cases - are idle while they wait for other modules to complete. Another method used to increase throughput is superscalar execution. By making the processor superscalar, hardware is added in parallel to allow for the processing of multiple instructions at one stage in the pipeline. While these methods greatly improve throughput, by themselves they only allow processors to process instructions in the order that they are received. This can create issues where future instructions are waiting long periods of time for another instruction to finish before they can be processed, which results in a decrease in throughput.

In order to help alleviate this issue, out of order execution has been introduced to processors. Using out of order execution, the processor is able to have certain stages of the pipeline be completed out of the order that the instructions were received in. However, precautions must be taken to prevent data hazards that could occur as a result of this out of order processing of instructions. If a future instruction has a source register which depends on the result of a previous instruction, then it would be incorrect to allow it to complete before the results of that previous instruction are known and reflected in the architectural register file. The most common out of order processors complete the fetching of instructions, decoding of them, and retiring of them - where the results are reflected in architectural states and memory - in order. The stage that is allowed to complete out of order - with certain restrictions - is the execution stage.

In order to accomplish this out of order processing in the execution stage of the pipeline while preventing data hazards, various hardware blocks are created. The concept of physical registers is introduced, which are a larger set of registers that the processor places the architectural register values used in incoming instructions inside of. This larger set allows for many instructions that are referencing the same architectural register to have dedicated physical registers, allowing for out of

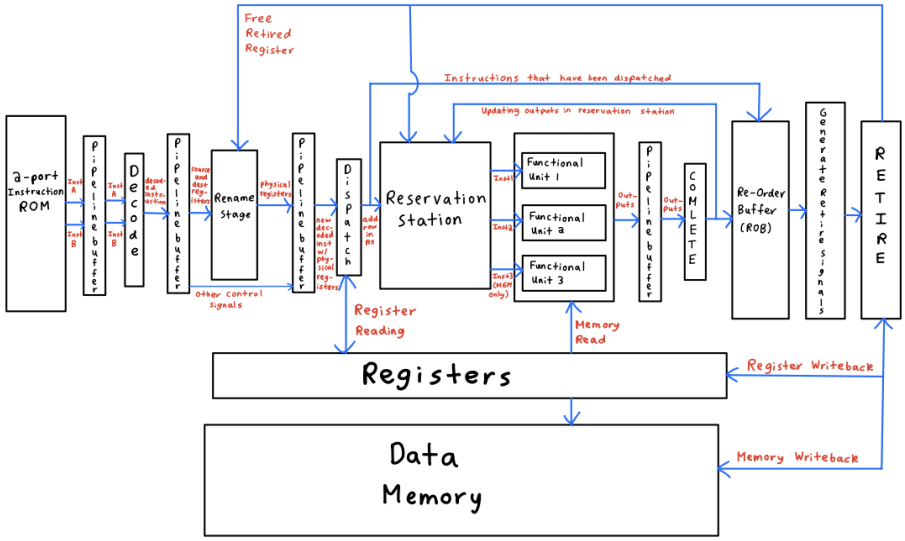

Figure 1: Block Diagram for Complete OoO Processor

order execution. A RAT (Register Alias Table), free pool, reservation station, ROB (Reorder buffer) are used to accomplish this. Table 2 provides a detailed description for each of these microarchitecture blocks. In their essence, they organize the incoming instructions - which have been decoded - in a manner that allows instructions to be executed out of order as long as their source registers don't depend on a previous instruction's result. They also keep track of which architectural registers have been mapped to which physical registers, so that they can be remapped in the correct order later in the pipeline. The result of the introduction of out of order execution to the processor architecture is an increase in throughput and overall efficiency.

2 Implementation

2.1 Module Descriptions

In order to implement this processor, the design was modularized and connected through a top module (see Appendix II) using the SystemVerilog HDL (Hardware Description Language). Interfaces were used to store information that could be accessed between various modules (see Appendix I). The modules that were created include Fetch, Decode, Rename, Dispatch, Fire, and Retire. Table 1 describes each module used in the design in detail.

2.2 Microarchitectures for Out of Order execution

Furthermore, within the modules there are various microarchitectural structures which were implemented in order to accomplish Out of Order (OoO) execution. Table 2 provides a more detailed description of each execution unit which was constructed.

2.3 Block Diagrams

The block diagrams for the complete OoO processor, RAT (Register Alias Table), Reservation Station, and ROB (Re-order Buffer), have been included in Figure 1, Figure 2, Figure 3, and Figure 4, respectively.

Module	Description
Fetch	Every clock cycle, two input instructions are stored from a text file containing hexadecimal values. Each input instruction corresponds to 4 bytes (32 bits), so 8 bytes (64 bits) are fetched each clock cycle. (See Appendix III)
Decode	Extracts necessary information from the two instructions which are fetched every clock cycle, including the destination register, source registers, immediate values, function values, and control signals such as AluOp, aluSrc, memRead, memWrite, and memtoReg. These values are stored and passed to the next stages in the pipeline for further processing. The 8 possible instructions for this processor are ADDI, ANDI, ADD, SUB, XOR, SRA, LW, and SW. (See Appendix IV)
Rename	Register renaming is enabled in order to implement out-of-order execution. Architectural destination registers are assigned to available physical registers, and architectural source registers are mapped to their assigned physical register names. Doing this prevents data hazards that could arise when completing out-of-order execution. A RAT (Register Alias Table) in addition to a "free pool" structure is utilized to keep track of which architectural registers are mapped to which physical registers. A physical register file is also created to hold physical register 32 bit values. (See Appendix V)
Dispatch	The information collected in the previous stages - the decoded signals as well as the renamed physical registers - are placed inside a reservation station. The reservation station contains "ready" bits for each (physical) source register in each instruction, as well as a ROB number and a functional unit number. A ROB (Reorder buffer) is created that stores the new physical destination registers, old physical destination registers (before renaming), and a "complete" bit which indicates whether the instruction has been completed. When both source registers are ready (and a functional unit is available), the instruction can be issued to the functional unit and removed from the reservation station. Up to three instructions can be issued to the three functional units at one time, if they are available. Two of the functional units are for R-type and I-type instructions (ALU) and the third function unit is for memory instructions (LW and SW). (See Appendix VI)
Fire	Once the reservation station issues an instruction in the table to a functional unit, it is processed based on the instruction's control signals (which were calculated in the decode stage), the physical register names, and the values in the physical register file. The final results of these calculations are stored in the physical register file (at the correct physical destination register location). (See Appendix VII)
Retire	While the execution of instructions can be completed out-of-order, the retiring (which actually updates the architectural registers and memory) must be done in-order to prevent data hazards. To accomplish this, the ROB is checked to see which instructions have completed (after the fire stage). It will retire instructions which have had all instructions above it (indicated by a program counter) also retired. This will assign the values stored in the physical register file to their corresponding architectural registers or memory. Up to two instructions can be retired each clock cycle. After retiring, they are removed from the ROB and the corresponding physical registers are freed for further usage by future instructions. (See Appendix VIII)

Table 1: Processor Modules and Descriptions

Mircroarchitecture Block	Description
RAT (Register Alias Table)	The RAT is used to keep track of which architectural registers have been renamed to which physical registers. The RAT is created in the Rename module.
Free Pool	The free pool is an array containing 64 one bit values. Each one bit value indicates whether the corresponding physical register is occupied or not. For example, if (free pool)[7] == 1'b1 is true, then this means physical register 7 is occupied (and hence should not be assigned to any other architectural register). In this microarchitecture, p0 is therefore reserved and to not be occupied / maintain a value of 0. This is to aid further calculations within the pipeline as well as to maintain symmetry with the architectural registers in RISC-IV, of which x0 is reserved to always keep a value of 0. The free pool is created in the Rename module.
Reservation Station	The reservation station holds all the decoded information for incoming instructions as well as their corresponding renamed physical registers in a 2D table. This table also contains signal bits that indicate whether the (physical) source registers of an instruction are ready, which will allow - assuming a relevant functional unit is also ready - for the instruction to be executed. All relevant control signals and physical registers are stored in this table, which paired with the physical register file allows for the instructions to be executed. The reservation station is first created in the Dispatch module, and subsequently accessed in the Fire module.
ROB (Reorder Buffer)	The reorder buffer keeps track of what order the instructions are originally in, so that they can be retired in-order. While the processor efficiency benefits greatly from out-of-order execution, the ROB ensures that those instructions are still retired in order to prevent data hazards. The reorder buffer is created in the dispatch module and is subsequently accessed and modified in the retire module.

Table 2: Microarchitecture Blocks Necessary for Out of Order Execution

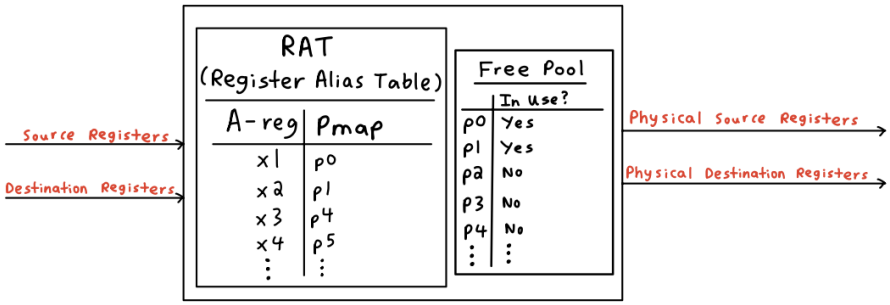

Figure 2: Block Diagram for RAT (Register Alias Table)

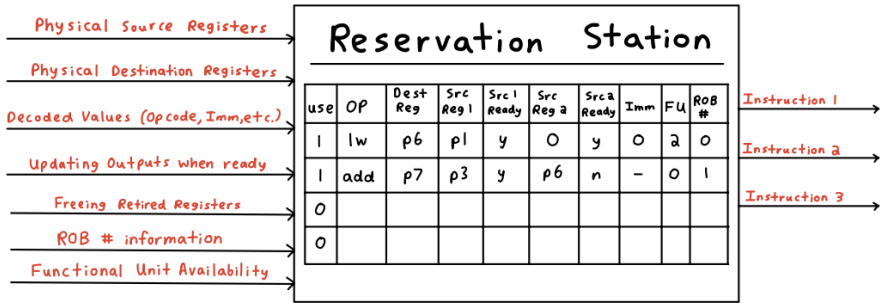

Figure 3: Block Diagram for Reservation Station

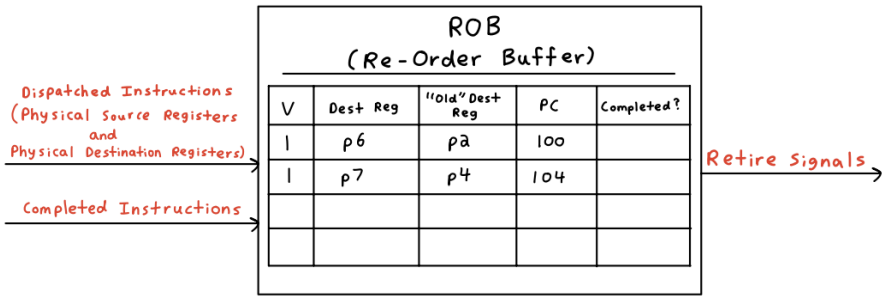

Figure 4: Block Diagram for ROB (Re-order Buffer)

3 Results

3.1 Functional Verification

Verification of digital processor designs is a complex task that must take into account a number of different possible cases to ensure correctness. Of the many test cases which were verified, two example test cases are shown. The first example involves instructions that do not require access to memory (ADDI, ANDI, ADD, SUB, XOR, SRA). The second example includes these instructions as well as LW and SW, both of which require access to memory in order to either load data or store data. In both cases, the input to the processor is a text file containing hexadecimal values, each corresponding to one byte of data (hence 4 creating a 32 bit instruction). Regarding the results, only the first 12 architectural registers are shown for brevity (rest are unused in these test cases).

Example 1. Instructions including ADDI, ANDI, ADD, SUB, XOR, SRA (Non memory):
These are the corresponding machine language instructions and calculations:

```
addi x2, x0, 6 -> x2 = 0 + 6 = 6 = 0110
addi x3, x0, 15 -> x2 = 0 + 15 = 15 = 1111
andi x5, x3, 1 -> x5 = 1111 & 0001 = 0001
addi x1, x0, 36 -> x1 = 0 + 36 = 36 = 100100
add x4, x5, x2 -> x4 = 0001 + 0110 = 0111
sub x0, x0, x0 -> x0 = 0 (by definition)
addi x9, x2, 23 -> x9 = x2 + 23 = 0110 + 10111 = 11101
xor x6, x2, x3 -> x6 = 0110 ^ 1111 = 1001
sub x2, x1, x6 -> x2 = 100100 - 1001 = 011011
sra x7, x3, x0 -> x7 = 1111 >> 0 = 1111
sra x3, x3, x5 -> x3 = 1111 >> 0001 = 0111
andi x9, x3, 23 -> x9 = 0111 & 10111 = 00110

Final register file values:
x0: 00000000000000000000000000000000
```

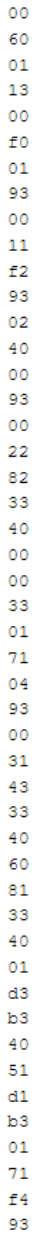

Figure 5: Test Case 1 Non-memory Instructions (hexadecimal values)

```
x1: 00000000000000000000000000100100
x2: 00000000000000000000000000011011
x3: 00000000000000000000000000000111
x4: 00000000000000000000000000000111
x5: 00000000000000000000000000000001
x6: 00000000000000000000000000001001
x7: 00000000000000000000000000001111
x8: 00000000000000000000000000000000
x9: 00000000000000000000000000000111
x10: 00000000000000000000000000000000
x11: 00000000000000000000000000000000
x12: 00000000000000000000000000000000
```

```
# Register File 0 (x0): 00000000000000000000000000000000
# Register File 1 (x1): 00000000000000000000000000100100
# Register File 2 (x2): 00000000000000000000000000011011
# Register File 3 (x3): 00000000000000000000000000000111
# Register File 4 (x4): 00000000000000000000000000000111
# Register File 5 (x5): 00000000000000000000000000000001
# Register File 6 (x6): 00000000000000000000000000001001
# Register File 7 (x7): 00000000000000000000000000001111
# Register File 8 (x8): 00000000000000000000000000000000
# Register File 9 (x9): 00000000000000000000000000000111
# Register File 10 (x10): 00000000000000000000000000000000
# Register File 11 (x11): 00000000000000000000000000000000
# Register File 12 (x12): 00000000000000000000000000000000
```

Figure 6: Final ModelSim architectural registers displaying correct values (Example 1)

Example 2. Instructions including LW, SW, ADDI, ANDI, ADD, SUB, XOR, SRA (Includes memory):

```
addi x2, x0, 6  -> x2 = 0 + 0110 = 0110
addi x3, x0, 15   -> x3 = 0 + 1111 = 1111
addi x1, x2, 36  -> x1 = 0110 + 36 = 101010
sw x3, x3, 9  -> 1111 + 1001 = 11000 (store 1111 in memory[11000])
add x4, x3, x2  -> x4 = 1111 + 0110 = 010101
add x0, x0, x0  -> x0 = 0 (by definition)
sw x1, x0, 0  -> 0 + 0 = 0 -> (store 101010 in memory[0])
lw x2, 24, x0  -> x2 = memory[0 + 11000] = memory[11000] = 1111
xor x6, x2, x3  -> x6 = 1111 ^ 1111 = 0
lw x3, xo, 0  -> x3 = memory[0 + 0] = 101010
sub x5, x3, x4  -> 101010 - 10101 = 010101
sra x7, x2, x9  -> x7 = 1111 >> 0 = 1111
andi x9, x3, 23  -> x9 = 101010 & 10111 = 000010
sw x5, x7, 1  -> 1111 + 1 = 10000 (store 10101 in memory[10000])
```

```
Final register file values:
x0: 00000000000000000000000000000000
x1: 00000000000000000000000000101010
x2: 00000000000000000000000000001111
x3: 00000000000000000000000000101010
x4: 00000000000000000000000000010101
x5: 00000000000000000000000000010101
x6: 00000000000000000000000000000000
x7: 00000000000000000000000000001111
x8: 00000000000000000000000000000000
x9: 00000000000000000000000000000010
x10: 00000000000000000000000000000000
```

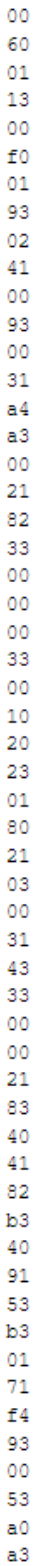

Figure 7: Test Case 2, All Instructions, including LW and SW (hexadecimal values)

```
x11: 00000000000000000000000000000000
x12: 00000000000000000000000000000000

memory[0]: 101010
memory[16]: 10101
memory[24]: 1111
```

```
# Register File 0 (x0): 00000000000000000000000000000000
# Register File 1 (x1): 00000000000000000000000000101010
# Register File 2 (x2): 00000000000000000000000000001111
# Register File 3 (x3): 00000000000000000000000000101010
# Register File 4 (x4): 00000000000000000000000000010101
# Register File 5 (x5): 00000000000000000000000000010101
# Register File 6 (x6): 00000000000000000000000000000000
# Register File 7 (x7): 00000000000000000000000000001111
# Register File 8 (x8): 00000000000000000000000000000000
# Register File 9 (x9): 00000000000000000000000000000010
# Register File 10 (x10): 00000000000000000000000000000000
# Register File 11 (x11): 00000000000000000000000000000000
# Register File 12 (x12): 00000000000000000000000000000000
```

Figure 8: Final ModelSim architectural registers displaying correct values (Example 2)

3.2 Area / Resource and Power Utilization

In addition to the verification of functionality, the area and power usage of the design must be taken into account. Since FPGA resources are fixed, the area usage is measured in terms of FPGA resource utilization. In Table 3, the area / resource utilization is shown. Table 4 displays the power usage for each module.

3.3 Additional Synthesis Results

The synthesis RTL Viewer results (from Quartus) for a the Decode module is displayed in Figure 5 as a reference. An example of the power output from the synthesis is also displayed in Figure 6.

4 Discussion

From a functional standpoint, the different modules each work to ensure their part of the pipeline is completed correctly before the instruction reaches the next module. The presence of the ROB, reservation station, free pool, and RAT ensure that out of order execution is completed properly without data hazards. The OoO execution aspect of the processor greatly helps in efficiently executing instructions without having to wait unnecessarily for previous instructions to finish. While the instructions must still be retired in-order, having the execution finish based on which instructions are available greatly helps in improving throughput. Pairing this with pipelining and being 2-issue superscalar further increases the processor throughput and overall performance. Combining the different optimizations has a strong effect; for example, the superscalar aspect of the processor causes more instructions to be stored in the reservation station each clock cycle. This in turn presents more options to the OoO processor to complete instructions in whatever order is most efficient, rather than simply waiting for instructions to complete.

The design was simulated and verified using ModelSim. In Quartus, the design was synthesized to find area and power results. In processor design, PPA (Power, Performance, and Area) values are of interest to use as an evaluation metric. Regarding the area results, the number of logic elements

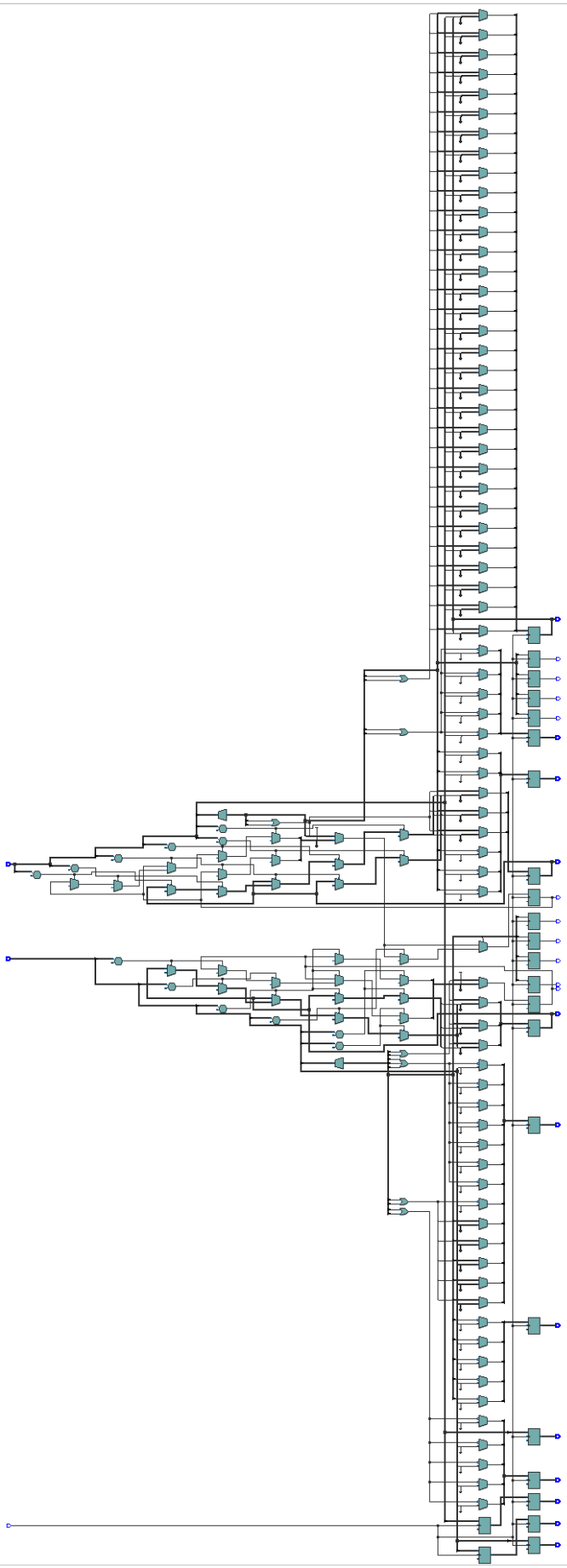

Figure 9: Decode Module RTL Viewer post synthesis results in Quartus

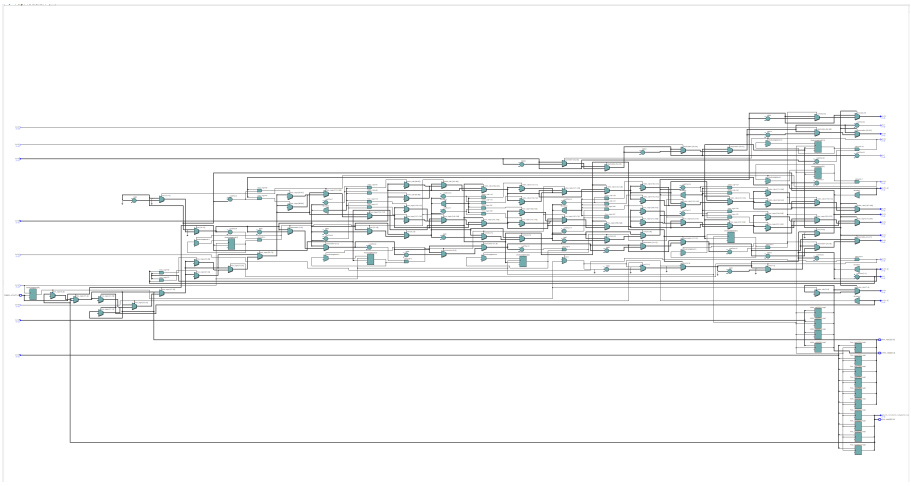

Figure 10: Retire Module Part 1 RTL Viewer post synthesis results in Quartus

Figure 11: Retire Module Part 2 RTL Viewer post synthesis results in Quartus

Power Analyzer Status	Successful - Tue Mar 26 14:33:19 2024
Quartus Prime Version	21.1.0 Build 842 10/21/2021 SJ Lite Edition
Revision Name	TopModule
Top-level Entity Name	Fetch2
Family	MAX 10
Device	10M08DAF484C8G
Power Models	Final
Total Thermal Power Dissipation	45.79 mW
Core Dynamic Thermal Power Dissipation	0.00 mW
Core Static Thermal Power Dissipation	39.76 mW
I/O Thermal Power Dissipation	6.03 mW

Figure 12: Fetch Module Power post synthesis results in Quartus

Fetch	Total Logic Elements: 1 Total Registers: 0 Total Modular Pins: 65 Total Combinational Functions: 1 Logic element usage by number of LUT inputs 4 input functions: 0 3 input functions: 0 ≤2 input functions: 1
Decode	Total Logic Elements: 183 Total Registers: 138 Total Modular Pins: 189 Total Combinational Functions: 101 Logic element usage by number of LUT inputs 4 input functions: 46 3 input functions: 29 ≤2 input functions: 26
Rename	Total Logic Elements: 423 Total Registers: 85 Total Modular Pins: 185 Total Combinational Functions: 412 Logic element usage by number of LUT inputs 4 input functions: 240 3 input functions: 150 ≤2 input functions: 23
Dispatch	Total Logic Elements: 1576 Total Registers: 1400 Total Modular Pins: 1922 Total Combinational Functions: 204 Logic element usage by number of LUT inputs 4 input functions: 179 3 input functions: 2 ≤2 input functions: 13
Fire	Total Logic Elements: 257472 Total Registers: 3212 Total Modular Pins: 4909 Total Combinational Functions: 257472 Logic element usage by number of LUT inputs: 4 input functions: 245651 3 input functions: 9029 ≤2 input functions: 2792
Retire	Total Logic Elements: 33098 Total Registers: 1060 Total Modular Pins: 36293 Total Combinational Functions: 33098 Logic element usage by number of LUT inputs: 4 input functions: 32789 3 input functions: 150 ≤2 input functions: 159

Table 3: Area / Resource Utilization of Modules on FPGA

Fetch	45.79 mW
Decode	46.20 mW
Rename	46.19 mW
Dispatch	46.12 mW
Fire	46.05 mW
Retire	45.96 mW

Table 4: Power Usage of Modules on FPGA

is lowest in the Fetch module, which intuitively makes sense due to the simple nature of the task (simply assigning values from a text file to memory). The Fire module uses the most amount of logic elements, which is to be expected as it contains the logic where multiple instructions are taken from the reservation station and executed in each of the functional units. Most of the logic elements are allocated for the more complex modules later in the pipeline, such as fire, retire, and dispatch. There are a number of loops in each of these modules which require many logic element resources to implement. The input pins to the Retire module seem high (36293), but the majority of them are unused memory bits. If desired, the memory available to the module can be lessened to reduce these modular pins. The default toggle rate used for input I/O signals for power calculations was set to 12.5%, which is the default in Quartus. The default toggle rate for the remaining signals - non I/O - was also set to 12.5%. In regards to the power usage of the design, the modules are similar in their estimated usage. The fetch module has the lowest estimated power usage which is to be expected due to its simple logic.

5 Conclusion

The design of processors can be complex with many tradeoffs needing to be taken into account. The processor built for this project successfully accomplishes OoO (out of order) execution, superscalar execution (2 issue), and pipelining. The processor achieves this by utilizing a reservation station, ROB (Re-order buffer), free pool, register renaming, forwarding, and a physical register file. In addition to functionality, processors should be designed in a manner that optimizes for throughput, efficiency, power, and area / resource utilization. Each module should be designed with minimal redundancy in logic and hardware in order to maximize the efficiency of the processor.

6 Acknowledgements

This project was primarily created with the knowledge given in Professor Nader Sehatbakhsh's UCLA ECE M116C (Computer Architecture) course. Here are links to the lecture videos from Fall 2023 which are most relevant to this project:

1. L10 - Superscalar and Intro to OoO
2. L11 - Out of Order Execution
3. L12 - Cache Design

References

[1] Nader Sehatbakhsh, *Superscalar and Intro to OoO*, https://www.youtube.com/watch?v=nrJxRNfmgbQ&ab_channel=NaderSehatbakhsh

[2] Nader Sehatbakhsh, *Out of Order Execution*, https://www.youtube.com/watch?v=db6UHxfw0C4&ab_channel=NaderSehatbakhsh

[3] Nader Sehatbakhsh, *Cache Design*, https://www.youtube.com/watch?v=RwdiDAaK6U8&ab_channel=NaderSehatbakhsh