2024학년도 컴퓨터구조 Lab Assignment #4-2
Pipelined CPU w/ control flow instructions

김도영, 선민수

2024년 5월 14일

1 Introduction

이 과제에서는 Verilog를 이용하여 이전 과제인 Multi Cycle CPU를 개선한 Pipelined CPU를 구현하는 것을 목적으로 한다. Pipelined CPU의 가장 큰 차이점은 기존 Single Cycle CPU나 Multi Cycle CPU의 경우 동시에 하나의 instruction만을 처리할 수 있었지만, Pipelined CPU는 Pipelining을 통하여 동시에 최대 5개의 instruction을 실행할 수 있는 점이다. 기존의 Multi Cycle CPU 대비 장점은 다음과 같다.

- Multi Cycle CPU에서도 개선되지 못하였던, Instruction 실행 시 모듈들이 작동하지 않고 유휴 상태로 낭비되는 것을 Pipelining을 통하여 막았고 효율적으로 사용한다.
- Multi Cycle CPU까지도 한 개의 Instruction이 끝나기 전까지는 이전의 Instruction이 절대 실행 혹은 처리되지 않았지만, Pipelining을 통해서 각 스테이지별 Instruction을 사용하여 Instruction의 Throughput을 증가시켰다.

이전 4-1과는 달리 control flow instruction을 지원하도록 구현하였으며, 이는 2-bit Saturation Branch Predictor를 통해 지원된다.(본문에서는 기존 4-1과 다른 점 위주로 서술한다.)

2 Design

본 과제에서 구현한 Pipelined CPU에서의 Control Flow Data Path는 교과서의 강의 교안을 기준으로 구현하였으며, Branch Prediction은 2-bit Saturation Branch Prediction을 적용하여 구현하였다.

Pipelined CPU는 이전 Multi Cycle CPU와 같이 IF - ID - EX - MEM - WB의 스테이지로 나뉘어서 실행된다. Multi Cycle CPU와의 차이점은 모든 스테이지가 동시에 사용된다는 점으로, Instruction 1이 ID 스테이지에서 처리될 때, 다른 Instruction 2는 IF 스테이지를 거치는 식으로 동시에 최대 5개의 instruction을 처리한다.

Data hazard가 발생하는 경우에는 Hazard Detection Module을 통해서 Hazard 발생을 탐지하고 Hazard가 발생된 원인(레지스터의 WB Stage, Store Instruction의 MEM Stage 등)이 모두 완료될 때까지 Stall하여 이를 해결한다.

레지스터가 아직 WB되지 않아 발생하는 Hazard는 Data Forwarding을 통하여 WB 이전에 값을 받아와 Stall되는 시간을 줄여, 최적화할 수 있다. 해당 과제에서는 Data Forwarding을 함께 구현한다.

구현한 Pipelined CPU는 아래와 같은 구조를 가진다. (단, 그림에서 Data Forwarding Unit과 Hazard Detection Module은 표현되지 않았다.) 구현에서의 `wire`나 `reg`의 이름은 아래 그림에서 제시된 이름을 사용한다.

Design of Pipelined CPU

Pipelined CPU의 Control Flow Data Path에 관여하는 세부 모듈들과 각각의 역할은 아래와 같다.

- PC: 현재의 program counter 값을 저장하는 모듈로, clock의 positive edge마다 `next_pc` 신호를 받아 `pc_write` 신호가 `1`일 때 program counter를 업데이트하는 동기 회로이다.
- Hazard Detection Module: 현재 실행되어야 하는 Instruction이 앞서 실행된 Instruction의 실행 종료를 기다려야 하는지에 대한 여부를 판단하는 모듈이다. 주어진 Instruction에 대해 이전 Instruction의 Dependency를 계산하는 비동기 회로이다.
- Forwarding Unit: Stall로 인한 딜레이를 최소화하고자 ALU의 피연산자를 MEM 혹은 WB Stage에서 바로 Fetch할 수 있는지에 대한 여부를 판단하여, MEM 혹은 WB Stage에서의 사용때문에 생기는 Stall을 줄일 수 있는 모듈이다. 주어진 Instruction에 대해 해당 피연산자들이 어느 단계에서 사용되고 있는지를 판단하는 비동기 회로이다.
- Control: 현재 명령어의 opcode를 받아 명령어의 실행 과정에 따라 해당하는 control 신호를 계산하는 비동기 회로이다.
- Registers: CPU의 programmer visible state 중 하나인 레지스터이다. `x0`부터 `x31`까지 총 32개를 가지고 있다. Register의 업데이트는 clock의 negative edge에서 업데이트되는 동기 회로이며, Register의 읽기의 경우 비동기 회로이다.
- Immediate Generator: 명령어를 받아 명령어에 따른 immediate 값을 계산하는 모듈로, 비동기 회로이다.
- ALU control: ALU가 수행해야 할 연산을 지정해주는 ALU control 신호를 계산하는 비동기 회로 모듈이다.
- ALU: 두 입력값과 ALU contorl 신호를 받아 해당하는 연산을 하는 모듈로, 입력이 바뀌면 출력도 곧바로 바뀌는 비동기 회로이다.
- Memory: 실제 CPU에 연결된 memory의 역할을 하는 모듈로, Instruction Memory와 Data Memory를 구분하여 사용한다. clock의 positive edge마다 입력에 따라 메모리 값을 변경하는 동기 회로이다.
- Pipeline Register: 각 Stage 별로 다음 Stage에 넘겨주어야 할 Control Signal이나 Register의 정보 등의 신호를 저장하는 Register이다. 스테이지의 사이마다 존재하여 값을 넘겨준다. Clock의 Positive Edge마다 Stage에서 계산된 결과를 Fetch하여 Register의 값을 변경하는 동기 회로이다.

3 Implementation

3.1 Program Counter

```
always @(posedge clk) begin
    if(reset)
        current_pc <= 0;
    else if(pc_write)
        current_pc <= next_pc;
end
```

PC.v

Program Counter는 주어진 clk에 따라서 current_pc를 next_pc로 업데이트하는 Synchronous 모듈이다.

3.2 Control Unit

```
always @(*) begin
    mem_read = (opcode == `LOAD);
    mem_to_reg = (opcode == `LOAD);
    mem_write = (opcode == `STORE) && (!is_hazard);
    alu_src = (opcode != `ARITHMETIC) && (opcode != `BRANCH);
    write_enable = (opcode != `STORE) && (opcode != `BRANCH) && (!is_hazard);
    alu_op = {
        (opcode == `ARITHMETIC) ||
        (opcode == `ARITHMETIC_IMM) ||
        (opcode == `BRANCH),
        1'b0
    };
    is_ecall = (opcode == `ECALL);
end
```

ControlUnit.v

Control Unit은 주어진 instruction에서 opcode에 따라 필요한 Control Value를 계산하여 산출하는 Asynchronous 모듈이다.

3.3 ALU Control Unit

```
always @(*) begin
    case(aluOp)
        2'b00: alu_op = `ALU_ADD;
        2'b01: alu_op = `ALU_SUB;
        default: begin
            case(instruction[6:0])
                `ARITHMETIC: begin
                    case(instruction[14:12])
                        `FUNCT3_ADD: alu_op = (instruction[30]) ? `ALU_SUB : `ALU_ADD;
                        `FUNCT3_SLL: alu_op = `ALU_SLL;
                        `FUNCT3_SRL: alu_op = `ALU_SLR;
                        `FUNCT3_AND: alu_op = `ALU_AND;
                        `FUNCT3_OR:  alu_op = `ALU_OR;
```

```
                        `FUNCT3_XOR: alu_op = `ALU_XOR;
                        default:     alu_op = 4'b0;
                    endcase
                end
                `ARITHMETIC_IMM: begin
                    case(instruction[14:12])
                        `FUNCT3_ADD: alu_op = `ALU_ADD;
                        `FUNCT3_SLL: alu_op = `ALU_SLL;
                        `FUNCT3_SRL: alu_op = `ALU_SLR;
                        `FUNCT3_AND: alu_op = `ALU_AND;
                        `FUNCT3_OR:  alu_op = `ALU_OR;
                        `FUNCT3_XOR: alu_op = `ALU_XOR;
                        default:     alu_op = 4'b0;
                    endcase
                end
                `BRANCH: begin
                    case(instruction[14:12])
                        `FUNCT3_BEQ: alu_op = `ALU_BEQ;
                        `FUNCT3_BNE: alu_op = `ALU_BNE;
                        `FUNCT3_BLT: alu_op = `ALU_BLT;
                        `FUNCT3_BGE: alu_op = `ALU_BGE;
                        default:     alu_op = 4'b0;
                    endcase
                end
                `LOAD:   alu_op = `ALU_ADD;
                `STORE:  alu_op = `ALU_ADD;
                `JAL:    alu_op = `ALU_ADD;
                `JALR:   alu_op = `ALU_ADD;
                `ECALL:  alu_op = `ALU_BEQ;
                default: alu_op = 4'b0;
            endcase
        end
    endcase
end
```

ALUControlUnit.v

ALU Control Unit은 주어진 funct3, funct7, opcode를 이용해 alu_op 신호를 생성하는 Asynchronous 모듈이다.

3.4 Immediate Generator

```
always @(*) begin
    case(opcode)
    `ARITHMETIC_IMM: imm_gen_out = {{20{part_of_inst[31]}}, part_of_inst[31:20]};
    `LOAD          : imm_gen_out = {{20{part_of_inst[31]}}, part_of_inst[31:20]};

    `STORE: begin
      imm_gen_out = {{20{part_of_inst[31]}}, part_of_inst[31:25], part_of_inst[11:7]};
    end

    `BRANCH: begin
      imm_gen_out = {
        {19{part_of_inst[31]}},
        part_of_inst[31],
```

```
          part_of_inst[7],
          part_of_inst[30:25],
          part_of_inst[11:8],
          1'b0
        };
      end

      `JAL: begin
        imm_gen_out = {
          {11{part_of_inst[31]}},
          part_of_inst[31],
          part_of_inst[19:12],
          part_of_inst[20],
          part_of_inst[30:21],
          1'b0
        };
      end

      `JALR  : imm_gen_out = {{20{part_of_inst[31]}}, part_of_inst[31:20]};
      `LUI   : imm_gen_out = {part_of_inst[31:12], 12'b0};
      `AUIPC : imm_gen_out = {part_of_inst[31:12], 12'b0};
      default: imm_gen_out = {32{1'b0}};
      endcase
  end
```

ImmediateGenerator.v

Immediate Generator는 주어진 instruction의 opcode에 따라서 immediate value를 추출하는 Asynchronous 모듈이다. Immediate value가 필요하지 않을 경우 32'b0으로 출력한다.

3.5 ALU

```
always @(*) begin
    case(alu_op)
        `ALU_ADD: begin
            alu_result = alu_in_1 + alu_in_2;
            alu_zero = 0;
        end
        `ALU_SUB: begin
            alu_result = alu_in_1 - alu_in_2;
            alu_zero = 0;
        end
        `ALU_AND: begin
            alu_result = alu_in_1 & alu_in_2;
            alu_zero = 0;
        end
        `ALU_OR: begin
            alu_result = alu_in_1 | alu_in_2;
            alu_zero = 0;
        end
        `ALU_XOR: begin
            alu_result = alu_in_1 ^ alu_in_2;
            alu_zero = 0;
        end
        `ALU_SLL: begin
```

```
            alu_result = alu_in_1 << alu_in_2;
            alu_zero = 0;
        end
        `ALU_SLR: begin
            alu_result = alu_in_1 >> alu_in_2;
            alu_zero = 0;
        end
        `ALU_BEQ: begin
            alu_result = 32'b0;
            alu_zero = alu_in_1 == alu_in_2;
        end
        `ALU_BNE: begin
            alu_result = 32'b0;
            alu_zero = alu_in_1 != alu_in_2;
        end
        `ALU_BLT: begin
            alu_result = 32'b0;
            alu_zero = alu_in_1 < alu_in_2;
        end
        `ALU_BGE: begin
            alu_result = 32'b0;
            alu_zero = alu_in_1 >= alu_in_2;
        end
        default: begin
            alu_result = 32'b0;
            alu_zero = 0;
        end
    endcase
end
```

ALU.v

ALU는 주어진 opcode, funct3, funct7과 alu_in_1, alu_in_2에 따라 연산 결과값을 계산하는 Asynchronous 모듈로 구현하였다. 필요한 instruction의 경우에는 alu_result와 alu_bcond를 0으로 처리하였다.

3.6 Hazard Detection Module

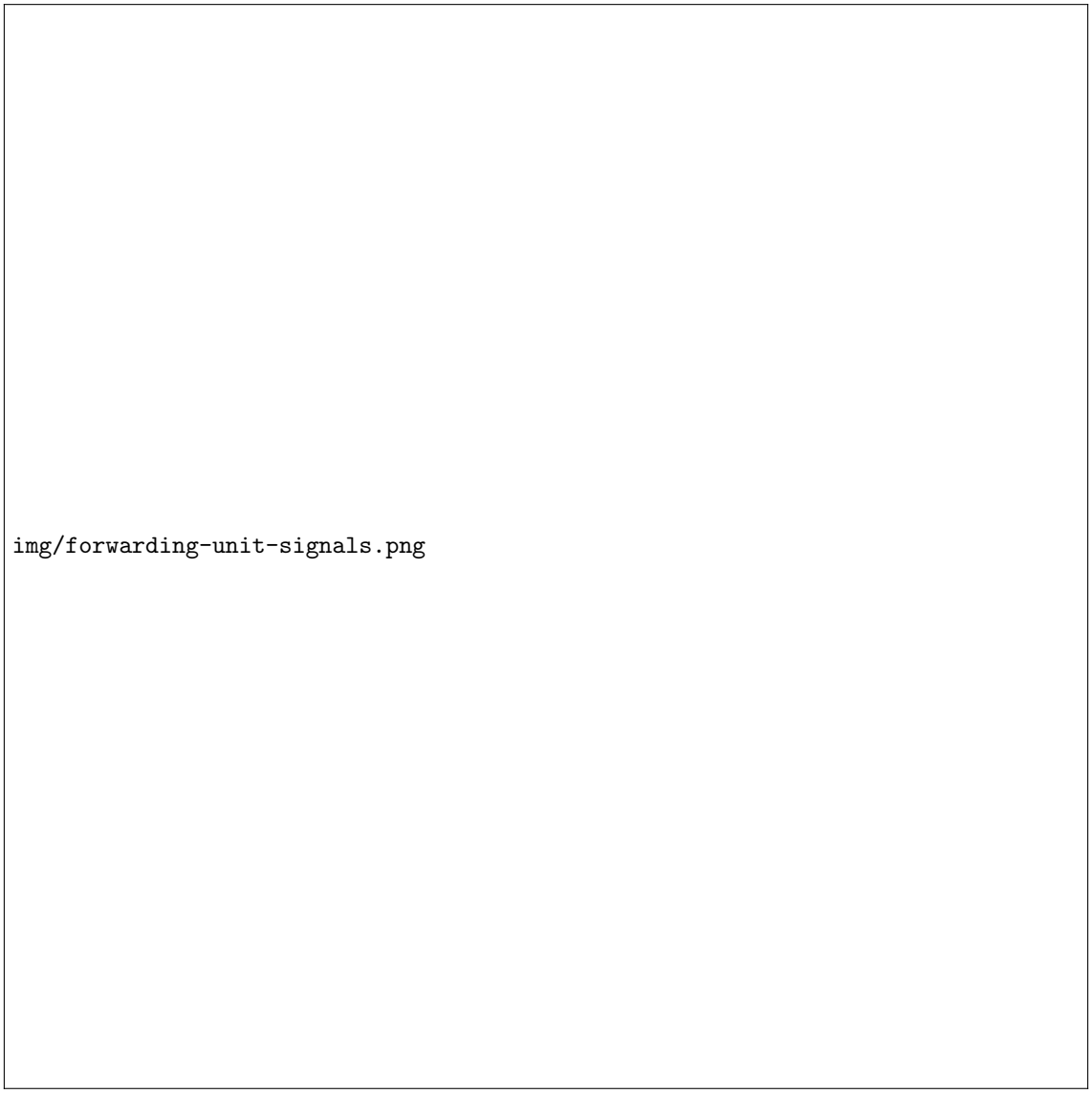
```
wire [6:0] ID_opcode = ID_inst[6:0];
wire is_ecall = (ID_opcode == `ECALL);
wire [4:0] ID_rs1 = is_ecall ? 17 : ID_inst[19:15];
wire [4:0] ID_rs2 = ID_inst[24:20];

wire use_rs1 = (
    (ID_opcode != `LUI) || (ID_opcode != `AUIPC) || (ID_opcode != `JAL)
) && ID_rs1 != 5'b0;

wire use_rs2 = (
    (ID_opcode == `ARITHMETIC) || (ID_opcode == `STORE) ||
    (ID_opcode == `BRANCH)
) && ID_rs2 != 5'b0;

assign is_hazard = (
    (ID_rs1 == EX_rd) && use_rs1 || (ID_rs2 == EX_rd) && use_rs2
) && EX_mem_read;
```

HazardDetection.v

Hazard Detection Module은 현재 Data Hazard가 발생하여 다른 Stage의 실행이 완료될 때까지 기다려야 되는 경우를 판단하는 모듈이다. Instruction의 종류와 사용되는 Register의 종류를 판단하여 `is_hazard` 신호를 통해 이를 알리는 비동기 회로이다.

3.7 Forwarding Unit

Forwarding Unit은 교재에서 제시하고 있는 기준 그대로 Signal을 지정하여 구현하였다. 사용된 Signal의 상세 내용은 아래의 그림과 같다.



Forwarding Unit Signals Specification

```
wire [6:0] ID_opcode = ID_inst[6:0];
wire is_ecall = (ID_opcode == `ECALL);
wire [4:0] ID_rs1 = is_ecall ? 17 : ID_inst[19:15];
wire [4:0] ID_rs2 = ID_inst[24:20];
```

```
5
6   wire use_rs1 = (
7       (ID_opcode != `LUI) || (ID_opcode != `AUIPC) || (ID_opcode != `JAL)
8   ) && ID_rs1 != 5'b0;
9
10  wire use_rs2 = (
11      (ID_opcode == `ARITHMETIC) || (ID_opcode == `STORE) ||
12      (ID_opcode == `BRANCH)
13  ) && ID_rs2 != 5'b0;
14
15  assign is_hazard = (
16      (ID_rs1 == EX_rd) && use_rs1 || (ID_rs2 == EX_rd) && use_rs2
17  ) && EX_mem_read;
18
19  always @(*) begin
20      if(is_ecall && (EX_rd == 17))
21          forward_1 = 2'b11;
22      else if((EX_rs1 != 5'b0) && (EX_rs1 == MEM_rd) && MEM_reg_write)
23          forward_1 = 2'b10;
24      else if((EX_rs1 != 5'b0) && (EX_rs1 == WB_rd) && WB_reg_write)
25          forward_1 = 2'b01;
26      else
27          forward_1 = 2'b00;
28
29      if((EX_rs2 != 5'b0) && (EX_rs2 == MEM_rd) && MEM_reg_write)
30          forward_2 = 2'b10;
31      else if((EX_rs2 != 5'b0) && (EX_rs2 == WB_rd) && WB_reg_write)
32          forward_2 = 2'b01;
33      else
34          forward_2 = 2'b00;
35  end
```

ForwardingUnit.v

Forwarding Unit은 현재 ALU에서 사용되어야할 피연산자들이 앞선 EX, MEM, WB Stage 등에서 사용되어 있는지를 판단하고 사용할 수 있을 경우 사전 정의된 `ForwardA`, `ForwardB`를 통해서 알리는 모듈이다. 주어진 Instruction을 이용하여 Signal을 생성하는 Asynchronous 모듈로 구현하였다.

3.8 CPU

(아래는 CPU Module의 구현으로, 위에서 설명된 Module의 선언에 대해서는 생략한다.)

```
1   // Update IF/ID pipeline registers here
2   always @(posedge clk) begin
3     if (reset) begin
4       IF_ID_inst <= 32'b0;
5     end
6     else if (!ID_is_hazard) begin
7       IF_ID_inst <= IF_inst;
8     end
9   end
10
11  // Update ID/EX pipeline registers here
12  always @(posedge clk) begin
13    if (reset) begin
```

```
14      ID_EX_alu_op <= 0;
15      ID_EX_alu_src <= 0;
16      ID_EX_mem_write <= 0;
17      ID_EX_mem_read <= 0;
18      ID_EX_mem_to_reg <= 0;
19      ID_EX_reg_write <= 0;
20      ID_EX_rs1_data <= 32'b0;
21      ID_EX_rs2_data <= 32'b0;
22      ID_EX_imm <= 32'b0;
23      ID_EX_ALU_ctrl_unit_input <= 0;
24      ID_EX_rs1 <= 5'b0;
25      ID_EX_rs2 <= 5'b0;
26      ID_EX_rd <= 5'b0;
27      ID_EX_is_halted <= 0;
28    end
29    else begin
30      ID_EX_alu_op <= ID_alu_op;
31      ID_EX_alu_src <= ID_alu_src;
32      ID_EX_mem_write <= ID_mem_write;
33      ID_EX_mem_read <= ID_mem_read;
34      ID_EX_mem_to_reg <= ID_mem_to_reg;
35      ID_EX_reg_write <= ID_reg_write;
36      ID_EX_rs1_data <= ID_rs1_data;
37      ID_EX_rs2_data <= ID_rs2_data;
38      ID_EX_imm <= ID_imm;
39      ID_EX_ALU_ctrl_unit_input <= ID_ALU_ctrl_unit_input;
40      ID_EX_rs1 <= ID_rs1;
41      ID_EX_rs2 <= ID_rs2;
42      ID_EX_rd <= ID_rd;
43      ID_EX_is_halted <= ID_is_halted;
44    end
45  end
46
47  // Update EX/MEM pipeline registers here
48  always @(posedge clk) begin
49    if (reset) begin
50      EX_MEM_mem_write <= 0;
51      EX_MEM_mem_read <= 0;
52      EX_MEM_is_branch <= 0;
53      EX_MEM_mem_to_reg <= 0;
54      EX_MEM_reg_write <= 0;
55      EX_MEM_alu_out <= 32'b0;
56      EX_MEM_dmem_data <= 32'b0;
57      EX_MEM_rd <= 5'b0;
58      EX_MEM_is_halted <= 0;
59    end
60    else begin
61      EX_MEM_mem_write <= EX_mem_write;
62      EX_MEM_mem_read <= EX_mem_read;
63      EX_MEM_is_branch <= EX_is_branch;
64      EX_MEM_mem_to_reg <= EX_mem_to_reg;
65      EX_MEM_reg_write <= EX_reg_write;
66      EX_MEM_alu_out <= EX_alu_out;
67      EX_MEM_dmem_data <= EX_dmem_data;
68      EX_MEM_rd <= EX_rd;
69      EX_MEM_is_halted <= EX_is_halted;
```

```
70    end
71  end
72
73  // Update MEM/WB pipeline registers here
74  always @(posedge clk) begin
75    if (reset) begin
76      MEM_WB_mem_to_reg <= 0;
77      MEM_WB_reg_write <= 0;
78      MEM_WB_mem_to_reg_src_1 <= 32'b0;
79      MEM_WB_mem_to_reg_src_2 <= 32'b0;
80      MEM_WB_is_halted <= 0;
81    end
82    else begin
83      MEM_WB_mem_to_reg <= MEM_mem_to_reg;
84      MEM_WB_reg_write <= MEM_reg_write;
85      MEM_WB_mem_to_reg_src_1 <= MEM_mem_to_reg_src_1;
86      MEM_WB_mem_to_reg_src_2 <= MEM_mem_to_reg_src_2;
87      MEM_WB_rd <= MEM_rd;
88      MEM_WB_is_halted <= MEM_is_halted;
89    end
90  end
91
92  always @(posedge clk) begin
93    is_halted <= MEM_WB_is_halted;
94  end
95
96  always @(*) begin
97    case(EX_forward_1)
98    2'b00: EX_ALU_in_1 = ID_EX_rs1_data;
99    2'b01: EX_ALU_in_1 = WB_rdin_data;
100   2'b10: EX_ALU_in_1 = EX_MEM_alu_out;
101   default: EX_ALU_in_1 = ID_EX_rs1_data;
102   endcase
103
104   case(EX_forward_2)
105   2'b00: EX_ALU_rs2_data = ID_EX_rs2_data;
106   2'b01: EX_ALU_rs2_data = WB_rdin_data;
107   2'b10: EX_ALU_rs2_data = EX_MEM_alu_out;
108   default: EX_ALU_rs2_data = ID_EX_rs2_data;
109   endcase
110 end
111
112 always @(*) begin
113   if(EX_forward_1 == 3)
114     ID_ecall_comp = EX_alu_out;
115   else
116     ID_ecall_comp = ID_rs1_data;
117 end
118
119 assign next_pc = current_pc + 4;
120
121 assign ID_PC = IF_ID_PC;
122 assign ID_ALU_ctrl_unit_input = IF_ID_inst;
123 assign ID_rd = IF_ID_inst[11: 7];
124 assign ID_rs1 = ID_is_ecall ? 17 : IF_ID_inst[19:15];
125 assign ID_rs2 = IF_ID_inst[24:20];
```

```
126 assign ID_is_halted = (ID_is_ecall && (ID_ecall_comp == 10));
127
128 assign EX_ALU_in_2 = ID_EX_alu_src ? ID_EX_imm : EX_ALU_rs2_data;
129 assign EX_reg_write = ID_EX_reg_write;
130 assign EX_mem_to_reg = ID_EX_mem_to_reg;
131 assign EX_is_branch = ID_EX_is_branch;
132 assign EX_mem_read = ID_EX_mem_read;
133 assign EX_mem_write = ID_EX_mem_write;
134 assign EX_shifted_imm = ID_EX_imm << 2;
135 assign EX_rd = ID_EX_rd;
136 assign EX_is_halted = ID_EX_is_halted;
137 assign EX_dmem_data = EX_ALU_rs2_data;
138
139 assign MEM_reg_write = EX_MEM_reg_write;
140 assign MEM_mem_to_reg = EX_MEM_mem_to_reg;
141 assign MEM_PCSrc = EX_MEM_is_branch & EX_MEM_alu_bcond;
142 assign MEM_mem_to_reg_src_2 = EX_MEM_alu_out;
143 assign MEM_rd = EX_MEM_rd;
144 assign MEM_is_halted = EX_MEM_is_halted;
145
146 assign WB_rdin_data = (MEM_WB_mem_to_reg) ? MEM_WB_mem_to_reg_src_1 : MEM_WB_mem_to_reg_src_2;
```

cpu.v

CPU 모듈은 위 Design에서 제시한 대로 다른 모듈 간의 Wiring을 진행하며, 각 Stage별 Pipeline Register의 업데이트를 구현한다. 구현에서 사용된 각 wire와 reg의 이름은 아래의 Convention을 따른다.

- Pipeline Register: `<FIRST_STAGE_NAME>_<SECOND_STAGE_NAME>_<REGISTER_OR_WIRE_NAME>`
- Stage Register: `<STAGE_NAME>_<REGISTER_OR_WIRE_NAME>`

ECALL에 따라서 종료될 수 있도록 ECALL을 한 후 x17 레지스터의 값 비교를 통해 `is_halted`로 데이터를 생성해 Pipelining을 하도록 구현하였다. Pipelining에 의해서 ECALL 이전의 Instruction이 모두 실행 완료되어야 하므로, [is_halted] 또한 Stage를 모두 거쳐서 마지막 WB 이후 종료될 수 있도록 구현하였다.

4 Discussion

4.1 How to Handle Branch Prediction

가장 기본적인 Branch Prediction은 Always Taken과 Always Not Taken 전략이 있다. 이는 다음과 같은 특성을 가진다.

- Always Taken: 프로그램 실행 중 마주치는 모든 Branch Instruction에 대해서 Branch가 Taken이 될 것이라 예상하여 Branch Taken 시의 `PC`에 대해서 미리 Instruction을 Fetch한다.
- Always Not Taken: 프로그램 실행 중 마주치는 모든 Branch Instruction에 대해서 Branch가 Not Taken이 될 것이라 예상하여 Branch Not Taken 시의 `PC`인 `PC+4`에 대해서 미리 Instruction을 Fetch한다.

가장 기본적인 전략이지만, 특정 경우에 편향하여 예측하기 때문에, 이를 보완하기 위해서 Counter 기반의 전략을 선택하게 된다. Counter 기반의 대표적인 예시는 다음과 같다.

- 1-Bit Counter: `Predict Take, 1'b1`, `Predict Not Take, 1'b0`의 상태를 가지는 Counter 기반의 전략이다.
- 2-Bit Counter: `Strongly Taken, 1'b11`, `Weakly Taken, 1'b10`, `Weakly Not Taken, 1'b01`, `Strongly Not Taken, 1'b00`의 4가지 상태를 가지는 Counter 기반의 전략이다. 상태의 전환에 따라서 Saturation Counter와 Hysteresis Counter로 나뉜다.

Counter에서 사용되는 bit 수에 따라서 Branch Taken/Not Taken Pattern History를 확인할 수 있다. 초반의 Counter 기반 Branch Predictor는 Branch의 위치(Program Counter)마다 해당 Pattern History를 저장하여 예측하는 방법으로 구현되었다.

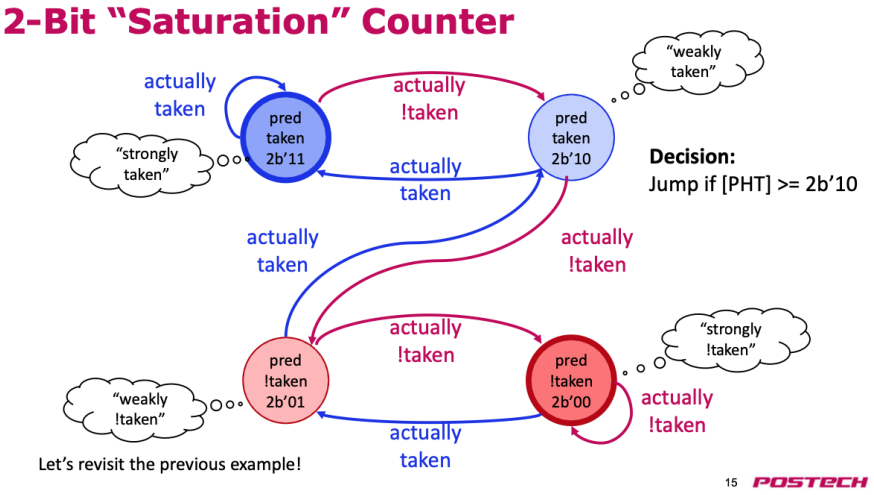
이에서 발전된 형태인 Global Path History를 적용하여 McFarling이 제시한 Gshare Branch Prediction는 기존의 PHT와 BTB의 원리처럼 각 PC에 해당하는 Pattern History도 저장하지만, 동시에 Global Path를 확인하여 Branch Prediction에 성능적 향상을 가져왔다.

이를 확장하여 2-Level Branch Direction Predictor 형태로 발전시켰는데, BHT(Branch History Table) + PHT(Pattern History Table)을 결합 형태이다. 각각의 BHT와 PHT의 형태가 Global, per-Set, per-Address인지에 따라서 다양한 형태의 Branch Predictor를 구성한다.

본 과제에서 우리는 2-Bit Saturation Counter 기반의 Branch Predictor를 사용하여 Branch Prediction을 진행하였다.

4.2 2-Bit Saturation Counter Branch Predictor의 구현

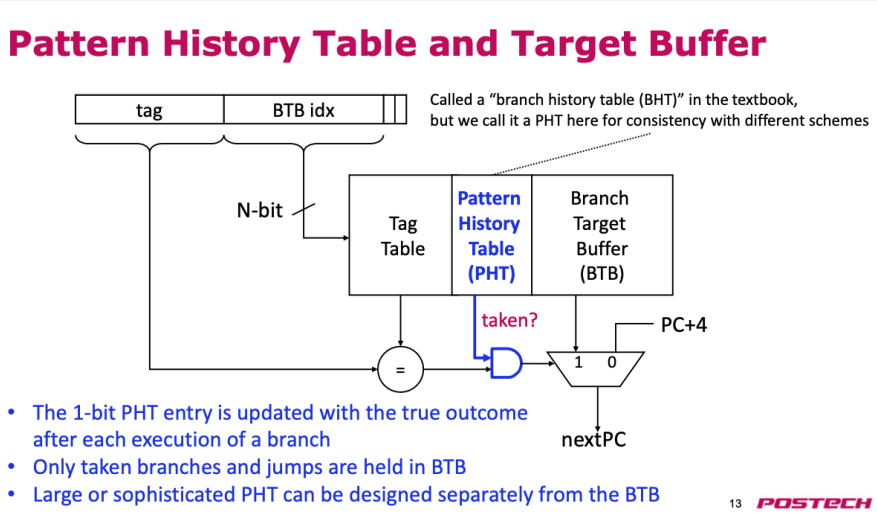
본 과제에서 구현한 2-Bit global branch predictor는 2-Bit Saturation Counter를 기반으로 한다. Branch Prediction에 의한 Counter의 유한 상태 기계의 Diagram은 다음과 같다.



FSM Diagram of 2-Bit Saturation Counter

Branch Predictor의 구조는 다음 그림과 같으며, BTB의 index bit는 N=4로 설정되어 전체 32개의 Branch Target Entry를 가지도록 디자인하였다.

2-Bit Saturation Counter는 Strongly Taken, Weakly Taken, Weakly Not Taken, Strongly Not Taken의 4가지 상태로 나타나며, 각 상태에 대한 Predictor의 Prediction과 실제 branch의 taken, not taken에 의한 상태 변화는 아래의 표와 같다. (실제 코드에서는 Implementation에서 확인할 수 있듯이, `BP_ST`, `BP_WT`, `BP_WN`, `BP_SN`으로 각 상태를 나타낸다.)



Branch Predictor with BTB and PHT

	Prediction	Taken(Real)	Not-Taken(Real)
Strongly Taken(2'b11)	Taken	Strongly Taken	Weakly Taken
Weakly Taken(2'b10)	Taken	Strongly Taken	Weakly Not Taken
Weakly Not Taken(2'b01)	Not Taken	Weakly Taken	Strongly Not Taken
Strongly Not Taken(2'b00)	Not Taken	Weakly Not Taken	Strongly Not Taken

Comparison on the Number of Cycles on Different Branch Predictor

4.3 다른 Branch Predictor와의 비교

다른 방식의 Branch Prediction(Always-Taken, Always-Not-Taken)과 본 과제에서 구현한 2-Bit Saturation Counter Branch Prediction 성능을 비교한 결과는 아래와 같다.

Input File	Always-Taken	Always-Not-Taken	Saturation Counter
basic	00	35	37
non-controlflow	00	45	46
ifelse	00	43	51
loop	00	300	357
recursive	00	1167	1165

Comparison on the Number of Cycles on Different Branch Predictor

Always-Taken, Always-Not-Taken과 본 과제에서 구현한 Saturation Counter 기반의 2-Bit Global Branch Predictor는 큰 차이를 보이지 않는 것을 확인할 수 있다. 그러나 이는 주어진 프로그램 상에서의 성능이기에 단언할 수 없으며, 상황에 따라 다른 성능을 보일 것이다.

5 Conclusion

이번 과제에서는 지난 4-1에 이어 Pipelined CPU의 Control Flow DataPath를 구현하는 것을 목적으로 하여, Pipelined CPU를 완성하였다. 기존 Multi-Cycle CPU와는 달리 Branch Taken/Not Taken 여부와 Branch Instruction 다음 Fetch하는 Instruction의 위치에 따라서 Stall이 발생할 수 있기 때문에, Branch Prediction이 중요하다. 2-Bit Saturation Counter 기반의 2-Bit Global Branch Predictor를 구현하여 기존 Always-Taken, Always-Not-Taken 전략과의 비교할 수 있었지만, 주어진 프로그램에 의존하는 특성상 정확한 비교를 할 수 없어 기대한 결과를 확인하지 못하였다. Single-Cycle CPU,

Multi-Cycle CPU, Pipelined CPU의 발전 동안 Data Memory 관점에서의 개선은 미비하였는데, Data Memory 접근성을 늘릴 수 있는 Cache의 필요성을 확인할 수 있었다.