ECE4700J Homework 1

Yiwen Yang

Q1

1. Since RISC has already proved to overwhelm CISC as time passed, we have entered the post-PC era where improvements on general-purpose processor ISA becomes quite difficult. At this moment, Moore's Law and Dennard Scaling both stopped, and we will have to look more on how the computer architecture implements ISA rather than ISA itself. The seemingly bottleneck in fact implies that we have already done very well and should look forward. Current challenge is about the inefficiency and wasted resources on mis-predictions of branches caused by ILP. Limited TDP pushes us to think more about energy saving, and in fact a small optimization in the compiler level may contribute greatly to the efficiency.

2. So far the industry has focused a lot on performance, but now it is time to think about hardware security. Rather than a correct implementation of ISA, we should design and produce a good implementation, where security is very important. In this era, we have the spare to make some encryptions on hardware level to fix the information leakage and potential vulnerabilities. Many attacks such as side-channel, Meltdown and Spectre are already known to us, which is not bad but instead helpful to figure out and fix problems at an early stage.

3. As an opportunity for future compiter architecture, we can now design specific-purpose processors to solve certain problems rather than sticking to general-purpose ones. The benefits of DSA and DSL are that complex designs are not necessary and very specific optimizations can be made with less precision yet greatly enhancing the processor performance in the domain, also parallelism problem can be solved partially. There are currently many new calculating hardwares to serve different needs, and the competition between large players will push the whole industry forward just like ISA competition in the last era.

Q2

1. CISC is more compiler-friendly, since instructions can be microcoded by hardware.

2. Addressing mode is more compound, such that memory access is more flexible.

Q3

CPI is calculated as

$$CPI = p_{load}C_{load} + p_{store}C_{store} + p_{ALU}C_{ALU} + p_{branch}C_{branch} + p_{jump}C_{jump}$$

Then we can get

	Method	Old CPI	New CPI	Speedup
sjeng	1	4.42	3.86	1.145
	2	4.42	4.09	1.080
mcf	1	4.8	4.51	1.064
	2	4.8	4.31	1.114

So **sjeng** will prefer the first method since it involves a large proportion of ALU operations, and **mcf** will prefer the second one since ALU speedup does not improve more than branch speedup.

Q4

Let p be the percentage that can be enhanced, then

$$\begin{aligned} \frac{p}{2.5} + 1 - p &= \frac{1}{1 + 10\%} \\ p &= 0.152 \end{aligned}$$

So at least 15.2% of the program should be run in enhanced mode.

Q5

1.

$$t = \frac{1}{f} = 3.70 \times 10^{-10} s$$

$$\begin{aligned} PDP &= P_{avg} \cdot t \\ &= 3.89 \times 10^{-8} J \end{aligned}$$

$$\begin{aligned} EDP &= PDP \cdot t \\ &= 1.44 \times 10^{-17} J \end{aligned}$$

$$\begin{aligned} EDDP &= EDP \cdot t \\ &= 5.33 \times 10^{-27} J \end{aligned}$$

2.

$$
\begin{aligned}
P_{dyn,old} &= 105W \times \frac{3}{4} = 78.75W \\
\frac{f_{new}}{f_{old}} &= \frac{P_{dyn,new}}{P_{dyn,old}} \\
&= 1.65 \\
f_{new} &= 1.65 f_{old} = 4.46GHz
\end{aligned}
$$

So the frequency can be sped up by 65% to 4.46 GHz.

Q6

From Bose-Einstein Formula

$$
\begin{aligned}
\text{Die yield} &= \text{Wafer yield} \times 1/(1 + \text{Defects per unit area} \times \text{Die area})^{\alpha} \\
&= 99\% \times 1/(1 + 0.016 \times 2.4)^{10} \\
&= 99\%
\end{aligned}
$$

Then

$$
\begin{aligned}
\text{Dies per wafer} &= \frac{\pi \times (\text{Wafer diameter}/2)^2}{\text{Die area}} - \frac{\pi \times \text{Wafer diameter}}{\sqrt{2 \times \text{Die area}}} \\
&= \frac{\pi \times (42/2)^2}{2.4} - \frac{\pi \times 42}{\sqrt{2 \times 2.4}} \\
&= 517
\end{aligned}
$$

$$
\begin{aligned}
\text{Cost of die} &= \frac{\text{Cost of wafer}}{\text{Dies per wafer} \times \text{Die yield}} \\
&= \frac{9000}{517 \times 99\%} \\
&= 17.58
\end{aligned}
$$

So expected cost of die is \$17.58.

Q7

$$
\begin{aligned}
S_{opt} &= \sqrt{\frac{(1-b)T}{bC}} \\
&= \sqrt{\frac{(1-0.2) \times \frac{1}{700 \times 10^6} \times (1 + 5\%)}{0.2 \times 500 \times 10^{-12}}} \\
&= 3.5
\end{aligned}
$$

$$f_{new} = 3.5 \times 700MHz = 2.45GHz$$

$$\begin{aligned} G &= (\frac{1}{1+(S-1)b}) \times (\frac{1}{(T/S)+C}) \\ &= (\frac{1}{1+(3.5-1)\times 0.2}) \times (\frac{1}{(1.5\times 10^{-9}/3.5)+500\times 10^{-12}}) \\ &= 7.18 \times 10^8 \end{aligned}$$

So the optimal number of stages is 3.5, clock frequency is 2.45 GHz, throughput is 7.18×10^8.

Q8

1) $2 \rightarrow 4$ (RAW)

2) $3 \rightarrow 4$ (RAW)

3) $4 \rightarrow 5$ (RAW, WAR)

4) $7 \rightarrow 1$ (WAR)

Q9

1. With more registers, the length of each instruction also has to be extended to address them.

2. The register file will have greater latency as the register number grows large, so that datapath and clock cycle will be longer.

3. Registers are expensive and consume much energy, having more registers means an increase in both heat and cost.