

第1部分 概述 习题集及参考答案

1.2、计算机体系结构中的八个伟大思想与其他领域的思想相似。请将计算机体系结构中的八个伟大思想：

- 1) 面向摩尔定律的设计
- 2) 使用抽象简化设计
- 3) 加速经常性事件
- 4) 通过并行提高性能
- 5) 通过流水线提高性能
- 6) 通过预测提高性能
- 7) 存储器层次
- 8) 通过冗余提高可靠性

与其他领域的下列思想进行匹配：

- 1) 汽车制造中的组装生产线
- 2) 吊桥缆索
- 3) 采用风向信息的飞机和船舶导航系统
- 4) 高楼中的高速电梯
- 5) 图书馆的存阅处
- 6) 通过增大 CMOS 晶体管的栅极面积来减少翻转时间
- 7) 增加电磁飞机弹射器（不同于当前蒸汽驱动模型，它采用电驱动），这可能通过新型反应堆技术增加的电能来实现
- 8) 制造自动驾驶汽车，其控制系统是安装在汽车上的传感器系统，例如车道偏离检测系统和智能导航控制系统。

参考答案：

- 1.2、
- 1) 通过流水线提高性能
 - 2) 通过冗余提高可靠性
 - 3) 通过预测提高性能
 - 4) 加速经常性事件
 - 5) 存储器层次
 - 6) 通过并行提高性能
 - 7) 面向摩尔定律的设计
 - 8) 使用抽象简化设计

1.3、请描述高级语言（例如 C）编写的程序转化为能够直接在计算机处理器上执行的表示的具体步骤。

参考答案：

1.3、源程序—(编译器)—汇编语言—(汇编器)—机器语言

1.4、一个彩色显示器中的每个像素由三种基色（红、绿、蓝）构成，每种基色用 8 位表示，帧大小为 1280*1024。

- 1) 为了保存一帧图像最小需要多大的帧缓冲（以字节计算）？
- 2) 在 100Mbps 的网络上传输一帧图像最少需要多长时间？

参考答案：

- 1.4、
- 1) $1280 \times 1024 \text{ 像素} \rightarrow 1280 \times 1024 \times 3 \text{ 字节} = 3,932,160 \text{ 字节}$
 - 2) $(3932160 \times 8 \text{ 位}) / (100 \times 10^6) = 0.31 \text{ 秒}$

1.5、有 3 种不同的处理器 P1、P2 和 P3 执行同样的指令系统。P1 的时钟频率为 3GHz，CPI 为 1.5；P2 的时钟频率为 2.5GHz，CPI 为 1.0；P3 的时钟频率为 4GHz，CPI 为 2.2。

- 1) 以每秒执行的指令数为标准, 哪个处理器性能最高?
- 2) 如果每个处理器执行一个程序都花费都花费 10 秒时间, 求它们的时钟周期数和指令数。
- 3) 我们试图把执行时间减少 30%, 但会引起 CPI 增加 20%。请问为达到时间减少 30%的目标, 时钟频率应达到多少?

参考答案:

1.5、同样的指令系统, 则“指令数”相同。公式: $T = \text{指令数} \times \text{CPI} \times \text{时钟周期长度}$

处理器	CPI	主频(GHz)
P1	1.5	3
P2	1.0	2.5
P3	2.2	4

1) 每秒钟执行的指令数: $1/(\text{CPI} \times \text{时钟周期长度}) = \text{主频}/\text{CPI}$

处理器	CPI	主频(GHz)	每秒钟执行的指令数
P1	1.5	3	$3/1.5=2\text{G}$
P2	1.0	2.5	$2.5/1.0=2.5\text{G}$
P3	2.2	4	$4/2.2=1.8\text{G}$

2) 依据公式求指令数: $\text{指令数} = 10/(\text{CPI} \times \text{时钟周期长度}) = 10 \times \text{主频}/\text{CPI}$; $\text{时钟周期数} = \text{指令数} \times \text{CPI} = 10 \times \text{主频}$

处理器	CPI	主频(GHz)	时钟周期数	指令数
P1	1.5	3	$10 \times 3 = 30\text{G}$	$30/1.5 = 20\text{G}$
P2	1.0	2.5	$10 \times 2.5 = 25\text{G}$	$25/1.0 = 25\text{G}$
P3	2.2	4	$10 \times 4 = 40\text{G}$	$40/2.2 = 18.18\text{G}$

3) 用“主频”替代“时钟周期长度”的公式: $\text{主频} = (\text{指令数} \times \text{CPI}) / T$
 新主频 = 原主频 * (CPI 比例) / (T 比例) = 原主频 * (120%) / (1-30%) = 1.714 * 原主频
 $f(P1) = 1.714 \times 3 = 5.14\text{GHZ}$
 $f(P2) = 1.714 \times 2.5 = 4.285\text{GHZ}$
 $f(P3) = 1.714 \times 4 = 6.856\text{GHZ}$

1.6、同一个指令系统体系结构有两种不同的实现方式。根据 CPI 的不同将指令分成四类 (A、B、C 和 D)。P1 的时钟频率为 2.5GHz, CPI 分别为 1、2、3 和 3; P2 时钟频率为 3GHz, CPI 分别为 2、2、2 和 2。给定一个程序, 有 1.0×10^6 条动态指令, 按如下比例分为 4 类: A, 10%; B, 20%; C, 50%; D, 20%。

- 1) 每种实现方式下的整体 CPI 是多少?
- 2) 计算两种情况下的时钟周期总数。

参考答案:

1.6、两类实现的 CPI:

处理器	A 类指令	B 类指令	C 类指令	D 类指令	主频(GHz)
P1	1	2	3	3	2.5
P2	2	2	2	2	3

程序有 1M 条指令, 四类指令类型对应的比例:

A 类指令	B 类指令	C 类指令	D 类指令
10%	20%	50%	20%

程序的周期数 = 指令数 * CPI

P1 的周期数 = $(10\% \times 1 + 20\% \times 2 + 50\% \times 3 + 20\% \times 3) \times 1\text{M} = 2.6\text{M}$
 则 $\text{CPI}(P1) = 2.6\text{M}/1\text{M} = 2.6$

P2 的周期数 = $(10\% \times 2 + 20\% \times 2 + 50\% \times 2 + 20\% \times 2) \times 1\text{M} = 2\text{M}$
 则 $\text{CPI}(P2) = 2\text{M}/1\text{M} = 2$

1.7、编译器对应用程序性能有极深的影响。假定对于一个程序, 如果采用编译器 A, 则动态指令数为 1.0×10^9 , 执行时间为 1.1s; 如果采用编译器 B, 则动态指令数为 1.2×10^9 , 执行时间为 1.5s。

- 1) 在给定处理器时钟周期长度为 1ns 时, 求每个程序的平均 CPI。
- 2) 假定被编译的程序分别在两个不同的处理器上运行。如果这两个处理器的执行时间相同, 求运行编译器 A 生成之代码的处理器时钟比运行编译器 B 生成之代码的处理器时钟快多少。
- 3) 假设开发了一种新的编译器, 只需 6.0×10^8 条指令, 程序平均 CPI 为 1.1。求这种新的编译器在原处理器环境下相对于原编译器 A 和 B 的加速比。

参考答案:

1.7、依据 CPU 执行时间公式。

- 1) 编译器 A: $1.1 = 1 * CPI(A) * 1$, 则 $CPI(A) = 1.1$
编译器 B: $1.5 = 1.2 * CPI(B) * 1$, 则 $CPI(B) = 1.25$
- 2) 假定两个处理器的时钟长度分别为 $C(1)$ 和 $C(2)$ 。则
 $1G * 1.1 * C(1) = 1.2G * 1.25 * C(2) \rightarrow C(1)/C(2) = 1.364$,
运行 B 编译器的代码的处理器是运行 A 编译器代码的处理器速度快 1.364 倍。
- 3) $T(new) = 0.6 * 1.1 * 1 = 0.66$ 秒
相对编译器 A 的加速比: $T(A)/T(new) = 1.1/0.66 = 1.67$
相对编译器 B 的加速比: $T(B)/T(new) = 1.5/0.66 = 2.27$

1.12、在 1.10 节中提到使用性能公式的子集来作为性能评价指标的陷阱。下面的习题将对其进行说明。考虑下面两种处理器。P1 的时钟频率为 4GHz，平均 CPI 为 0.9，需要执行 $5.0 * 10^9$ 条指令；P2 的时钟频率为 3GHz，平均 CPI 为 0.75，需要执行 $1.0 * 10^9$ 条件指令。

- 1) 一个常见的错误是，认为时钟频率最高的计算机具有最高的性能。这种说法正确吗？请用 P1 和 P2 来验证这一说法是否正确。
- 2) 另一个错误是，认为执行指令最多的处理器需要更多的 CPU 时间。考虑 P1 执行 $1.0 * 10^9$ 条指令序列所需的时间，假定 P1 和 P2 的 CPI 不变，计算一下 P2 用同样的时间可以执行多少条指令。
- 3) 一个常见的错误是用 MIPS（每秒百万条指令数）来比较两台不同的处理器的性能，并认为 MIPS 数值大的处理器具有最高的性能。这种说法正确吗？请用 P1 和 P2 验证这一说法是否正确。
- 4) 另一个常见的性能标志是 MFLOPS（每秒百万条浮点指令），其定义为

$$MFLOPS = (\text{浮点操作的数目}) / (\text{执行时间} * 10^6)$$

但此指标与 MIPS 有同样的问题。假定 P1 和 P2 上执行的指令有 40% 的浮点指令，求出各处理器的 MFLOPS 的数值。

参考答案：

$$1.12、T(P1) = 5G * 0.9 / 4G = 1.125 \text{ 秒}$$

$$T(P2) = 1G * 0.75 / 3G = 0.25 \text{ 秒}$$

结论：P2 性能是 P1 性能的 2.22 倍。

- 1) 不正确。f(P1) > f(P2)，但 P1 的执行时间较 P2 要长。
- 2) 执行 P1 的 1G 条指令的时间是：1G * 0.9 / 4G = 0.225 秒。若在此段时间，P2 可执行的指令数是 n，则有 $0.225 = n * 0.75 / 3G$ ， $n = 0.9G$ 。结论：P1 较 P2 执行较多条指令，但用时相同。
- 3) 主频/CPI：单位时间内执行指令的数量。
 $MIPS(P1) = 4 * 10^3 / 0.9 = 4.44 * 10^3$
 $MIPS(P2) = 3 * 10^3 / 0.75 = 4.0 * 10^3$
结论：MIPS(P1) > MIPS(P2)，与实际性能相背。
- 4) $MFLOPS(P1) = MIPS(P1) * 40\% = 1.78 * 10^3$
 $MFLOPS(P2) = MIPS(P2) * 40\% = 1.60 * 10^3$
结论：MFLOPS(P1) > MFLOPS(P2)，与实际性能相背。

1.13、在 1.10 节中提到的另一个易犯的错误是希望通过只改进计算机的一个方面来改进计算机的总体性能。假如一台计算机上运行一个程序需要 250 秒，其中 70 秒用于执行浮点指令，85 秒用于执行 L/S 指令，40 秒用于执行分支指令，其它时间用于执行整型操作指令。

- 1) 如果浮点操作的时间减少 20%，总时间将减少多少？
- 2) 如果将总时间减少 20%，整型操作时间应减少多少？
- 3) 如果只减少分支指令时间，总时间能否减少 20%？

参考答案：

1.13、Amdahl 公式：总时间 = (改进部分的时间 / 改进的比例) + 未改进部分的时间

当前，总时间(250) = 浮点(70) + L/S(85) + 分支(40) + 其它(55)

- 1) 浮点操作时间减少 20%，则加速比为 $100/80 = 1.25$ ，则
 $T(\text{浮点}) = (70/1.25) + 85 + 40 + 55 = 236$ 秒，改进后减少：250 - 236 = 14 秒，改进 14/250 = 5.6%
- 2) 假定整型操作时间为 T，则
 $250 * 80\% = 70 + 85 + 40 + T \rightarrow T = 5$ 。则，整型操作的时间需减少：55 - 5 = 50 秒。
- 3) 总时间减少 20%，则新的总执行时间为 200 秒。分支指令时间为 40 秒，其它指令时间为 210 秒。所以，仅减少分支指令时间，总时间极限也需要 210 秒，不可能。

1.14、假定一个程序需要执行 50×10^6 条浮点指令、 110×10^6 条整型指令、 80×10^6 条 L/S 指令和 10×10^6 条分支指令。每种类型指令的 CPI 分别是 1、1、4 和 2。假定处理器的时钟频率为 2GHz。

- 1) 如果我们要将程序运行速度提高到原来的 2 倍，浮点指令的 CPI 应如何改进？
- 2) 如果我们要将程序运行速度提高到原来的 2 倍，L/S 指令的 CPI 应如何改进？
- 3) 如果整数和浮点指令的 CPI 减少 40%，L/S 和分支指令的 CPI 减少 30%，程序的执行时间能改进多少？

参考答案：

1.14、

	浮点	整型	L/S	分支
指令数(M)	50	110	80	16
CPI	1	1	4	2

$$T_{old} = (50 \times 1 + 110 \times 1 + 80 \times 4 + 16 \times 2) \times 10^6 / (2 \times 10^9) = 512 \times 10^6 / (2 \times 10^9)$$

- 1) 假定改进“浮点”的 CPI 为 C_1

$T_{old}/T_{new} = 2$ 。则有

$$2 = 512 / (50 \times C_1 + 110 \times 1 + 80 \times 4 + 16 \times 2)$$

$$= 512 / (50 \times C_1 + 462)$$

则， $C_1 = (256 - 462) / 50 < 0$ ，不可能实现！

- 2) 假定改进“L/S”的 CPI 为 C_2

$T_{old}/T_{new} = 2$ 。则有

$$2 = 512 / (50 \times 1 + 110 \times 1 + 80 \times C_2 + 16 \times 2)$$

$$= 512 / (80 \times C_2 + 192)$$

则， $C_2 = (256 - 192) / 80 = 0.8$ 。

- 3) $T_{new} = (50 \times 1 \times 60\% + 110 \times 1 \times 60\% + 80 \times 4 \times 70\% + 16 \times 2 \times 70\%) \times 10^6 / (2 \times 10^9) = 342.4 \times 10^6 / (2 \times 10^9)$

性能加速比： $512 / 342.4 = 1.495$ 。

$T_{old} = 0.256$ 秒

$T_{new} = 0.171$ 秒