

1.14.

1.8 节引证了一个用性能公式的一个子集去计算性能的陷阱。为了说明它，下表是在不同的处理器中执行 10^6 条指令序列的有关数据。

处理器	时钟频率	CPI
P_1	4GHz	1.25
P_2	3GHz	0.75

1.14.1 [5] <1.8> 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 P_1 和 P_2 来验证这一说法。

1.14.2 [10] <1.8> 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 P_1 执行 10^6 条指令序列所需的时间， P_1 和 P_2 的 CPI 不变，计算一下 P_2 用同样的时间可以执行多少条指令？

1.14.3 [10] <1.8> 一个常见的错误是用 MIPS（每秒百万条指令）来比较 2 台不同的处理器的性能，并认为 MIPS 最大的处理器具有最高的性能。这种说法正确吗？请用 P_1 和 P_2 验证这一说法。

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 = 浮点操作的个数 / (执行时间 $\times 10^6$)。它与 MIPS 有同样的问题。

考虑下表所示的程序，在时钟频率为 3GHz 的处理器上运行。

	指令数	读/写	浮点	分支	CPI (读/写)	CPI (浮点)	CPI (分支)
程序 a	10^6	50%	40%	10%	0.75	1	1.5
程序 b	3×10^6	40%	40%	20%	1.25	0.70	1.25

1.14.4 [10] <1.8> 求程序的 MFLOPS 值。

1.14.5 [10] <1.8> 求程序的 MIPS 值。

1.14.6 [10] <1.8> 求程序的性能，并与 MFLOPS 和 MIPS 值作比较。

解：(1) 这种说法不正确。计算机性能还与 CPI 有关。下面计算这两个 CPU 处理 10^6 条指令所需要的时间：

$$T_{P1} = \frac{N_{P1} * CPI_{P1}}{f_{P1}} = \frac{10^6 * 1.25}{4 * 10^9} = 3.125 * 10^{-4} s$$
$$T_{P2} = \frac{N_{P2} * CPI_{P2}}{f_{P2}} = \frac{10^6 * 0.75}{3 * 10^9} = 2.5 * 10^{-4} s$$

可见，虽然 P_1 的主频高于 P_2 ，但由于其 CPI 较大，处理相同数量的指令耗时却更多。

(2) 上面已经算过， $T_{P1} = 3.125 * 10^{-4} s$ 。在相同的时间内， P_2 可执行的指令数为：

$$N_{P2} = \frac{T_{P2} * f_{P2}}{CPI_{P2}} = \frac{T_{P1} * f_{P2}}{CPI_{P2}} = \frac{3.125 * 10^{-4} * 3 * 10^9}{0.75} = 1.25 * 10^6 \text{ (条)}$$

(3) 分别计算两个 CPU 的 MIPS：（单位：百万条指令/秒）

$$MIPS_{P1} = \frac{f_{P1}}{10^6 * CPI_{P1}} = \frac{4 * 10^9}{10^6 * 1.25} = 3200$$
$$MIPS_{P2} = \frac{f_{P2}}{10^6 * CPI_{P2}} = \frac{3 * 10^9}{10^6 * 0.75} = 4000$$

MIPS 大的 CPU 不一定有更好的性能。但在本例中，确实是 P_2 CPU 性能更好，同时 MIPS 高。

(4) 先计算程序 a 的 MFLOPS 值：

$$MFLOPS_a = \frac{10^6 * 40\%}{\frac{10^6 * 40\% * 1}{3 * 10^9} * 10^6} = 3000$$

再计算程序 b 的 MFLOPS 值：

$$MFLOPS_b = \frac{3 * 10^6 * 40\%}{\frac{3 * 10^6 * 40\% * 0.7}{3 * 10^9} * 10^6} \approx 4286$$

(5) 先计算程序 a 的 MIPS 值，为此先算程序运行时间：

$$T_a = \frac{\sum N_i * CPI_i}{f} = \frac{10^6 (50\% * 0.75 + 40\% * 1 + 10\% * 1.5)}{3 * 10^9} = 3.083 * 10^{-4}$$

所以该程序的 MIPS 为：

$$MIPS_a = \frac{N_a}{T_a * 10^6} = \frac{10^6}{3.083 * 10^{-4} * 10^6} = 3243$$

再计算程序 b 的 MIPS 值，为此先算程序运行时间：

$$T_b = \frac{3 * 10^6 (40\% * 1.25 + 40\% * 0.7 + 20\% * 1.25)}{3 * 10^9} = 1.03 * 10^{-3}$$

所以该程序的 MIPS 为：

$$MIPS_b = \frac{N_b}{T_b * 10^6} = \frac{3 * 10^6}{1.03 * 10^{-3} * 10^6} = 2913$$

(6) 运行程序 a 的性能 = $\frac{1}{T_a} = \frac{1}{3.083 * 10^{-4}} = 3243$

运行程序 b 的性能 = $\frac{1}{T_b} = \frac{1}{1.03 * 10^{-3}} = 971$

可见程序 b 在运行时，MFLOPS 比 a 程序高，MIPS 却没有 a 程序好，总的性能也不如 a。所以各个指标需要综合起来看，才能科学的评价计算机的性能。

1.16

还有一个易犯的错误是有关在多处理器系统中运行， 希望只改进一部分例行程序来改进整体性能。

下表表示某个程序的 5 个例程在不同数目处理器中的执行时间。

	处理器数	例程 A(ms)	例程 B(ms)	例程 C(ms)	例程 D(ms)	例程 E(ms)
a	2	20	80	10	70	5
b	16	4	14	2	12	2

1. 16. 1 [10] <1.8> 求总的执行时间。如果例程 A, C 和 E 的时间改进 15%，总的执行时间能减少多少？

1. 16. 2 [10] <1.8>如果例程 B 的时间改进 10%，总的执行时间能减少多少？

1. 16. 3 [10] <1.8>如果例程 D 的时间改进 10%，总的执行时间能减少多少？

多处理器系统中的执行时间可分成 例程计算时间 加 处理器之间的通信时间。下表给出了例程计算时间和通信时间。在这种情况下，通信时间是总时间的重要组成部分。（时间单位：ms）

处理器数	例程 A	例程 B	例程 C	例程 D	例程 E	通信
2	20	78	9	65	4	11
4	12	44	4	34	2	13
8	1	23	3	19	3	17
16	4	13	1	10	2	22
32	2	5	1	5	1	23
64	1	3	0.5	1	1	26

1. 16. 4 [10] <1.8> 每当处理器数最加倍时，求新的计算时间与旧的计算时间之比和新与旧的通信时间比。

1. 16. 5 [5] <1.8> 用比值的几何平均值，推算在 128 台处理器的系统中的计算时间和通信时间。

1. 16. 6 [10] <1.8>求在 1 个处理器系统中的计算时间和通信时间。

解：(1) 处理器 a: $t_a = \sum \text{各例程执行时间} = 20 + 80 + 10 + 70 + 5 = 185\ ms$

处理器 b: $t_b = \sum \text{各例程执行时间} = 4 + 14 + 2 + 12 + 2 = 34\ ms$

经过例程 A, C 和 E 的时间改进 15%之后，缩短的运行时间为：

$\Delta t_a = 15\% * (20 + 10 + 5) = 5.25\ ms ; \quad \frac{\Delta t_a}{t_a} = \frac{5.25\ ms}{185\ ms} = 2.84\%$

$\Delta t_b = 15\% * (4 + 2 + 2) = 1.2\ ms ; \quad \frac{\Delta t_b}{t_b} = \frac{1.2\ ms}{34\ ms} = 3.53\%$

(2) 对程序的改动仅有把 B 的执行时间缩短 10%。

$\Delta t_a = 10\% * 8 = 8\ ms ; \quad \frac{\Delta t_a}{t_a} = \frac{8\ ms}{185\ ms} = 4.32\%$

$\Delta t_b = 10\% * 14 = 1.4\ ms ; \quad \frac{\Delta t_b}{t_b} = \frac{1.4\ ms}{34\ ms} = 4.11\%$

(3) 对程序的改动仅有把 D 的执行时间缩短 10%。

$\Delta t_a = 10\% * 70 = 7\ ms ; \quad \frac{\Delta t_a}{t_a} = \frac{7\ ms}{185\ ms} = 3.78\%$

$\Delta t_b = 10\% * 12 = 1.2\ ms ; \quad \frac{\Delta t_b}{t_b} = \frac{1.2\ ms}{34\ ms} = 3.53\%$

(4) 将所需数据与所求数据列为下表：（单位：ms）

处理器数	计算总时	通信用时	计算用时比 λ	通信用时比 μ
2	176	11	/	/
4	96	13	0.5455	1.1818
8	49	17	0.5104	1.3077
16	30	22	0.6122	1.2941
32	14	23	0.4667	1.0455
64	6.5	26	0.4643	1.1304

(5) 以上得到五个比值，即为有五组数据，对其进行几何平均：（各数据权值相等，为 1.）

$\overline{\lambda} = \sqrt[5]{\lambda_1\lambda_2\lambda_3\lambda_4\lambda_5} = \sqrt[5]{0.5455 * 0.5104 * 0.6122 * 0.4667 * 0.4643} = 0.5170$

$\overline{\mu} = \sqrt[5]{\mu_1\mu_2\mu_3\mu_4\mu_5} = \sqrt[5]{1.1818 * 1.3077 * 1.2941 * 1.0455 * 1.1304} = 1.1877$

估计对于 128 个处理器的情况：

$t_{\text{计}128} = t_{\text{计}64} * \overline{\lambda} = 6.5 * 0.5170 = 3.36\ ms ; \quad t_{\text{通}128} = t_{\text{通}64} * \overline{\mu} = 26 * 1.1877 = 30.88\ ms ;$

(6) 对于单处理器的情况，同上估计：而单核处理器不需要通信，故通信用时为 0.

$t_{\text{计}1} = \frac{t_{\text{计}2}}{\overline{\lambda}} = \frac{176}{0.5170} = 340.43\ ms ;$

$t_{\text{通}1} = 0$