1

Solutions

1.1 **Personal computer:** Computer that emphasizes delivery of good performance to a single user at low cost and usually executes third-party software.

Server: Computer used for large workloads and usually accessed via a network.

Embedded computer: Computer designed to run one application or one set of related applications and integrated into a single system.

1.2

a. Performance via Pipelining

b. Dependability via Redundancy

c. Performance via Prediction

d. Make the Common Case Fast

e. Hierarchy of Memories

f. Performance via Parallelism

g. Use Abstraction to Simplify Design

1.3 The program is compiled into an assembly language program, which is itself assembled into a machine-language program.

1.4

a. 1280×1024 pixels $= 1,310,720$ pixels $=> 1,310,720 \times 3 = 3,932,160$ bytes/frame.

b. $3,932,160$ bytes $\times$ (8 bits/byte) /100E6 bits/second $= 0.31$ seconds

1.5

Desktop Processor	Year	Tech	Max. Clock Speed (GHz)	Integer IPC/core	Cores	Max. DRAM Bandwidth (GB/s)	SP Floating Point (Gflop/s)	L3 cache (MiB)
Westmere i7-620	2100	32	3.33	4	2	17.1	107	4
Ivy Bridge i7-3770K	2013	22	3.90	6	4	25.6	250	8
Broadwell i7-6700K	2015	14	4.20	8	4	34.1	269	8
Kaby Lake i7-7700K	2017	14	4.50	8	4	38.4	288	8
Coffee Lake i7-9700K	2019	14	4.90	8	8	42.7	627	12
Imp./year		20%	4%	7%	15%	10%	19%	12%
Doubles every		4 years	18 years	10 years	5 years	7 years	4 years	6 years

1.6

a. performance of P1 (instructions/sec) $= 3 \times 10^9/1.5 = 2 \times 10^9$

performance of P2 (instructions/sec) $= 2.5 \times 10^9/1.0 = 2.5 \times 10^9$

performance of P3 (instructions/sec) $= 4 \times 10^9/2.2 = 1.8 \times 10^9$

b. cycles(P1) $= 10 \times 3 \times 10^9 = 30 \times 10^9$ s

cycles(P2) $= 10 \times 2.5 \times 10^9 = 25 \times 10^9$ s

cycles(P3) $= 10 \times 4 \times 10^9 = 40 \times 10^9$ s

c. No. instructions(P1) $= 30 \times 10^9/1.5 = 20 \times 10^9$

No. instructions(P2) $= 25 \times 10^9/1 = 25 \times 10^9$

No. instructions(P3) $= 40 \times 10^9/2.2 = 18.18 \times 10^9$

$CPI_{new} = CPI_{old} \times 1.2$, then CPI(P1) $= 1.8$, CPI(P2) $= 1.2$, CPI(P3) $= 2.6$

$f = $ No. instr. $\times$ CPI/time, then

$f(P1) = 20 \times 10^9 \times 1.8/7 = 5.14$ GHz

$f(P2) = 25 \times 10^9 \times 1.2/7 = 4.28$ GHz

$f(P3) = 18.18 \times 10^9 \times 2.6/7 = 6.75$ GHz

1.7

a. Class A: 10^5 instr. Class B: 2×10^5 instr. Class C: 5×10^5 instr. Class D: 2×10^5 instr.

Time $=$ No. instr. $\times$ CPI/clock rate

Total time P1 $= (10^5 + 2 \times 10^5 \times 2 + 5 \times 10^5 \times 3 + 2 \times 10^5 \times 3)/(2.5 \times 10^9) = 10.4 \times 10^{-4}$ s

Total time P2 $= (10^5 \times 2 + 2 \times 10^5 \times 2 + 5 \times 10^5 \times 2 + 2 \times 10^5 \times 2)/(3 \times 10^9) = 6.66 \times 10^{-4}$ s

CPI(P1) $= 10.4 \times 10^{-4} \times 2.5 \times 10^9/10^6 = 2.6$

CPI(P2) $= 6.66 \times 10^{-4} \times 3 \times 10^9/10^6 = 2.0$

b. clock cycles(P1) $= 10^5 \times 1 + 2 \times 10^5 \times 2 + 5 \times 10^5 \times 3 + 2 \times 10^5 \times 3 = 26 \times 10^5$

clock cycles(P2) $= 10^5 \times 2 + 2 \times 10^5 \times 2 + 5 \times 10^5 \times 2 + 2 \times 10^5 \times 2 = 20 \times 10^5$

1.8

a. $CPI = T_{exec} \times f/No.\ instr.$

Compiler A CPI = 1.1

Compiler B CPI = 1.25

b. $f_B/f_A = (No.\ instr.(B) \times CPI(B))/(No.\ instr.(A) \times CPI(A)) = 1.37$

c. $T_A/T_{new} = 1.67$

$T_B/T_{new} = 2.27$

1.9

1.9.1 $C = 2 \times DP/(V^2 \times F)$

Pentium 4: $C = 3.2E{-}8F$

Core i5 Ivy Bridge: $C = 2.9E{-}8F$

1.9.2 Pentium 4: $10/100 = 10\%$

Core i5 Ivy Bridge: $30/70 = 42.9\%$

1.9.3 $(S_{new} + D_{new})/(S_{old} + D_{old}) = 0.90$

$D_{new} = C \times V_{new}2 \times F$

$S_{old} = V_{old} \times I$

$S_{new} = V_{new} \times I$

Therefore:

$V_{new} = [D_{new}/(C \times F)]1/2$

$D_{new} = 0.90 \times (S_{old} + D_{old}) - S_{new}$

$S_{new} = V_{new} \times (S_{old}/V_{old})$

Pentium 4:

$S_{new} = V_{new} \times (10/1.25) = V_{new} \times 8$

$D_{new} = 0.90 \times 100 - V_{new} \times 8 = 90 - V_{new} \times 8$

$V_{new} = [(90 - V_{new} \times 8)/(3.2E8 \times 3.6E9)]^{1/2}$

$V_{new} = 0.85\ V$

Core i5:

$S_{new} = V_{new} \times (30/0.9) = V_{new} \times 33.3$

$D_{new} = 0.90 \times 70 - V_{new} \times 33.3 = 63 - V_{new} \times 33.3$

$V_{new} = [(63 - V_{new} \times 33.3)/(2.9E8 \times 3.4E9)]^{1/2}$

$V_{new} = 0.64\ V$

p	# arith inst.	# L/S inst.	# branch inst.	cycles	ex. time	speedup
1	2.56E9	1.28E9	2.56E8	1.92E10	9.60	1.00
2	1.83E9	9.14E8	2.56E8	1.41E10	7.04	1.36
4	9.14E8	4.57E8	2.56E8	7.68E9	3.84	2.50
8	4.57E8	2.29E8	2.56E8	4.48E9	2.24	4.29

1.10

1.10.1

p	ex. time
1	41.0
2	29.3
4	14.6
8	7.33

1.10.2

1.10.3   3

1.11

1.11.1 die area$_{15cm}$ = wafer area/dies per wafer = $\pi \times 7.5^2/84 = 2.10\,cm^2$

yield$_{15cm}$ = $1/(1 + (0.020 \times 2.10/2))^2 = 0.9593$

die area$_{20cm}$ = wafer area/dies per wafer = $\pi \times 10^2/100 = 3.14\,cm^2$

yield$_{20cm}$ = $1/(1 + (0.031 \times 3.14/2))^2 = 0.9093$

1.11.2 cost/die$_{15cm}$ = $12/(84 \times 0.9593) = 0.1489$

cost/die$_{20cm}$ = $15/(100 \times 0.9093) = 0.1650$

1.11.3 die area$_{15cm}$ = wafer area/dies per wafer = $\pi \times 7.5^2/(84 \times 1.1) = 1.91\,cm^2$

yield$_{15cm}$ = $1/(1 + (0.020 \times 1.15 \times 1.91/2))^2 = 0.9575$

die area$_{20cm}$ = wafer area/dies per wafer = $\pi \times 10^2/(100 \times 1.1) = 2.86\,cm^2$

yield$_{20cm}$ = $1/(1 + (0.03 \times 1.15 \times 2.86/2))^2 = 0.9082$

1.11.4 defects per area$_{0.92}$ = $(1-y^{.5})/(y^{.5} \times die\_area/2) = (1 - 0.92^{.5})/(0.92^{.5} \times 2/2) = 0.043\,defects/cm^2$

defects per area$_{0.95}$ = $(1-y^{.5})/(y^{.5} \times$ die_area/2) = $(1 - 0.95^{.5})/$
$(0.95^{.5} \times 2/2) = 0.026$ defects/cm^2

1.12

1.12.1 CPI = clock rate $\times$ CPU time/instr. count

clock rate = 1/cycle time = 3 GHz

CPI(bzip2) = $3 \times 10^9 \times 750/(2389 \times 10^9) = 0.94$

1.12.2 SPEC ratio = ref. time/execution time

SPEC ratio(bzip2) = 9650/750 = 12.86

1.12.3 CPU time = No. instr. $\times$ CPI/clock rate

If CPI and clock rate do not change, the CPU time increase is equal to the increase in the number of instructions, that is 10%.

1.12.4 CPU time(before) = No. instr. $\times$ CPI/clock rate

CPU time(after) = $1.1 \times$ No. instr. $\times 1.05 \times$ CPI/clock rate

CPU time(after)/CPU time(before) = $1.1 \times 1.05 = 1.155$. Thus, CPU time is increased by 15.5%.

1.12.5 SPECratio = reference time/CPU time

SPECratio(after)/SPECratio(before) = CPU time(before)/CPU time(after) = 1/1.1555 = 0.86. The SPECratio is decreased by 14%.

1.12.6 CPI = (CPU time $\times$ clock rate)/No. instr.

CPI = $700 \times 4 \times 10^9/(0.85 \times 2389 \times 10^9) = 1.37$

1.12.7 Clock rate ratio = 4 GHz/3 GHz = 1.33

CPI @ 4 GHz = 1.37, CPI @ 3 GHz = 0.94, ratio = 1.45

They are different because, although the number of instructions has been reduced by 15%, the CPU time has been reduced by a lower percentage.

1.12.8 700/750 = 0.933. CPU time reduction: 6.7%

1.12.9 No. instr. = CPU time $\times$ clock rate/CPI

No. instr. = $960 \times 0.9 \times 4 \times 10^9/1.61 = 2146 \times 10^9$

1.12.10 Clock rate = No. instr. $\times$ CPI/CPU time.

Clock rate$_{new}$ = No. instr. $\times$ CPI/0.9 $\times$ CPU time = 1/0.9 clock rate$_{old}$ = 4.44 GHz

1.12.11 Clock rate = No. instr. $\times$ CPI/CPU time.

Clock rate$_{new}$ = No. instr. $\times$ 0.85 $\times$ CPI/0.80 CPU time = 0.85/0.80, clock rate$_{old}$ = 3.18 GHz

1.13

1.13.1 $T(P1) = 5 \times 10^9 \times 0.9/(4 \times 10^9) = 1.125$ s

$T(P2) = 10^9 \times 0.75/(3 \times 10^9) = 0.25$ s

clock rate(P1) > clock rate(P2), performance(P1) < performance(P2)

1.13.2 $T(P1) =$ No. instr. $\times$ CPI/clock rate

$T(P1) = 2.25\ 3\ 1021$ s

$T(P2)\ 5\ N \times 0.75/(3 \times 10^9)$, then $N = 9 \times 10^8$

1.13.3 MIPS = Clock rate $\times 10^{-6}$/CPI

MIPS(P1) = $4 \times 10^9 \times 10^{-6}/0.9 = 4.44 \times 10^3$

MIPS(P2) = $3 \times 10^9 \times 10^{-6}/0.75 = 4.0 \times 10^3$

MIPS(P1) > MIPS(P2), performance(P1) < performance(P2) (from 11a)

1.13.4 MFLOPS = No. FP operations $\times 10^{-6}$/T

MFLOPS(P1) = $.4 \times 5E9 \times 1E\text{-}6/1.125 = 1.78E3$

MFLOPS(P2) = $.4 \times 1E9 \times 1E\text{-}6/.25 = 1.60E3$

MFLOPS(P1) > MFLOPS(P2), performance(P1) < performance(P2) (from 11a)

1.14

1.14.1 $T_{fp} = 70 \times 0.8 = 56$ s. $T_{new} = 56 + 85 + 55 + 40 = 236$ s. Reduction: 5.6%

1.14.2 $T_{new} = 250 \times 0.8 = 200$ s, $T_{fp} + T_{l/s} + T_{branch} = 165$ s, $T_{int} = 35$ s. Reduction time INT: 58.8%

1.14.3 $T_{new} = 250 \times 0.8 = 200$ s, $T_{fp} + T_{int} + T_{l/s} = 210$ s. NO

1.15

1.15.1 Clock cycles = $CPI_{fp} \times$ No. FP instr. + $CPI_{int} \times$ No. INT instr. + $CPI_{l/s} \times$ No. L/S instr. + $CPI_{branch} \times$ No. branch instr.

T_{CPU} = clock cycles/clock rate = clock cycles/2 $\times 10^9$

clock cycles = 512×10^6; $T_{CPU} = 0.256$ s

To have the number of clock cycles by improving the CPI of FP instructions:

$CPI_{improved\ fp} \times$ No. FP instr. + $CPI_{int} \times$ No. INT instr. + $CPI_{l/s} \times$ No. L/S instr. + $CPI_{branch} \times$ No. branch instr. = clock cycles/2

$CPI_{improved\,fp} = ($clock cycles$/2 - (CPI_{int} \times$ No. INT instr. $+ CPI_{l/s} \times$ No. L/S instr. $+ CPI_{branch} \times$ No. branch instr.$)) /$ No. FP instr.

$CPI_{improved\,fp} = (256 - 462)/50 < 0 ==>$ not possible

1.15.2 Using the clock cycle data from a.

To have the number of clock cycles improving the CPI of L/S instructions:

$CPI_{fp} \times$ No. FP instr. $+ CPI_{int} \times$ No. INT instr. $+ CPI_{improved\,l/s} \times$ No. L/S instr. $+ CPI_{branch} \times$ No. branch instr. $=$ clock cycles$/2$

$CPI_{improved\,l/s} = ($clock cycles$/2 - (CPI_{fp} \times$ No. FP instr. $+ CPI_{int} \times$ No. INT instr. $+ CPI_{branch} \times$ No. branch instr.$)) /$ No. L/S instr.

$CPI_{improved\,l/s} = (256 - 198)/80 = 0.725$

1.15.3 Clock cycles $= CPI_{fp} \times$ No. FP instr. $+ CPI_{int} \times$ No. INT instr. $+ CPI_{l/s} \times$ No. L/S instr. $+ CPI_{branch} \times$ No. branch instr.

$T_{CPU} =$ clock cycles/clock rate $=$ clock cycles$/2 \times 10^9$

$CPI_{int} = 0.6 \times 1 = 0.6; CPI_{fp} = 0.6 \times 1 = 0.6; CPI_{l/s} = 0.7 \times 4 = 2.8; CPI_{branch} = 0.7 \times 2 = 1.4$

T_{CPU} (before improv.) $= 0.256\,s; T_{CPU}$ (after improv.) $= 0.171\,s$

1.16

processors	exec. time/ processor	time w/overhead	speedup	actual speedup/ideal speedup
1	100			
2	50	54	100/54 = 1.85	1.85/2 = .93
4	25	29	100/29 = 3.44	3.44/4 = 0.86
8	12.5	16.5	100/16.5 = 6.06	6.06/8 = 0.75
16	6.25	10.25	100/10.25 = 9.76	9.76/16 = 0.61