SUHAS JAIN

19CS30048

classmate

Date

Page

Q1) When the carry out is 1 and is discarded the unsigned value of the sum becomes less than both the original numbers.

Let original numbers be a and b.

$$a, b < 2^{32}$$

$$S = a+b-2^{32}$$

↓

discarding carry.

$$a+b-2^{32} < a+b-b$$

$$S < a$$

Similarly

$$S < b$$

MIPS CODE:

```
addu  $t3, $t1, $t2
sltu  $t3, $t3, $t1
```

If $t3 stores 1 then there is overflow otherwise there is no overflow.

SUHAS JAIN
19CS30048

Q2) $\Rightarrow$ $34A+B = 32A + 2A + B$

$32A$: left shift A 5 times

$2A$: left shift A 1 time

$\Rightarrow$ To do $\frac{x}{7}$

$$z = \frac{s + \frac{s + \frac{s + \cdots}{8}}{8}}{8}$$

[division by 8 $\Rightarrow$ right shift 3 times]

$$z = \frac{s+z}{8}$$

$$s = 7z$$

$$z = \frac{s}{7}$$

$34A = 01100000 + 00000110$
$= 01100110$

$34A+B = 01100110 + 00011001$
$= 01111111$

$\Rightarrow$ After doing 3 right shifts atmost 3 times all the 1's will be exhausted.

So.

$$z = \frac{s + \frac{s + \frac{s}{8}}{8}}{8} \Rightarrow z = \left\lfloor \frac{s}{7} \right\rfloor$$

$\Rightarrow$ Doing this for 01111111

$S = 01111111$

$\frac{S}{8} = 00001111$ (3 right shift)

$S + \frac{S}{8} = 10001110$

$S + \frac{S + \frac{S}{8}}{8} = 01111111 + 00010001$ (3 right shift)

$= 10010000$ right shift

$\frac{S + \frac{S + \frac{S}{8}}{8}}{8} = \boxed{00010010}$ Ans

$\rightarrow$ $34A + B = 127$

$\left\lfloor \frac{127}{7} \right\rfloor = 18$

$00010010 = 18$

Hence confirmed.

SUHAS JAIN
19CS30048

Total no. of operations.

(i) $S = 32A + 2A + B$

5 shifts + 1 shift + 2 additions.

= 8 operations.

(ii)
$$z = \frac{s + \frac{s + s/8}{8}}{8}$$

= 3x 3 + 2 additions
↓
3 right shift

= 11 operations

Total = 19 operations.

SUHAS JAIN

19CS30048

Q3) (a) M = 10101 01 01111 0011 01111 01110 111 0110

N = 0101 01 0000 1101 11011001 0111 0111 011

For M = 01 pairs = 8

10 pairs = 9

For N = 01 pairs = 9

10 pairs = 9

We will choose M as our multiplicand.

No. of additions = 8 (01 pairs)

No. of subtractions = 9 (10 pairs)

Bit pair tests. = 32 (on every step)

Shift operations = 32 (on every step)

(b) (i) CCT = max of all delays.

= 0.25ns.

Total operations = 32+32+9+8 = 81

Total time = 81 X 0.25ns.

= $\boxed{20.25\text{ ns}}$

(ii) Now CCT = 0.05ns (GCD of all)

Total Time = 0.05(17X5 + 32 + 32X2)

0.25 = 0.05×5   0.10 = 0.05×2.

= $\boxed{11.5\text{ns}}$

SUMAS JAIN
19CS30048

Q4) A = 0 11111110 1111 0000 0000 0000 0000 001
B = 0 11111011 1100 0000 0000 0000 0001 001

Biased FP exponent of A = 254 - 127 = 127
Biased FP exponent of B = 251 - 127 = 124.

a) We shift right the number with smallest exponent
So, B will shift right 3 times.
Guard bit of B = 0
Round bit of B = 0
Sticky bit of B = 1

b) After shifting is performed

Mantissa of A = 1111 0000 0000 0000 0000 001
Mantissa of B = 0011 1000 0000 0000 0000 001

A + B = 11.0010(01 000 000 000 000 000 010 × 2^{127}

Renormalising we get.
Exponent = 128 + 127 = 255.

A + B = 0 1111 1111 10010100 0000 0000 0000 001

5) Exponent $\Rightarrow$ $t_1
Mantiss $\Rightarrow$ $t_2

- $t_1 is biased so subract 127

- $t_3 = $t_1 - 127 , $t_3 must be positi (≥0) else number cannot be represented.

- for $(1.M) \times 2^{exp}$ for this if we right shift exp times, then we should only have 0s after the decimal.

- So we left shift $t_2 9+ $t_3 times, if remaining all numbers are 0 then we can ~~represent~~ represent otherwise not.

9 → because of padding.

SUMAS JAIN

19CS30048

Q6) a) $(0.0011 0111)_2$

Normalised form $= 1.10111 \times 2^{-3}$

Biased exponent $= -3+7 = 4 = 0100$

Positive so sign bit $= 0$

1.10111

↳ Even rounding to 3 bits of the decimal

$= 1.110$

~~Mat~~ Mantissa $= 110$

S	E	M
0	0100	110

⟹ Ans.

8-bit representation

b) $(16.0)_{10}$

$\Rightarrow (10000.0)_2 = (1.000 \times 2^{4})$

Normalised form.

Biased Exponent $= 4+7 = 11$

$= 1011$

Sign bit $= 0$. (positive number)

S	E	M
0	1011	000

Ans

8-bit representation of $(16.0)_{10}$

Q7) a) Maximum difference is

$= 2^{-23} \times [\text{Max exponent}]$
$= 2^{-23} \times 2^{127}$
$= 2^{104}$

b) B is max underflow no.

$B = 1 \underbrace{0000 \quad 0000}_{8 \text{ zeros}} \underbrace{1111 \ 1111 \ 1111111111 \ 1111}_{23 \text{ ones}}$

c)
⇒ On multiplication two normalized significands (≥ 1) we always get significand ≥ 1 so only right. shift

⇒ But in addition we may get significand ≤ 1 so there can be left shift also.

8) a) Total delay for 32 rows = 32 ns.

Delay for 1 8bit RCA = 8 ns.

Total delay = 32 + 8 = 40 ns.

b) For first 32 rows = 32 · x 32

= 256 .

Last row = 7 x 8 = 56 ns.

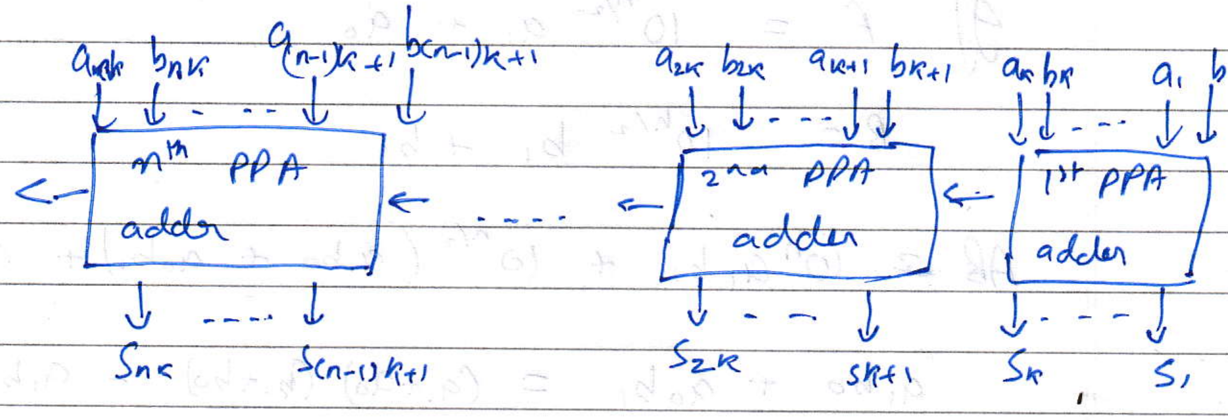
Total FAs used. = 256 + 56

= $\boxed{312}$

SUHAS JAIN
19CS30048

Q a)

a)



For each PPA adder

delay $= O(\log k)$

cost $= O(k \log k)$

Since they are connected in series in a RCA fashion, these will add up over n units.

$$\boxed{\begin{array}{l}\text{Total delay} = O(n \log k) \\ \text{Total cost} = O(nk \log k)\end{array}}$$

b) $A = 57 = 5 \times 10^1 + 7$

$B = 58 = 5 \times 10^1 + 8$

Using Karatsuba Algo.

$A \times B = 10^{1+1} \times 5 \times 5 + 10^1((5+7)(5+8) - 5 \times 5 - 8 \times 7) + 7 \times 8$

$= 2500 + 10(12 \times 13 - 25 - 56) + 56$

$= 2500 + 750 + 56$

$= \boxed{3306}$

g) $A = 10^{n/2} a_1 + a_0$

$B = 10^{n/2} b_1 + b_0$

$AB = 10^n a_1 b_1 + 10^{n/2}(a_1 b_0 + a_0 b_1) + a_0 b_0$

$a_1 b_0 + a_0 b_1 = (a_1 + a_0)(b_1 + b_0) - a_1 b_1 - a_0 b_0$

$T(n) = T(\lceil n/2 \rceil) + T(\lfloor n/2 \rfloor) + T(1 + \lceil n/2 \rceil) + O(n)$

$\Rightarrow T(n) = O(n^{\log 3}) = O(n^{1.585})$