EE 3755

Computer Arithmetic

Handout # 9

Multiplicative Division Technique

<u>The Problem:</u> A = dividend; B = divisor.
We want to get an approximation of the quotient Q where $Q \cong \frac{A}{B}$.

- Here A, B are n-bit positive fractions.

∴ B = <u>normalized</u> fraction = .1xx⋯xx.

If B is not normalized then both A and B must be adjusted (ie; $\frac{A}{B} = \frac{.01010}{.01110} = \frac{.10100}{.11100}$).

<u>The main concept:</u> Find a factor F such that $B \times F \cong 1$. Then $Q = \frac{A}{B} = \frac{A \times F}{B \times F} \cong A \times F$

or $\boxed{Q \cong A \times F}$

The factor F is going to be the result of several iterations ($F = F_1 \times F_2 \times F_3 \times \cdots \times F_k$)

or

$$Q = \frac{A}{B} = \frac{A \times F_1}{B \times F_1} = \frac{(A \times F_1) \times F_2}{(B \times F_1) \times F_2} = \frac{(A \times F_1 \times F_2) \times F_3}{(B \times F_1 \times F_2) \times F_3} =$$

$$= \cdots\cdots = \frac{(A \times F_1 \times F_2 \times \cdots \times F_{k-1}) \times F_k}{(B \times F_1 \times F_2 \times \cdots \times F_{k-1}) \times F_k}.$$

For some value of k, $B \times \underbrace{F_1 \times F_2 \times \cdots \times F_k}_{F} = .\underbrace{11 \cdots 11}_{n \text{ ones}} \rightarrow 1$

and thus $A \times \underbrace{F_1 \times F_2 \times \cdots \times F_k}_{F} \rightarrow Q$. Therefore, the technique then stops.

The various iterations give:

Iteration 1: $Q_1 = \dfrac{A \times F_1}{B \times F_1} = \dfrac{A_1}{B_1}$ or $\boxed{\begin{array}{l} B_1 = B \times F_1 \\ A_1 = A \times F_1 \end{array}}$

Iteration 2: $Q_2 = \dfrac{A_1 \times F_2}{B_1 \times F_2} = \dfrac{A_2}{B_2}$ or $\boxed{\begin{array}{l} B_2 = B_1 \times F_2 \\ A_2 = A_1 \times F_2 \end{array}}$

Iteration 3: $Q_3 = \dfrac{A_2 \times F_3}{B_2 \times F_3} = \dfrac{A_3}{B_3}$ or $\boxed{\begin{array}{l} B_3 = B_2 \times F_3 \\ A_3 = A_2 \times F_3 \end{array}}$

$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$

Iteration k: $Q_k = \dfrac{A_{k-1} \times F_k}{B_{k-1} \times F_k} = \dfrac{A_k}{B_k}$ or $\boxed{\begin{array}{l} B_k = B_{k-1} \times F_k \\ A_k = A_{k-1} \times F_k \end{array}}$

In the above, the sequence $B, B_1, B_2, \ldots, B_k$ must be such that $B < B_1 < B_2 < B_3 \cdots < B_k$, (or in general $B_{i-1} < B_i$). Also, $B_k \rightarrow 1$; (the system's precision dictates how close is B_k to 1).

Question: How do we choose the factors $F_1, F_2, F_3, \ldots, F_k$ so that $B \times F \to 1$; $(F = F_1 \times F_2 \times \cdots F_k)$?

Answer: $B =$ normalized fraction $= 0.1xx\ldots xx$ or $\frac{1}{2} \leq B < 1$. Thus $\boxed{B = 1 - \delta \text{ where } 0 < \delta \leq \frac{1}{2}}$

Iteration 1: Choose $F_1 = 1 + \delta$. Then $A_1 = A \times F_1 = A(1+\delta)$ and $\boxed{B_1 = B \times F_1 = (1-\delta)(1+\delta) = 1 - \delta^2}$

Since $\delta^2 < \delta$ then $B < B_1$. Observe that $0 < \delta^2 \leq \frac{1}{4}$ and thus $B_1 = 1 - \delta^2 \geq \frac{3}{4}$ or

$\boxed{B_1 = 0.11xx\ldots xx}$; B_1 has at least two leading ones.

Iteration 2: Next choose $F_2 = 1 + \delta^2$. Then $A_2 = A_1 \times F_2 = A_1 \times (1+\delta^2) = A \times (1+\delta) \times (1+\delta^2)$ and

$\boxed{B_2 = B_1 \times F_2 = (1-\delta^2)(1+\delta^2) = 1 - \delta^4}$.

Since $\delta^4 < \delta^2$, $(\text{or } 1-\delta^2 < 1-\delta^4)$, it is implied that $B_1 < B_2$. Observe that $0 < \delta^4 \leq \frac{1}{2^4}$ and thus $B_2 = 1 - \delta^4 \geq 1 - \frac{1}{2^4} = \frac{2^4-1}{2^4} = .1111$ or

$\boxed{B_2 = 0.1111xx\ldots xx}$; B_2 has at least four leading ~~ones~~ ones.

Iteration 3: Next choose $F_3 = 1+\delta^4$. Then

$A_3 = A_2 \times F_3 = A(1+\delta)(1+\delta^2)(1+\delta^4)$ and

$$\boxed{B_3 = B_2 \times F_3 = (1-\delta^4)(1+\delta^4) = 1-\delta^8}$$

Obsiously $\delta^8 < \delta^4$ and thus $1-\delta^4 < 1-\delta^8$ or $B_2 < B_3$. Observe that $0 < \delta^8 \leq \frac{1}{2^8}$ and thus

$$B_3 = 1-\delta^8 \geq 1 - \frac{1}{2^8} = \frac{2^8-1}{2^8} = .\underbrace{11111111}_{\text{eight ones}} \text{ or}$$

$$\boxed{B_3 = .\underbrace{11111111}_{\text{eight ones}}\,xx\cdots xx}$$; B_3 has at least eight leading ones.

Iteration 4: The factor here is $F_4 = 1+\delta^8$ etc, ...

* If the hardware allows precision of n fractional bits, (fractions A, B are n-bit long) the technique stops whenever

$$B_k = .\underbrace{111\cdots111}_{n \text{ ones}}$$

Problem: The presented technique requires computing $1+\delta$ given that we have $B = 1-\delta$;
computing $1+\delta^2$ after we get $B_1 = 1-\delta^2$;
computing $1+\delta^4$ after we get $B_2 = 1-\delta^4$;
computing $1+\delta^8$ " " " $B_3 = 1-\delta^8$;
etc, etc ...

Question: How do we get these $1+\delta$, $1+\delta^2$, $1+\delta^4$, $1+\delta^8$, etc ... ?

Answer:

The problem here is: "given $1-x$ being a positive fraction, compute x or $1+x$".

Let $1-x$ be $1-x = (\cdot\, y_{n-1}\, y_{n-2} \cdots y_0)_2 \neq 0$

Then $x = (2\text{'s complement of } (1-x)) = (\cdot\, z_{n-1} \cdots z_0)_2$

and (of course) $1+x = 1 \cdot z_{n-1} \cdots z_1\, z_0$

Proof :

$$\boxed{1-x+x=1} \quad (1)$$

$$1-x+2\text{'s compl. of }(1-x) = \begin{array}{r} .\,y_{n-1}\,y_{n-2}\cdots y_1\,y_0 \\ .\,\overline{y_{n-1}}\,\overline{y_{n-2}}\cdots\overline{y_1}\,\overline{y_0} \\ +)\qquad\qquad\qquad 1 \\ \hline (1.\,0\;0\cdots 0\;0)_2 \end{array}$$

or

$$\boxed{1-x+(2\text{'s compl. of }(1-x))=1} \quad (2)$$

Eqs. (1), (2) imply

$$\boxed{x = 2\text{'s complement of }(1-x)}$$

Problem: Consider the multiplicative division technique for computing $Q=\frac{A}{B}$. Here A, B are n-bit positive fractions with B being normalized. If the hardware allows precision of 32 fractional bits, what is the maximum number of iterations that need to be performed before the technique stops? Justify your answer. What is the overall factor F in this case?

<u>Solution:</u> The maximum # of iterations to be performed (before the technique stops) is <u>five</u>. Since B is normalized fraction, then $\frac{1}{2} \leq B < 1$ and therefore $B = 1 - \delta$ where $0 < \delta \leq 1/2$. The factors involved are $F_1 = 1 + \delta$, $F_2 = 1 + \delta^2$, $F_3 = 1 + \delta^4$, $F_4 = 1 + \delta^8$, $F_5 = 1 + \delta^{16}$.

1st iteration gives $B_1 = B \times F_1 = (1-\delta)(1+\delta) = 1 - \delta^2$.
2nd " " $B_2 = B_1 \times F_2 = (1-\delta^2)(1+\delta^2) = 1 - \delta^4$.
3rd " " $B_3 = B_2 \times F_3 = (1-\delta^4)(1+\delta^4) = 1 - \delta^8$.
4th " " $B_4 = B_3 \times F_4 = (1-\delta^8)(1+\delta^8) = 1 - \delta^{16}$.
5th " " $B_5 = B_4 \times F_5 = (1-\delta^{16})(1+\delta^{16}) = 1 - \delta^{32}$.

Since $0 < \delta \leq \frac{1}{2}$ then $0 < \delta^{32} \leq \frac{1}{2^{32}}$ and

thus $B_5 = 1 - \delta^{32} \geqslant 1 - \frac{1}{2^{32}} = \frac{2^{32}-1}{2^{32}} = .\underbrace{111\cdots111}_{\text{32 ones}}$.

Therefore, after at most five iterations the technique stops. The overall factor F is
$F = F_1 \times F_2 \times F_3 \times F_4 \times F_5 = (1+\delta)(1+\delta^2)(1+\delta^4)(1+\delta^8)(1+\delta^{16})$,

and $Q \cong A \times F$.