Guard Digits I

- $p = 3, \beta = 10$
- Calculate $2.15 \times 10^{12} - 1.25 \times 10^{-5}$:
- Compute and then round

$$
\begin{aligned}
x &= 2.15 \times 10^{12} \\
y &= 0.0000000000000000125 \times 10^{12} \\
x - y &= 2.1499999999999999875 \times 10^{12}
\end{aligned}
$$

round to 2.15×10^{12}

Here we assume that computation is exactly done

Guard Digits II

- Round and then compute

$$
\begin{aligned}
x &= 2.15 \times 10^{12} \\
y &= 0.00 \times 10^{12} \\
x - y &= 2.15 \times 10^{12}
\end{aligned}
$$

  Answer is the same

  Reasonable as $x \approx x - y$

- Another example: $10.1 - 9.93 = 0.17$

Guard Digits III

- Round and then compute

$$10.1 - 9.93 = 1.01 \times 10^1 - 0.99 \times 10^1 = 0.02 \times 10^1$$
$$= 2.00 \times 10^{-1}$$
$$\text{error} = 2.00 \times 10^{-1} - 0.17 = 0.03$$
$$\text{ulps} = 0.01 \times 10^{-1} = 10^{-3}$$
$$\text{error} = 0.03 = 30\text{ulps}$$

Relative error is

$$0.03/0.17 = 3/17$$

Guard Digits IV

The error is quite large

- Compute and round

$$10.1 - 9.93 = 0.17 = 1.7 \times 10^{-1}$$

error $= 0$

We can afford to exactly store the answer

- The problem is that we cannot compute and then round

Guard Digits V

- How big can the error be? (if round and then compute)

Theorem

Using p digits with base β for $x - y$, the relative error can be as large as $\beta - 1$

Proof:

- Let

$$x = 1.0 \ldots 0, y = 0.\underbrace{\eta \ldots \eta}_{p \text{ digits}}, \eta = \beta - 1$$

Guard Digits VI

Correct solution $x - y = \beta^{-p}$

Under some rounding scheme, the computed solution is

$$1.0\ldots0 - 0.\underbrace{\eta\ldots\eta}_{p-1 \text{ digits}} = \beta^{-p+1}$$

Relative error

$$\frac{|\beta^{-p} - \beta^{-p+1}|}{\beta^{-p}} = \beta - 1$$

- Example: $p = 3$, $\beta = 10$

Guard Digits VII

$x = 1.00, y = 0.999, x - y = 0.001 = 10^{-3}$

Computed solution $= 1.00 \times 10^0 - 0.99 \times 10^0$
$= 0.01 \times 10^0 = 0.01$

Relative error

$$\frac{|0.01 - 0.001|}{0.001} = 9$$

Such large errors occur if x and y are close

- Single guard digit

  p increased by 1 in the device for addition and subtraction

Guard Digits VIII

round and then compute

$1.010 \times 10^1 - 0.993 \times 10^1 = 0.017 \times 10^1$

Note $0.017 \times 10^1 = 1.70 \times 10^{-1}$ can be stored as $p = 3$

- That is, one additional digit in the process of subtraction. All values are still stored using $p = 3$
- So in the device for subtraction, we should put additional digits

Guard Digits IX

- Another example:

$$
\begin{aligned}
& 110 - 8.59 \\
= & 1.100 \times 10^2 - 0.085 \times 10^2 \\
= & 1.015 \times 10^2 \approx 1.02 \times 10^2
\end{aligned}
$$

Correct answer 101.41

Relative error around 0.006

$\epsilon = \frac{1}{2}\beta^{-p+1} = \frac{1}{2}10^{-2} = 0.005$