Verificarlo, Verrou, Interflop: Floating point computing verification on new architecture and large scale systems

Devan Sohier\textsuperscript{1}, Pablo Oliveira\textsuperscript{1,2}, François Févotte\textsuperscript{3}, Bruno Lathuilière\textsuperscript{3}, Éric Petit\textsuperscript{2,4}, Olivier Jamond\textsuperscript{5}

\textsuperscript{1}UVSQ; \textsuperscript{2}Exascale Computing Research Lab; \textsuperscript{3}EDF R&D Pericles; \textsuperscript{4}Intel DCG E&G and \textsuperscript{5}CEA

September 27, 2019
Section 1

Toward a common theoretical framework
Statistical Analysis of Stochastic Arithmetic: Motivation

- **Stochastic Arithmetic**
  - Numerical errors modeled by introducing random perturbations.
  - Estimate significance of result by collecting many samples.

- **Motivation for statistical analysis**
  - How many stochastic samples should be run?
  - What is the probability of over-estimating the number of significant digits?
  - Can we give a sound confidence interval for the number of significant digits?
Example: Kahan 2x2 System

- Ill-conditioned linear system (condition number $2.5 \times 10^8$).
- We solve it with the Cramer’s formula.

$$
\begin{pmatrix}
0.2161 & 0.1441 \\
1.2969 & 0.8648
\end{pmatrix}
\times
= 
\begin{pmatrix}
0.1440 \\
0.8642
\end{pmatrix}
$$

(1)

$$
\begin{pmatrix}
2 \\
-2
\end{pmatrix}
\quad 
\begin{pmatrix}
1.9999999958366637 \\
-1.9999999972244424
\end{pmatrix}
$$

(2)

- The IEEE-754 result has 8 significant decimal digits.
- $x_{IEEE}[0]$ has 28.8 significant bits.
With Verificarlo, we collect 10000 $t = 52$ FULL MCA samples.

$$s_{\text{PARKER}} = -\log_2 \frac{\hat{\sigma}}{|\hat{\mu}|} \approx 28.5.$$  

But how confident are we that it is a good estimate? Could we have used a smaller number of samples and still get a reliable estimation of the results quality?
Some notations

- $x_{\text{ieee}}$ is the IEEE-754 result
- $X_1, X_2, \ldots, X_n$ are the values returned by $n$ runs of the program using stochastic arithmetic. These are seen as realizations of a random variable $X$.
- $\hat{\mu}$ and $\hat{\sigma}$ are the empirical average and standard deviation.
- $\mu$ and $\sigma$ are the mean and std. deviation of the random variable $X$. 

![Graph showing histogram and normal distribution with labels $x_{\text{ieee}}$, $\hat{\mu}$, $\mu$, and # occurrences vs. $x$.]
Choosing a reference value

- We require a reference value against which accuracy is measured.
- Examples of common reference values,
  - \( x_{\text{real}} \), if the exact solution is known.
  - \( x_{\text{IEEE}} \), when the program is deterministic.
  - \( \hat{\mu} \), if the program is non-deterministic.
  - \( Y \), a random variable, to compare two implementations of an algorithm or measuring significance between runs of the same program.
Four kind of scenarios are studied in our paper.
In each case the error is modeled by a random variable $Z$.
For simplicity, in the following we consider the relative precision with scalar reference.

<table>
<thead>
<tr>
<th></th>
<th>reference $x$</th>
<th>reference $Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>absolute precision</td>
<td>$Z = X - x$</td>
<td>$Z = X - Y$</td>
</tr>
<tr>
<td>relative precision</td>
<td>$Z = X/x - 1$</td>
<td>$Z = X/Y - 1$</td>
</tr>
</tbody>
</table>

With no error, the expected result of $Z$ is 0.
Significant bits

- Stott Parker defines the significant digits between \(x\) and \(y\) as the largest \(s\) that satisfies \(|x/y - 1| \leq 2^{-s}\).
- Or put more simply, the error is less than \(2^{-s}\).
- We naturally extend this definition to \(Z\) the random variable modeling the stochastic error.

Significant bits

The number of significant digits with probability \(p_s\) can be defined as the largest number \(s\) such that

\[
P(|Z| \leq 2^{-s}) \geq p_s. \tag{3}
\]

\[0 \quad 1 \quad 2 \quad \ldots \quad s \quad \ldots \quad 45 \quad 46 \quad 47 \quad 48 \quad 49 \quad 50 \quad 51 \quad 52\]

\(s\) significant \(\times\) error satisfies \(|Z| \leq 2^{-s}\)
Bits after s still can encode useful information about the result.
  ▶ Even if bits on its left are wrong, they can improve the accuracy...
  ▶ ...if they are correct on average ($p_c > 51\%$).
  ▶ Keeping these bits improves the rounded result on average.

A bit $k$ after $s$ contributes to the result with probability $p_c$ iff the $k$-th bit of $Z$ is 0 (no error in this bit) with probability $p_c$.

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>s</th>
<th>...</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>...</th>
<th>c</th>
<th>...</th>
<th>51</th>
<th>52</th>
</tr>
</thead>
</table>

significant at $p_s = .99$  \[\leftarrow \rightarrow\]
contributing at $p_c = .51$  \[\leftarrow \rightarrow\]
random noise
Results

1. Probability for significance and contribution for Normal Centered Distributions.
2. Probability for significance and contribution for General Distributions.

Normality of the Kahan 2x2 System

Figure: Normality of 10000 samples of $X[0]$ with $t = 52$ and FULL MCA

- We take as reference the empirical mean $\hat{\mu}$.
Centered Normal Hypothesis: Significant bits

\[ \mathcal{N}(0,1) \text{ Cumulative distribution function} \]

\[ F_{0,1}(x) \]

\[ P(Z < 2^{-s}) = F(2^{-s}/\sigma) \]

\[ P(|Z| < 2^{-s}) = 2F(2^{-s}/\sigma) - 1 \]

\[ P(Z < -2^{-s}) = 1 - F(2^{-s}/\sigma) \]
Centered Normal Hypothesis: Significant bits

**Theorem**

For a normal centered error distribution $Z \sim \mathcal{N}(0, \sigma)$, the $s$-th bit is significant with probability

$$p_s = 2F \left( \frac{2^{-s}}{\sigma} \right) - 1,$$

with $F$ the cumulative function of the normal distribution with mean 0 and variance 1.

By inverting this formula, we can provide a formula for the number of significant digits that only depends on $\sigma$ and $p_s$,

$$s = -\log_2 (\sigma) - \log_2 \left( F^{-1} \left( \frac{p_s + 1}{2} \right) \right).$$
Figure: Profile of the significant bit curve $p_s = 2F \left( \frac{2^{-s}}{\sigma} \right) - 1$

- If we take the empirical average as reference value, we fall back into Stott Parker definition of significant bits assuming a large number of samples $- \log_2(\sigma) = - \log_2(\frac{\sigma_X}{|\hat{\mu}|})$
- The digit of Stott Parker's formula has 68 % chances of being significant. (1-sigma rule)
- If we substract 1.37 bits from Stott Parker's formula, the resulting bit has 99 % chances of being significant.
\[ s = - \log_2 (\sigma) - \log_2 \left( F^{-1} \left( \frac{p_s + 1}{2} \right) \right). \]

- Why is this formula independent of the number of samples \( n \)?
- \( \sigma \) is unknown; we can only estimate it from \( \hat{\sigma} \).
- For normal distributions, the following confidence interval with confidence \( 1 - \alpha \) based on the \( \chi^2 \) distribution with \( (n - 1) \) degrees of freedom is sound [?]:

\[
\frac{(n - 1)\hat{\sigma}^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n - 1)\hat{\sigma}^2}{\chi^2_{1 - \alpha/2}}. \tag{4}
\]

- In the following we choose a confidence of \( 1 - \alpha = 95\% \).
By combination, we produce a sound lower bound on the significant bits,

\[
s \geq -\log_2(\hat{\sigma}) - \left[\frac{1}{2} \log_2 \left( \frac{n - 1}{\chi^2_{1 - \alpha/2}} \right) + \log_2 \left( F^{-1} \left( \frac{p + 1}{2} \right) \right) \right]
\]

\[
\delta_{\text{CNH}}
\]

\[
\hat{s}_{\text{CNH}}
\]

For \( n = 30 \) samples and \( p = 99\% \) \( s \geq -\log_2\hat{\sigma} - 1.792 \)

For \( n = 15 \) samples and \( p = 99\% \) \( s \geq -\log_2\hat{\sigma} - 2.023 \)

(\( \log_2\hat{\sigma} \) is Stott Parker’s formula when the reference is \( \hat{\mu} \))
Figure: Normal curve; the gray zones correspond to the area where the $k$-th bit contributes to make the result closer to 0 (whatever the preceding digits).

$$\exists i, \left\lfloor 2^k |Z| \right\rfloor = 2i$$

$$\Leftrightarrow \quad 2i \leq 2^k |Z| < 2i + 1$$

$$\Leftrightarrow \quad 2^{-k}(2i) \leq |Z| < 2^{-k}(2i + 1).$$

(6)
Theorem

For a normal centered error distribution $Z \sim \mathcal{N}(0, \sigma)$, when $\frac{2^{-c}}{\sigma}$ is small, the $c$-th bit contributes to the result accuracy with probability

$$p_c \sim \frac{2^{-c}}{2\sigma\sqrt{2\pi}} + \frac{1}{2}.$$ 

If we wish to keep only bits improving the result with a probability greater than $p$, then we will keep $c$ contributing bits, with

$$c = -\log_2(\sigma) - \log_2(p_c - \frac{1}{2}) - \log_2(2\sqrt{2\pi}).$$ \hspace{1cm} (7)

Again, this formula only depends on $\sigma$ and the probability $p_c$. As previously, a sound lower or upper bound can be computed with the Chi-2 confidence interval of $\sigma$. 

18 / 38
Figure: Profile of the contribution bit curve: The shaded area represents the bound on the error. The approximation is very tight for probabilities less than 70%.
Results: Significant bits

Figure: Significant bits for Cramer $x[0]$ variable computed under the normal hypothesis using 30 and 10000 samples. The Confidence Interval (CI) lower bound is computed by using the probability of theorem 1 and bounding $\sigma$ with a 95% Chi-2 confidence interval.
Results: Contributing bits

**Figure:** Contributing bits for Cramer $x[0]$ variable computed under the normal hypothesis using 30 and 10000 samples.
Summary: Significant and Contributing bits in the CNH (1/2)

\[-\log_2 \sigma \geq 28.45\]  
\[-\log_2 \left( F^{-1} \left( \frac{p_s+1}{2} \right) \right) \approx -1.37\]

\[-\log_2 (p_c - \frac{1}{2}) - \log_2 (2\sqrt{2\pi}) \approx +4.32\]

| 0 | ... | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | ... | 52 |

significant at \( p = .99 \)

random noise

contributing at \( p' = .51 \)

1. We estimate a lower bound for

\[-\log \sigma \geq 28.45 \approx -\log_2 \hat{\sigma} - \frac{1}{2} \log_2 \left( \frac{n-1}{\chi^2_{1-\alpha/2}} \right)\]

2. We apply a **shift left** (computed with \( p_s = 99\% \)) to get a safe significant bits lower-bound.

3. We apply a **shift right** (computed with \( p_c = 51\% \)) to get a safe contributing bits lower-bound.
Summary: Significant and Contributing bits in the CNH (2/2)

\[
- \log_2 \sigma \geq 28.45
\]

\[
- \log_2 \left( F^{-1} \left( \frac{p_s + 1}{2} \right) \right) \approx -1.37
\]

\[
- \log_2 (p_c - \frac{1}{2}) - \log_2 (2\sqrt{2\pi}) \approx +4.32
\]

<table>
<thead>
<tr>
<th>0</th>
<th>...</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>...</th>
<th>52</th>
</tr>
</thead>
</table>

- significant at \( p = .99 \)
- contributing at \( p' = .51 \)
- random noise

- Contributing bits help decide how many digits to print or store during a check-point restart.

- Only keeping contributing bits can help reducing storage and database sizes!
General Distributions

► What if the distribution is not centered normal?

**Figure:** Non normality of buckling samples on z axis and node 1. Shapiro Wilk rejects the normality hypothesis.
Let us choose a single $k$ in the mantissa and single sample $i$ among the $n$ samples.

We can define two binary tests,

- $S^k_i = \{ |Z_i| \leq 2^{-k} \}$, true iff for the $i$-th sample the $k$-th first bits are significant.
- $C^k_i = \{ \lfloor 2^k |Z_i| \rfloor \text{ is even} \}$, true iff for the $i$-th sample the $k$-th bit is contributing.

With $n$ samples we have $n$ Bernoulli Trials.

The trials are realizations of two Bernoulli random variables $S^k$ and $C^k$. 
We choose a given $k$.

Sample $X_1$:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>k</th>
<th>...</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>51</th>
<th>52</th>
</tr>
</thead>
</table>

$S_1^k$ Success

$|Z_1| \leq 2^{-k}$

Sample $X_2$:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>k</th>
<th>...</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>51</th>
<th>52</th>
</tr>
</thead>
</table>

$S_2^k$ Failure

$|Z_2| > 2^{-k}$

Sample $X_3$:

<table>
<thead>
<tr>
<th>0</th>
<th>1</th>
<th>2</th>
<th>...</th>
<th>k</th>
<th>...</th>
<th>48</th>
<th>49</th>
<th>50</th>
<th>51</th>
<th>52</th>
</tr>
</thead>
</table>

$S_3^k$ Success

$|Z_3| \leq 2^{-k}$

Out of three samples: 2 success and 1 failure; $n_s = 2$.

Can we estimate the Bernoulli distribution of $S^k$?
[?] gives the following lower-bound for the success probability of a Bernoulli distribution at 95% confidence,

\[
\frac{n_s + 2}{n + 4} - 1.65 \sqrt{\frac{(n_s + 2)(n - n_s + 2)}{(n + 4)^3}}
\]

By counting for \( S_i^k \) the number of successes \( n_s \) (where the first \( k \) digits are significant) we can derive a safe lower-bound probability.
We want to estimate the probability $p_s$ of the $s$-th bit being significant.

Suppose $p_s = \frac{1}{3}$ is the true parameter of the Bernoulli distribution.

We do $m$ different samplings of $n = 10$ values:

- 1st sampling: $n_s = 3 \rightarrow p_s \in [.15, .57]$
- 2nd sampling: $n_s = 8 \rightarrow p_s \notin [.52, .91]$
- 3rd sampling: $n_s = 2 \rightarrow p_s \in [.08, .48]$
- ...  

The confidence $1 - \alpha$ measures the proportion of samplings that produce an interval containing $p_s$.

Increasing the number of samples $n$ reduces the probability of a biased interval and therefore increases the confidence.
Example of Bernoulli Estimator on Kahan’s system

Figure: Significance and contribution per bit for variable \( X \) of the Cramer's system with 30 and 10000 samples.
Let us consider the largest $k$ so that $S_i^k$ is true for all $i$. In other words, $k$ is significant in all the collected samples.

In that case, $\Pr(S^k) > p$ with confidence $1 - \alpha$ if we have

$$n = n_s \geq \left\lceil \frac{\ln(\alpha)}{\ln(p)} \right\rceil$$

This formula gives us a simple criterion for choosing a minimal number of samples depending on the required confidence level.

1. Choose a probability and confidence level that are acceptable for your experiment: eg. $p = 90\%$ and $1 - \alpha = 95\%$
2. Compute and collect the required number of samples, here $n = 29$.
3. Find the largest $k$ that is significant for all samples; that $k$ is significant with $p = 90\%$ at confidence level 95\%.
How many samples are required?

<table>
<thead>
<tr>
<th>Confidence level $1 - \alpha$</th>
<th>0.66</th>
<th>0.75</th>
<th>0.8</th>
<th>0.85</th>
<th>0.9</th>
<th>0.95</th>
<th>0.99</th>
<th>0.995</th>
<th>0.999</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.66</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>22</td>
<td>108</td>
<td>216</td>
<td>1079</td>
</tr>
<tr>
<td>0.75</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>14</td>
<td>28</td>
<td>138</td>
<td>277</td>
<td>1386</td>
</tr>
<tr>
<td>0.8</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>16</td>
<td>32</td>
<td>161</td>
<td>322</td>
<td>1609</td>
</tr>
<tr>
<td>0.85</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>12</td>
<td>19</td>
<td>37</td>
<td>189</td>
<td>379</td>
<td>1897</td>
</tr>
<tr>
<td>0.9</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>22</td>
<td>45</td>
<td>230</td>
<td>460</td>
<td>2302</td>
</tr>
<tr>
<td>0.95</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>19</td>
<td>29</td>
<td>59</td>
<td>299</td>
<td>598</td>
<td>2995</td>
</tr>
<tr>
<td>0.99</td>
<td>12</td>
<td>17</td>
<td>21</td>
<td>29</td>
<td>44</td>
<td>90</td>
<td>459</td>
<td>919</td>
<td>4603</td>
</tr>
<tr>
<td>0.995</td>
<td>13</td>
<td>19</td>
<td>24</td>
<td>33</td>
<td>51</td>
<td>104</td>
<td>528</td>
<td>1058</td>
<td>5296</td>
</tr>
<tr>
<td>0.999</td>
<td>17</td>
<td>25</td>
<td>31</td>
<td>43</td>
<td>66</td>
<td>135</td>
<td>688</td>
<td>1379</td>
<td>6905</td>
</tr>
</tbody>
</table>

**Table:** Number of samples necessary to obtain a given confidence interval with probability $p$, according to the Bernoulli estimator (i.e. without any assumption on the probability law).
Figure: Significant bits on the $z$ axis distribution. Bernoulli estimation captures precisely the behavior (except for node 2). Normal formula overestimates the number of digits, this is expected since the distribution is strongly non normal.
Figure: Relative error between the samples and the mean of the $z$-axis distribution. The blue envelope corresponds to the computed confidence interval with 30 samples. Black dots are samples that fall inside the CI. Red crosses are outliers that fall outside the CI. In the Bernoulli case, only 3 samples out of 70 fall outside of the interval; which is compatible with the 90% probability threshold.
Limits and Discussion

- These confidence intervals estimate the error of over-estimating $s$ due to sampling errors
  - not enough samples taken or biased sampling
- These confidence intervals do not account for model errors
  - Changes in the dataset
  - Failures of MCA or CESTAC to correctly model FP errors (thread scheduling, model corner-cases, etc.)
Conclusion on Confidence Intervals for Stochastic Arithmetic

- For normal centered distributions:
  - Simple probability formulations for significance and contribution that only depend on $\hat{\sigma}$, $n$ and $1 - \alpha$.
  - Applying a left or right shift to the pivotal $-\log_2(\sigma)$ Stott Parker’s estimator produces a lower-bound on the number of significant and contributing bits.

- For general distributions:
  - Model each mantissa bit as a separate Bernoulli distribution.
  - When only interested in the significant bits, a simple formula computes how many samples are needed to reach a given probability level.

- How can I apply these results to my studies?
  - Tables for the CNH shifts and number of required samples are available in the preprint.
  - A jupyter notebook implementing the formulas is also available.
Table: Summary of the numerical quality assessment of 4 versions of code_aster, using Verrou and the standard MCA estimator with 6 samples.

<table>
<thead>
<tr>
<th>Implementation</th>
<th>( \hat{s}_{\text{MCA}} )</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( a )</td>
<td>( e )</td>
</tr>
<tr>
<td>version0</td>
<td>Fail</td>
<td>Fail</td>
</tr>
<tr>
<td>version1</td>
<td>30.89</td>
<td>19.73</td>
</tr>
<tr>
<td>version2</td>
<td>30.96</td>
<td>19.80</td>
</tr>
<tr>
<td>version3</td>
<td>32.82</td>
<td>21.65</td>
</tr>
</tbody>
</table>

With version3 the accuracy seems improved, but we need confidence intervals to update reference value. We choose \( p = (1 - \alpha) = 0.995 \):

\[ N_{\text{sample}} = 1058 \]
<table>
<thead>
<tr>
<th>Implementation</th>
<th>$\hat{s}_B$</th>
<th>$\hat{s}_{\text{IEEE}}$</th>
<th>$\hat{s}_{\text{CNH}}$ (normality test p-value)</th>
<th>$\hat{s}_{\text{MCA}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>version1</td>
<td>28</td>
<td>28</td>
<td>29.01 (0.10)</td>
<td>30.59</td>
</tr>
<tr>
<td>version2</td>
<td>29</td>
<td>29</td>
<td>29.55 (0.89)</td>
<td>31.13</td>
</tr>
<tr>
<td>version3</td>
<td>30</td>
<td>31</td>
<td>31.22 (0.52)</td>
<td>32.79</td>
</tr>
</tbody>
</table>

(a) quantity $a$

<table>
<thead>
<tr>
<th>Implementation</th>
<th>$\hat{s}_B$</th>
<th>$\hat{s}_{\text{IEEE}}$</th>
<th>$\hat{s}_{\text{CNH}}$ (normality test p-value)</th>
<th>$\hat{s}_{\text{MCA}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>version1</td>
<td>17</td>
<td>17</td>
<td>17.85 (0.10)</td>
<td>19.43</td>
</tr>
<tr>
<td>version2</td>
<td>18</td>
<td>18</td>
<td>18.39 (0.89)</td>
<td>19.97</td>
</tr>
<tr>
<td>version3</td>
<td>19</td>
<td>19</td>
<td>20.05 (0.52)</td>
<td>21.63</td>
</tr>
</tbody>
</table>

(b) quantity $e$

Table: Comparison of stochastic estimators for 3 version of code_aster, with 1058 samples.
### Table: Comparison of stochastic estimators for 3 version of code_aster, with 1058 samples.

<table>
<thead>
<tr>
<th>Implementation</th>
<th>$\hat{s}_{B}^\mu$</th>
<th>$\hat{s}_{B}^{\text{IEEE}}$</th>
<th>$\hat{s}_{\text{CNH}}$ (normality test $p$-value)</th>
<th>$\hat{s}_{\text{MCA}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>version 1</td>
<td>28.89</td>
<td>28.57</td>
<td>29.01 (0.10)</td>
<td>30.59</td>
</tr>
<tr>
<td>version 2</td>
<td>29.33</td>
<td>29.35</td>
<td>29.55 (0.89)</td>
<td>31.13</td>
</tr>
<tr>
<td>version 3</td>
<td>30.91</td>
<td>31.00</td>
<td>31.22 (0.52)</td>
<td>32.79</td>
</tr>
</tbody>
</table>

- **(a) quantity $a$**

<table>
<thead>
<tr>
<th>Implementation</th>
<th>$\hat{s}_{B}^\mu$</th>
<th>$\hat{s}_{B}^{\text{IEEE}}$</th>
<th>$\hat{s}_{\text{CNH}}$ (normality test $p$-value)</th>
<th>$\hat{s}_{\text{MCA}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>version 1</td>
<td>17.73</td>
<td>17.41</td>
<td>17.85 (0.10)</td>
<td>19.43</td>
</tr>
<tr>
<td>version 2</td>
<td>18.16</td>
<td>18.19</td>
<td>18.39 (0.89)</td>
<td>19.97</td>
</tr>
<tr>
<td>version 3</td>
<td>19.75</td>
<td>19.84</td>
<td>20.05 (0.52)</td>
<td>21.63</td>
</tr>
</tbody>
</table>

- **(b) quantity $e$**
version3 is our new reference.

**version0 analysis**

\[
\left| \frac{a_{\text{ieee}}^{\text{version0}} - a_{\text{ieee}}^{\text{version3}}}{a_{\text{ieee}}^{\text{version3}}} \right| = 4.29 \times 10^{-10}
\]

\[
\left| \frac{e_{\text{ieee}}^{\text{version0}} - e_{\text{ieee}}^{\text{version3}}}{e_{\text{ieee}}^{\text{version3}}} \right| = 9.84 \times 10^{-7}
\]

Need to analyze other test cases related to these 2 corrections before integration