Aug. 10, 2021

Floating-point Square Root Calculation Arithmetic Based on Taylor-Series Expansion and Region Division

Jianglin Wei, A. Kuwana, H. Kobayashi
K. Kubo, Y. Tanaka

Division of Electronics and Informatics, Gunma University
Oyama National College of Technology
Division of Mechanical Science and Technology, Gunma University
Japan

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Tradeoff
- Conclusion

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Results
- Hardware Implementation Tradeoff
- Conclusion

Research Background

◆High-speed high-precision floating-point arithmetic

→ Scientific computing
Embedded systems
Mobile applications

◆Square Root calculation ⇒ Very tough !

Multiple usage of basic arithmetic operations

- ➢ Addition / Subtraction → Relatively easy
- ➢ Multiplication → Modestly complicated

Research Objective

- Floating-point square root calculation
  - Simple circuit
  - High-speed
- Application of Taylor-series expansion with divide-and-conquer of mantissa region
- Clarification of
  - calculation algorithm
  - design tradeoff among accuracy, number of operations and LUT size.

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Tradeoff
- Conclusion

Taylor Series Expansion

Re-write a smooth function
as an infinite sum of polynomial terms.

Function $f(x)$ for a point $x = a$ is given by :

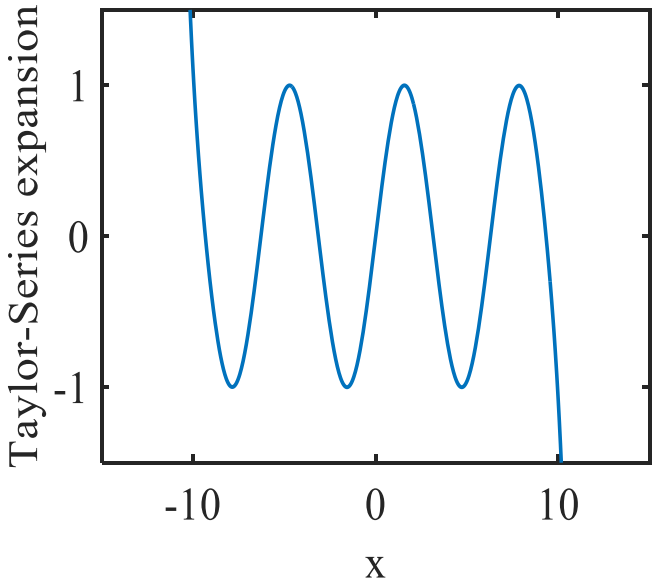
$$f(x) = f(a) + f'(a)(x-a) + \frac{(f)''(a)}{2!}(x-a)^2 + \cdots + \frac{(f)^n(a)}{n!}(x-a)^n + \cdots$$

Convergence range $\alpha < x < \beta$

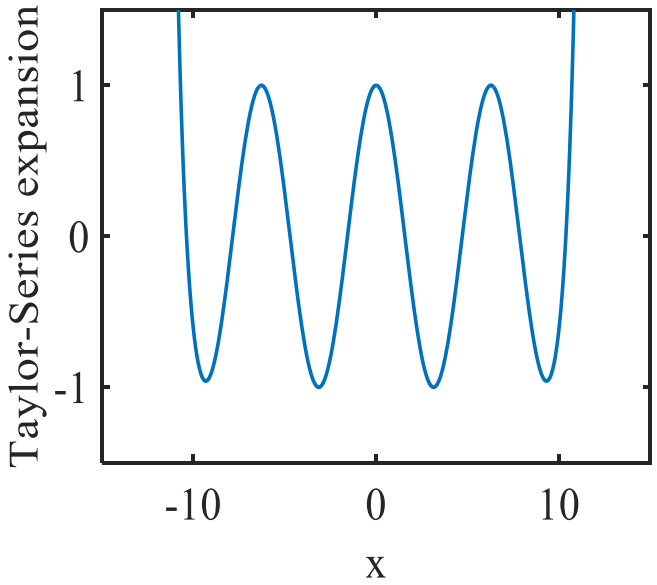
Taylor-Series of sin(x), cos(x)

Taylor-series expansion with 20 terms.

for center value : a=0



$\sin(x)$



$\cos(x)$

Convergence range : $-\infty < x < +\infty$

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Tradeoff
- Conclusion

Normalized Floating-Point Representation

Floating-point representation in binary :

Mantissa : M　　(1 ≦ M < 2)　　Exponent : E

$$M \times 2^{E}$$

Decimal point

$$1.abcdef\cdots \times 2^{E}$$

Mantissa　　Exponent

$a, b, c, d, e, f, \cdots$: 0 or 1

Square Root Calculation

Floating-point binary : $S = M \times 2^E$

Square root calculation of S

$$\sqrt{S} = \sqrt{M} \times \sqrt{2^E}$$

E: even $\rightarrow$ Let $E=2k$

$$\sqrt{S} = \sqrt{M} \times 2^k$$

E: odd $\rightarrow$ Let $E=2k+1$

$$\sqrt{S} = \sqrt{M} \times \sqrt{2} \times 2^k$$

$\sqrt{M}$ calculation by Taylor-series expansion for a specified accuracy.

Usage of Taylor Expansion

Calculate square root of mantissa : $\sqrt{M}$ ($1 \leq M < 2$)

$x = M$

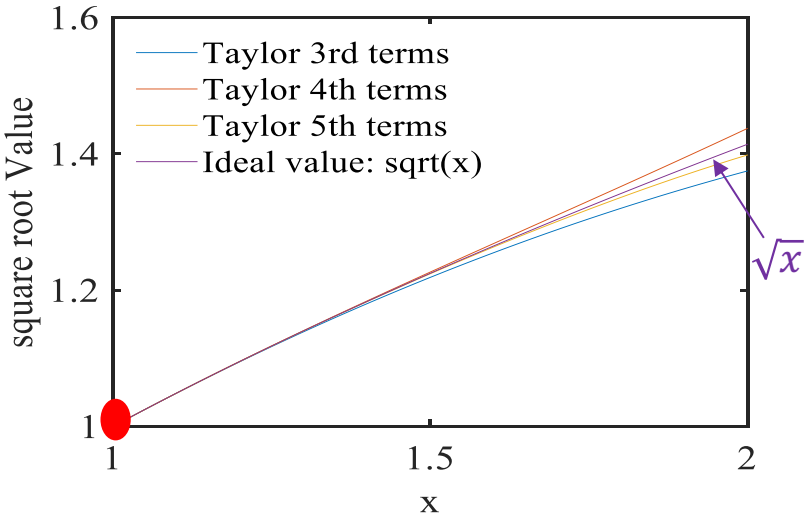
$f(x) = \sqrt{x}$ by Taylor expansion at $x = a$ $(1 \leq a < 2)$

$f(x) = \sqrt{x}$ Taylor expansion at $x = a$:

$$f(x) = \sqrt{a} \times \left\{ 1 + \frac{x - a}{2} - \frac{(x - a)^2}{8 \times a} + \frac{(x - a)^3}{16 \times a^2} - \frac{5 \times (x - a)^4}{128 \times a^3} + \frac{7 \times (x - a)^5}{256 \times a^4} - \cdots \right\}$$

Coefficient values are stored in LUT in advance.

Mantissa Region Division



Taylor series expansion of $\sqrt{x}$ at center value a = 1

Divide and Conquer of Mantissa Region

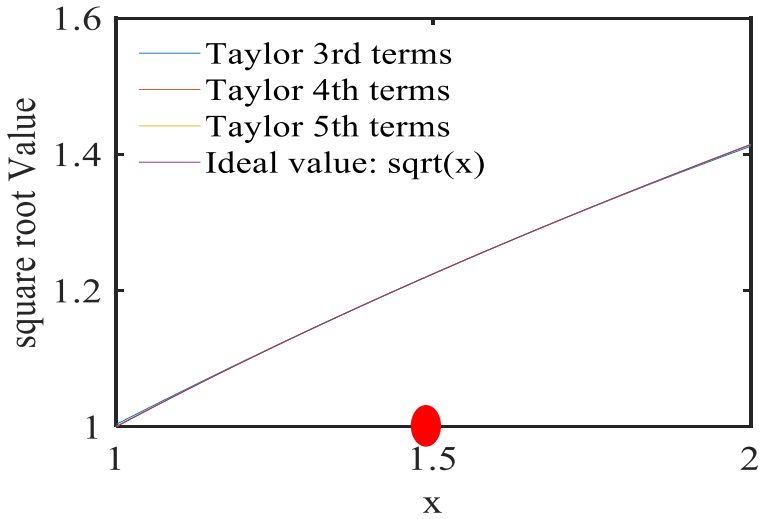
1 region :

a = 1.5 $1 \leq x < 2$

2 regions :

a = 1.25 $1 \leq x < 1.5$

a = 1.75 $1.5 \leq x < 2$



Taylor series expansion of $\sqrt{x}$ at center value a = 1.5

4 regions :

a = 1.125 $1 \leq x < 1.25$

a = 1.375 $1.25 \leq x < 1.5$

a = 1.625 $1.5 \leq x < 1.75$

a = 1.875 $1.75 \leq x < 2$

⋮

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Tradeoff
- Conclusion

Definition of Accuracy

Ex : $\frac{1}{2^{16}}$ accuracy

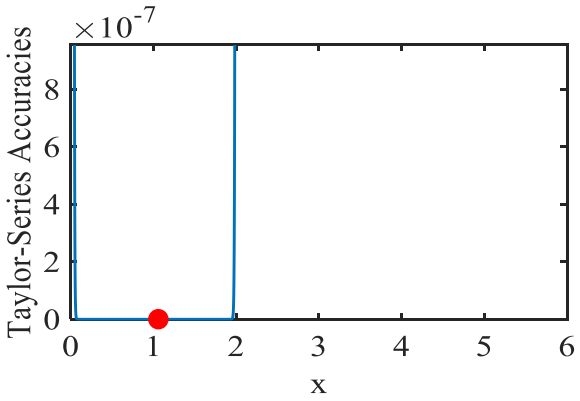
$$max \left| \frac{f(x) - t(n,x)}{f(x)} \right| \leq \frac{1}{2^{16}}$$

$f(x)$: Original function (Ideal value)

$t(n,x)$: Taylor expansion of n terms

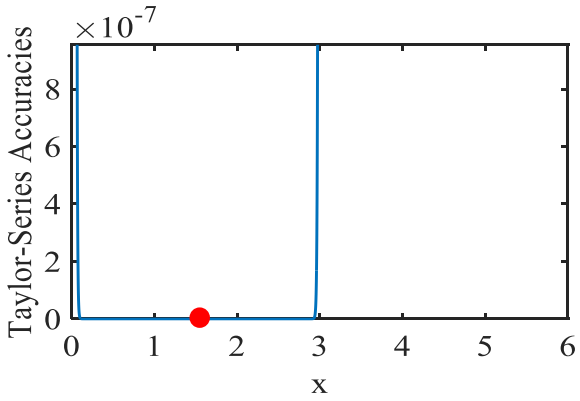
Graph of $\sqrt{x}$ Taylor Expansion

Number of Taylor expansion terms : 200



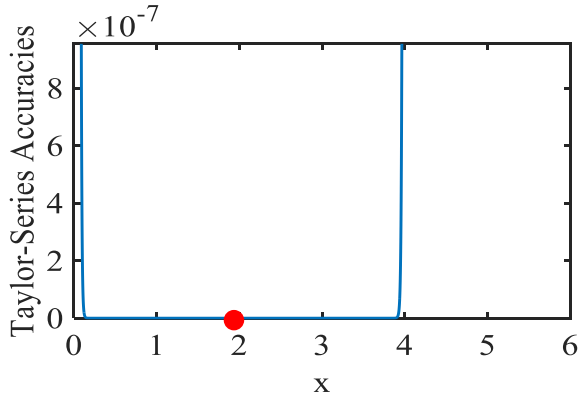
a = 1

Convergence range : 0 < *x* < 2



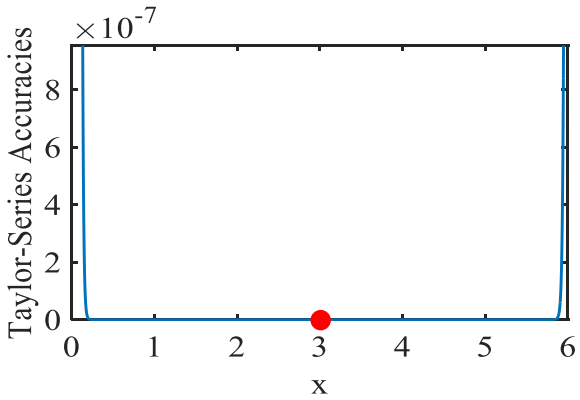
a = 1.5

Convergence range : 0 < *x* < 3



a = 2

Convergence range : 0 < *x* < 4



a = 3

Convergence range : 0 < *x* < 6

One-Region Case

Use Taylor series expansion equation :

$f(x) = \sqrt{x}$ $\quad (1 \leq x < 2)$

Mantissa represented by binary decimal point.

Specified accuracy

Taylor-series expansion \ precision		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) $M =$ 1.xxxxxx… $1 \leq M < 2$	$a = 1.5$	3	7	9	12	15

Taylor series expansion at center value $a = 1.5$

Number of Taylor expansion terms to meet specified accuracy.

Two-Region Case

Use Taylor series expansion equation :

$f(x) = \sqrt{x}$ $\quad (1 \leq x < 2)$

We check value (0 or 1)
of the first decimal place of Mantissa.

Taylor-series expansion / precision		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) $M_D = 1.0xxxxx\cdots$ $1 \leq M_D < 1.5$	$a = 1.25$	3	5	7	8	11
(ii) $M_D = 1.1xxxxx\cdots$ $1.5 \leq M_D < 2$	$a = 1.75$	2	5	6	7	10

Four-Region Case

Use Taylor series expansion equation :

$f(x) = \sqrt{x} \qquad (1 \leq x < 2)$

We check values (00, 01, 10 or 11)
of the first two decimal places of Mantissa.

Taylor-series expansion \ precision		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M = 1.00xxxx… $1 \leq M_D < 1.25$	a=1.125	2	4	5	6	9
(ii) M = 1.01xxxx… $1.25 \leq M_D < 1.5$	a=1.375	2	4	5	6	8
(iii) M = 1.10xxxx… $1.5 \leq M_D < 1.75$	a=1.625	2	4	5	6	8
(iv) M = 1.11xxxx… $1.75 \leq M_D < 2$	a=1.875	2	4	5	5	7

Eight-Region Case

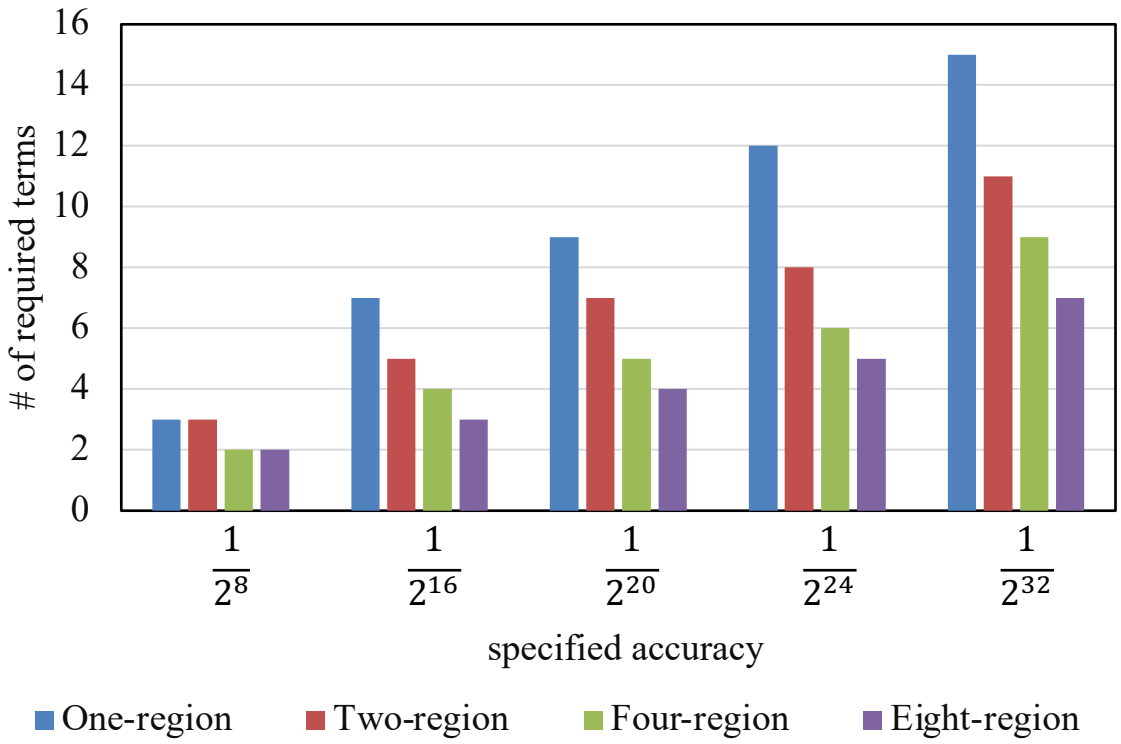
Check the values (000, 001,…, 111)
of the first three decimal places of Mantissa.

Taylor-series expansion \ precision		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M = 1.000xxxx… $1 \le M_D < 1.125$	a = 1.0625	2	3	4	5	7
(ii) M = 1.001xxxx… $1.125 \le M_D < 1.25$	a = 1.1875	2	3	4	5	7
(iii) M = 1.010xxxx… $1.25 \le M_D < 1.375$	a=1.3125	2	3	4	5	7
(iv) M = 1.011xxxx… $1.375 \le M_D < 1.5$	a=1.4375	2	3	4	5	6
(v) M = 1.100xxxx… $1.5 \le M_D < 1.625$	a=1.5625	2	3	4	5	6
(vi) M = 1.101xxxx… $1.625 \le M_D < 1.75$	a=1.6875	2	3	4	5	6
(vii) M = 1.110xxxx… $1.75 \le M_D < 1.875$	a=1.8125	2	3	4	4	6
(viii) M = 1.111xxxx… $1.875 \le M_D < 2$	a=1.9375	2	3	4	4	6

Comparison of Number of Required Terms

Comparison of
required number of terms
for different number of region division

Taylor-series expansion \ precision		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) $M_D = 1.0xxxxx\cdots$ $1 \le M_D < 1.5$	$a = 1.25$	3	5	7	8	11
(ii) $M_D = 1.1xxxxx\cdots$ $1.5 \le M_D < 2$	$a = 1.75$	2	5	6	7	10



Number of divided regions
becomes larger

Number of terms reduced

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Calculation Complexity

- In case of Taylor expansion 5 terms :

$$f_5(x) = \sqrt{a} \times \left\{1 + \frac{x-a}{2} - \frac{(x-a)^2}{8\times a} + \frac{(x-a)^3}{16\times a^2} - \frac{5\times(x-a)^4}{128\times a^3}\right\}$$
$$= \alpha_0[2 + (x-a)] - \alpha_2(x-a)^2 + \alpha_3(x-a)^3 - \alpha_4(x-a)^4$$

Here, $\alpha_0 = \frac{\sqrt{a}}{2}, \alpha_2 = \frac{\sqrt{a}}{8\times a}, \alpha_3 = \frac{\sqrt{a}}{16\times a^2}, \alpha_4 = \frac{5\sqrt{a}}{128\times a^3}$.

***E* (exponent) Even case.**

$$g_5(x) = \sqrt{2} \times f_5(x)$$
$$= \beta_0[2 + (x-a)] - \beta_2(x-a)^2 + \beta_3(x-a)^3 - \beta_4(x-a)^4$$

Here, $\beta_0 = \sqrt{2}\alpha_0,\ \beta_2 = \sqrt{2}\alpha_2,\ \beta_3 = \sqrt{2}\alpha_3,\ \beta_4 = \sqrt{2}\alpha_4$.

***E* (exponent) Odd case.**

$\alpha_0, \alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_2, \beta_3, \beta_4$ values : Stored in LUT and read.

y = x−a Subtraction: 1 time

z = y^2 Multiplication: 1 time

$f_5(x) = \alpha_0(2 + y) - z(\alpha_2 - \alpha_3 y + \alpha_4 z)$

Multiplication: 4 times
Addition / Subtraction: 4 times

Total : Multiplication: 5 times
Addition / Subtraction: 5 times

Number of Operations

Number of terms versus number of operations in Taylor expansion

Taylor expansion of $f(x) = \sqrt{x}$ can be calculated with a small number of Mul/Add/Sub operations.

Terms of Taylor expansion	multiplication	Addition or subtraction
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8

In case of 5 terms

LUT Size

$$f_5(x) = \sqrt{a}\times\left\{1+\frac{x-a}{2}-\frac{(x-a)^2}{8\times a}+\frac{(x-a)^3}{16\times a^2}-\frac{5\times(x-a)^4}{128\times a^3}\right\}$$
$$= \alpha_0[2+(x-a)]-\alpha_2(x-a)^2+\alpha_3(x-a)^3-\alpha_4(x-a)^4$$

E (exponent) Even case.

Here, $\alpha_0=\frac{\sqrt{a}}{2}, \alpha_2=\frac{\sqrt{a}}{8\times a}, \alpha_3=\frac{\sqrt{a}}{16\times a^2}, \alpha_4=\frac{5\sqrt{a}}{128\times a^3}$.

$$g_5(x) = \sqrt{2}\times f_5(x)$$
$$= \beta_0[2+(x-a)]-\beta_2(x-a)^2+\beta_3(x-a)^3-\beta_4(x-a)^4$$

E (exponent) Odd case.

Here, $\beta_0=\sqrt{2}\alpha_0,\ \beta_2=\sqrt{2}\alpha_2,\ \beta_3=\sqrt{2}\alpha_3,\ \beta_4=\sqrt{2}\alpha_4$.

$\alpha_0, \alpha_2, \alpha_3, \alpha_4, \beta_0, \beta_2, \beta_3, \beta_4$ values : Stored in LUT and read.

4 - region case → LUT size of 32 words

Address (M=1.*ab*⋯)	LUT data
00	$\alpha_0, \alpha_2, \alpha_3, \alpha_4,\ \beta_0, \beta_2, \beta_3, \beta_4$ for $a = 1.125$
01	$\alpha_0, \alpha_2, \alpha_3, \alpha_4,\ \beta_0, \beta_2, \beta_3, \beta_4$ for $a = 1.357$
10	$\alpha_0, \alpha_2, \alpha_3, \alpha_4,\ \beta_0, \beta_2, \beta_3, \beta_4$ for $a = 1.625$
11	$\alpha_0, \alpha_2, \alpha_3, \alpha_4,\ \beta_0, \beta_2, \beta_3, \beta_4$ for $a = 1.875$

Outline

- Research Background and Objective
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Conclusion

- Square root calculation of mantissa in binary floating format using Taylor expansion has been investigated.
- Proposed mantissa region division method can effectively control the number of Taylor series expansion terms.
- Number of divided regions becomes larger :
  - ➢Number of Taylor expansion terms ➡ smaller
  - ➢LUT size ➡ larger.

Divide-and-Conquer is effective.

Thank you for listening !

Comment

Form Prof. Robert Brennan

https://schulich.ucalgary.ca/contacts/robert-brennan

Comment:

1. It would be more efficient to calculate 1/sqrt(x) using fixed point iteration followed by multiplying by x. This would converge in 2 iterations.
2. you need 4 bit accuracy for convergence to 16 bits. A second order polynomial will give you that. Global convergence is actually excellent given a 4-bit accurate input. Derivative is well behaved.