2021 IEEE 14th International Conference on ASICON

Oct. 27, 2021

Divide and Conquer: Floating-Point Exponential Calculation Arithmetic Based on Taylor-Series Expansion

Jianglin Wei, A. Kuwana, H. Kobayashi

K. Kubo

Division of Electronics and Informatics, Gunma University

Oyama National College of Technology

Japan

Outline

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

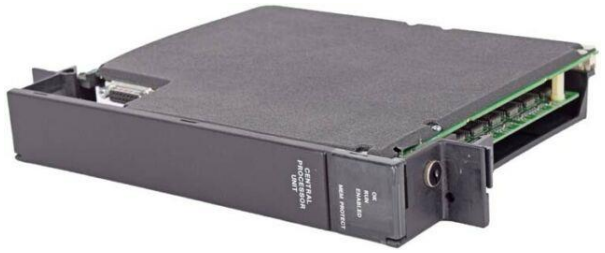
Outline

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Results
- Hardware Implementation Consideration
- Conclusion

Research Background

◆High-speed high-precision floating-point arithmetic

→ Scientific computing
Embedded systems
Mobile applications



Floating-Point Processor

◆Exponential exp(x) calculation ⇒ Very tough !

Multiple usage of basic arithmetic operations

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

Research Objective

- Floating-point exponential calculation
  - Simple circuit
  - High-speed

Divide the difficulties.

René Descartes

- 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

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Taylor Series Expansion

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

Function $f(x)$ for $x = a$

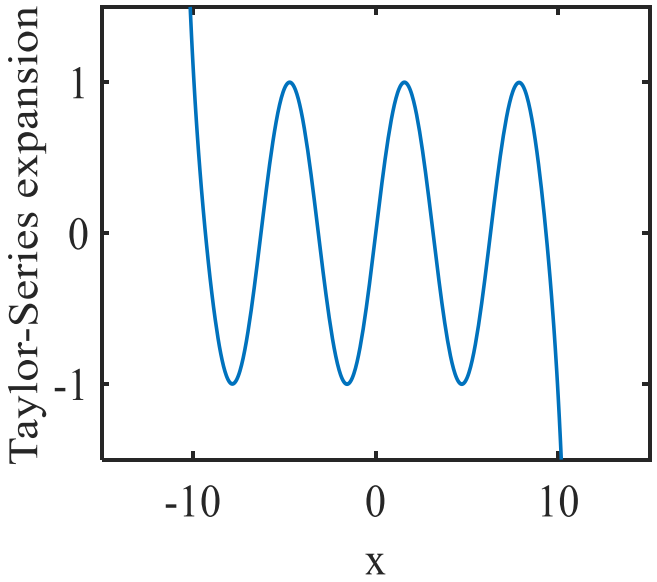
$$f(x) = f(a) + f'(a)(x-a) + \frac{(f)''(a)}{2!}(x-a)^2 + \cdots + \frac{(f)^n(a)}{n!}(x-a)^n + \cdots$$

Convergence range $\alpha < x < \beta$

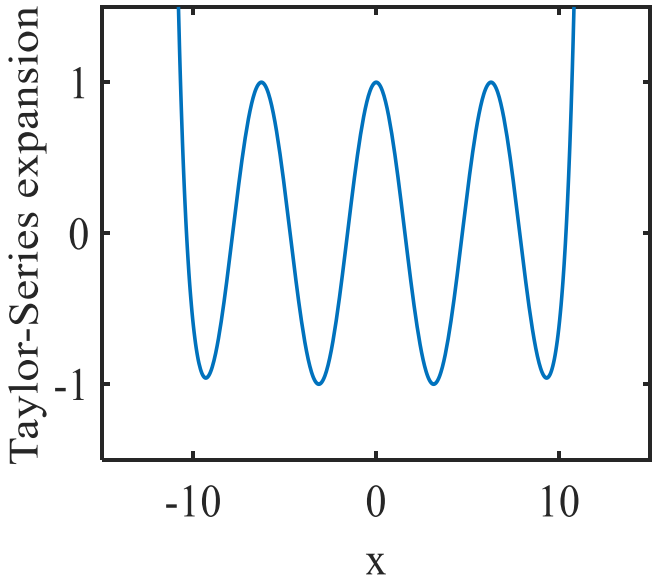
Taylor-Series of sin(x), cos(x)

Taylor-series expansion with 20 terms.

for center value : a=0



$\sin(x)$



$\cos(x)$

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

Outline

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Normalized Floating-Point Representation

Floating-point representation in binary :

Mantissa : M $(1 \leqq M < 2)$ Exponent : E

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

Decimal point

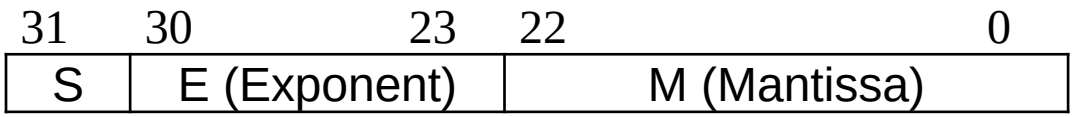
$$1.abcdef\cdots \times 2^{E}$$

Mantissa Exponent

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

IEEE-754 standard

- Half-precision 16-bit
- Single-precision 32-bit
- Double-precision 64-bit



IEEE-754 single-precision floating-point format

Exponential Calculation

Floating-point binary : $\mathrm{X} = M \times 2^{E}$

Exponential calculation of X.

$$EXP = exp(X) = exp(M \times 2^{E})$$

$$(exp(M))^{2^{E}}$$

$exp(M)$ calculation by Taylor-series expansion for specified accuracy.

Analysis of Taylor Expansion

Calculate exponential of mantissa : $exp(M)$ $(1 \leq M < 2)$

$x = M$

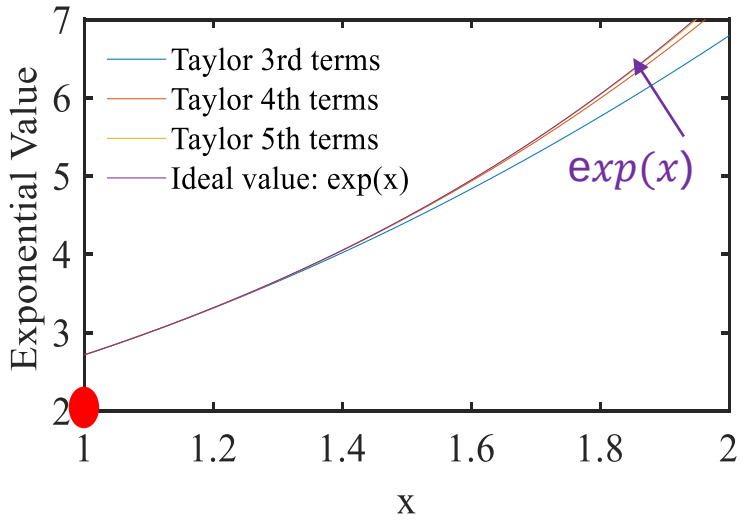
$f(x) = exp(x)$ by Taylor expansion at $x = a$ $(1 \leq a < 2)$

$$f(x) = exp(a) \times \left\{1 + q + \frac{1}{2}q^2 + \boxed{\frac{1}{6}}q^3 + \boxed{\frac{1}{24}}q^4 + \boxed{\frac{1}{120}}q^5 + \boxed{\frac{1}{720}}q^6 + \cdots\right\}$$

$q = x - a.$

Coefficient values are stored in LUT in advance.

Mantissa Region Division



Taylor series expansion of $exp(x)$ at center value a = 1

Divide and Conquer Method

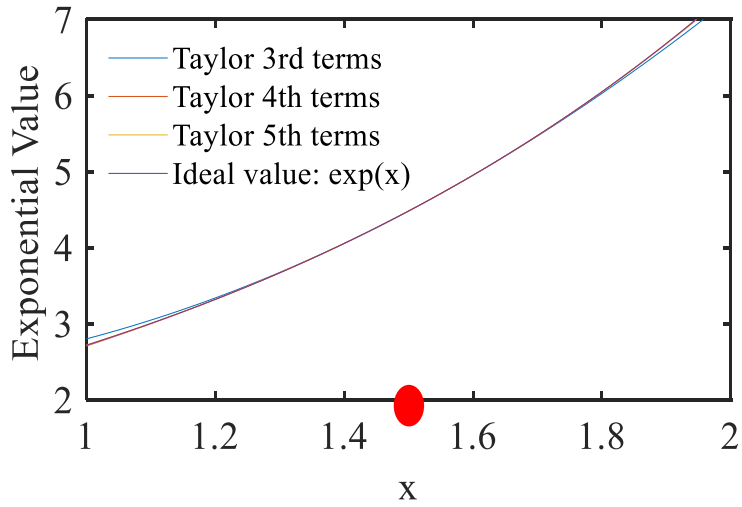
1 region :

a = 1.5 $\quad 1 \leq x < 2$

2 regions :

a = 1.25 $\quad 1 \leq x < 1.5$

a = 1.75 $\quad 1.5 \leq x < 2$



4 regions :

a = 1.125 $\quad 1 \leq x < 1.25$

a = 1.375 $\quad 1.25 \leq x < 1.5$

a = 1.625 $\quad 1.5 \leq x < 1.75$

a = 1.875 $\quad 1.75 \leq x < 2$

⋮

Taylor series expansion of $exp(x)$ at center value a = 1.5

Outline

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Definition of Accuracy

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

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

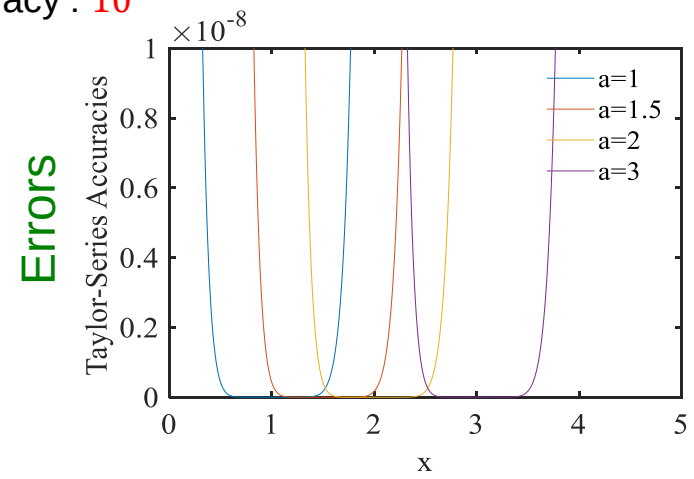
$f(x)$: Original function

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

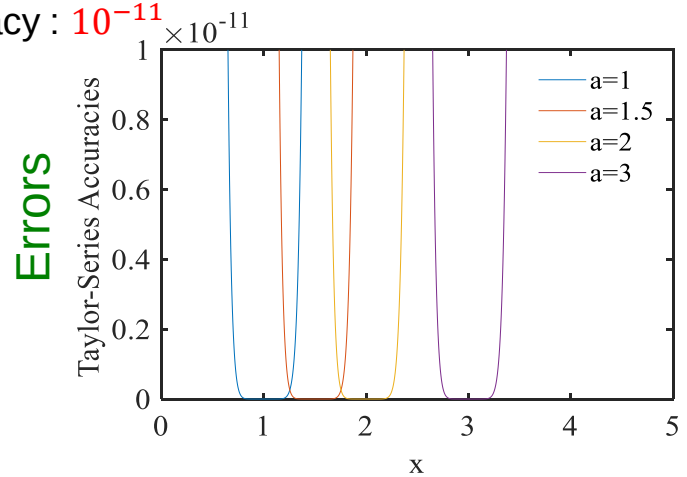
Accuracy of $exp(x)$ Taylor Expansion

Number of Taylor expansion terms : 10

Accuracy : 10^{-8}

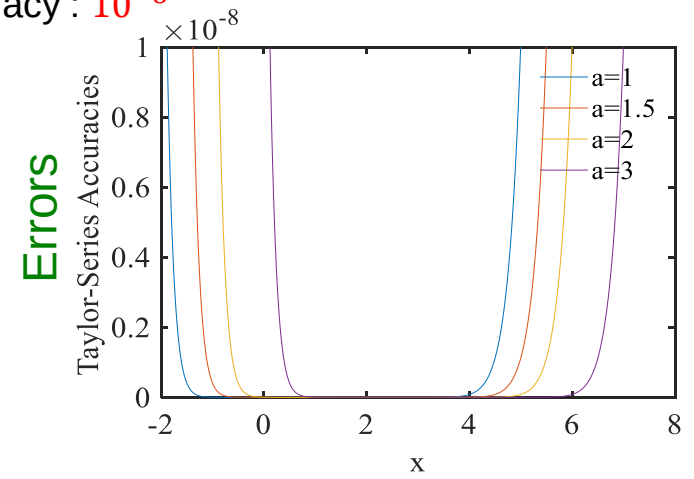


Accuracy : 10^{-11}

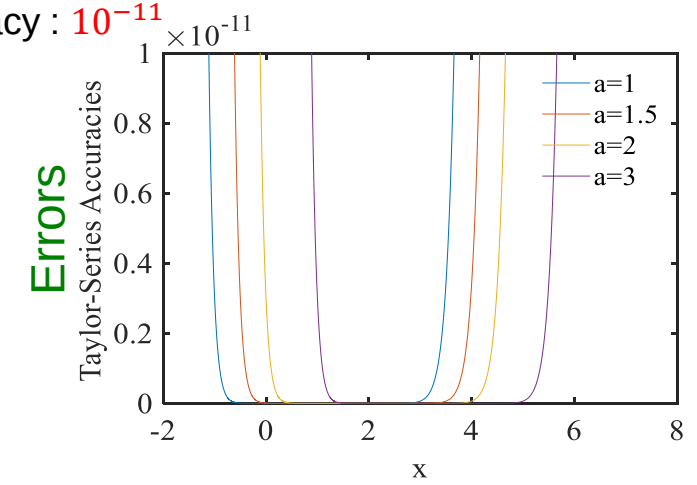


Number of Taylor expansion terms : 20

Accuracy : 10^{-8}



Accuracy : 10^{-11}



One-Region Case

Use Taylor series expansion equation :

$f(x) = exp(x)$ $\quad (1 \leq x < 2)$

Mantissa represented by binary decimal point.

Specified accuracy

Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M = 1.xxxxxx… $1 \leq M < 2$	$a = 1.5$	4	7	8	9	11

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 :

$f(x) = exp(x) \quad (1 \leq x < 2)$

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

Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M_D = 1.0xxxxx… $1 \leq M_D < 1.5$	$a = 1.25$	3	5	6	7	9
(ii) M_D = 1.1xxxxx… $1.5 \leq M_D < 2$	$a = 1.75$	3	5	6	7	9

Four-Region Case

Use Taylor series expansion equation :

$f(x) = exp(x) \quad (1 \leq x < 2)$

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

Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M = 1.00xxxx… $1 \leq M_D < 1.25$	a=1.125	3	4	5	6	7
(ii) M = 1.01xxxx… $1.25 \leq M_D < 1.5$	a=1.375	3	4	5	6	7
(iii) M = 1.10xxxx… $1.5 \leq M_D < 1.75$	a=1.625	3	4	5	6	7
(iv) M = 1.11xxxx… $1.75 \leq M_D < 2$	a=1.875	3	4	5	6	7

Eight-Region Case

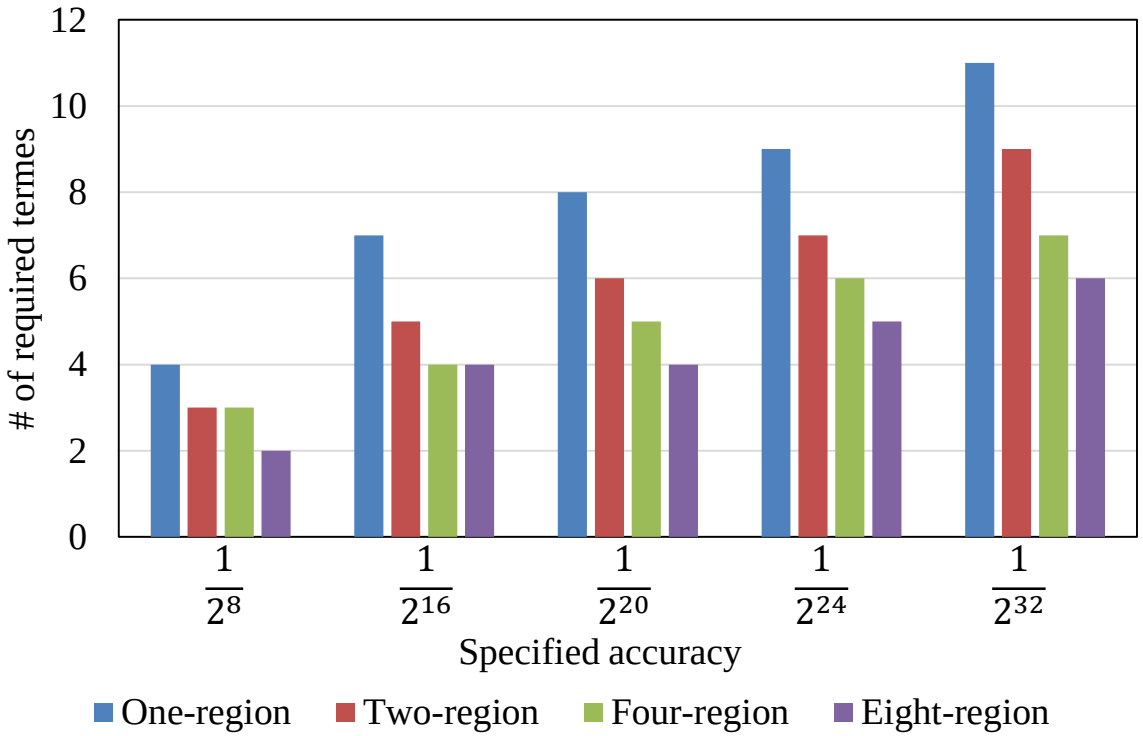
Check the values (000, 001,…, 111)
of the first three decimal places of Mantissa.

Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) M = 1.000xxxx… $1 \le M_D < 1.125$	a=1.0625	2	4	4	5	6
(ii) M = 1.001xxxx… $1.125 \le M_D < 1.25$	a = 1.1875	2	4	4	5	6
(iii) M = 1.010xxxx… $1.25 \le M_D < 1.375$	a=1.3125	2	4	4	5	6
(iv) M = 1.011xxxx… $1.375 \le M_D < 1.5$	a=1.4375	2	4	4	5	6
(v) M = 1.100xxxx… $1.5 \le M_D < 1.625$	a=1.5625	2	4	4	5	6
(vi) M = 1.101xxxx… $1.625 \le M_D < 1.75$	a=1.6875	2	4	4	5	6
(vii) M = 1.110xxxx… $1.75 \le M_D < 1.875$	a=1.8125	2	4	4	5	6
(viii) M = 1.111xxxx… $1.875 \le M_D < 2$	a=1.9375	2	4	4	5	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⋯ $1 \leq M_D < 1.5$	$a = 1.25$	3	5	6	7	9
(ii) M_D = 1.1xxxxx⋯ $1.5 \leq M_D < 2$	$a = 1.75$	3	5	6	7	9



Number of divided regions
becomes larger

Number of terms reduced

Exponential Calculation in Different Ranges

$-2 \leq x < -1$ case:

Use Taylor series expansion equation : $f(x) = \exp(x)$ $(-2 \leq x < -1)$

$-2 \leq x < -1$ in One-region case						
Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) $M = 1.xxxxxx\cdots$ $-2 \leq M < -1$	$a = -1.5$	4	7	8	9	10

$0.5 \leq x < 1$ case:

Use Taylor series expansion equation : $f(x) = \exp(x)$ $(0.5 \leq x < 1)$

$0.5 \leq x < 1$ in One-region case						
Taylor-series expansion \ Accuracy		$\frac{1}{2^8}$	$\frac{1}{2^{16}}$	$\frac{1}{2^{20}}$	$\frac{1}{2^{24}}$	$\frac{1}{2^{32}}$
(i) $M = 1.xxxxxx\cdots$ $-2 \leq M < -1$	$a = 0.75$	3	5	6	7	9

Outline

- Motivation
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Calculation Complexity

- In case of Taylor expansion 5 terms :

$$f_5(x) = exp(a) \times \left\{1 + (x-a) + \frac{(x-a)^2}{2} + \frac{(x-a)^3}{6} + \frac{(x-a)^4}{24}\right\}$$

◆ $exp(a)$ values : Stored in LUT and read.

y = x−a Subtraction: 1 time z = y^2 Multiplication: 1 time

$$f_5(x) = \exp(a) \times \left(1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \frac{y^4}{24}\right)$$

$$= \exp(a) \times \left\{1 + y + \frac{z}{2} \times \left(1 + \frac{y}{3} + \frac{z}{12}\right)\right\}$$

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 $f(x) = exp(x)$ can be calculated with a small number of Mul/ Add/Sub operations.

Terms of Taylor expansion	Multiplication	Addition or subtraction
3	3	3
4	5	4
5	6	5
6	8	6
7	9	7
8	10	8

In case of 5 terms

LUT Size

$$f_5(x) = \boxed{exp(a)} \times \left\{1 + (x-a) + \frac{(x-a)^2}{2} + \frac{(x-a)^3}{6} + \frac{(x-a)^4}{24}\right\}$$

Stored in LUT

LUT : Look-Up Table

4 - region case ⟶ LUT size of **4** words

Address (M=1.*ab*⋯)	LUT data
00	$Exp(a)$ for a = 1.125
01	$Exp(a)$ for a = 1.357
10	$Exp(a)$ for a = 1.625
11	$Exp(a)$ for a = 1.875

Outline

- Motivation
- Taylor-Series Expansion
- Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration
- Conclusion

Conclusion

- Exponential calculation of mantissa in binary floating using Taylor expansion has been investigated.

- Number of divided mantissa regions becomes larger

  ⬇

  ➢ Number of Taylor expansion terms ➡ smaller

  ➢ LUT size ➡ larger

  ⬇

  Optimal hardware configuration

Final Statement

Proposal of Taylor-series

Brook Taylor

English mathematician

1685 – 1731

Thought of Divide-and-Conquer

Yú Yuè (俞 樾)

Qing dynasty scholar

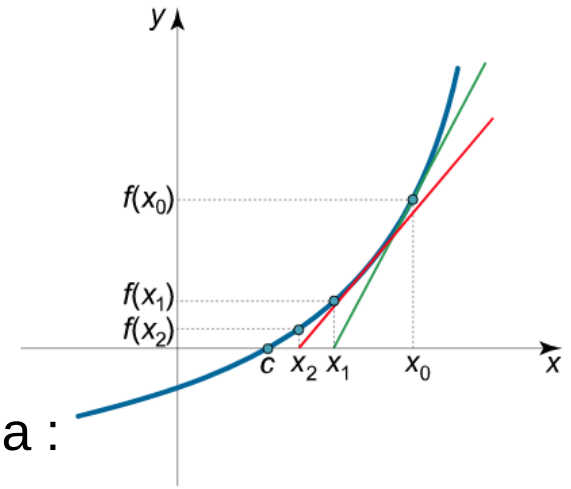
1821 - 1907

Thank you for listening !

Appendix

Newton's method

Newton's method step:



Fist, Start with an initial approximation x_0 close to c.

Second, Determine the next approximation by the formula :

$$x_1 = x_0 - \frac{f(x_0)}{f(\acute{x}_0)}$$

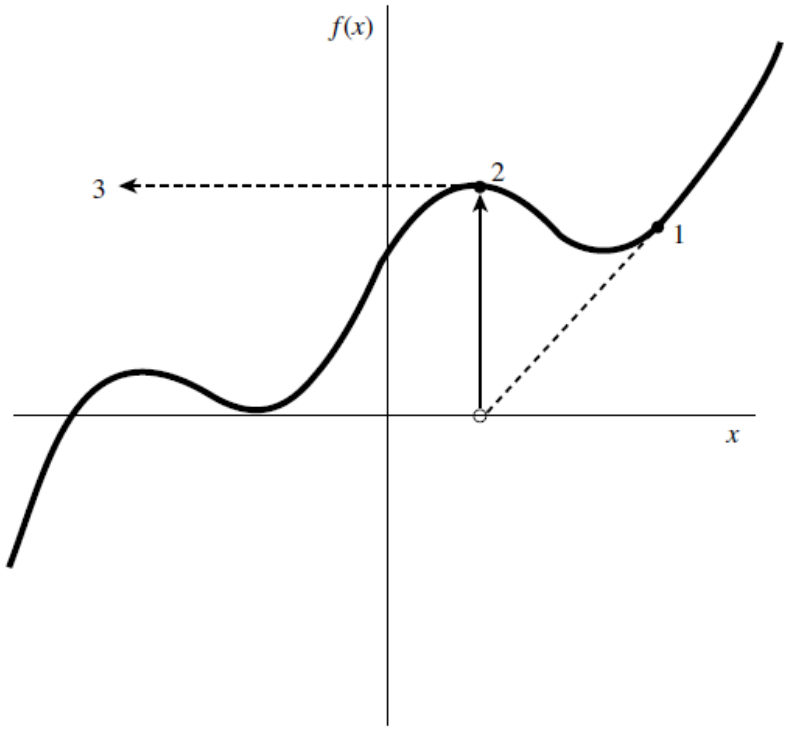
Third, Continue the iterative process using the formula :

$$x_{n+1} = x_n - \frac{f(x_n)}{f(\acute{x}_n)}$$

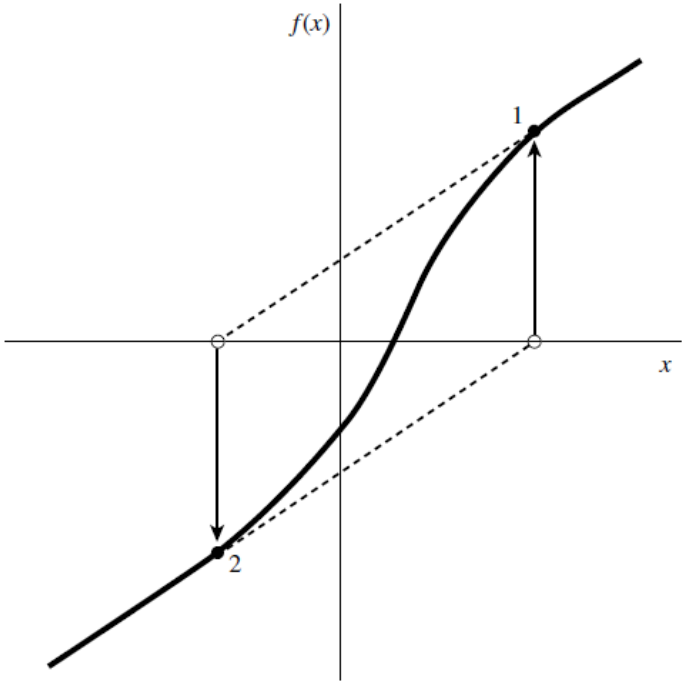
Last, until the root is found to the desired accuracy.

- Poor global convergence properties
- Dependent on initial guess
  - May be too far from local root
  - May encounter a zero derivative
  - May loop indefinitely

Examples of disadvantages



On the left, we have Newton’s Method finding a local maxima, in such cases the method will shoot off into negative infinity.



Newton's Method has entered an infinite cycle. Better initial guesses may be able to alleviate this problem.

Another Decimal Point Position

Change the decimal point position of the mantissa

Mantissa: M Exponent: E

$M \times 2^E$

Original decimal point

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

Mantissa Exponent

$1 \leqq M < 2$

$M \times 2^E$

New decimal point

$0.1abcdef \cdots \times 2^E$

Mantissa Exponent

$0.5 \leqq M < 1$

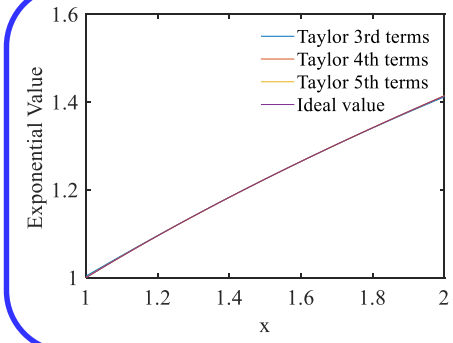
Ex :1011001 (binary) = 89 (decimal)

Binary representation : 0.1011001×2^{111}

Decimal representation : $0.6953125 \times 2^7 = 89$

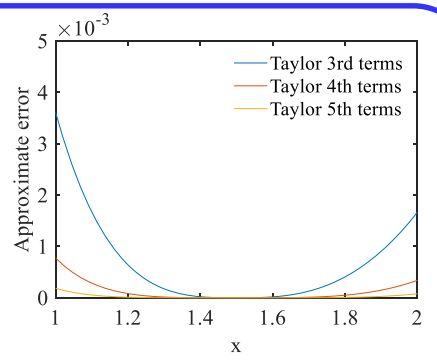
One and Two region cases

One region case

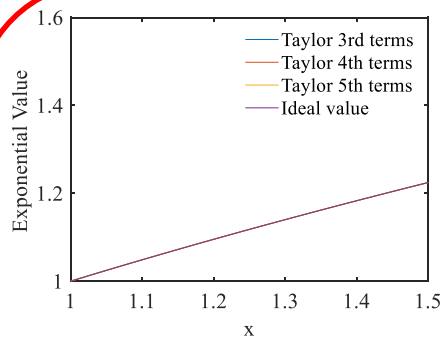


Center value a=1.5

$1 \leq x < 2$

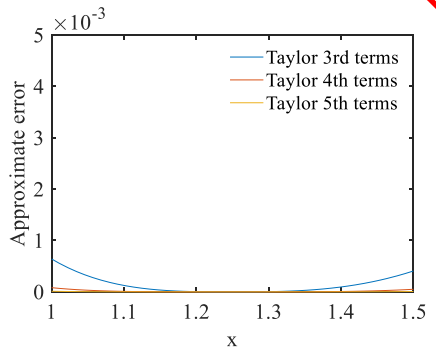


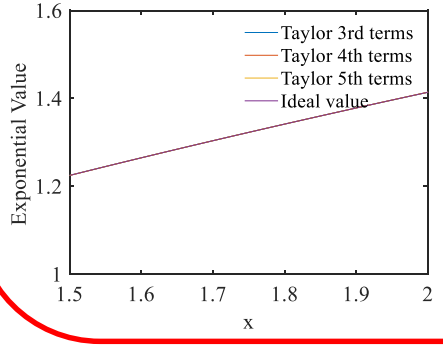
Two regions case



Center value a=1.25

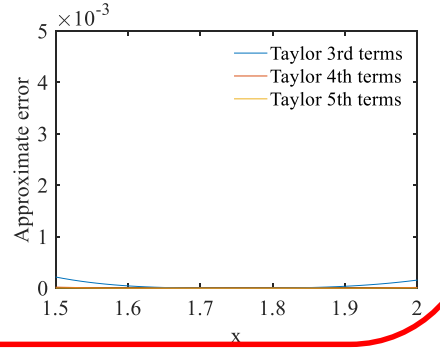
$1 \leq x < 1.5$





Center value a=1.75

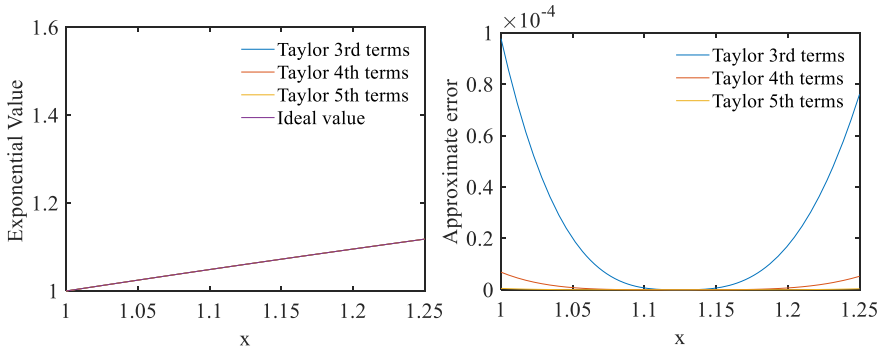
$1.5 \leq x < 2$



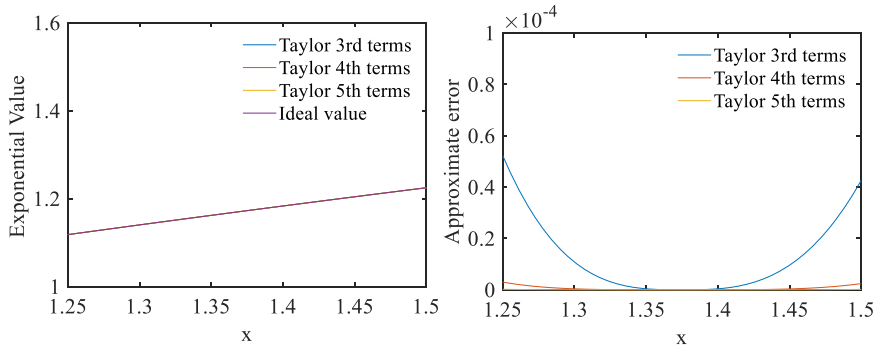
Approximate error = $\frac{f(x)-t(n,x)}{f(x)}$

Taylor series expansion of $\sqrt{x}$.

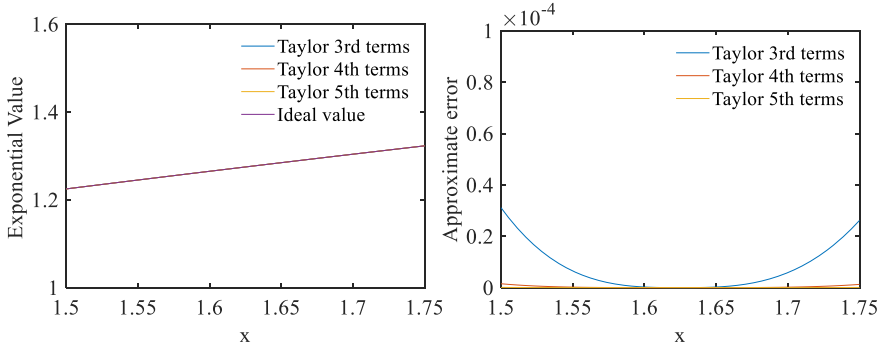
4 regions case



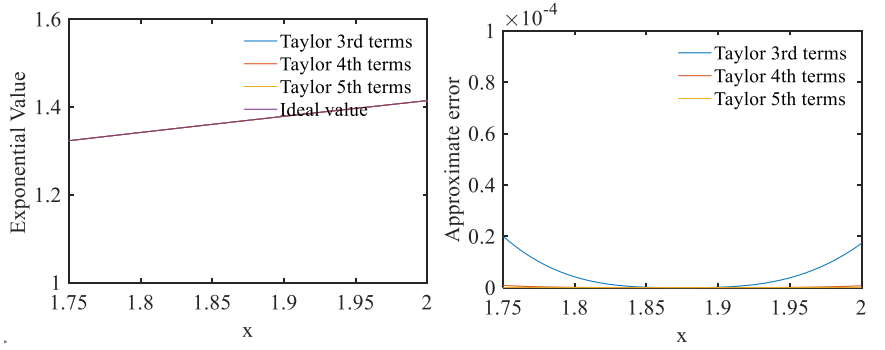
Center value a=1.125 $1 \leq x < 1.25$



Center value a=1.375 $1.25 \leq x < 1.5$



Center value a=1.625 $1.5 \leq x < 1.75$



Center value a=1.875 $175 \leq x < 2$

Taylor series expansion of $\sqrt{x}$.

Number of Taylor expansion terms：200