

# ATS Doctoral Thesis Award



Virtual Event Hosted by Japan, Nov. 22-24, 2021

Semi-Final of 2022 TTTC's E. J. McCluskey Doctoral Thesis Award

---

## $\Delta\Sigma$ ADC Linearity Testing Technology and Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

Student: **Jiang-Lin Wei**

Supervisor: Prof. **Haruo Kobayashi**

Division of Electronics and Informatics,  
Gunma University, Japan



# Outline

---

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# Outline

---

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# Research Background

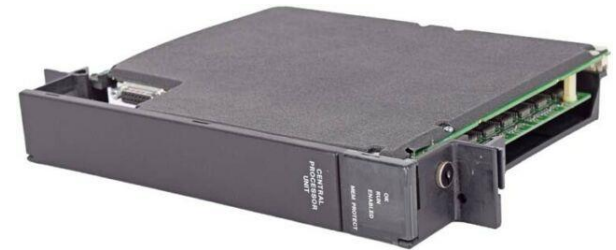
- IoT devices are becoming important.
- Their high quality and low cost testing is required.
- High-speed, high-precision floating-point DSP in ATE systems.



Efficient algorithms are required.



ATE (Automatic Test Equipment)



Floating-point processor

# Research Objective (1)

## High resolution, low speed $\Delta\Sigma$ ADC

- Sensor interface
- Mass production test

**Linearity test takes a long time.**

→ Usually omitted.



High reliability requirements



- ✓ Its linearity test in short time
- ✓ Develop its algorithm

# Research Objective (2)

## ◆ Floating-point **arithmetic** calculation

➤ **Simple circuit**

➤ **High-speed**

Divide 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.

**LUT** : Look-Up Table

# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# $\Delta\Sigma$ ADC Testing Challenge

**Sensor + amplifier +  $\Delta\Sigma$ ADC + microcomputer**

- **Complicated ADC output signal processing** → **Reproduction circuit**
- **Low speed sampling, High resolution** → **Long test time**
- **High linearity analog input signal** → **High precision signal generator**

**1** US dollar chip

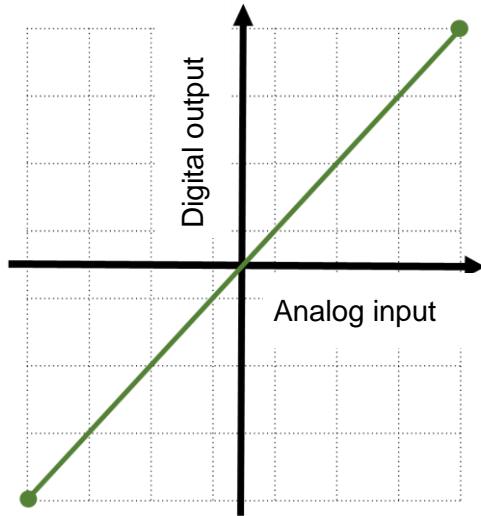


Test time should be less than **1** second

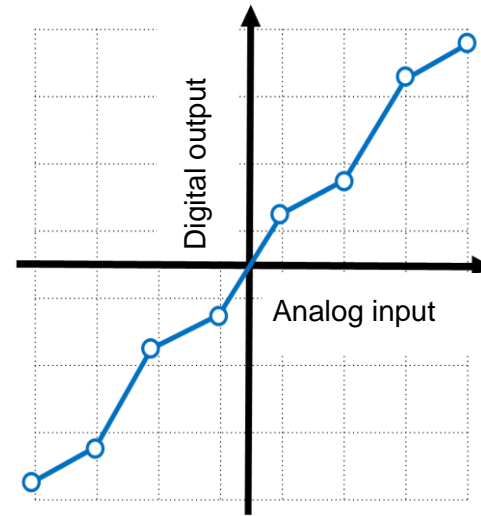


# Linearity of $\Delta\Sigma$ ADC

Ideal characteristic (**linear**)



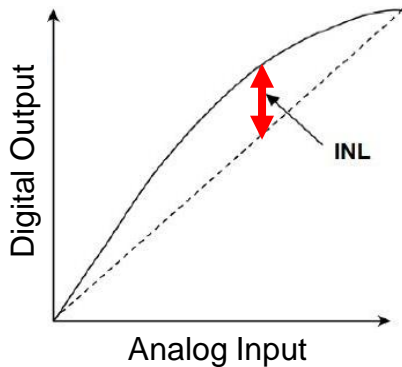
Actual (**nonlinear**)



Circuit imperfection, variation



**Nonlinear**



**Large INL**



- Missing codes
- Lack of monotonicity

**INL**: Integral Non-Linearity

# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

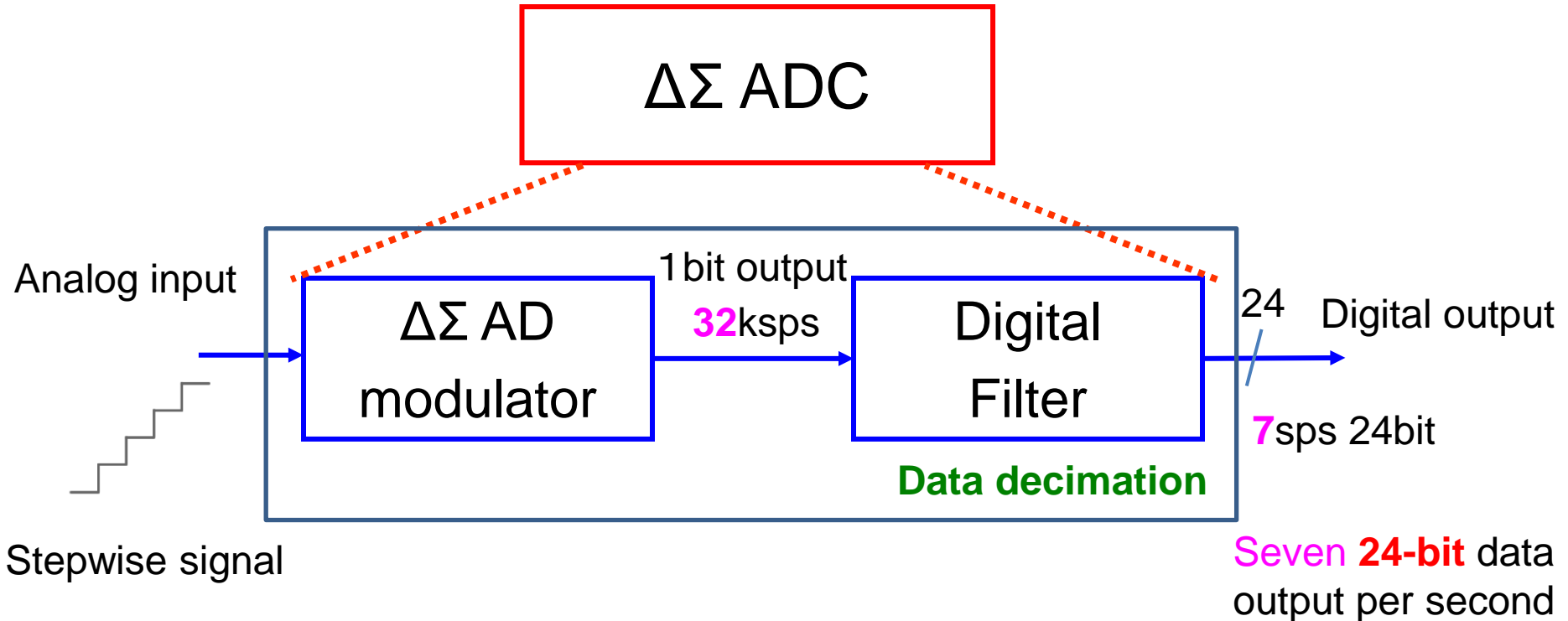
- $\Delta\Sigma$  ADC Testing challenge and Linearity
- **Proposed linearity test method**
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# Problem of Direct Linearity Test



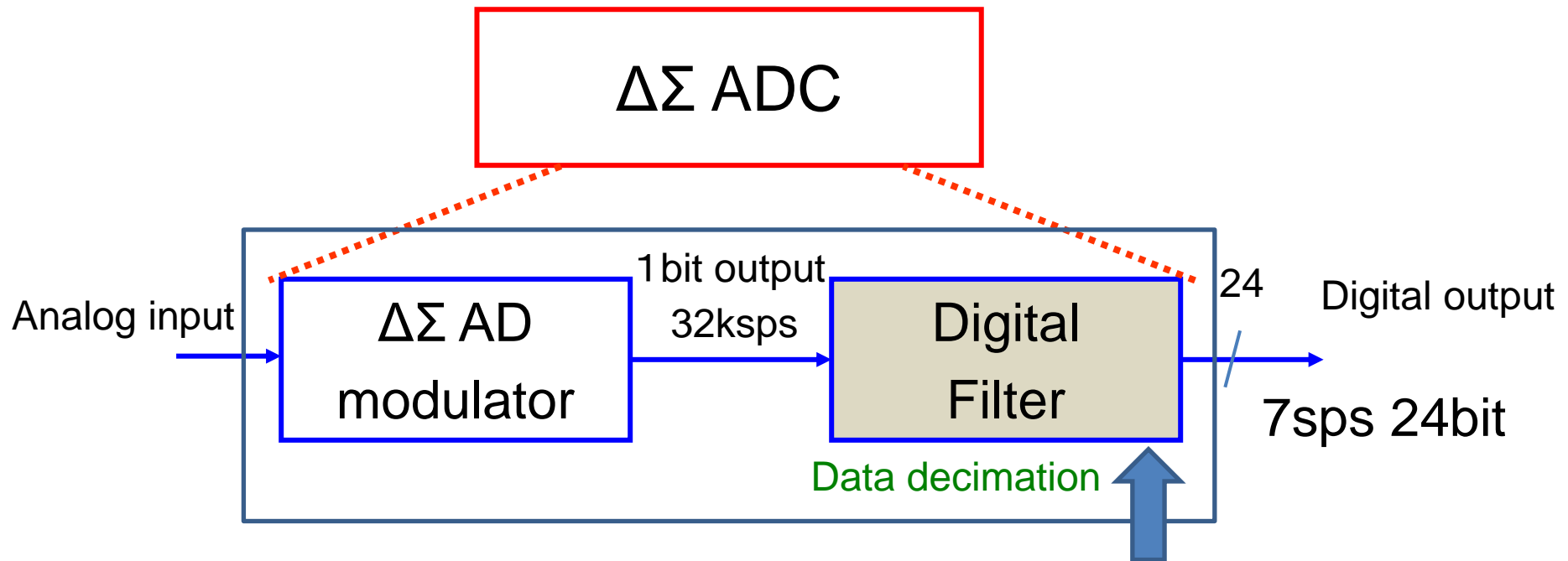
## Linearity test time:

4 point per code  
 $(1/7) \times 2^{24} \times 4$  seconds = 104 day



**Totally unrealistic**

# Proposed: Digital Filter Test



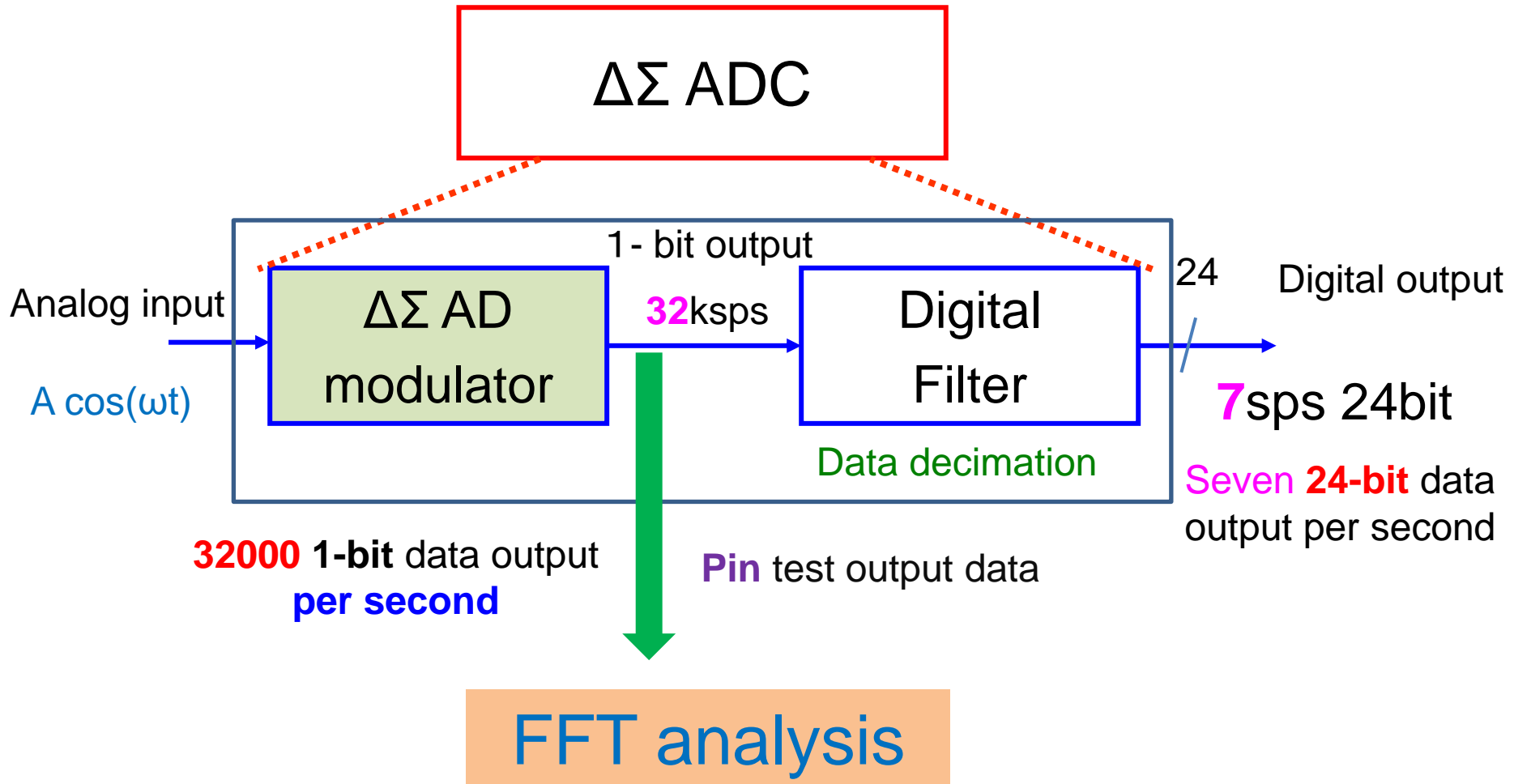
Digital filter does **NOT** affect the **linearity**.

Only **pass** or **fail**.

Test with  
**scan path** method

# Proposed: $\Delta\Sigma$ AD Modulator Test

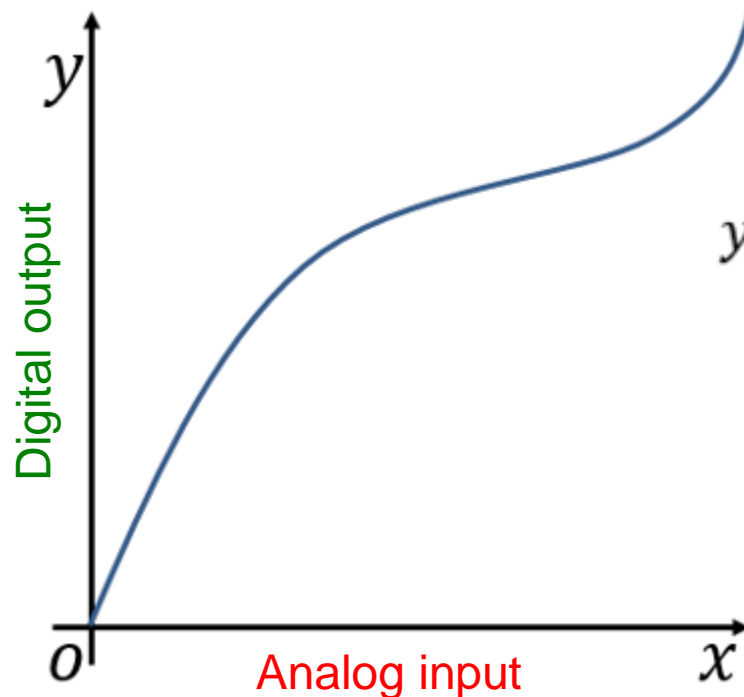
## Cosine Input & FFT Analysis



# I/O Characteristic Modeling of Modulator

## Modeling by polynomial approximation

- ✓ **Assumption:** I/O characteristics are continuous in the AD modulator.



$$y = a_0 + a_1x + a_2x^2 + \dots + a_nx^n$$

**3<sup>rd</sup> order model** for simplicity :

$$y(t) = a_1x(t) + a_3x(t)^3$$

# Polynomial Coefficient Estimation

Analog cosine input :

$$x(t) = A \cos(\omega t)$$



Modulator **1-bit** output stream

**FFT**

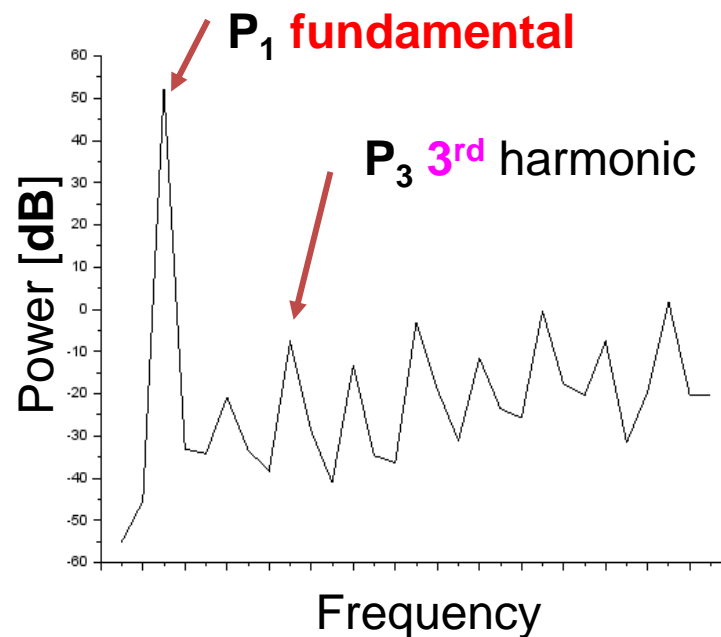


Measure **fundamental** &  
**3<sup>rd</sup> harmonic** power



Estimate  **$a_1$** ,  **$a_3$**  :

$$y(t) = a_1 x(t) + a_3 x(t)^3$$



# Fundamental / HD3 and Polynomial Coefficients

Cosine input :

$$x(t) = A \cos \omega t$$

Output characteristic model :

$$y(t) = a_1 x(t) + a_3 x(t)^3$$

$$y(t) = a_1 A \cos \omega t + a_3 (A \cos \omega t)^3$$



$$(a_1 A + \frac{3}{4} a_3 A^3) \cos \omega t + \frac{1}{4} a_3 A^3 \cos 3\omega t$$



**Fundamental**

$$a_1 A + \frac{3}{4} a_3 A^3$$



**HD3**

$$\frac{1}{4} a_3 A^3$$



# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- **Simulation configuration and results**

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

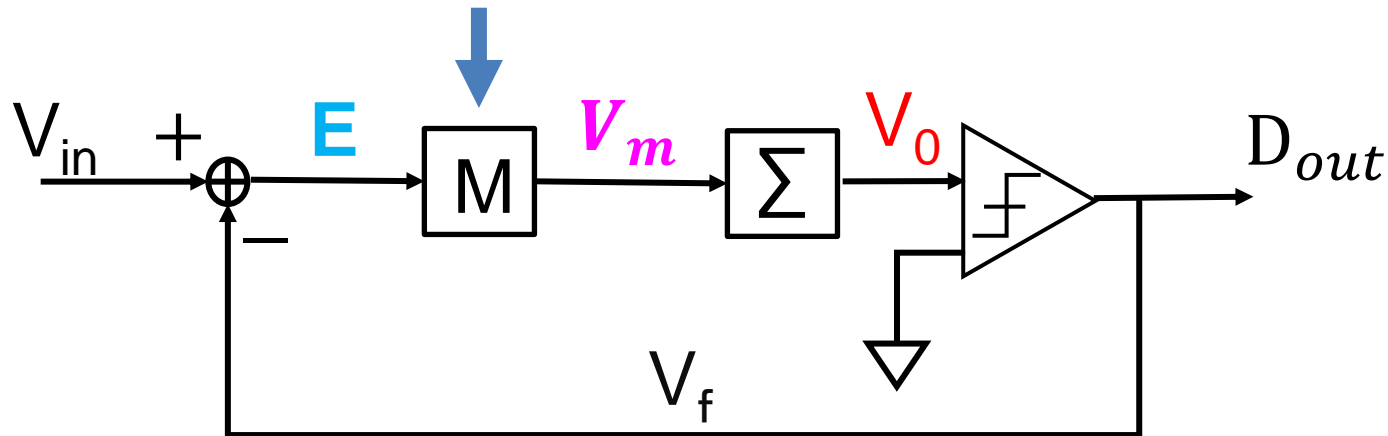
- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# Simulation Model

**3<sup>rd</sup>-order** nonlinearity model

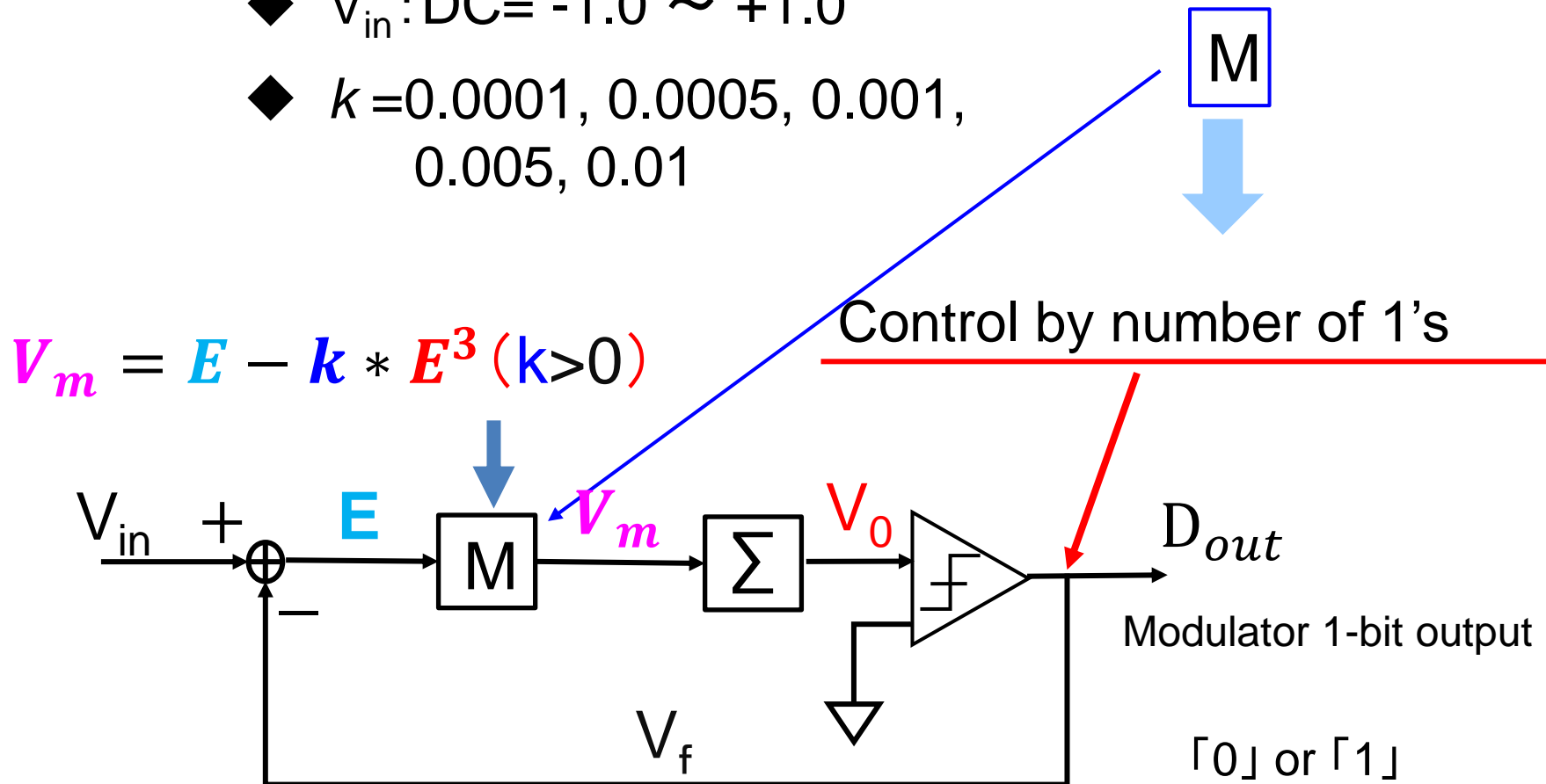
$$V_m = E - k * E^3 \quad (k > 0)$$



**1<sup>st</sup>-order** modulator

# DC Input Simulation

- ◆ Number of data :  $N=2^{20}$
- ◆  $V_{in}$ : DC = -1.0 ~ +1.0
- ◆  $k=0.0001, 0.0005, 0.001, 0.005, 0.01$



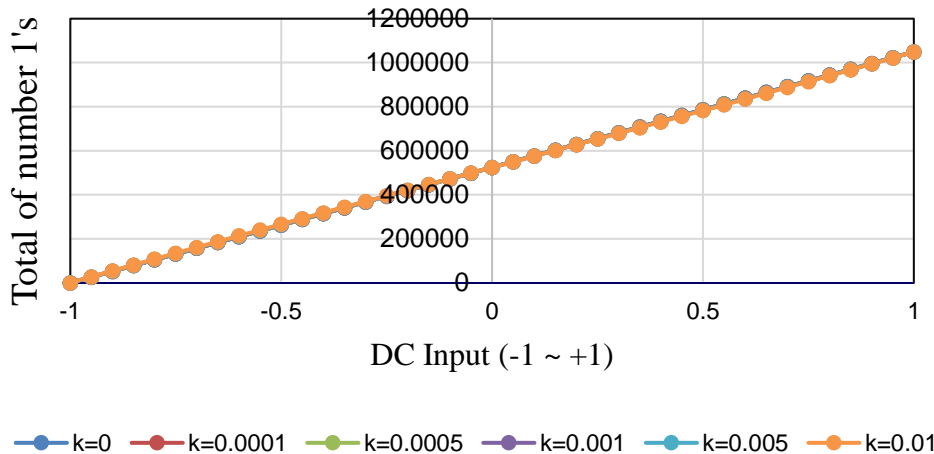
**1<sup>st</sup> -order** modulator

# DC Input Simulation Result

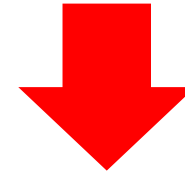
Number of modulator output:

$$N = 2^{20}$$

Number of output 1's

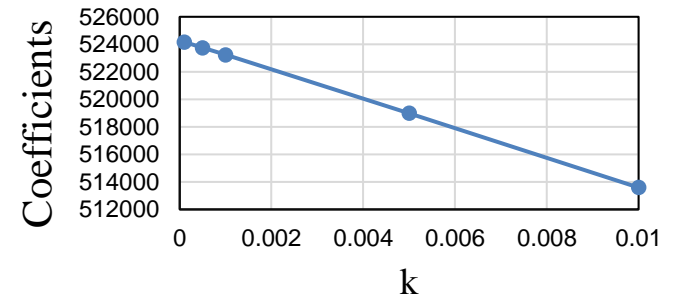


DC characteristic curve fitting

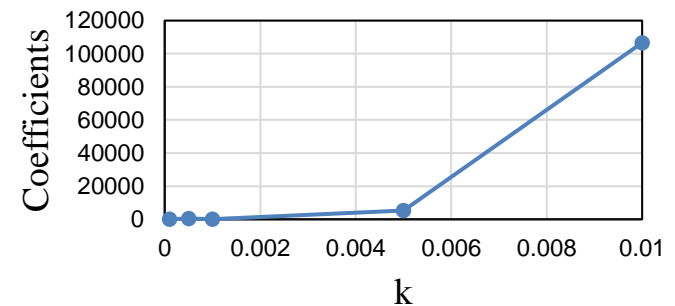


$$y = a_3 \times x^3 + a_1 \times x$$

$a_1$



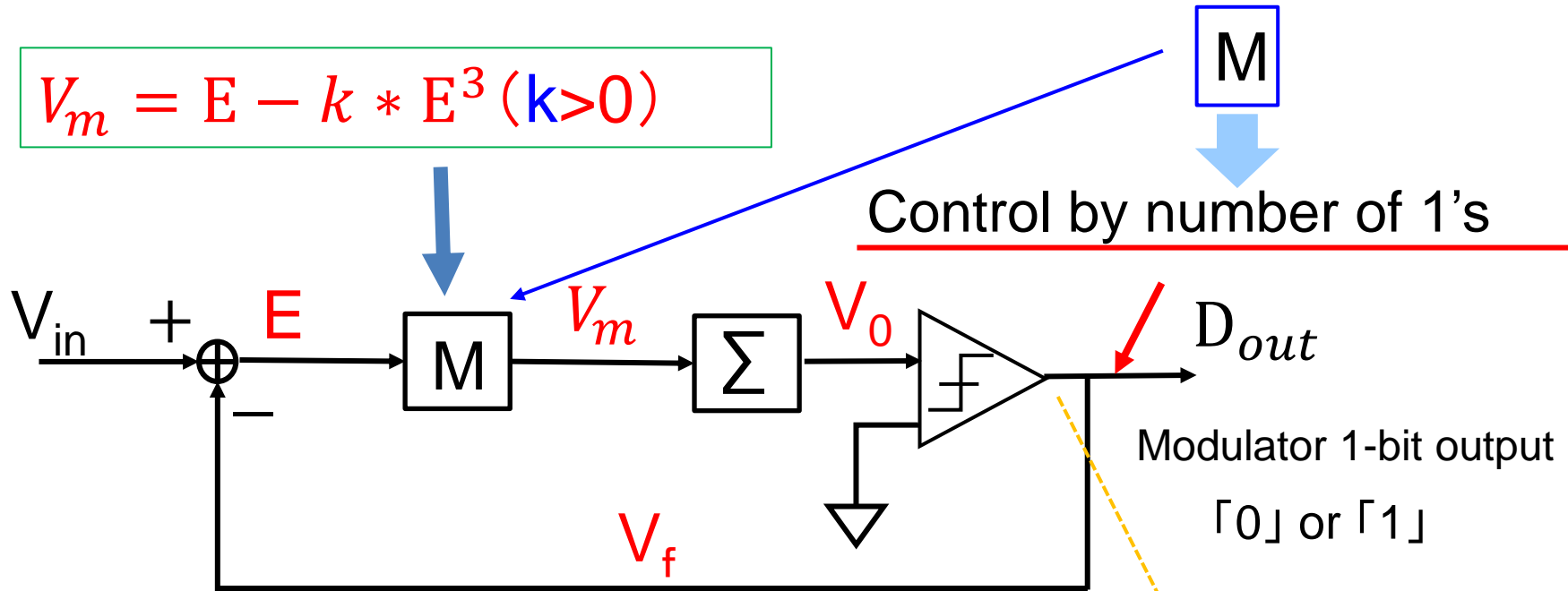
$a_3$



| $k$    | $a_3$   | $a_1$  |
|--------|---------|--------|
| 0.0001 | 104.84  | 524180 |
| 0.0005 | 524.48  | 523760 |
| 0.0010 | 1050.5  | 523240 |
| 0.0050 | 5282.5  | 519000 |
| 0.0100 | 10643.0 | 513610 |

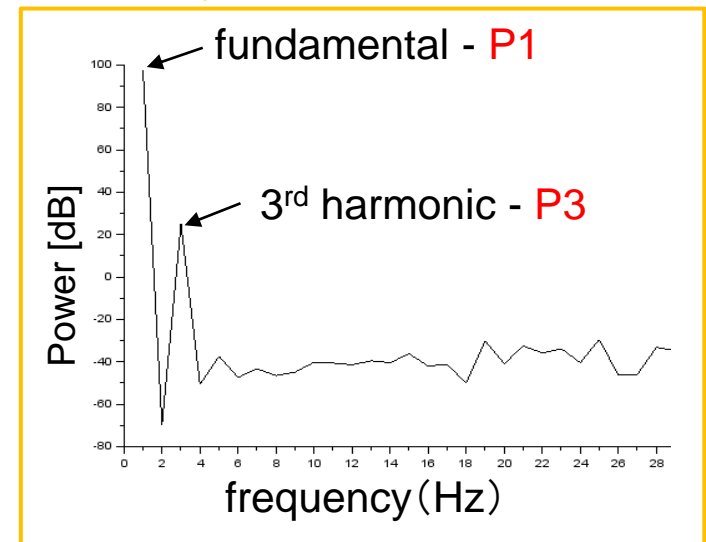
# Cosine Input Simulation Configuration

$$V_m = E - k * E^3 \quad (k > 0)$$

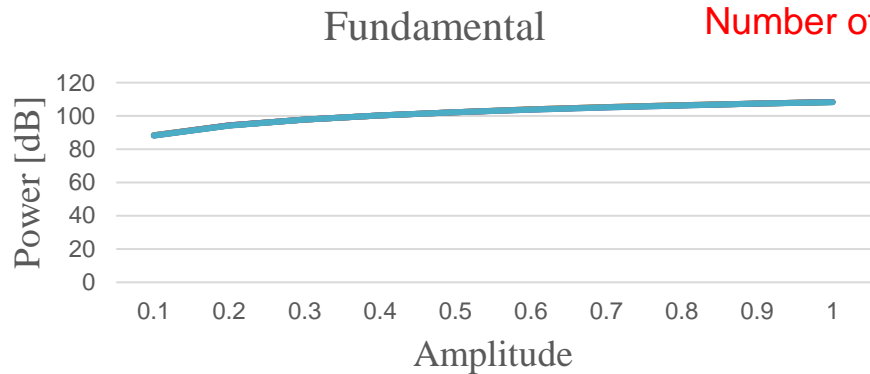


1<sup>st</sup>-order modulator

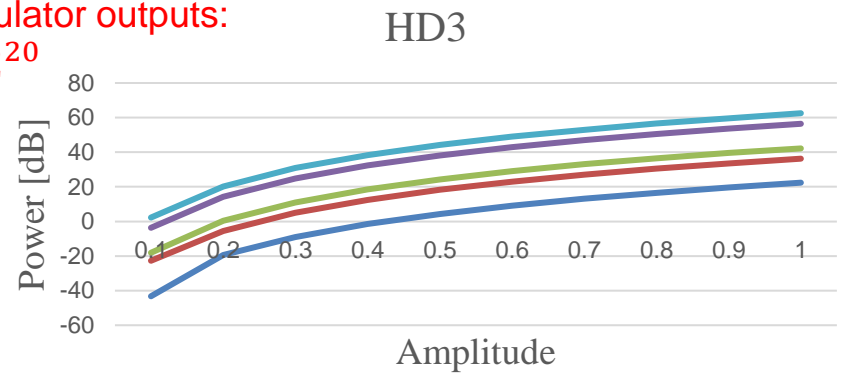
- ◆ Number of data:  $N=2^{20}$
- ◆  $V_{in}: A\cos(\omega t)$  ( $A = 0.1 \sim 1$ )
- ◆  $k=0.0001, 0.0005, 0.001, 0.005, 0.01$



# Cosine Input Simulation Result



Number of modulator outputs:  
 $N=2^{20}$

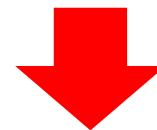


— p1=0.0001 — p1=0.0005 — p1=0.001 — p1=0.005 — p1=0.01

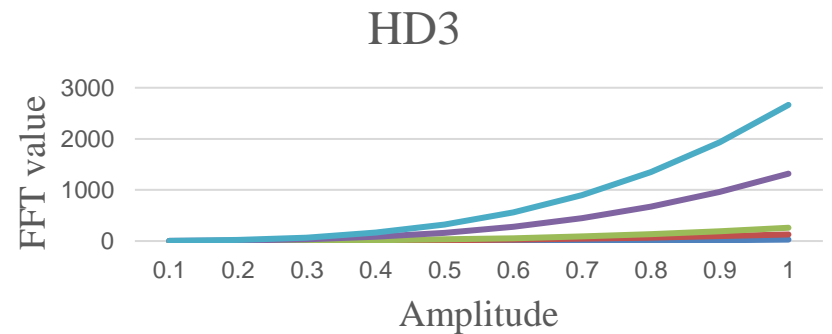
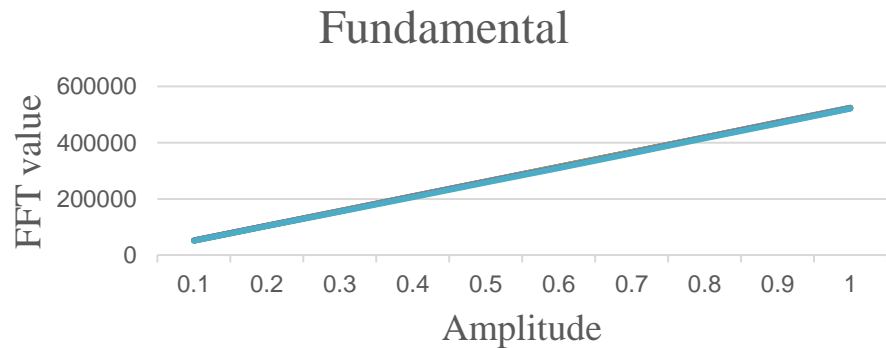
— p3=0.0001 — p3=0.0005 — p3=0.001 — p3=0.005 — p3=0.01



$$\text{Power} = 20\log(\text{FFT}_{\text{value}}) - 6.02 \text{ [dB]}$$



**FFT result**



— Q1=0.0001 — Q1=0.0005 — Q1=0.001 — Q1=0.005 — Q1=0.01

— p3=0.0001 — p3=0.0005 — p3=0.001 — p3=0.005 — p3=0.01

# Find Spectrum Power from DC Characteristics

- ◆ **1<sup>st</sup> - order** modulator
- ◆ Number of 1-bit output data :  $N = 2^{20}$

By nonlinearity

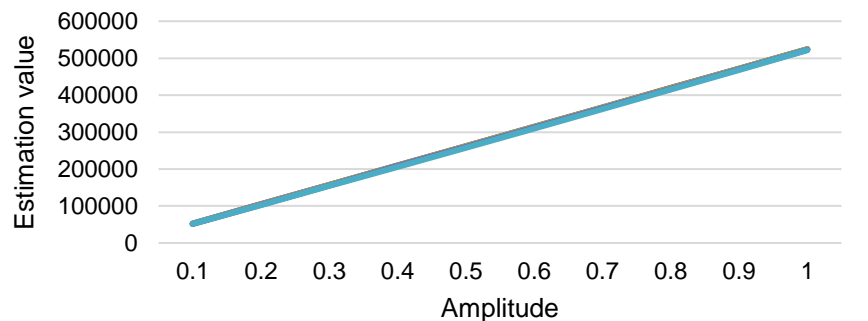
$$\text{Fundamental : } a_1 A + \frac{3}{4} a_3 A^3$$

$$\text{HD3 : } \frac{1}{4} a_3 A^3$$

A: amplitude

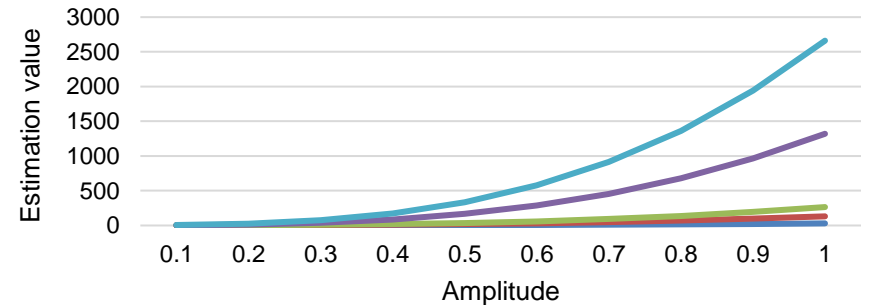
| k      | $a_3$   | $a_1$  |
|--------|---------|--------|
| 0.0001 | 104.84  | 524180 |
| 0.0005 | 524.48  | 523760 |
| 0.0010 | 1050.5  | 523240 |
| 0.0050 | 5282.5  | 519000 |
| 0.0100 | 10643.0 | 513610 |

Fundamental estimation value



— Q1=0.0001 — Q1=0.0005 — Q1=0.001 — Q1=0.005 — Q1=0.01

HD3 estimation value

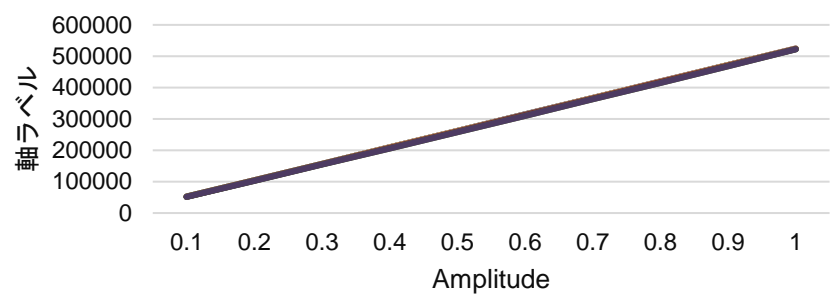


— Q3=0.0001 — Q3=0.0005 — Q3=0.001 — Q3=0.005 — Q3=0.01

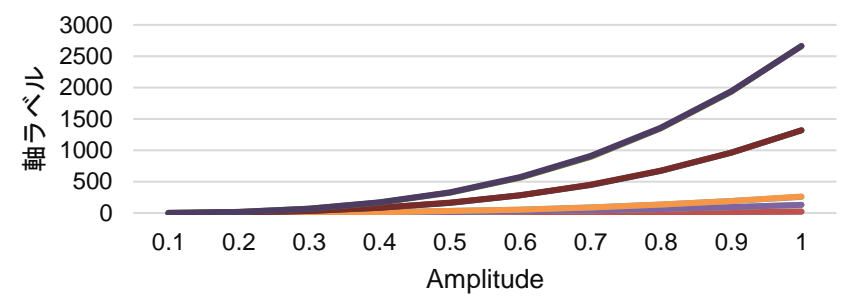
$N=2^{20}$

# Comparison between Estimated and FFT Values

Fundamental values



3rd harmonic values



- p1=0.0001
- Q1=0.0001
- p1=0.0005
- Q1=0.0005
- p1=0.001
- Q1=0.001
- p1=0.005
- Q1=0.005
- p1=0.01
- Q1=0.01

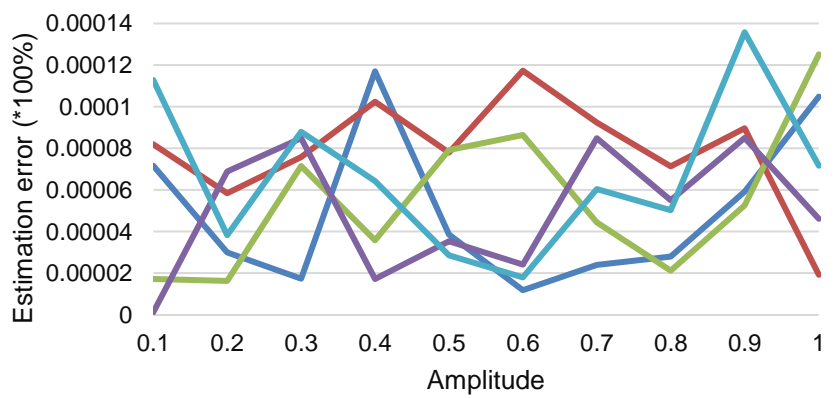
- p3=0.0001
- Q3=0.0001
- p3=0.0005
- Q3=0.0005
- p3=0.001
- Q3=0.001
- p3=0.005
- Q3=0.005
- p3=0.01
- Q3=0.01

$P_1$  : fundamental obtained by FFT  
 $Q_1$  : estimated fundamental

$$\text{Error} = \frac{Q_{\text{values}} - P_{\text{values}}}{Q_{\text{values}}}$$

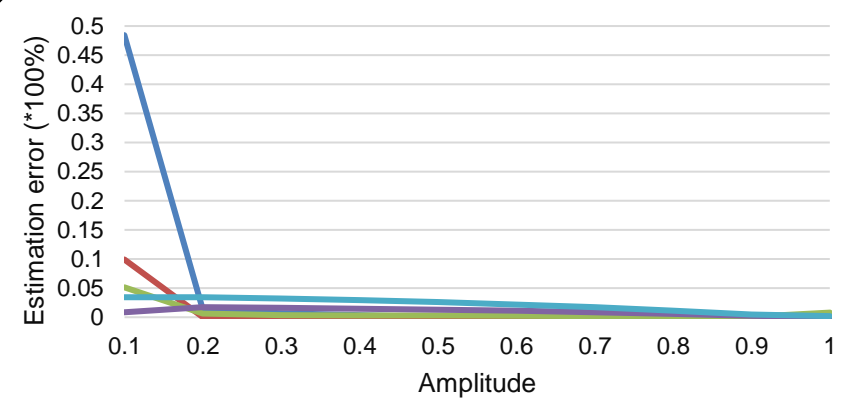
$P_3$  : HD3 obtained by FFT  
 $Q_3$  : estimated HD3

Estimation error of fundamental



- k=0.0001
- k=0.0005
- k=0.001
- k=0.005
- k=0.01

Estimation error of 3rd harmonic



- k=0.0001
- k=0.0005
- k=0.001
- k=0.005
- k=0.01

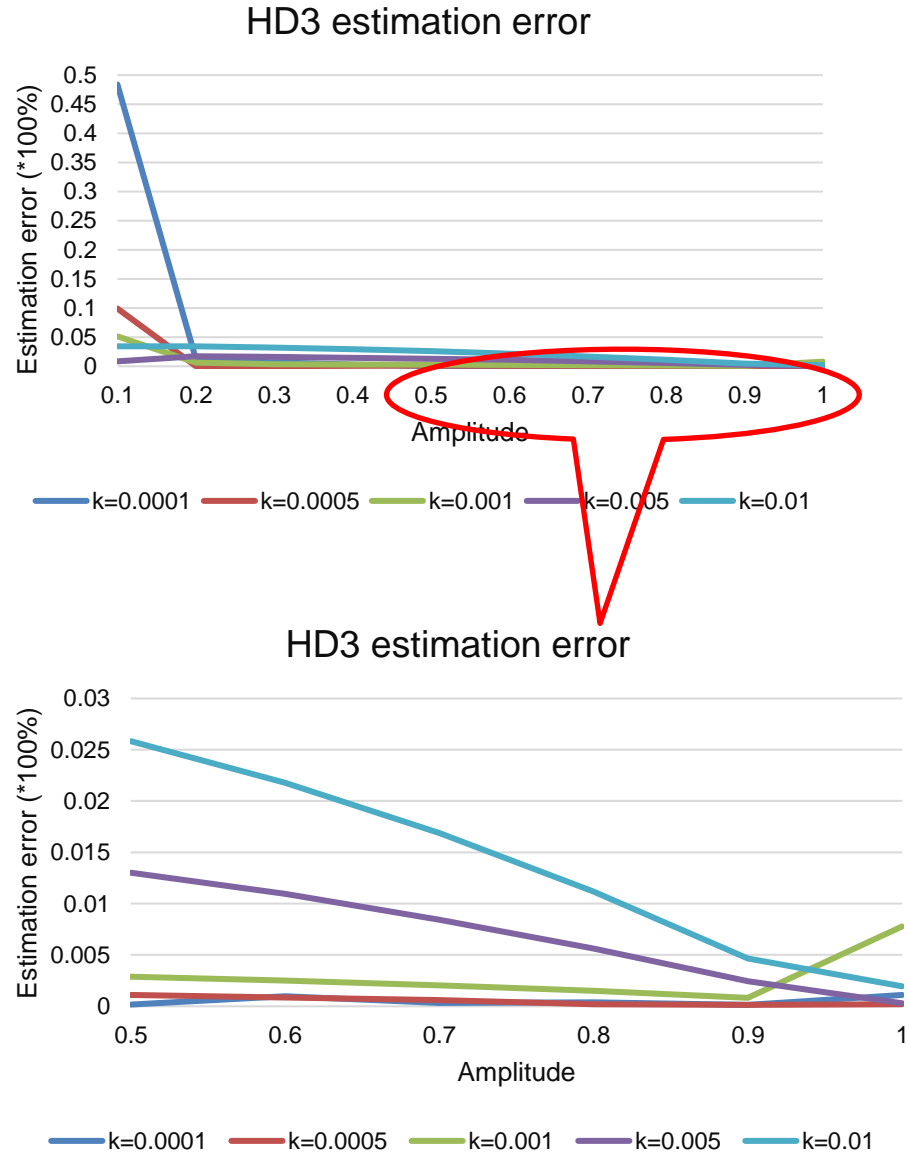


# Accurate Estimation Condition for HD3

1<sup>st</sup> -order modulator

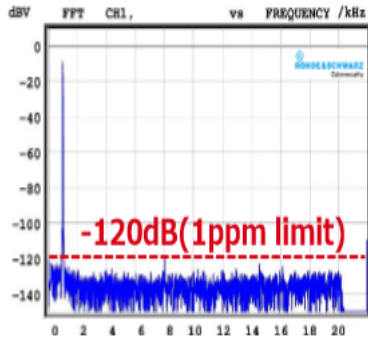
**Good** condition  
HD3

Amplitude = 0.9  
Error = 0.0123%  
(k = 0.0005)



# DUT Measurement Result

## Measurements results from ROHM semiconductor company



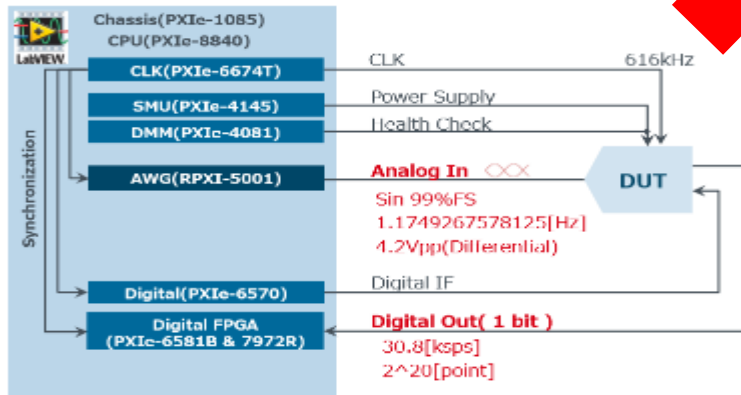
Output : 1kHz 44.1ksps  
 THD : 122dB( ~Fifth Harmonics)  
 SN : 132dB(Filter:20kHzLPF)

Meet the requirements

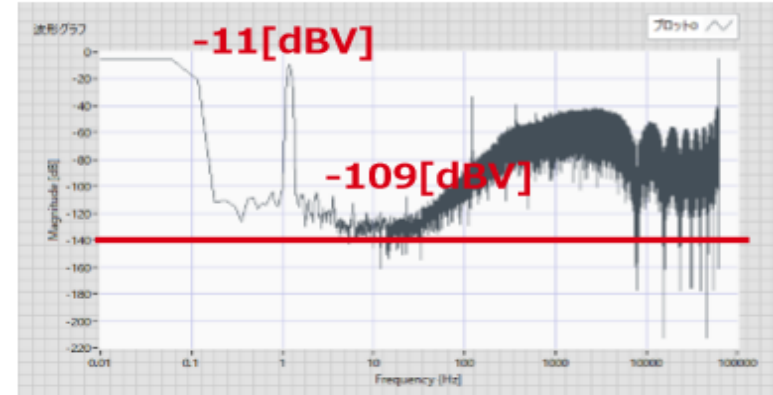
Signal from our developed **AWG**  
 (AWG: Arbitrary Waveform Generator)



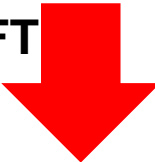
Test System Diagram and Test Condition



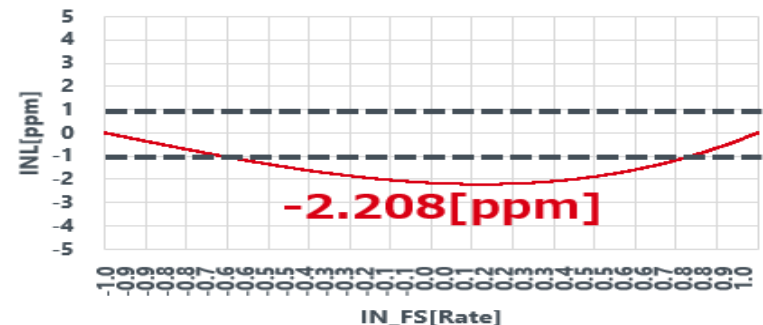
Use of NI PXI system  
 for experiment



Experimental result of  
 the modulator output FFT



Predicted INL



Obtained INL prediction  
 with the proposed method

# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

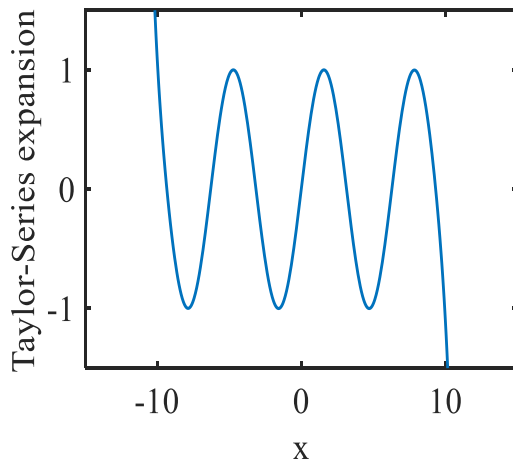
# Taylor Series Expansion

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

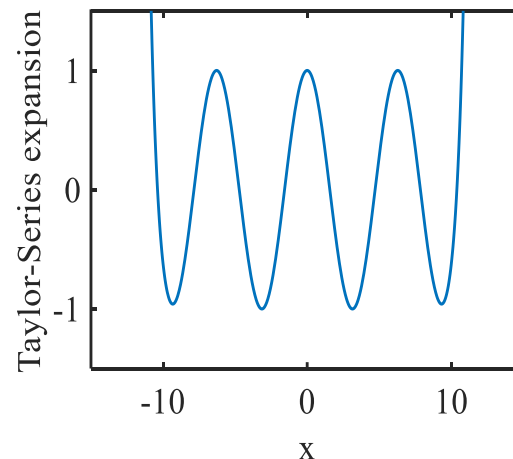
Function  $f(x)$  for  $x = a$

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

Convergence range:  $\alpha < x < \beta$



$\sin(x)$



$\cos(x)$

Central value:  $a = 0$

Number of Taylor-series expansion: 20.

# Floating-Point Representation

Mantissa : **M** ( $1 \leq M < 2$ )

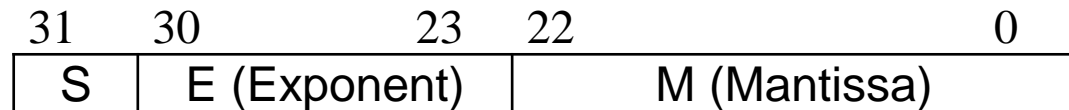
Exponent : **E**

$$\begin{array}{c}
 \underline{M} \times 2^E \\
 \downarrow \text{Decimal point} \\
 \underline{1.abcdef \dots} \times 2^E \\
 \text{Mantissa} \qquad \text{Exponent}
 \end{array}$$

$a, b, c, d, e, f, \dots : 0 \text{ or } 1$

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 :  $X = M \times 2^E$

**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 :  $\mathit{exp}(M)$  ( $1 \leq M < 2$ )


$$x = M$$

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

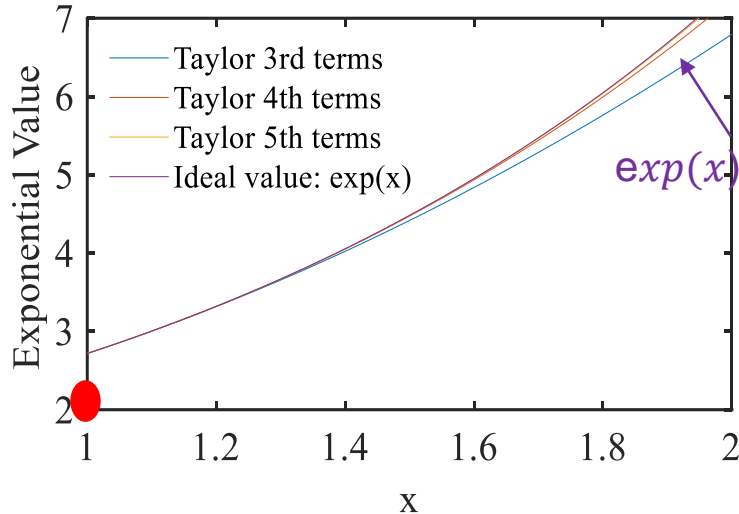


$$f(x) = \mathit{exp}(a) \times \left\{ 1 + q + \frac{1}{2}q^2 + \frac{1}{6}q^3 + \frac{1}{24}q^4 + \frac{1}{120}q^5 + \frac{1}{720}q^6 + \dots \right\}$$

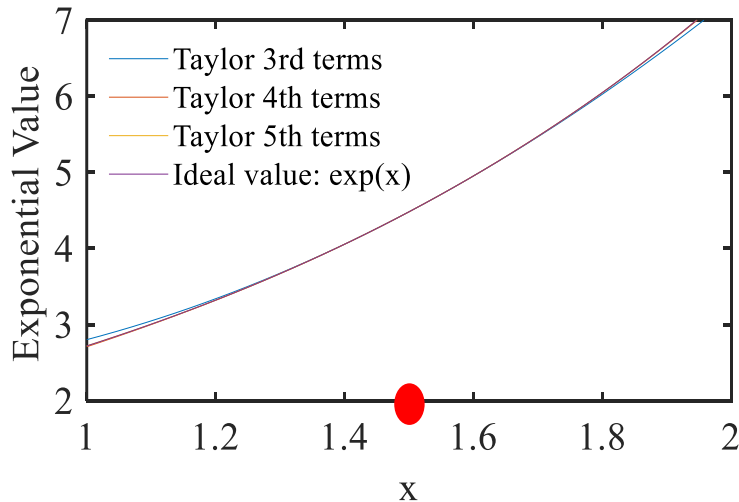
$q = x - a$

**Coefficient values:** stored in **LUT** in advance.

# Mantissa Region Division



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



at center value  $a = 1.5$

## 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 :

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

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

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

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

⋮



# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- **Simulation Verification**
- Hardware Implementation Consideration

## 4. Conclusion

# Definition of Accuracy

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

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

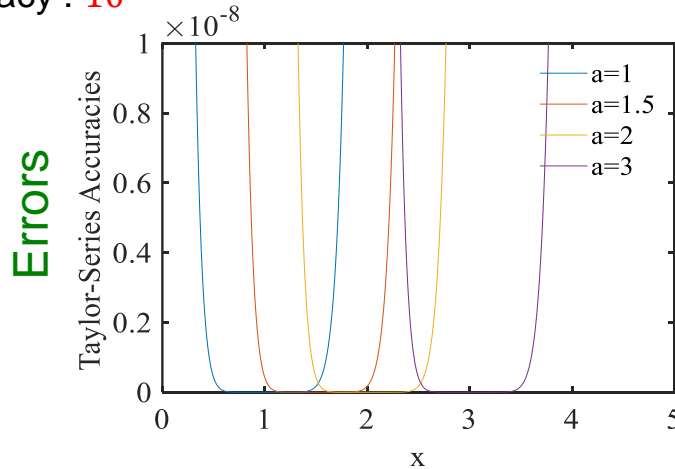
$f(x)$  : Original function

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

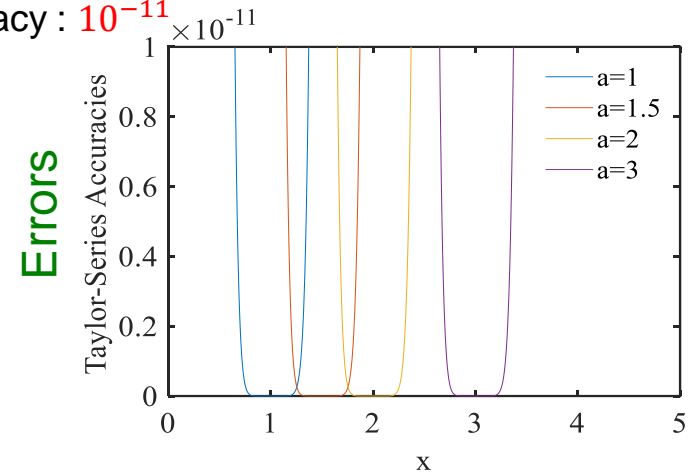
# Accuracy of $\exp(x)$ Taylor Expansion

Number of Taylor expansion terms: **10**

Accuracy:  $10^{-8}$

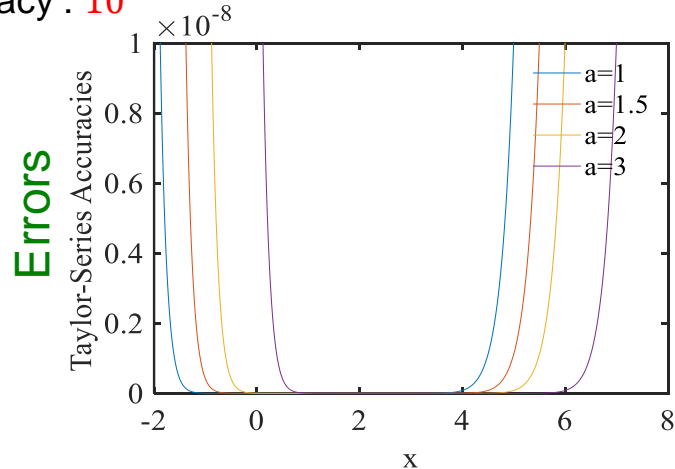


Accuracy:  $10^{-11}$

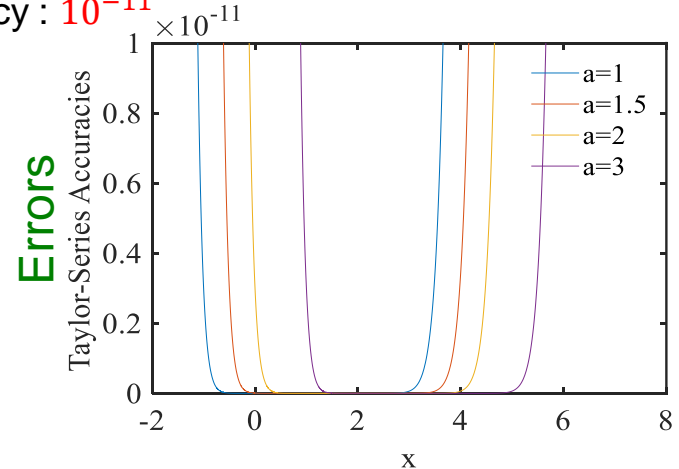


Number of Taylor expansion terms: **20**

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) $M = 1.xxxxxx\dots$<br>$1 \leq M < 2$ | $a = 1.5$ | 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) $M_D = 1.0xxxxx\dots$<br>$1 \leq M_D < 1.5$  | $a = 1.25$ | 3               | 5                  | 6                  | 7                  | 9                  |
| (ii) $M_D = 1.1xxxxx\dots$<br>$1.5 \leq M_D < 2$ | $a = 1.75$ | 3               | 5                  | 6                  | 7                  | 9                  |

# Four-Region Case

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

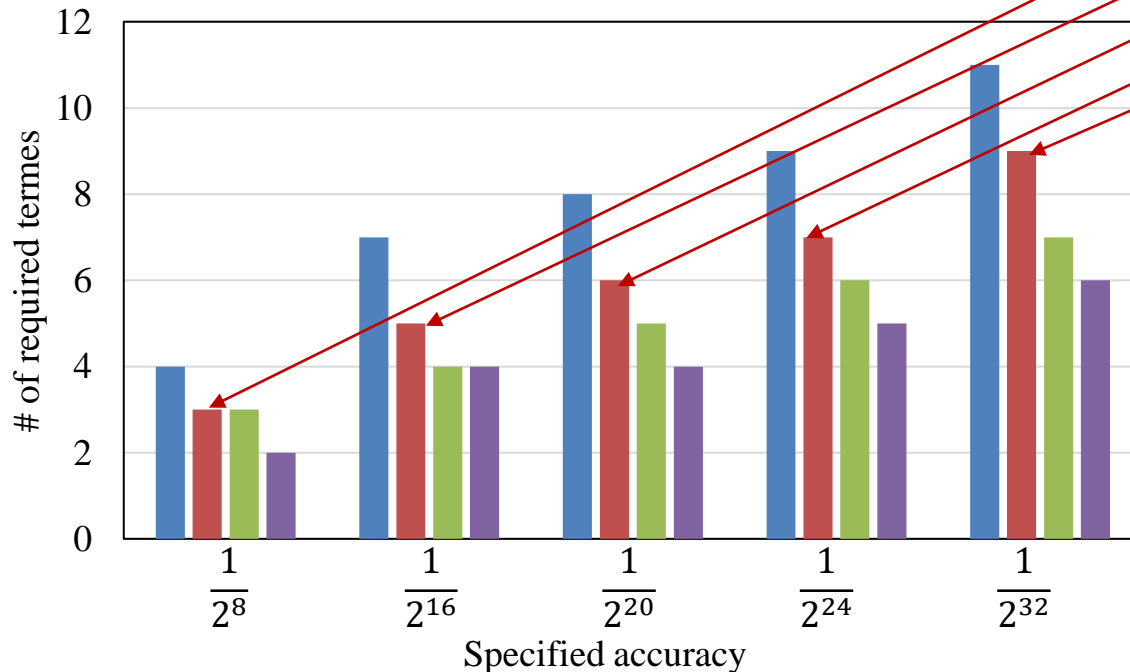
(00, 01, 10 or 11)  
of the **first two decimal** places of Mantissa.

| 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) $M = 1.00xxxx\dots$<br>$1 \leq M_D < 1.25$     | $a=1.125$ | 3               | 4                  | 5                  | 6                  | 7                  |
| (ii) $M = 1.01xxxx\dots$<br>$1.25 \leq M_D < 1.5$  | $a=1.375$ | 3               | 4                  | 5                  | 6                  | 7                  |
| (iii) $M = 1.10xxxx\dots$<br>$1.5 \leq M_D < 1.75$ | $a=1.625$ | 3               | 4                  | 5                  | 6                  | 7                  |
| (iv) $M = 1.11xxxx\dots$<br>$1.75 \leq M_D < 2$    | $a=1.875$ | 3               | 4                  | 5                  | 6                  | 7                  |

# Comparison of Number of Required Terms

Comparison of **required number of terms** for **different number of region divisions**

| 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) $M_D = 1.0xxxxx\dots$<br>$1 \leq M_D < 1.5$  | $a = 1.25$ |          | 3               | 5                  | 6                  | 7                  | 9                  |
| (ii) $M_D = 1.1xxxxx\dots$<br>$1.5 \leq M_D < 2$ | $a = 1.75$ |          | 3               | 5                  | 6                  | 7                  | 9                  |



Larger number of divided regions



Number of terms **reduced**;  
**LUT size** becomes **larger**.

■ One-region ■ Two-region ■ Four-region ■ Eight-region

# Outline

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- **Hardware Implementation Consideration**

## 4. 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**.

$$y = x - a \quad \text{Subtraction: } \mathbf{1 \text{ time}} \quad z = y^2 \quad \text{Multiplication: } \mathbf{1 \text{ time}}$$

$$\begin{aligned} f_5(x) &= \exp(a) \times \left( 1 + y + \frac{y^2}{2} + \frac{y^3}{6} + \frac{y^4}{24} \right) \\ &= \exp(a) \times \left\{ 1 + y + \frac{z}{2} \times \left( 1 + \frac{y}{3} + \frac{z}{12} \right) \right\} \end{aligned}$$

Multiplication: **5** times  
Addition / Subtraction : **4** times

**Total** : Multiplication: **6** times  
Addition / Subtraction : **5** times

# Number of Operations

## Number of terms versus number of operations in Taylor expansion

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

| Terms of Taylor expansion | Multiplication | Addition / Subtraction |
|---------------------------|----------------|------------------------|
| 3                         | 3              | 3                      |
| 4                         | 5              | 4                      |
| 5                         | 6              | 5                      |
| 6                         | 8              | 6                      |
| 7                         | 9              | 7                      |
| 8                         | 10             | 8                      |

# LUT Size

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

Stored in **LUT**

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

| Address (M=1. <i>ab</i> ...) | LUT data                    |
|------------------------------|-----------------------------|
| <b>00</b>                    | <i>Exp(a)</i> for a = 1.125 |
| <b>01</b>                    | <i>Exp(a)</i> for a = 1.357 |
| <b>10</b>                    | <i>Exp(a)</i> for a = 1.625 |
| <b>11</b>                    | <i>Exp(a)</i> for a = 1.875 |

# Outline

---

## 1. Research Background

## 2. $\Delta\Sigma$ ADC Linearity Testing Technology

- $\Delta\Sigma$  ADC Testing challenge and Linearity
- Proposed linearity test method
- Simulation configuration and results

## 3. Floating-Point Arithmetic Algorithms with Taylor-Series Expansion

- Taylor-Series Expansion and Proposed Algorithm
- Simulation Verification
- Hardware Implementation Consideration

## 4. Conclusion

# Conclusion

- High resolution, low speed  $\Delta\Sigma$  ADC linearity  
short time testing algorithm
- Polynomial modeling of modulator input / output characteristics
- FFT of modulator 1-bit output stream for cosine input
- Estimate polynomial coefficients from fundamental and harmonic powers
- Verified by simulation and experiments

Drastic testing time reduction:

104 days → 32 seconds

# Conclusion

- **Mantissa calculation** of **exponential function** with Taylor-expansion

Divide and Conquer Method

- Number of **divided mantissa regions** becomes **larger**



➤ Number of Taylor expansion terms ➔ **smaller**

➤ **LUT size** ➔ **larger**



**Optimal hardware configuration**

---

Thank you for listening !

---

# Appendix



# Newton's method

Newton's method step:

**First**, Start with an initial approximation  $x_0$  close to  $c$ .

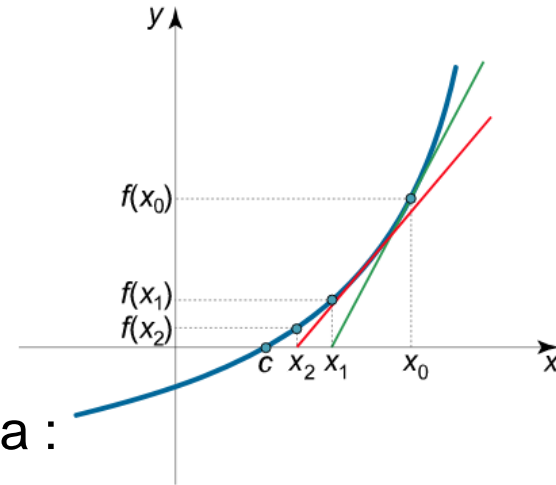
**Second**, Determine the next approximation by the formula :

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

**Third**, Continue the iterative process using the formula :

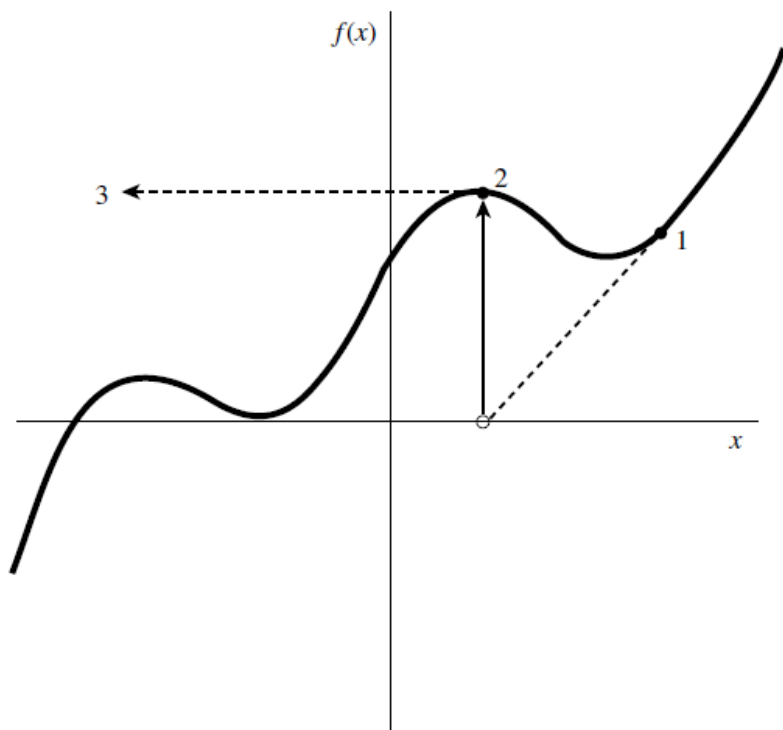
$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

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

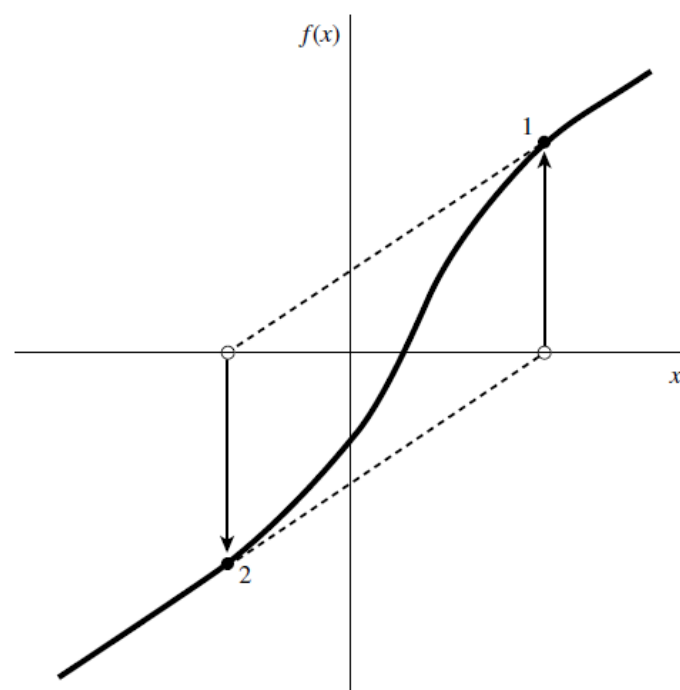


- Poor global convergence properties
- Dependent on initial guess
  - May be too far from local root
  - May encounter a zero derivative
  - May loop indefinitely

# 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:  $M$

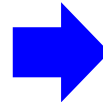
Exponent:  $E$

Original decimal point

$$\underbrace{1 \downarrow abcdef \dots}_{\text{Mantissa}} \times 2^{\underline{E}}_{\text{Exponent}}$$

$M \times 2^E$

$$1 \leq M < 2$$



New decimal point

$$\underbrace{0 \downarrow 1abcdef \dots}_{\text{Mantissa}} \times 2^{\underline{E}}_{\text{Exponent}}$$

$M \times 2^E$

$$0.5 \leq M < 1$$

Ex :  $1011001$  (binary) = 89 (decimal)

Binary representation :  $0.1011001 \times 2^{111}$



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

# 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)                     | $M = 1.000_{xxxx}\dots$<br>$1 \leq M_D < 1.125$    | $a=1.0625$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (ii)                    | $M = 1.001_{xxxx}\dots$<br>$1.125 \leq M_D < 1.25$ | $a = 1.1875$ | 2               | 4                  | 4                  | 5                  | 6                  |
| (iii)                   | $M = 1.010_{xxxx}\dots$<br>$1.25 \leq M_D < 1.375$ | $a=1.3125$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (iv)                    | $M = 1.011_{xxxx}\dots$<br>$1.375 \leq M_D < 1.5$  | $a=1.4375$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (v)                     | $M = 1.100_{xxxx}\dots$<br>$1.5 \leq M_D < 1.625$  | $a=1.5625$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (vi)                    | $M = 1.101_{xxxx}\dots$<br>$1.625 \leq M_D < 1.75$ | $a=1.6875$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (vii)                   | $M = 1.110_{xxxx}\dots$<br>$1.75 \leq M_D < 1.875$ | $a=1.8125$   | 2               | 4                  | 4                  | 5                  | 6                  |
| (viii)                  | $M = 1.111_{xxxx}\dots$<br>$1.875 \leq M_D < 2$    | $a=1.9375$   | 2               | 4                  | 4                  | 5                  | 6                  |

# 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

|                                             |            | Accuracy        |                    |                    |                    |                    |
|---------------------------------------------|------------|-----------------|--------------------|--------------------|--------------------|--------------------|
| Taylor-series expansion                     |            | $\frac{1}{2^8}$ | $\frac{1}{2^{16}}$ | $\frac{1}{2^{20}}$ | $\frac{1}{2^{24}}$ | $\frac{1}{2^{32}}$ |
| (i) $M = 1.xxxxxx\dots$<br>$-2 \leq M < -1$ | $a = -1.5$ | 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

|                                             |            | Accuracy        |                    |                    |                    |                    |
|---------------------------------------------|------------|-----------------|--------------------|--------------------|--------------------|--------------------|
| Taylor-series expansion                     |            | $\frac{1}{2^8}$ | $\frac{1}{2^{16}}$ | $\frac{1}{2^{20}}$ | $\frac{1}{2^{24}}$ | $\frac{1}{2^{32}}$ |
| (i) $M = 1.xxxxxx\dots$<br>$0.5 \leq M < 1$ | $a = 0.75$ | 3               | 5                  | 6                  | 7                  | 9                  |