DIGITAL CALIBRATION OF NONLINEAR MEMORY ERRORS IN SIGMA-DELTA ADCS

BY

SEUNGCHUL LEE

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering in the Graduate College of the University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Doctoral Committee:

Adjunct Assistant Professor Yun Chiu, Chair
Professor Milton Feng
Professor Martin D. F. Wong
Associate Professor Deming Chen
ABSTRACT

Background digital calibration techniques based on an output-referred error model are proposed to linearize sigma-delta (ΣΔ) modulators. A sequential power series (a special form of Volterra series) is found sufficient to model the nonlinear memory errors in a discrete-time integrator (DTI), which entails the application of adaptive polynomial transversal filtering for DTI error correction. For the calibration of feedback digital-to-analog converters, an analog domain pseudorandom noise (PN) removal technique is devised to resolve the input signal dynamic range loss resulting from the PN circulation in modulator loops. Error model identification is accomplished by correlating various moments of the digital output with a one-bit PN by exploiting the independence between the input and injected PN.

A 1-0 multi-stage noise shaping ΣΔ analog-to-digital converter (ADC) demonstrates the effectiveness/limitation of the proposed digital linearization techniques treating both amplifier distortion and capacitor mismatch in one frame. The design perspectives of low gain two-stage amplifiers are studied with other practical design issues in the implementation. The ADC employing 29 dB gain amplifiers achieves 85 dBC spurious-free dynamic range and 67 dB signal-to-noise ratio for a −1 dBFS (1.1 Vpp) 5 MHz sinusoidal input at 240 MS/s and 8
oversampling ratio with the support of digital linearization in which the calibration time is around 7.5 msec. With –6.5 dBFS two-tone signal at 14.9 MHz and 15.1 MHz, the average third order intermodulation after calibration is 87.1 dBc, which is more than 30 dB better than that without calibration. The core ADC consumes 37 mW from a 1.25 V supply and occupies 0.28 mm² in a 65 nm CMOS low leakage digital process in which the transistor threshold voltages are around 0.5 V.
To my family
ACKNOWLEDGMENTS

I would like to express my gratitude to my research adviser, Professor Yun Chiu, for his valuable advice in my Ph.D study. I have benefited greatly from his keen insight and perceptiveness on circuit analysis and innovation during hundreds of meetings. I am grateful for his encouragement and support. I would also like to thank Prof. Milton Feng, Prof. Martin Wong, and Prof. Deming Chen for serving on my doctoral committee and providing valuable suggestions.

I would like to give thanks to my colleagues, Pingli Huang, Bei Feng, Wenbo Liu, Daehynn Kwon, Hao Li, Peiyuan Wan, Richard Tseng, Brian Elies, Foti Kacani, Yuan Zhou, Sudipta Sarkar, Bo Wu, and Benway Xu, for enlightening discussion. Since I left my home country five years ago, I got a lot of help from many people. I still remember the moment that Prof. Yuchun Chang gave me an electric heater when he returned to China, and Richard and Daehyun helped me to settle in new surroundings. Especially, I will never forget the great support during my research at Beijing Embedded System Key Lab in China from Prof. John Yu, Prof. Ping Lin, Lang Wei, and Zhang Zejun, and Huang Guanzhong.

I thank Intel, ADI, and Agilent for funding support and Texas Instruments, Dallas, for chip fabrication. I also thank many people in TXACE for the support during the last stage of my Ph.D. program.
Most importantly, I would like to express my deepest gratitude to my wife and daughter for their love and support. I am grateful to my parents and all other friends for their encouragement.
# TABLE OF CONTENTS

## CHAPTER 1  INTRODUCTION ................................................................. 1
1.1 Background .................................................................................. 1
1.2 Limitations of ΣΔ ADC ................................................................. 6
1.3 Research Goals ............................................................................ 7
1.4 Thesis Organization ..................................................................... 10

## CHAPTER 2  EXISTING LINEARIZATION TECHNIQUES ................. 12
2.1 Analog Domain Linearization Techniques ................................... 12
2.2 Digital Calibration Techniques .................................................... 15

## CHAPTER 3  OUTPUT-REFERRED DISTORTION ANALYSIS .......... 20
3.1 First-Order Modulator ................................................................. 22
3.2 Second-Order Modulator ............................................................. 29
3.3 DAC Error .................................................................................. 32
3.4 Reversed Path Method ................................................................. 34

## CHAPTER 4  CALIBRATION BASED ON INDEPENDENT COMPONENT ANALYSIS ................................................................. 40
4.1 First-Order Modulator ................................................................. 41
4.2 Second-Order Modulator ............................................................. 47
4.3 Behavioral Simulation ................................................................. 50
4.4 Independent Component Analysis .............................................. 56

## CHAPTER 5  CAPACITOR MISMATCH CALIBRATION ..................... 58
5.1 Memory Error Model for Capacitor Mismatch ......................... 59
5.2 Calibration Approach ................................................................. 65
5.3 PN Removal Technique ............................................................... 71
5.4 Behavioral Simulation ................................................................. 80
5.5 Extension of Selective PN Injection and Cancellation ................ 85
CHAPTER 6  PROTOTYPE 1-0 MASH ADC DESIGN ...............................  87
  6.1 Nonlinear Calibration of MASH Structure .................................  87
  6.2 Prototype 1-0 MASH Architecture ..........................................  92
  6.3 Circuit Design .........................................................................  98
     6.3.1 DAC and PN Injection .......................................................  99
     6.3.2 Integrator ....................................................................... 102
     6.3.3 Low-Gain Two-Stage Amplifier .................................  106
     6.3.4 Comparator ..............................................................  112
  6.4 Measurement ........................................................................... 114

CHAPTER 7  PIECEWISE LINEAR MODEL ............................................... 128
  7.1 Polynomial Model .................................................................... 129
  7.2 Piecewise Linear Model .......................................................... 131
  7.3 Comparison ............................................................................. 136

CHAPTER 8  CONCLUSION ................................................................. 138

APPENDIX A  ANALYSIS OF FIXED-POINT ERROR
    IN CALIBRATION ................................................................. 140
     A.1 First-Order Modulator ...................................................... 140
     A.2 Second-Order Modulator .................................................. 141

APPENDIX B  ANALYSIS OF CAPACITOR MISMATCH
    IN PN INJECTION ............................................................... 144
     B.1 First-Order Modulator ...................................................... 144
     B.2 Second-Order Modulator .................................................. 147

REFERENCES .................................................................................. 153
CHAPTER 1

INTRODUCTION

1.1 Background

The development of process technology has driven the change of circuit topologies. While technology scaling offers high-level integration of low-power, high-speed digital circuits, the accompanying low supply voltages degrade the on-chip signal integrity and constrain the choice of analog circuit structures. At such low supply voltages, the conventional stacked cascode amplifier structures need to be replaced by multi-stage cascade ones to accommodate an adequately high dc-gain and large signal swings compromising the operating speed [1]. Due to the low output impedance of the short channel devices, it is getting more and more difficult to design highly linear circuits based on a high-gain amplifier in a feedback configuration. In addition, unlike in digital circuits, the device scaling in high-performance analog circuits is usually dictated by the thermal noise and device mismatch, so the performance has been limitedly enhanced in the process development.

For analog-to-digital converters (ADCs), by virtue of oversampling and noise shaping, the sigma-delta (ΣΔ) conversion technique makes a favorable
The tradeoff between the signal bandwidth and signal-to-noise ratio (SNR) when compared to Nyquist-rate counterparts. The comparison of these two different structures helps to comprehend the strengths and weaknesses of ΣΔ ADCs. In Figure 1.1, the major difference between these two types of ADCs is the feedback path from the digital output to the analog input. The Nyquist rate ADCs process each analog input sample independently one by one, so the input signal bandwidth can be as large as the Nyquist’s criterion. Without the feedback loop, the overall ADC structure is unconditionally stable, and the analysis of the circuit non-ideality is relatively straightforward. However, the quantization noise directly appears at the output. Thus, to reduce the quantization noise, the Nyquist structure usually consists of multiple stages and the first stage quantization noise is further quantized by the following stages as shown in Figure 1.1(a). This analog signal processing relies on the highly accurate inter-stage gain, which demands high gain amplifiers and accurate passive element matching.

![Figure 1.1. Conceptual diagram of (a) Nyquist rate and (b) ΣΔ ADCs.](image-url)
On the other hand, the oversampling architecture utilizes the previous output information for the present input data processing through the feedback path. With the feedback, the in-band quantization noise at the ADC output is attenuated by the loop. In the first-order sigma-delta architecture of Figure 1.1(b), the signal transfer function (STF) is just one sample delay ($z^{-1}$), but the quantization noise transfer function (NTF) is a high pass filter ($1-z^{-1}$). At the low frequency, $z$ is near 1, so the quantization noise is attenuated by the loop, but at the high frequency, which means $z$ is near $-1$, the noise is magnified by two. In turn, the maximum input signal bandwidth is restricted below the Nyquist rate. The noise shaping and oversampling allow a high SNR even with a low resolution quantizer in a single stage structure where the amplifier gain and passive element matching requirements are much relaxed in terms of the SNR.

Although ΣΔ ADCs need a far higher sampling speed than Nyquist rate ones for a given input bandwidth, in practice, many applications do not adopt the Nyquist rate sampling. Especially, in communication systems, an oversampling ratio (OSR) greater than two is necessary to facilitate analog and digital filter design. Figure 1.2 exhibits the approximated ADC quantization noise power $P_q$ for a given quantizer resolution (bit) and noise shaping order ($n$) [2], and the signal-to-quantization noise ratio (SQNR) derived for a single sinusoidal input. In both Nyquist/oversampling architectures, we can get around 6 dB SQNR improvement whenever we increase one bit resolution of the quantizer (part (i)).
However, when we adapt the noise shaping, we can get the additional SQNR depending on the noise shaping order $n$ (part (ii)) (without noise shaping (Nyquist rate), $n$ is 0).

For example, in a first-order $\Sigma\Delta$ ADC, we achieve additional 22 dB SQNR by an OSR of 8, which means 13 times more accurate digitization is achieved by 8 times faster sampling speed. Lastly, with oversampling, the in-band thermal noise power becomes $1/\text{OSR}$ of the total noise power, so we can scale down the device size according to the OSR. As a result, $\Sigma\Delta$ ADCs gain more benefit from the high processing speed and continue to improve their figure-of-merit (FoM) of conversion [3]–[7].

$$P_q = \frac{1}{12 \cdot 2^{bit}} \cdot \frac{\pi^{2n}}{OSR^{2n+1}} \cdot (2n+1)$$

$$(i) \quad \quad \quad (ii) \quad \quad \quad (iii)$$

$$SQNR = 6.02 \times bit - 1.8 \quad \quad \quad \quad \quad (i)$$

$$+ \quad 6.02 \times (n + 0.5) \times \log_2 \text{OSR} \quad \quad \quad \quad \quad (ii)$$

$$- \quad 9.94 \times n + 10 \log_{10} (2n + 1) \quad \quad \quad \quad \quad (iii)$$

Figure 1.2. Quantization noise of a $\Sigma\Delta$ ADCs.

$\Sigma\Delta$ ADCs are usually implemented with either continuous-time (CT) or discrete-time (DT) structures. The comparison of these two architectures has been presented in the literature [2],[8],[9]. In CT structures, the nonlinearity of the
integrators is attenuated by the effective loop gain at the output signal frequency which is mainly dependent on the input signal bandwidth and noise shaping order. Thus, the required gain-bandwidth product (GBW) of the amplifier increases proportionally to the linearity and input bandwidth specifications, whereas it is not dominantly affected by the sampling clock speed. For instance, to achieve 6 dB better linearity for the same input, the GBW should be doubled, and so is the power consumption. In other words, as the linearity requirement is relaxed or the output signal swing is small, the CT structure is more advantageous in implementing high-speed sampling with a large OSR.

On the other hand, the nonlinearity of DT integrators consists of two parts, the static and dynamic errors. The static error results from the insufficient dc gain of the amplifier, which is not directly related to the power consumption. When the amplifier dc-gain is guaranteed, the GBW is determined by the desired settling accuracy $\varepsilon$ of the integrator output for a clock frequency $f_{ck}$ [10]:

$$GBW = \frac{f_{ck}}{\pi \beta} \ln\left(\frac{A_{tr}}{\varepsilon}\right)$$

(1.1)

where $\beta$ is the feedback factor in a closed-loop configuration, and $A_{tr}$ is the transition between two consecutive integrator outputs. Although the required GBW increases in proportion to the clock frequency $f_{ck}$, it is log-proportional to the output transition $A_{tr}$. Considering that $A_{tr}$ is inversely proportional the OSR, the DT structure is more efficient to implement a highly linear circuit (small $\varepsilon$).
with a low clock frequency (low OSR), which means that the Nyquist rate is the best choice to process a given bandwidth signal with the DT structure in terms of the circuit linearity.

1.2 Limitations of \( \Sigma \Delta \) ADC

As mentioned, to maintain a low distortion level for large input signals and to ensure an overall good signal-to-noise and distortion ratio (SNDR), the ADC nonlinearity needs to be properly dealt with in circuit design. There are two main functional blocks determining the linearity in \( \Sigma \Delta \) ADCs: an integrator and feedback digital-to-analog converter (DAC). The integrator linearity in a single-stage \( \Sigma \Delta \) structure is not deteriorated from the finite amplifier gain if the gain is constant regardless of the signal amplitude. However, in a multi-stage noise shaping (MASH) structure which is preferred to process a wideband signal with a high SNR, the gain mismatch between stages resulting from the finite amplifier gain significantly degrades the linearity. Moreover, in practice, the amplifier gain is signal dependent, which causes the harmonic distortion in both single- and multi-stage structures. Coupled with the oversampling and low supply voltage, high-performance amplifiers present a fundamental design challenge for wideband high-resolution \( \Sigma \Delta \) ADCs in advanced CMOS technologies.

The DAC in its feedback path is classified into single-bit (or 1.5-bit in differential forms) and multi-bit structures. Single-bit DACs with one unit
element are inherently linear, which is instrumental to realizing intrinsically linear \(\Sigma \Delta\) ADCs without trimming or calibration. However, a high OSR is usually necessary to achieve high SNR using single-bit architectures; it is therefore mostly useful in narrow-band applications, e.g., instrumentation and audio. For broadband digitization, multi-bit DACs often allow better trade-offs between OSR and ADC resolution, leading to improved conversion FoM. Potential mismatch among constituent unit elements in a multi-bit structure, however, becomes a dominant linearity performance limiter. On the other hand, once the error is generated from the circuits, it remains in the modulators loop for a long time and impairs the signal integrity in the latter data conversion process (memory effect).

In summary, the \(\Sigma \Delta\) architecture possibly achieves better FoM by more efficiently trading off the operation speed for the SNR, especially in communication systems. The SQNR performance is not much affected by the non-ideality of the analog circuits in a single stage structure, but it still needs high gain amplifiers and accurate device matching for a high SNDR, which limits the maximum achievable operating speed of this structure.

### 1.3 Research Goals

The goal of this work is to propose a digital calibration technique handling both active and passive device nonlinearities in the presence of the memory effect and thereby to demonstrate the feasibility of the performance enhancement of \(\Sigma \Delta\)
ADCs with high-speed but less accurate analog circuits. For this goal, the following topics are studied:

- Analyzing nonlinear memory errors in ΣΔ modulators and proposing the universal model representing a nonlinear ΣΔ modulator.
- Investigating the limitations of the conventional PN based calibration techniques and suggesting the identification method by using a one-bit PN.
- Studying the PN circulation in the modulator loop and devising the technique to alleviate the loss of the conversion dynamic range.
- Implementation of the prototype ADC adapting the proposed techniques and demonstration of the effectiveness/limitation.
- Presenting the practical issues and design perspectives of the analog circuits with the support of the digital compensation.
- Suggesting the possible solutions to overcome the limitation of the proposed techniques.

The major contribution of this work is divided into two parts, the circuit error modeling and identification. In the modeling part, each signal-path distortion is represented as an additive signal directly derived from the modulator output, so it is manifested that an FIR form can represent the long-term nonlinear memory errors in modulators. For the identification, the information of the input signal in conjunction with the injected PN, both contained in the digital output of a ΣΔ
modulator, are utilized to identify the nonlinear memory errors in \( \Sigma \Delta \) conversion. Especially, the moment correlation functions of the input signal and PN are exploited to extract the nonlinear model of a discrete-time integrator (DTI), which is the most critical component of a \( \Sigma \Delta \) modulator. By this approach, each nonlinear error source of a DTI is identified with a one-bit PN signal.

To identify multiple nonlinear error sources, multiple PN signals potentially degrade the input signal dynamic range. In a system with feedback paths, like a \( \Sigma \Delta \) modulator which is our main interest, the injected PN signal circulates in the loop and sometimes saturates the signal swing. To resolve this problem, the selective PN injection with PN cancellation is proposed, particularly for the compensation of capacitor mismatch in a feedback DAC.

The proposed techniques are implemented in a prototype 1-0 MASH ADC. The practical issues and new design perspectives in analog circuits supported by the digital linearization are investigated in detail. The simulation and measurement results show the effectiveness/limitation of the techniques.

The nonlinear error model in this work is mostly based on the power series expansion. Comparing different nonlinear models reveals the merit and demerit of each model and helps to achieve efficient implementation of the digital calibration according to the application. The last part introduces a piecewise linear model and suggests possible future research topics to further enhance the efficiency of the digital calibration techniques.
1.4 Thesis Organization

This dissertation is organized as follows:

Chapter 2 introduces the conventional linearization techniques in analog and digital domain and describes the limitation of each existing technique.

Chapter 3 proposes the output-referred distortion analysis of the $\Sigma\Delta$ modulator, which takes into account the error circulation in the DTI. From this analysis, the universal model including multiple nonlinear error sources is proposed. The similarity with the exiting method is provided to give more insight into the proposed approach. The mismatch effect of PN injection in modeling is also considered in this chapter.

Chapter 4 presents the identification procedure and digital compensation scheme in detail. The simulation results of first- and second-order modulators are shown with a brief discussion of the effectiveness/limitations of the proposed calibration approach. It is explained that the identification method can be comprehended as the independent component analysis (ICA).

Chapter 5 proposes an error model and digital compensation scheme specialized for DAC calibration. The analog PN removal technique is devised to solve the dynamic range loss problem in the PN injection. Behavioral simulation results of first- and second-order modulators follow.

Chapter 6 demonstrates a 1-0 MASH structure employing the digital linearization techniques. The circuit design issues in the advanced digital process
are described in detail. The measurement results of the ADC are shown with a discussion of the effectiveness/limitations of the proposed techniques.

Chapter 7 describes the piecewise linear model as an alternative approach to alleviate the limitations of the polynomial based nonlinear model. The advantages and disadvantages of these two models are compared and future work is suggested.

Chapter 8 concludes and summarizes this research.
ΣΔ ADCs can achieve a high SNR by trading off the input signal bandwidth, but, at high operating speed, the input signal dynamic range of this structure is eventually limited by the circuit nonlinearity. This chapter describes the conventional linearization approaches and the limitations of each approach to clarify the unsolved problems which we tackle in this research.

2.1 Analog Domain Linearization Techniques

Due to the difficulty of analyzing the memory effect, the conventional linearization techniques have focused on dealing with the nonlinearity in analog domain. The linearization techniques are usually developed for two main functional blocks, an integrator and feedback DAC.

To enhance the amplifier linearity in the integrators, the low distortion structure in Figure 2.1 has been widely employed [11]–[15]. In this structure, the input signal is directly coupled at the quantizer input through the passive feed-forward path, such that the input is cancelled out by itself at the front stage of the
loop filter. In turn, the integrator theoretically does not process the input signal, and it only handles the shaped quantization noise, which greatly reduces the signal swing in the main signal path. The low-distortion structure would relieve the requirement of the high-gain amplifier, but still the fairly high gain, additional hardware for the summation, and delay-free feedback DAC in this topology have been recognized as challenges in highly-linear wideband modulator design [13].

The signal summation at the quantizer input can be implemented by either passive or active adders. Without active components (buffer), the passive summation inevitably entails the attenuation of the signal gain and the susceptibility to the interference from other circuitry, which increases the complexity and power consumption of the following blocks (quantizer). With an active adder, the additional amplifier stage requires extra power consumption and timing. In both cases, the gain mismatch between the input sampling network and

![Figure 2.1. Low distortion feed-forward structure.](image-url)
the signal path starting from the feed-forward path through the feedback DAC to the loop filter input, which are highlighted with the blue and red paths in Figure 2.1, should be carefully matched to minimize the signal component in the main signal path.

By manipulating the signal flow graph, the recent improved structures eliminate the necessity of the additional amplifier [15] or relax the timing overhead [13]–[15]. In [15], the summation node is moved from the last stage integrator output to the input, so that the integrator can support the summation function, such that it allows the timing for the DAC operation. On the contrary, an approach to utilize this additional amplifier for coupling the quantization noise is also considered [12]. The basic concept of the noise coupling is that the summation amplifier performs an additional role (quantization noise storage) and increases the order of the noise shaping, so the existence of the summation amplifier is justified. The timing overhead is relaxed by the additional delay in the signal path by using the multiple input sampling [13],[14]. However, with multiple sampling capacitors, the mismatch between the sampling networks and the sampling clock, which is an integer fraction of the main clock, can generate unwanted image tones at the ADC output spectrum.

For DAC linearization, dynamic element matching (DEM) techniques have been widely explored to treat capacitor mismatch errors [16]–[19]. The fundamental concept of DEM is to convert signal-dependent distortions into noise
by a random selection (shuffling) of the constituent unit elements in DAC operation. Since the shuffling has to be done for all unit elements simultaneously, the circuit complexity increases dramatically as a function of the number of involved capacitors.

In general, $\Sigma\Delta$ modulators allow maximally one clock cycle for each integrator operation. A complicated DEM operation often occupies a considerable amount of loop-delay budget, which, otherwise, can be used for analog processing. The timing overhead is particularly undesirable in the low-distortion $\Sigma\Delta$ structures, in which the linearity of modulation is improved while necessitating additional settling time for the signal summation. Although simple DEM structures with minimum delay can mitigate the timing problem [18],[19], these approaches often result in signal-dependent, in-band tones and/or SNR degradation when the OSR is low due to the insufficient noise shaping. In addition, the design effort to minimize the DEM delay and switching noise in modulator loops is still significant in high-speed applications. As results, these analog linearization approaches slow down the maximum achievable operating speed for linearity.

## 2.2 Digital Calibration Techniques

One possibility to address this issue in advanced CMOS processes is to apply digital calibration, such that the design overhead is displaced out of the loop. The nonlinear memory error is generalized as a Volterra function in the digital
signal processing, which contains all possible combinations of the sequential output and its higher-order terms [20]. The number of error parameters in a Volterra model greatly increases according to the required precision, so it is not practically useful to calibrate high-resolution ADCs. For this reason, conventionally, the linearization of Nyquist-rate and oversampling converters have been reported in literature addressing either linear memory [21]–[24] or memoryless nonlinear errors [25]–[40].

The linear memory error model pertains to the compensation of the quantization noise leakage in a MASH structure where the infinite impulse response (IIR) form of noise transfer function (NTF) mismatch between the analog and digital domains is approximated to a finite impulse response (FIR) model. On the other hand, recent works directly adapt a memoryless nonlinear model to a single stage ΣΔ structure ignoring the memory effect in the modulator loops [39],[40]. When the OSR is large, the nonlinear error originating from the input sequence changes slowly, so the errors adjacent in time can be summed together and approximated as a function of a single input. However, this simplified model does not accurately represent the nonlinearity of a wideband signal, particularly when the loop gain of the modulator is not sufficiently large. As a result, a general model treating the nonlinearities in the presence of memory effect has yet to be developed.
The other portion in the digital calibration is the identification of a given error model. The error parameter identification can be classified into two categories depending on whether the technique employs a test signal or not [41]. Without a test signal, the compensation usually derives the error statistics directly from the conversion results assisted by a reference path [25]–[27],[39],[40] or a second ADC [29],[30], known as the parallel- and split-path ADC, respectively. When a test signal—typically a one-bit pseudorandom noise (PN)—is applied, the error parameters are often identified from the long-term correlation of the digital output and PN [31]–[38]. In this approach, the existence of the input signal actually results in the fluctuation of the identified parameters in steady state [30].

The goal of the calibration is to identify the adaptive filter with an inverse transfer function of the ADC in Figure 2.2. If the signal transfer functions of these two blocks are matched, the residue error after removing input signal $X$ should not include any moment of $X$. Thus, the adaptive filter is trained by the correlation with the moment of $X$. As stated, conventionally, the $X$ removal block in Figure 2.2 is implemented by either input statistics or a second ADC with different nonlinear characteristics.

By replacing all $X$ in Figure 2.2 by the known signal $PN$, it becomes the PN signal based calibration. Then, only the moment of the injected $PN$ is utilized in the correlation and the input signal $X$ is considered as large noise perturbing the parameter learning. When the $PN$ is one bit, higher order moments, i.e., $PN$, $PN^2$, ...
and $PN^3\ldots$ are all the same as the first-order moment. In turn, the error information obtained from the ADC output is used for learning only one parameter, which is generally a linear term; estimation of multiple nonlinear coefficients thus demands multiple PN injection, potentially degrading the input signal dynamic range and complicating the analog circuitry involved in the injection [37].

![Figure 2.2. Conceptual diagram of the conventional ADC calibration.](image)

In terms of the error observation, the digital techniques in sigma-delta ADCs are focused on the quantization noise leakage from the NTF mismatch in a multi-stage structure [7],[23]. The concept behind these approaches is that the in-band noise power increases according to the amount of the component mismatch. Thus, these approaches find the digital coefficients that minimize the in-band noise by using the steepest descent method. However, during the normal operation mode, the quantization noise and signal are indistinguishable, so a pre-assigned calibration period is required (foreground calibration). The opposite
approach is considered in [42], where the digital coefficients maximizing the out-of-band noise are selected. This approach uses the out-of-band quantization noise as an indicator of the component matching, which enables background calibration, but it is sensitive to the existence of the blocker signals or switching noise. The calibration based on the test signal injection has been applied only for the case when the modulator loop gain [21]–[24] is sufficiently large, such that the elimination of the injected PN can be achieved with a simplified model, i.e. linear memory or nonlinear memoryless model.
CHAPTER 3

OUTPUT-REFERRED DISTORTION ANALYSIS

The primary technical challenge for the calibration of ΣΔ modulators resides in the memory introduced by the ΣΔ loop. As shown in Figure 3.1(a), a typical treatment of the finite gain error in integrators leads to the pole and zero movements embodied in a z-domain model of the modulator given by

\[
D(z) = \frac{\alpha z^{-1}}{1 + (\alpha - \beta)z^{-1}} X(z) + \frac{1 - \beta z^{-1}}{1 + (\alpha - \beta)z^{-1}} E_q(z). \tag{3.1}
\]

In a linear formulation, coefficients \( \alpha \) and \( \beta \) are considered signal independent; thus the pole relocation can be compensated by a truncated finite impulse response (FIR) filter approximating the denominator of Equation (3.1) [21], while the zero relocation, i.e., \( \beta < 1 \) in the numerator of Equation (3.1), leads to quantization noise leakage. However, these linear models based on the assumption of a constant amplifier gain are insufficient to capture the nonlinear distortions in the DTI shown in Figure 3.1(b), where the gain of the amplifier, \( A \), is a function of the output signal \( Y \), resulting in the signal dependence of the modulator coefficients \( \alpha(Y) \) and \( \beta(Y) \).
Figure 3.1. (a) Conventional $z$-domain model of a first-order $\Sigma\Delta$ modulator, (b) switched-capacitor realization of the discrete-time integrator, and (c) the proposed nonlinear model of (a).

For weakly nonlinear amplifiers, the gain nonlinearity can be well approximated by an $N$th-order power series [43] given by

$$\frac{y(n)}{A(n)} \approx \frac{\sum_{i=1}^{N} k_i y^i(n)}{A_{\text{max}}},$$

(3.2)
where $A_{\text{max}}$ is the maximum gain, $k_1 = 1$, and the higher-order coefficients are determined by the amplifier nonlinearity (Figure 3.2). This is equivalent to stating that the amplifier gain varies from cycle to cycle dependent on $Y$. The nonlinearity of other circuit elements, e.g., switches and capacitors, is assumed to be sufficiently attenuated by other techniques [44],[45] in the subsequent analyses.

![Figure 3.2. Polynomial model of the output signal-dependent amplifier gain.](image)

### 3.1 First-Order Modulator

The $Y$-dependence of the coefficients $\alpha$ and $\beta$ motivates an output-dependent error model of the modulator shown in Figure 3.1(c), where an additive signal $E_a(Y)$ is inserted after the delay element. The modulator is otherwise ideal. The output digital signal $D$ can be derived as

$$D(z) = z^{-1}X(z) + E_a(z) + (1 - z^{-1})E_y(z).$$

(3.3)

Essentially, $E_a$ can be regarded as the difference between the actual and ideal integrator outputs. Since $E_a$ is a function of $Y$, we expect that $D$, the
quantized version of $Y$, includes information needed to estimate $E_a$. Thus, if $E_a$ can be derived from $D$, it can then be eliminated in the digital domain. The relationship between the integrator output $Y$ (or $D$) and the error $E_a$ will be explicated in the following analyses.

The error model can also be explained using a switched-capacitor realization of the DTI in Figure 3.1(b), where a standard non-overlapping two-phase sampling/integration operation is assumed. In the $\phi_2$ phase of the clock cycle $n-1$, the integrator output $y(n-1)$ stored in $C_F$ contains distortions dependent on the amplifier gain $A(n-1)$, while, at the same time, the input signal $w(n-1)$ is sampled by $C_S$. In this phase, the signal path between $W$ and $Y$ is open due to the off-switch $SW_1$; thus the integrator does not generate any cross products of these two signals. The error generated in this phase is solely a function of $y(n-1)$. In the next clock phase $\phi_1$ of cycle $n$, the previous distorted $y(n-1)$ is integrated with $w(n-1)$ in a nonlinear fashion, resulting in a distorted $y(n)$ dependent on the present gain $A(n)$. In this phase, the integrator does generate cross terms of $w(n-1)$ and $y(n-1)$. However, the distortion can still be expressed solely as a function of $y(n)$. This observation indicates that the error added in each clock cycle, i.e., $e_a(n)$, is a function of two consecutive integrator outputs, $y(n-1)$ and $y(n)$. A rigorous analysis of the DTI shown in Figure 3.1(b) using the charge conservation law reveals that
\[ y(n) = \alpha_n w(n-1) + \beta_n y(n-1), \]

\[ \alpha_n = \frac{C_S}{C_S + C_p + C_F}, \quad \beta_n = \frac{C_F + C_p + C_F}{A(n) + C_S + C_p + C_F}. \]  

(3.4)

Note in Equation (3.4) that the feed-forward gain \( \alpha_n \) is dependent on \( A(n) \), whereas the feedback gain \( \beta_n \) is a function of both \( A(n) \) and \( A(n-1) \). The ideal output of the integrator, \( y_i(n) \), in Figure 3.1(c) is

\[ y_i(n) = w(n-1) + y(n-1). \]  

(3.5)

From Equations (3.4) and (3.5), the additive error is derived as

\[ e_a(n) = y(n) - y_i(n) = (\alpha_n - 1) w(n-1) + (\beta_n - 1) y(n-1). \]  

(3.6)

Substituting \( w(n-1) \) using Equation (3.4), Equation (3.6) is rewritten as

\[ e_a(n) = \frac{\alpha_n - 1}{\alpha_n} y(n) + \frac{\beta_n - \alpha_n}{\alpha_n} y(n-1) \]

\[ = -\left[ \frac{C_F - C_S}{C_S} + \frac{C_S + C_F + C_P}{C_S A(n)} \right] y(n) \]

\[ + \left[ \frac{C_F - C_S}{C_S} + \frac{C_F + C_P}{C_S A(n-1)} \right] y(n-1). \]  

(3.7)

Hence, \( e_a(n) \) is a recursive function of the integrator output and the cycle-dependent gain (a similar formulation was reported in [46]). Note that each of the two terms in Equation (3.7) is instantaneous with respect to the corresponding output, i.e., a transversal structure of memory. Using Equation (3.2), Equation (3.7) is rewritten as
\[ e_a(n) = - \sum_{i=1}^{N} a_i y_i(n) + \sum_{i=1}^{N} b_i y_i(n-1), \]

\[
a_i = \frac{C_F - C_S}{C_S} + \frac{C_F + C_P + C_P}{A_{\text{max}} C_S}, \quad b_i = \frac{C_F - C_S}{C_S} + \frac{C_F + C_P}{A_{\text{max}} C_S}, \tag{3.8}
\]

\[
a_i(i \neq 1) = \frac{k_i}{A_{\text{max}}} \cdot \frac{C_S + C_F + C_P}{C_S}, \quad b_i(i \neq 1) = \frac{k_i}{A_{\text{max}}} \cdot \frac{C_F + C_P}{C_S}.
\]

As expected, the additive error \( E_a \) is represented as a sequential power series of the integrator output without any cross terms. If capacitor mismatch is neglected, i.e., \( C_F = C_S \), then the coefficients \( a_i \)'s and \( b_i \)'s will have a fixed ratio, \((C_S + C_F + C_P) / (C_F + C_P)\), regardless of the index \( i \). Considering that \( a_i \)'s and \( b_i \)'s represent the nonlinearity of the same amplifier in two different closed-loop configurations, i.e., in \( \phi_1 \) and \( \phi_2 \), the \( a_i \)'s and \( b_i \)'s are related through the feedback factors in these clock phases. The coefficients can be learned using a gradient-descent algorithm, which will be detailed in the next chapter.

It is noticeable that the transverse structure of Equation (3.7) enables an efficient formulation without cross error terms between the sequential data. In contrast, a similar effort with the input signal would introduce numerous cross terms with memory, necessitating a full Volterra treatment [20]. To further elaborate this point, we use the simple integrator model in Figure 3.1(a) as an example. The linear \( z \)-domain transfer function from \( W \) to \( Y \) is

\[
Y(z) = \frac{\alpha z^{-1}}{1 - \beta z^{-1}} W(z), \tag{3.9}
\]

and the same function expressed from \( Y \) to \( W \) is
The former takes an IIR form while the latter assumes an FIR one. In time domain, Equations (3.9) and (3.10) can be written as

$$y(n) = \alpha w(n-1) + \beta \alpha w(n-2) + \beta^2 \alpha w(n-3) + ..., \quad (3.11)$$

$$w(n-1) = \alpha^{-1} y(n) - \beta \alpha^{-1} y(n-1). \quad (3.12)$$

It becomes clear that Equation (3.11) exhibits a long memory of $W$ while Equation (3.12) only shows a two-tap memory of $Y$. For analog-to-digital conversion, it can be argued that the formulation of Equation (3.11) is not of ultimate interest, whereas Equation (3.12) is of interest since it is the input signal that needs to be eventually digitized. Therefore, Equation (3.7) not only embodies a concise integrator-error formulation, it is also more appropriate for use in our context. In Equation (3.8), $y(n)$ can be further replaced by $d(n)$ and $e_q(n)$:

$$e_a(n) \approx -\sum_{i=1}^{N} a_i \left[ b(n) - e_q(n) \right] + \sum_{i=1}^{N} \hat{b}_i \left[ d(n-1) - e_q(n-1) \right]. \quad (3.13)$$

In a digital-domain treatment, the quantization noise is dropped in Equation (3.13), and $D$ is regarded as $Y$. Therefore, the estimated (digital) version of $e_a(n)$ is

$$e_a(n) = -\sum_{i=1}^{N} a_i d'(n) + \sum_{i=1}^{N} \hat{b}_i d'(n-1). \quad (3.14)$$

Since $E_q$ is determined by the resolution of the quantizer, which is typically low in $\Sigma\Delta$ ADCs, neglecting $E_q$ inevitably introduces fixed-point errors in estimating $E_a$. 

$$W(z) = \frac{1 - \beta z^{-1}}{\alpha z^{-1}} Y(z). \quad (3.10)$$
In turn, this limits the calibration accuracy. Comparing Equations (3.13) and (3.14), the estimation error can be defined as

\[ e_c(n) = e_q(n) - e_o(n) \]

\[ = -\sum_{i=1}^{N} a_i [d_i'(n) - y_i'(n)] + \sum_{i=1}^{N} b_i [d_i'(n-1) - y_i'(n-1)] \]

\[ \approx -\sum_{i=1}^{N} a_i [y_i^{-1}(n)e_q(n)] + \sum_{i=1}^{N} b_i [y_i^{-1}(n-1)e_q(n-1)] \]  

(3.15)

where \(|Y| >> |E_q|\) is assumed. In general, \(|a_1| >> |a_i|\) and \(|b_1| >> |b_i|\) for \(i \neq 1\) are true, and the higher-order terms in Equation (3.15) can be neglected, yielding (Appendix A.1)

\[ \overline{E_e}\approx \frac{1}{12\text{OSR}A_{\max}^2} \left[ \frac{1}{A_{\max}^2} + \frac{\gamma(\gamma - 1)\pi^2}{3\text{OSR}^2} \right]. \]  

(3.16)

for an \(N\)-bit quantizer, where \(\gamma = (C_S + C_F + C_P) / C_S\).

In Equation (3.16), the power of the fixed-point error is inversely proportional to \(\text{OSR}A_{\max}^2\), whereas the in-band quantization noise of the first-order modulator is inversely proportional to \(\text{OSR}^3\). Thus, an \(A_{\max}\) sufficiently larger than \(\text{OSR}\) is required to avoid degrading the SNR of the modulator. If this degradation is allowed to be no more than 3 dB, i.e.,

\[ \overline{E_e}\approx \frac{1}{12\text{OSR}A_{\max}^2} \left[ \frac{1}{A_{\max}^2} + \frac{\gamma(\gamma - 1)\pi^2}{3\text{OSR}^2} \right] < \frac{1}{12 \cdot 4^N \text{OSR}^3}, \]

(3.17)

then the minimum \(A_{\max}\) is obtained as

\[ A_{\max} \geq \sqrt[3]{\frac{3\text{OSR}^2}{\pi^2} + \gamma(\gamma - 1)} \approx \frac{\sqrt{3}\text{OSR}}{\pi}, \]  

(3.18)
where $\gamma \approx 2$ ($C_F \approx C_S \gg C_P$) and $\text{OSR} \gg \pi$ are assumed. For example, when OSR is 16, a gain of around 9 is needed to guarantee less than 3-dB SNR loss.

Further investigation reveals that the fixed-point error of the calibration is the same as the input-referred quantization noise leakage due to an insufficient (constant) amplifier gain (Appendix A.1). This can be explained in the aspect of the signal-dependent pole-zero movement. By multiplying both sides of Equation (3.1) with the inverse signal transfer function, we obtain

$$\frac{1}{\alpha} d(n) + \frac{\alpha - \beta}{\alpha} d(n-1) = x(n-1) + \frac{e_q(n) - \beta e_q(n-1)}{\alpha}.$$  \hspace{1cm} (3.19)

In contrast, from Equations (3.3), (3.7), and (3.13), the calibrated output $D_c$ after eliminating the estimated error $E_d$ from $D$ is given by

$$d_c(n) = d(n) - e_q(n) = \frac{1}{\alpha} d(n) + \frac{\alpha - \beta}{\alpha} d(n-1),$$  \hspace{1cm} (3.20)

where, as explained before, $1/\alpha$ and $(\alpha - \beta)/\alpha$ are power series of $d(n)$ and $d(n-1)$, respectively. By comparing Equations (3.19) and (3.20), it becomes clear that the elimination of $E_d$ effectively calibrates the signal-dependent pole movement, thus the distortion in $X$. However, the quantization noise leakage from the zero movement (the right-hand side of Equation (3.19)) still remains, resulting in the fixed-point error. It can be argued that the minimum gain set by Equation (3.18) is quite relaxed for a low-to-medium OSR, practically not limiting the performance of the calibration.
3.2 Second-Order Modulator

The same analysis is extended to the second-order modulator with two delaying integrators shown in Figure 3.3 [47]. The non-ideal modulator in Figure 3.3(a) with a signal-dependent transfer function is replaced by the ideal modulator with two additive error signals, $E_{a1}$ and $E_{a2}$, in Figure 3.3(b).

![Figure 3.3. (a) Conventional and (b) output-referred distortion models of a second-order $\Sigma\Delta$ modulator.](image)

In this case, the modulator transfer function of in Figure 3.3(b) is given by

$$D(z) = z^{-2}X(z) + z^{-1}E_{a1}(z) + (1 - z^{-1})E_{a2}(z) + (1 - z^{-1})^2 E_y(z). \quad (3.21)$$
Similar to the first-order case, the integrator errors in the second-order modulator are also functions of the integrator outputs $Y_1$ and $Y_2$, respectively. Therefore, from Equation (3.21), the total error at the modulator output is

$$e_a(n) = e_{a1}(n-1) + e_{a2}(n) - e_{a2}(n-1)$$

$$\approx \sum_{i=1}^{N} \left\{ -a_{i1}y_i^1(n-1) + b_{i1}y_i^1(n-2) - a_{i2}y_i^2(n) + (b_{i2} + a_{i2})y_i^2(n-1) - b_{i2}y_i^2(n-2) \right\},$$

(3.22)

where $Y_1$ and $Y_2$ are derived from the state-space equations of the $\Sigma\Delta$ modulator as

$$Y_2(z) = D(z) - E_q(z),$$

$$Y_1(z) = \frac{(1 + z^{-1})D(z) - E_{a1}(z) - (1 - z^{-1})E_q(z)}{z^{-1}},$$

(3.23)

Note the dependence on $E_{a2}$ in the $Y_1$ expression in Equation (3.23), which dictates a sequential error estimation procedure in calibration, wherein $E_{a2}$ is determined first and its result is applied in determining $E_{a1}$ subsequently.

Since the quantization noise is again neglected in a digital-domain treatment, from Equations (3.22) and (3.23), the fixed-point error of the calibration can be derived as (Appendix A.2)

$$E_c^2 \approx \frac{1}{12 \text{OSR}} \left[ \frac{1}{A_1^2 A_2^2} + \left( \frac{1}{A_1} + \frac{1}{A_2} \right)^2 \frac{\pi^2}{3 \text{OSR}^2} \right],$$

(3.24)

where similar assumptions in deriving Equation (3.16) are used. Again, the fixed-point error is identical to the input-referred quantization noise leakage (Appendix A.2). Thus, Equation (3.24) implies that the second-stage amplifier gain is as important as that of the first-stage one in terms of the fixed-point error.
For an SNR loss of less than 3 dB, the fixed-point error in Equation (3.24) should be less than or equal to the in-band quantization noise power of an ideal second-order modulator:

\[
\frac{1}{12\text{OSR}^{4\pi}} \left[ \frac{1}{A_{i}^{2}} + \left( \frac{1}{A_{1}} + \frac{1}{A_{2}} \right)^{2} \frac{\pi^{2}}{3\text{OSR}^{2}} \right] \leq \frac{1}{12 \cdot 4^{\pi}} \frac{\pi^{4}}{5\text{OSR}^{2}} \tag{3.25}
\]

Assuming that the amplifier gains of both stages are the same, the minimum gain \(A_{\text{max}}\) is determined as

\[
A_{\text{max}} \geq \frac{\sqrt{7.3 \text{ OSR}}}{\pi} \tag{3.26}
\]

For example, when OSR is 16, \(A_{\text{max}}\) of around 14 is needed.

It now becomes rather interesting to determine the gain requirement for each stage in a second-order modulator. As \(E_{a_{2}}\) is first-order shaped compared to \(E_{a_{1}}\) in Equation (3.21), the amplifier gain of the second stage can be more relaxed than that of the first stage. Since \(E_{a_{1}}\) and \(E_{a_{2}}\) capture the nonlinearity of the modulator, this is a distortion-oriented argument. In contrast, Equation (3.24) suggests that both stage gains are equally important when the fixed-point error, or the input-referred quantization noise leakage, is under consideration. This is the noise perspective. Following these arguments, a low OSR seems to be desirable in this approach, which suggests that the algorithm is more suitable for treating wideband \(\Sigma\Delta\) modulators. This is in line with the general understanding that wideband \(\Sigma\Delta\) ADCs usually utilize low-order modulators with a low OSR.
Figure 3.4. Generalized output-referred distortion models incorporating both integrator nonlinearity and DAC mismatch errors: (a) first- and (b) second-order ΣΔ modulators.

3.3 DAC Error

The above output-referred error analysis/model of the first- and second-order modulators can be generalized and applied to single-loop ΣΔ modulators of any order and even the cascaded (MASH) architectures. Specifically, the approach is also useful in analyzing the capacitor mismatch errors of multibit feedback DACs as shown in Figure 3.4(a), where $E_c$ is the DAC mismatch error dependent on the output code $D$ and can be approximated by a polynomial function:
\[ e_i(n) = \sum_{i=1}^{N} c_i d_i'(n). \]  

(3.27)

With \( E_c \), the modulator transfer function is revised to be

\[ D(z) = z^{-1} \left[ X(z) - E_i(z) \right] + E_d(z) + (1 - z^{-1}) E_q(z). \]  

(3.28)

Equation (3.28) indicates that \( E_c \) can be lumped into \( E_d \) in the treatment. Neglecting the quantization noise, the total error \( E_d \) is revised to be

\[ e_d(n) = -\sum_{i=1}^{N} a_i d_i'(n) + \sum_{i=1}^{N} (b_i - c_i) d_i'(n-1). \]  

(3.29)

Theoretically, Equation (3.29) implies that there is no need to differentiate between the integrator and DAC errors. For the second-order modulator in Figure 3.4(b), the transfer function and error model can be derived in a similar manner:

\[ D(z) = z^{-2} \left[ X(z) - E_{i1}(z) \right] + z^{-1} E_{d1}(z) + (1 - z^{-1}) \left[ E_{d2}(z) - 2z^{-1} E_{i2}(z) \right] + (1 - z^{-1})^2 E_q(z), \]  

(3.30)

\[ e_d(n) = \sum_{i=1}^{N} \begin{pmatrix} -a_i d_i'(n-1) + (b_i - c_i) d_i'(n-2) \\ -a_2 d_i'(n) + (b_{i2} + a_{i2} - 2c_{i2}) d_i'(n-1) \\ -(b_{i2} - 2c_{i2}) d_i'(n-2) \end{pmatrix}. \]  

(3.31)

In practice, capacitor mismatch tends to cause fairly high-order nonlinearities in Equation (3.27), while lower-order distortions dominate in weakly nonlinear feedback amplifiers of Equation (3.2). Therefore, it may not be sensible to lump the two together as is done in Equations (3.29) and (3.31). In Chapter 4, the proposed digital calibration will be restricted to treating the integrator nonlinearity,
and the multibit DAC will be considered ideal. For the DAC, a dedicated model and calibration scheme will be presented Chapter 5.

3.4 Reversed Path Method

So far, the circuit model is derived from the individual circuit analysis, i.e., the charge conversation law. In this section, the generalization of the proposed error modeling method is provided by introducing top-down viewpoints.

The substantial argument suggested in this chapter is that the long term memory error in all-pole systems can be modeled as an FIR form of the output; that is, when the system has zeros, the error model becomes an IIR form. For better insight of this approach, we can think about the ideal integrator in Figure 3.5. The integrator output $Y$ contains all the history of the input $X$, which means that $Y$ is an IIR form of $X$. However, the sum of the previously stored input in the integrator is same as the previous output $Y(n-1)$, so the input can be easily calculated from the present and previous output without knowing the input history.

The idea developed in the output-referred error analysis is related to the nonlinear system identification techniques named ‘reverse path’ identification in mechanical engineering [48],[49]. One of the factors increasing the complexity in the conventional circuit analysis is that the derivation of the model initiated as an input function leads to an IIR form when the system includes a feedback. The
conversion of the input function into the output function needs to be followed for the calibration purpose. The reverse path method reveals that the modeling becomes much simpler by changing the roles of input and output and hence directly representing the error as an output function; that is, a feedback circuit as an input function is translated as a feed-forward one as an output function. As a result, in the case shown in Figure 3.5, it can be safely said that the input of any non-ideal integrator can be figured out from the present and previous output if it has only a one-tap memory.

\[ y[n] = x[n-1] + x[n-2] + \ldots + x[0] \]
\[ y[n] = x[n-1] + y[n-1] \]

Figure 3.5. Output-referred error analysis with a feedback.

Figure 3.6 illustrates another case when the output-referred error model can be effective. With a single error source in Figure 3.6(a), the error model represented by either input or output does not show much difference in the analysis. However, with multiple error sources in Figure 3.6(b), the error as an input function requires iterative calculation to reflect the interaction between the error sources, whereas the model derived from the output naturally includes the effect of the other error sources and simplifies the analysis.
Figure 3.6. Error modeling with (a) a single source and (b) multiple error sources.

For the extension of the reverse path approach to a DT circuit, a circuit component with multiple finite steady states is divided into multiple components with a single state (memoryless). For example, an integrator with integrating and holding states is conceived as two components, integrating and holding blocks, with a single state. After this decomposition, the nonlinearity of each constituent block is dependent on only the corresponding state and expressed as a polynomial function (Taylor series), which well explains the two-tap polynomial DTI model derived from the circuit analysis in Section 3.1.

To further facilitate the analysis, each single-input single-output (SISO) nonlinear component is transformed into a linear component with an additive input as shown in Figure 3.7. In the example of Figure 3.8, DT circuits with two nonlinear components dependent on the state $Y_1$ and $Y_2$, respectively, are replaced by the ideal components with two separate error sources $E_1$ and $E_2$. In turn, the nonlinear SISO circuit is replaced by a linear multi-input single-output (MISO)
one. Once $E_1$ and $E_2$ are analyzed, the effect of each error at the output is easily calculated by the ideal transfer function.

$$E = f(Y) - X$$

Figure 3.7. Transformation of a SISO nonlinear component into two-input single-output linear component.

Since this approach represents the circuit non-ideality based on the states, the identification of the states has to be preceded. This process will be detailed in the next chapter, so the argument here will be focused on the modeling of the non-ideality related to the identification.

As the name implies, the reverse path method identifies the system from the output to the input backward. In Figure 3.8, once $Y_2$ is estimated from $Y$ using
a linear transfer function, $E_2$ is estimated from $Y_2$. Then, $X_2$ is estimated from $Y_2$ and $E_2$. Again, $Y_1$ is estimated from $X_2$ and so on. Conventionally, to estimate the error sources $E_1$ and $E_2$, this approach requires the measurement at the location of nonlinearity, which is often not so practical if the system has multiple nonlinear components. Particularly, to measure the analog circuit without affecting the probing point, highly accurate buffers or/and ADCs are needed. In this work, each state is identified by injecting a known two-level PN signal at the location of nonlinearity instead of the direct measurement.

The use of a two-level PN has an advantage in modeling of non-ideality related to the identification. The inherent linearity of the two-level signal does not cause the nonlinear error in PN injection. However, it can still incur the gain error because the exact amount of PN injected in the analog domain is unknown. When the PN signal is scaled, it can be equivalently translated as a gain error at the corresponding signal path as illustrated in Figure 3.9. Since the digital calibration compensates for the gain error, the net result of the PN mismatch becomes the overcompensated residue gain error of the input signal of each nonlinear block (Appendix II details the effect of the PN mismatch in $\Sigma\Delta$ modulator modeling). If multiple one-bit PN signals identify a single path gain with different values, we can take one PN as a master and scale the other PN signals according to the master PN in calibration. Owing to the insensitivity of the gain error in nonlinear calibration, the injected PN can even be cancelled out just after passing through
the nonlinear component to maximize the signal dynamic range, which motivates the PN cancellation concept in Chapter 5.

\[ X_{\alpha} = \frac{X}{\alpha} + E \]

After Calibration

Figure 3.9. PN mismatch and equivalent error model.

Although the nonlinear memory error model is mainly applied in the context of the \(\Sigma\Delta\) conversion, it can be extensible to the calibration of other types of mixed-signal processing. In the pipeline ADC, the settling of the amplifier output is usually affected by the previous output, or more precisely, the digitized output stored in the next stage DAC capacitor, which cannot be reset in every clock cycle. Thus, the similar two-tap model is applicable to addressing the memory effect. It is also considered when analog blocks are shared in multi-channel signal process [50],[51].
CHAPTER 4

CALIBRATION BASED ON INDEPENDENT COMPONENT ANALYSIS

The previous chapter describes modeling of nonlinear $\Sigma\Delta$ modulators by the output-referred distortion analysis. The remaining task of the calibration is to determine the coefficients of the model. System identification is one of the most extensively studied topics in digital signal processing [52]. In conventional ADC calibration approaches, a known test signal is injected into the input to identify the conversion error parameters. Alternatively, accurate information of the input signal can be provisioned by a reference path to facilitate the model adaptation [25]–[27]. When the input signal is band-limited, e.g., in $\Sigma\Delta$ ADCs, blind compensation based on the spectral spreading effect of nonlinear distortion can also be applied to direct the learning process [53],[54]. Recently, one system identification method using low-level PN injection has become widely popular due to the background nature of the approach and the ease of injection due to the existence of a DAC in nearly every multistage ADC architecture [21],[22],[31]–[39]. In this technique, a single one-bit PN is used to determine one error parameter, which is often a constant inter-stage gain error. The procedure was
recently extended to nonlinear calibration [37]. Since the odd moment of a one-bit PN is simply itself and the even moments are a constant, without additional information, the technique requires the employment of multiple PNs to identify various nonlinear coefficients such as those in Equation (3.8), complicating the analog circuitry involved in the calibration.

4.1 First-Order Modulator

Considering that the nonlinear distortion of an amplifier is dependent on its output signal level, a joint nonlinear system identification procedure can be devised incorporating both the PN and the output signal information. In other words, based on the independence between the PN and any other signals, the signal-dependent variation of the correlation between the PN and the output signal (and its moments) potentially provides more information about the distortion mechanism. In turn, a single PN can be used, and the analog complexity associated with calibration can be minimized.

The nonlinear model parameter identification of a first-order $\Sigma\Delta$ modulator is facilitated by injecting a single one-bit PN at the modulator input as shown in Figure 4.1(a). With the sum of $T$ and $X$ being the input, $E_o$ can be expressed as follows by combining Equations (3.3) and (3.13):
When expanded, the power series will generate numerous cross terms between $X$ and $T$. Out of these terms, $-x^{i-1}(n-1)t(n-1)$ and $x^{i-1}(n-2)t(n-2)$, which are proportional to the coefficients $a_i$ and $b_i$, respectively, are of particular interest.

If $X$ is known, we can identify $a_i$ and $b_i$ by correlating the PN with various moments of the input signal ($X^{i-1}T$). In this correlation, a higher-order term contains additional nonlinear information which is not presented in a lower-order term. For example, the term $XT$ includes extra independent signal components than $T$ alone does. The nonlinear distortion generates other cross terms such as $X^2T^2$, $X^3T$, etc., as well in Equation (4.1). However, the power of $X$ is typically larger than that of $T$, so $X^{i-1}T$ is the most appropriate term for observing the distortion in such a joint estimation.

One question that arises is how to obtain accurate information of $X$ in this procedure. We hereby propose a bootstrapped method of parameter identification, where the calibrated output digital code $D_c$ is utilized in lieu of $X$ in the calculation of the various moments of $X$. When the modulator input is $X + T$, using Equation (3.3), $X$ can be written as

$$x(n-1) = d(n) - t(n-1) - e_d(n) - e_q(n) + e_q(n-1) \approx d_c(n).$$  (4.2)
In other words, $D_c$ faithfully represents $X$ when quantization noise is neglected and the modulator distortion is removed. Therefore, although $D_c$ initially may differ significantly from $X$, as the adaptation procedure iterates, it approaches $X$ asymptotically. The process will reach steady state when $D_c$ becomes uncorrelated with $T$, which only occurs when all error parameters are properly identified and the overall system, i.e., the cascade of the modulator and the PTF becomes linear.

Figure 4.1(a) illustrates the proposed calibration scheme in detail. A one-bit PN sequence $T$ is injected at the input using a small, dedicated capacitor. The corrected output $D_c$ is obtained by subtracting out $T$ (exact digital version) and $E_d$ (estimated modulator distortion) from the raw output $D$. The out-of-band quantization noise in $D_c$ is filtered out by a low-pass filter (LPF) before being applied to the moment correlator to determine the PTF coefficients; i.e., the filter output $E$ is substituted for $X$. Since it is the same modulator nonlinearity that causes the in-band as well as the out-of-band distortions, the effort of minimizing the in-band error will cancel out the out-of-band one at the same time. Owing to the noise shaping of the $\Sigma\Delta$ modulator, the high-level quantization noise is efficiently removed with a relatively low-order LPF. In turn, the convergence of the learning process is expedited without the interference from the large out-of-band quantization noise. A gradient-descent algorithm is utilized in the learning process and the update equations are
Figure 4.1. Proposed adaptive digital calibration scheme of a first-order ΣΔ modulator: (a) architecture, (b) error filter, and (c) LMS update block.

\[ a_i^{m+1} = a_i^m - \mu_{ai} \left[ e^{i-1}(n)D(n-1) \right] e(n), (i = 1, 2, \ldots, N), \quad (4.3) \]

\[ b_i^{m+1} = b_i^m + \mu_{bi} \left[ e^{i-1}(n)D(n-2) \right] e(n), (i = 1, 2, \ldots, N), \quad (4.4) \]

where the LPF latency is neglected for simplicity and \( m = kn \) \((k \geq 1 \text{ is an integer determined by the LMS block size})\).
A practical issue that arises is how to choose the step sizes in Equations (4.3) and (4.4). Since higher-order distortion terms typically have lower amplitudes than lower-order terms, larger step sizes for higher-order coefficients are needed for fast training. However, as the various polynomial terms are not orthogonal, the interaction between these terms presents practical difficulties in adaptation, often manifested as long convergence time or even instability. Since the distortion dealt with in this work is static (stationary), the step sizes were not aggressively optimized in the subsequent simulations, which helped to minimize the mean square error (MSE) in the steady state.

Another issue related to the quantizer resolution needs to be further discussed here. As a ΣΔ modulator typically employs low-resolution quantizers, the out-of-band quantization noise may have large power due to noise shaping. This motivates the introduction of an LPF in Figure 4.1(a). The LPF can actually be considered as a form of the decimation filter (without down-sampling) to derive a better approximation of $X$. If the quantizer resolution decreases further (potentially to as low as one-bit in ΣΔ ADCs), the resulting large out-of-band noise (especially in a first-order modulator, in which the quantization noise is highly correlated with the input) dictates a narrower pass-band of the LPF or a larger PN to be used. The former technique also decreases the in-band PN-to-signal ratio, which slows down the convergence of the algorithm; in the latter case, a larger PN implies higher loss in dynamic range. In contrast, when the
quantization noise is small compared to the PN (e.g., a high-resolution quantizer is used), the LPF is not necessary, or can even be replaced by a high-pass filter (HPF) to accelerate the calibration. When this is the case, $E$ does not include $X$; thus, $D_e^{i-1}T$ instead of $E^iT$ should be used in the correlator in Figure 4.1(c) to approximate $X^{i-1}T$. As a result, a higher resolution quantizer is preferred in terms of convergence speed and calibration accuracy. In this aspect, MASH structures [55],[56], which offer a high effective resolution, are worth consideration.

One unique feature of this calibration approach is that a single one-bit PN is utilized to simultaneously identify multiple coefficients in a nonlinear system model. The PN injection can be accomplished by the introduction of a separate, small capacitor at the summing node of the integrator, which involves negligible analog effort. In addition, as a scaling of the PN is equivalent to a corresponding linear scaling in $X$ without any impact on linearity, the calibration is insensitive to any potential mismatch of this capacitor in practice (Appendix B.1). When an accurate signal gain is required, the PN injection can be incorporated into the feedback DAC. Furthermore, as the calibration is insensitive to the amplitude variation in $T$, the convergence time can be easily shortened by trading off the signal dynamic range. Meanwhile, considering that $X$ is generally larger than $T$, the proposed calibration based on the correlation of $X^{i-1}T$ potentially converges faster than alternative methods such as the one described in [57], which is based on the correlation between various PNs, $T_1$, $T_2$, ..., $T_i$. On the other hand, a small
$X (X \approx 0)$ decreases the convergence speed seriously, which may bring about some practical issues. However, theoretically, it can be argued that the effect of nonlinear error is not critical for a small $X$.

### 4.2 Second-Order Modulator

In the second-order modulator shown in Figure 4.2(a), two PN sequences, $T_1$ and $T_2$, are injected at the summing nodes of the integrators using separate, small capacitors to estimate the integrator errors. Similar to the first-order case, the errors $E_{a1}$ and $E_{a2}$ are expressed as functions of the integrator outputs $Y_1$ and $Y_2$, respectively; the error model is constructed by replacing $Y_1$ and $Y_2$ with their digital versions $D_1$ and $D_2$, respectively. While $D_2$ can be obtained directly from the quantizer output, $D_1$ is calculated using Equation (3.23):

$$D_1(z) = \frac{(1+z^{-1})D_2(z) - z^{-1}T_2(z) - E_{a2}(z)}{z^{-1}}.$$  \hspace{1cm} (4.5)

The distortion terms are estimated as

$$e_{d2}(n) = -\sum_{i=1}^{N} a_{2i} d'_2(n) + \sum_{i=1}^{N} b_{2i} d'_2(n-1),$$

$$e_{d1}(n) = -\sum_{i=1}^{N} a_{1i} d'_1(n) + \sum_{i=1}^{N} b_{1i} d'_1(n-1).$$  \hspace{1cm} (4.6)
An intuitive justification of obtaining $D_1$ from $D_2$ resides in the fact that the inner loop of the modulator can be viewed as a first-order $\Sigma\Delta$ ADC digitizing $Y_1$; thus its in-band quantization noise (estimation error) is greatly suppressed by
noise shaping. Also, in Equation (4.5), the mismatch of $T_2$ causes a linear scaling of $D_1$ (Appendix B.2) without impacting the nonlinear error correction. If an accurate signal gain of $D_1$ is required, $T_2$ can be injected into the DAC of the second stage instead of using a dedicated capacitor. The ideal error-corrected digital output is

$$D_{e}(z) = D_2(z) - z^{-2}T_1(z) - z^{-1}(1-z^{-1})T_2(z) - z^{-1}E_{d1}(z) - (1-z^{-1})E_{d2}(z),$$  \hspace{1cm} (4.7)

which is realized in Figure 4.2(b). Note that a mismatch tap 1+c for $T_1$ is also included in Figure 4.2(b). The 1+c tap is necessary to accurately cancel the $T_1$ term in the digital domain if the mismatch in $T_2$ is considered as a scaling of $X$ (Appendix B.2).

The coefficients $a_{il}$’s and $b_{il}$’s of the error model are trained again using a bootstrapped, nonlinear system identification technique based on signal independency. The moment calculation requires the error-corrected versions of $D_1$ and $D_2$, which are given as

$$d_{a(i)}(n) = d_i(n) - t_i(n-1) - e_{a(i)}(n).$$  \hspace{1cm} (4.8)

It should be pointed out that $D_{c2}$ is different from the ultimate calibrated output $D_c$ of Equation (4.7), i.e., $D_{c2}$ is uncorrelated with $T_2$ after the calibration is complete; however, it still contains $T_1$. In contrast, $D_c$ will have both $T_1$ and $T_2$ completely eliminated. The LMS update equations to determine $a_{il}$’s and $b_{il}$’s are

$$a_{il}^{m+1} = a_{il}^m - \mu_{a(i)} \left[ d_{a(i)}^{l-1}(n-2) \right] e(n),$$  \hspace{1cm} (4.9)

- 49 -
\[ b_{li}^{m+1} = b_{li}^m + \mu_{li}(1) \left[ d_{e1}^{l-1}(n-2)\tau_1(n-3) \right] e(n), \]  
\[ (4.10) \]

where the latency of the LPF is ignored and \( m \) is defined the same way as in Equations (4.3) and (4.4). As the second-stage error \( E_{d2} \) is noise-shaped by \( 1-z^{-1} \) in Equation (4.7), the LMS update equations for \( a_{2i} \)’s and \( b_{2i} \)’s are given by

\[
a_{2i}^{m+1} = a_{2i}^m - \mu_{a(2i)} \cdot \left[ d_{i2}^{l-1}(n)\tau_2(n-1) - d_{i2}^{l-1}(n-1)\tau_2(n-2) \right] e(n),
\]
\[ (4.11) \]

\[
b_{2i}^{m+1} = b_{2i}^m + \mu_{b(2i)} \cdot \left[ d_{i2}^{l-1}(n-1)\tau_2(n-2) - d_{i2}^{l-1}(n-2)\tau_2(n-3) \right] e(n).
\]
\[ (4.12) \]

The additional tap \( c \) in Figure 4.2(b) is trained using the memory effect, i.e., the residual \( T_1 \) existing in \( E \). Thus, the correlation terms between \( E \) and \( T_1 \) not used in Equations (4.9) or (4.10) can be used for training \( c \). For example,

\[
c^{m+1} = c^m + \mu_c\tau_1(n-4) e(n).
\]
\[ (4.13) \]

### 4.3 Behavioral Simulation

The behavioral simulations of first- and second-order \( \Sigma\Delta \) modulators with a 5-bit quantizer for a sinusoidal input are performed using MATLAB. The OSR is 16 and the input frequency is set to approximately 1\% of the clock frequency. The input signal power with respect to the full swing is −2.6 dB for the first modulator and −2.9 dB for the second. Multibit DAC and switch nonlinearities are
excluded in the simulation. However, the mismatch between the feedback and total sampling capacitors is accounted for.

Figure 4.3. MATLAB simulation results of a 5-bit, first-order \( \Sigma \Delta \) modulator (OSR = 16) before calibration: (a) \( A_{\text{max}} = 10^8 \) and (b) \( A_{\text{max}} = 20 \), (c) the same as (a) with an injected PN, and after calibration: (d) \( A_{\text{max}} = 20 \) (without capacitor mismatch), (e) \( A_{\text{max}} = 20 \) (with 5 % PN and \( C_F \) mismatch), (f) \( a_i \)'s, (g) \( b_i \)'s, and (h) SNDR and SFDR after calibration.

The PN mismatch effect is also considered. The model of the nonlinear amplifier observes Equation (3.2) with \( k_1 = 1, k_2 = 0.04, k_3 = 0.4, k_4 = 0.02, k_5 = 0.2 \), and a fixed \( A_{\text{max}} \). \( A(n) \) varies in each cycle according to \( y(n) \). To precisely determine \( y(n) \) in each cycle using Equation (3.2), an iterative procedure is
followed with a predetermined error bound of $5 \times 10^{-5}$. A fifth-order PTF is adopted in the simulation.

The simulation results of a first-order modulator are shown in Figure 4.3. Figures 4.3(a) and (b), corresponding to an $A_{\max}$ of $10^8$ and 20, respectively, illustrate the normalized reconstructed spectra of the modulator output without calibration. The signal-dependent gain variation leads to gross harmonic distortions evident by the low SNDR and spurious-free dynamic range (SFDR). In this case, the distortion causes around 40% gain degradation for the maximum output signal. Figures 4.3(c) and (d) show the output spectra with PN injection for the same setup as in Figures 4.3(a) and (b), respectively (the injected PN is removed in digital domain). A 2.5-dB loss in SNDR results from the PN injection comparing Figures 4.3(a) and (c). A one-bit PN with an amplitude of ±0.25 (25% of the full swing) is utilized considering the tradeoff between the calibration time and signal dynamic range. Figure 4.3(d) clearly shows the effectiveness of the proposed scheme. Note that a superior SFDR is observed in Figure 4.3(d) due to the dithering effect of the injected PN that randomizes the pattern noise otherwise clearly visible in the output spectrum of a first-order modulator (Figure 4.3(a)). For comparison, a similar linearity ($\geq 70$-dB SFDR) can be achieved with an $A_{\max}$ of at least 1000 without calibration, whereas it is set to 20 in this simulation.

Comparing Figures 4.3(c) and (d), the fixed-point error in calibration causes an additional drop in SNDR of about 3 dB. If SNDR or power efficiency is
critical, a smaller PN can be used in the calibration at the cost of a longer adaptation time. The feedback capacitor $C_F$ and the capacitor $C_T$ for PN injection are set to 5% less than their nominal values, and the performance of the modulator including the mismatch is shown in Figure 4.3(e). The SNDR result is slightly better than that in Figure 4.3(d) since a smaller $C_F$ leads to lower values of $a_1$ and $b_1$ to begin with according to Equation (3.8). More intuitively, this is due to the larger integrator gain with a smaller $C_F$ that partially compensates for the effect of integrator leakage. The mismatch of $C_T$ results in, as expected, a linear scaling of the input signal in calibration (Appendix B.1). Mismatches of ±5% do not seem to limit the performance of the calibration in this simulation.

Figures 4.3(f) and (g) illustrate the learning curves of the PTF coefficients without capacitor mismatch. The value of $b_1$ is around 60% of that of $a_1$ as expected from Equation (3.8) for $C_P \approx 0.25C_S$ and $C_S \approx C_F$. As mentioned before, the non-orthogonality of the polynomial and the resulting correlation between the various nonlinear terms in Equation (3.8) dictate small step sizes in the adaptation, which result in several hundred million iterations to reach steady state. However, as evident in Figure 4.3(h), the performance of the modulator, e.g., SNDR and SFDR, has settled even before all PTF coefficients have completely settled, with a ≥ 70-dB SFDR achieved within one hundred million samples.

A simple 6th-order FIR LPF with a 30-dB out-of-band rejection designed using the Parks-McClellan algorithm [52] is used in this simulation. Large out-of-
band quantization noise of quantizers with even lower resolutions (≤ 5-bit) can be rejected efficiently. The steady-state SFDR variation in Figure 4.3(h) mostly results from the commensurate periodicity of the sinusoidal input and the PN sequence. As mentioned before, further acceleration of convergence is possible by replacing the LPF by an HPF when high-resolution quantizers are used.

![SNDR and SFDR](image)

Figure 4.4. MATLAB simulation results of a 5-bit, second-order ΣΔ modulator (OSR = 16) before calibration: (a) \( A_{\text{max}} = 10^8 \), and (b) \( A_{\text{max}} = 40 \), (c) the same as (a) with an additional injected PN, and after calibration: (d) \( A_{\text{max}} = 40 \) (without capacitor mismatch), (e) \( A_{\text{max}} = 40 \) (with 2% PN and \( C_F \) mismatch), and (f) SNDR and SFDR after calibration.
Figure 4.4 summarizes the simulation results of a second-order modulator. $A_{max}$’s of 4000 and 400 are required for the first- and second-stage amplifiers, respectively, to guarantee a $\geq 80$-dB SFDR without calibration. Figure 4.4(d) shows the calibrated modulator performance with an $A_{max}$ of 40 set for both stages, demonstrating the effectiveness of the APTF. This gain is chosen according to Equation (3.26) considering the fixed-point errors. One-bit PN sequences, $T_1$ and $T_2$, with an amplitude of $\pm 0.125$ (12.5% of the full swing) are used. The PN injection results in a 2.8-dB SNDR loss, while the fixed-point error causes an additional 2.5-dB loss. In this simulation, the PN-injection and feedback capacitors are also set to 2% less than their nominal values, and the calibration is found insensitive to these variations as depicted in Figure 4.4(e). A fifth-order APTF model for both first- and second-stage integrators is selected; thus the complete model contains 21 variables ($a_{11}$–$a_{15}$, $a_{21}$–$a_{25}$, $b_{11}$–$b_{15}$, $b_{21}$–$b_{25}$, and $c$) to be trained. Figure 4.4(f) shows the learning curves of the SNDR and SFDR. Compared to the first-order modulator, it achieves a significantly larger SNDR using the same 5-bit quantizer, which in turn requires much smaller step sizes in the LMS update. In addition, the smaller amplitude of the PNs further elongates the calibration. In general, harmonic distortion is found very sensitive to coefficient variations in the PTF, which explains the residual spurs in Figures 4.4(d) and (e), and the SFDR variations in Figures 4.4(f) and 4.3(h). Fortunately, the quantization noise is less structured in a second-order modulator and further
randomized by the two independent PNs, so the correlation between the quantization noise and the PN diminishes. Therefore, an HPF is applicable even with relatively low amplitude test signals to speed up the adaptation. In the simulation, an 8th-order HPF attenuates the in-band signal by around 40 dB.

### 4.4 Independent Component Analysis

The proposed identification approach can be explained by the independent component analysis (ICA) [58]–[60]. The uniqueness of the proposed approach is that both input and PN information is used together in correlation based on the independence of PN. Similarly to the conventional case, if the adaptive filter is the exact inverse of the ADC in Figure 4.5, the calibrated data $\hat{X}$ after subtracting $PN$ does not include any inter-modulation terms between $X$ and $PN$, i.e., the $\hat{X}$ is independent of any cross moments between $X$ and $PN$. Thus, the adaptive filter can be identified by the correlation with the high-order moments of $X$ and $PN$. Since $X$ is unknown, $X$ is recursively estimated from $\hat{X}$.

The input $X$ includes multi-bit information, so many high-order moment terms can be generated even with a single one-bit $PN$. At the same time, the correlation relying on the $PN$ signal is capable of calibrating the memory error by exploiting the delayed moment of the $PN$, which is independent of the original PN.
Moreover, $X$ independent to $PN$ does not necessarily need to be removed before correlation.

![Proposed calibration scheme based on ICA.](image)

Figure 4.5. Proposed calibration scheme based on ICA.

One of the obstacles limiting the calibration accuracy and speed in the PN based calibration is the implementation of the $X$ removal block in Figure 4.5. In the oversampling structure, an HPF can be employed for this purpose. To apply the proposed identification scheme to near Nyquist rate signal processing, a more efficient way to attenuate the perturbation caused by the input signal in correlation needs to be considered [61]. In the same aspect, the recently published method [62], which limits the variance of the input signal by the selective PN injection, is also worthy of notice.
CHAPTER 5

CAPACITOR MISMATCH CALIBRATION

To date, only a limited number of digital background calibration techniques have been published [28],[45],[63]–[66] in multi-bit DACs, especially in the context of ΣΔ conversion. Most previous works are based on randomization of the capacitor switching sequence in a multi-bit DAC, such that the signature of the randomization embedded in the DAC output can be utilized to identify the associated capacitor mismatch errors. Typically operating in the background, the procedure converges when the DAC output becomes uncorrelated with the sequence of the randomization after treatment, regardless of the input signal characteristics. The approaches in [28],[45],[63] simultaneously scramble the switching of all capacitors, which still necessitates a complicated DEM encoder and thus limits the maximum operating speed. The encoder complexity associated with calibration can be somewhat relaxed in [64]–[66] by scrambling one pair of capacitors at a time, but at the cost of more complicated digital control circuits. In addition, some of these methods [64],[65] are based on the commutated feedback capacitor switching technique [67], which is not directly applicable to ΣΔ ADCs.
Lastly, a test signal injection into the analog signal path is used to calibrate the DAC gain errors in [45] and [66], resulting in a loss of conversion dynamic range.

Another unique issue related to ΣΔ loops is the error circulation due to feedback. In this section, we will extend the memory error model to the background calibration of multi-bit DACs in ΣΔ ADCs. For this purpose, a signal-driven PN injection scheme is devised with minimum alteration to the analog circuitry. In addition, the injected signal is immediately removed after the DAC, leaving only the mismatch error circulating in the modulator. As a result, the loss of conversion dynamic range is minimized.

5.1 Memory Error Model for Capacitor Mismatch

In conventional ΣΔ modulator analyses, the gain errors in the feedback paths are considered as pole-zero movement of the signal transfer function. As a result, the capacitor error model is often focused on the relative mismatch among $M$ unit capacitors, shown in Figure 5.1(a), as a function of the DAC input $D_i$:

\[
\begin{align*}
  e_c(n) &= \sum_{i=1}^{M} e_{i,d_i}(n), \\
  e_i &= \frac{C_{S(i)}}{\sum_{k=1}^{M} C_{S(k)}} - \frac{1}{M}
\end{align*}
\]  

(5.1)
where $\varepsilon_i$ is the deviation of each capacitor from the mean value of the $M$ capacitors with $\sum \varepsilon_i = 0$, and $d(n)$ is the thermometer code of the modulator output.
with \( d(n) = \Sigma d(n)/M \). The remaining mismatch between \( C_S\) (= \( \Sigma C_{S(i)} \)) and \( C_F \) is handled by introducing a gain error \( \gamma \) in the feedback path in Figure 5.1(b). Thus, the resulting error in the modulator output is expressed in an IIR form with respect to \( E_z \):

\[
E_d(z) = \frac{-z^{-1}}{1+ (\gamma-1)z^{-1}} E_z(z), \quad \left( \gamma = \sum_{i=1}^{M} \frac{C_{S(i)}}{C_F} \right)
\]

(5.2)

Equation (5.2) reveals that the DAC error circulating in the modulator loop causes long-term memory error. Conventionally, Equation (5.2) is treated by an FIR approximation [21], in which the length of the FIR taps determines the accuracy of the treatment. This section explains how a simple modification of the conventional approach leads to a more efficient nonlinear memory error model. In Figure 5.1(b), the input-referred gain error \( \alpha \) is assumed to be linear for simplicity.

When the sampling capacitors are shared with the DAC capacitors as shown in Figure 5.1(a), \( \alpha \) is equal to \( \gamma \).

An alternative way to express the normalized integrator update equation of the modulator in Figure 5.1(a) is

\[
y(n+1) = y(n) + \sum_{i=1}^{M} \left[ (x(n)-d_i(n)) \left( \frac{1}{M+\delta_i} \right) \right],
\]

(5.3)

where the \( \delta \)-th DAC capacitor \( C_{S(\delta)} \) is equal to \( (1/M+\delta)C_F \). This leads to the following expression for the modulator output

\[
D(z) = z^{-1} [\alpha X(z) - E_z(z)] + (1-z^{-1})E_d(z),
\]

(5.4)
The block diagram for Equation (5.4) is shown in Figure 5.1(c). The modulator output error is obtained as

\[ E_d(z) = -z^{-1}E_c(z). \]  

(5.7)

Therefore, once \( \delta_i \)'s are identified, the signal-dependent DAC error \( E_c \) can be removed from \( D \):

\[ d_c(n) = d(n) - e_d(n) = d(n) + \sum_{i=1}^{M} \delta_i d_i(n-1). \]  

(5.8)

Between Equation (5.6) and the conventional model of Equation (5.1), the only difference is whether the non-ideal integrator gain \( \gamma \) is absorbed into the model or not. By including \( \gamma \) in Equation (5.6), the IIR from of \( E_d \) in Equation (5.2) is converted to the single-tap FIR form in Equation (5.7). In other words, the long-term memory error of the modulator can be expressed as a concise function of its output. The equivalence between Figures 5.1(b) and (c) can be elaborated by deriving the transfer function directly from Figure 5.1(b):

\[
D(z) = \frac{z^{-1}}{1 + (\gamma - 1)z^{-1}} \left[ \alpha X(z) - E_c(z) + (1 - z^{-1})E_d(z) \right],
\]  

(5.9)

which is equivalent to the following time-domain expression (with quantization noise neglected):
\[ a\alpha(n-1) = d(n) + (\gamma-1)d(n-1) + e_r(n). \quad (5.10) \]

Comparing Equations (5.4) and (5.9), it is evident that
\[ e_r(n) = e_r(n) + (\gamma-1)d(n), \quad (5.11) \]
i.e., a single tap FIR model sufficiently represents both linear and nonlinear DAC errors. Equation (5.11) can be intuitively explained using Figure 5.1(b), in which the gain difference between the two feedback paths of the modulator, i.e., the DAC path with a gain of \( \gamma \) and the integrator path with a gain of unity (ideal), results in the error circulation of \( E_r \) in Equation (5.2). This contrasts to Figure 5.1(c), where \( \gamma \) is absorbed into \( E_c \), and the simple expression of Equation (5.7).

A similar analysis can be derived for a second-order modulator. In Figure 5.2(a), \( \gamma_1 \) and \( \gamma_2 \) are feedback gain errors, \( E_{r_1} \) and \( E_{r_2} \) are relative capacitor mismatch errors, and \( \alpha_1 \) and \( \alpha_2 \) are feed-forward gain errors. In Figure 5.2(b), the non-ideal feedback paths are replaced by ideal ones with two additive DAC errors, \( E_{c_1} \) and \( E_{c_2} \). In addition, since \( \alpha_2 \) is assumed to be linear, it can be input-referred, which results in the equivalent diagram of Figure 5.2(c). With this treatment, the modulator becomes ideal except for an input scaling and two additive error sources. The modulator transfer function is given by
\[ D(z) = z^{-2} \left[ \alpha_1 \alpha_2 X(z) - E_{c_1}'(z) \right] - 2z^{-1} \left( 1 - z^{-1} \right) E_{c_2}(z) + \left( 1 - z^{-1} \right)^2 E_q(z). \quad (5.12) \]
Figure 5.2. (a) Conventional $z$-domain model, (b) proposed capacitor mismatch model of a second-order $\Sigma\Delta$ modulator, and (c) an equivalent diagram of (b).
The time-domain expressions for $E_{c1}'$ and $E_{c2}$ are given by

$$e_{c1}'(n) = \alpha_2 [d(n) + e_{c1}(n)] - d(n)$$

$$= \sum_{i=1}^{M} \left( \frac{\Delta_2}{M} + (1 + \Delta_2) \delta_{ii} \right) d_i(n), \quad (5.13)$$

$$e_{c2}(n) = \sum_{i=1}^{M} \delta_{ii} d_i(n),$$

where $\Delta_2 = \alpha_2 - 1$ and the definitions of other parameters follow those of the first-order case. Again, the capacitor mismatch errors are expressed as a concise FIR form of the modulator digital output with non-ideal integrator gain errors accounted for. In addition, the second DAC error term $E_{c2}$ in Equation (5.12) is first-order noise shaped, which implies that the capacitor matching accuracy of the second DAC can be much relaxed compared to that of the first.

### 5.2 Calibration Approach

A compensation using the proposed error model requires the identification of both the relative mismatch among the unit capacitors and the integrator gain error. Correlation-based calibration techniques have been utilized for this purpose [28],[45],[63]–[66]. In terms of the relative mismatch, conventionally, one PN is used to randomize one element switching sequence to identify the corresponding mismatch coefficient; thus, multiple PN correlations are necessary for the treatment of a multi-bit structure. To correct the integrator gain error, an additional test signal traversing the signal path is also necessary. In contrast, the
main feature of the technique described in this section lies in its simplicity of differential DAC encoder implementation suitable for multi-bit ΣΔ modulator calibration using a single PN signal.

![Conceptual diagrams of (a) the conventional and (b) proposed DAC encoders.](image)

**Figure 5.3.** Conceptual diagrams of (a) the conventional and (b) proposed DAC encoders.

Figures 5.3(a) and (b), respectively, are the conceptual diagrams of the conventional differential DAC encoder and the proposed one with PN injection. In the conventional differential encoder, each unit capacitor is connected to –$V_r$, $V_r$, or 0 through three different switches when the differential thermometer code $D_i$ is –1, 1, or 0, respectively (the two adjacent comparator outputs $Q_{2N}$ and $Q_{2N-1}$ in Figure 5.3(a) are 00, 11, or 01, respectively). Thus, $M$ unit capacitors in a DAC can represent $2M+1$ levels differentially as indicated in Table 5.1, where the encoding scheme ensures that only one of the $D_i$’s can be 0 for a given modulator output $D$. 

- 66 -
Table 5.1  Bit encoding scheme for a 9-level DAC.

<table>
<thead>
<tr>
<th>$D$</th>
<th>$-1$</th>
<th>$-0.75$</th>
<th>$-0.5$</th>
<th>$-0.25$</th>
<th>$0$</th>
<th>$0.25$</th>
<th>$0.5$</th>
<th>$0.75$</th>
<th>$1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_1$</td>
<td>$-1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$D_2$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$D_3$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$0$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$D_4$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$0$</td>
<td>$1$</td>
</tr>
</tbody>
</table>

Table 5.2  DAC input code $D_{s(i)}$ and mismatch error $E_c$ when $D_3 = 0$ ($D = 0.25$).

<table>
<thead>
<tr>
<th>PN</th>
<th>$D_{s1}$</th>
<th>$D_{s2}$</th>
<th>$D_{s3}$</th>
<th>$D_{s4}$</th>
<th>$E_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$-1$</td>
<td>$-1$</td>
<td>$\hat{\sum} \delta - \hat{\sum} \delta$</td>
</tr>
<tr>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$1$</td>
<td>$-1$</td>
<td>$\hat{\sum} \delta - \hat{\sum} \delta$</td>
</tr>
</tbody>
</table>

The proposed encoding scheme is identical to the conventional scheme when $D_i$ is $-1$ or $1$. However, when $D_i$ is $0$, the encoder connects the unit capacitor to either $-V_r$ or $V_r$ instead of 0 depending on the value of a PN. It is equivalent to stating that a PN signal $T_c$ taking on the value of $-1/M$ or $1/M$ is injected whenever one of the $D_i$’s is 0 in Figure 5.4(a), and this can be accomplished using a simple logic gate with minimal change to the original encoder. In addition, the switch connected to 0 can be saved as the PN is injected whenever $D_i$ is 0, resulting in a simpler (and potentially faster) switching configuration.

With this PN injection scheme, the DAC error is dependent on $D_i (= D + T_c)$ instead of $D$, and Equations (5.4) and (5.6) are rewritten as
\[ D(z) = z^{-1}[\alpha X(z) - T_c(z) - E_c(z)] + (1 - z^{-1})E_s(z), \] (5.14)

\[ e_c(n) = \sum_{i=1}^{M} \delta_i d_{\pi(i)}(n), \] (5.15)

where \( D_{\pi(i)} \), taking on a value of \(-1\) or \(1\), is the differential thermometer code of \( D_i \) with \( d_i(n) = \sum_{i=1}^{M} d_{\pi(i)}(n)/M \). Owing to \( T_c \), the DAC has two different configurations for a certain \( D \) when one of its \( D_i \)'s is 0. Assuming that \( T_c \) is subtracted out from \( D \) in the digital domain, the modulator output would be indistinguishable between the two configurations if \( E_c = 0 \) (no mismatch error), i.e., \( D \) is statistically independent of \( T_c \). Thus, any nonzero correlation will provide a clue to the mismatch of \( C_{\pi(i)} \). For example, as shown in Table 5.2, whenever \( D_3 \) is 0, i.e., \( D = 0.25 \), the DAC input \( D_{\pi(3)} \) is assigned to either \(-1\) or \(1\) depending on the value of the PN; all other \( D_{\pi(i)} \)'s are still identical to \( D_i \)'s, respectively. If any subset of a PN can be considered as another PN with similar statistical properties (this property can be guaranteed by choosing a PN with the length of a large prime number), from Equations (5.14) and (5.15), the average value of \( d(n) \cdot t_c(n-1) \) when \( d_3(n-1) = 0 \) yields the information of \( \delta_3 \).

Some digital processing can be performed on \( D_c \) in Figure 5.4(a) to improve the calibration performance. For example, the corrected output \( D_c \) can be low-pass filtered to attenuate the out-of-band quantization noise, which is possibly correlated to both the input and PN signal when the quantizer resolution...
is low. In contrast, high-pass filtering can be applied to eliminate the in-band signal to speed up the convergence when the quantization noise level is low.

![Block diagram of the proposed first-order ΣΔ modulator calibration and (b) LMS update block.](image)

Figure 5.4. (a) Block diagram of the proposed first-order ΣΔ modulator calibration and (b) LMS update block.

Because capacitor mismatch causes both in-band and out-of-band errors, minimizing one using the same error model will remove the other at the same
time. The filtered data $E_c$ is multiplied by $T_c$, as shown in Figure 5.4(b), and separately accumulated according to the thermometer code $\overline{D}_i$ ($= 1$ only when $D_i = 0$). In turn, the following LMS update equation for $\delta_i$ is obtained:

$$\hat{\delta}_{i}^{m+1} = \hat{\delta}_{i}^{m} + \mu e_c(n)\overline{D}_i(n-1)\mu_c(n-1),$$

(5.16)

where $i = 1, 2, \ldots, M$, the latency of the filter is neglected for simplicity, and $m = kn$ ($k \geq 1$ is an integer that determines the LMS block size). The DAC error is estimated by

$$\hat{e}_c(n) = \sum_{i=1}^{M} \hat{\delta}_{i}^{d}d_{i}^{(i)}(n).$$

(5.17)

As illustrated in Figure 5.4(b), $\overline{D}_i$, $T_c$, and $D_{t(i)}$ are all binary in Equations (5.16) and (5.17), so $2M$ adders and multiplexers are needed for the $M$-coefficient calibration (the multiplication of a fixed step size $\mu$ in Equation (5.16) can be implemented by bit shifting). Considering that the $M$ coefficients ($M$ capacitors) represent $2M+1$ DAC levels differentially, the expected hardware complexity is around one adder and multiplexer per DAC level.

The conceptual block diagram of a second-order modulator calibration is shown in Figure 5.5, in which two PN signals $T_{c1}$ and $T_{c2}$ are injected into the two DAC inputs, respectively, to identify the corresponding capacitor mismatch errors. In practice, the injected PN signals will occupy some signal dynamic range, especially for high-order modulators, in which each DAC will require an injected PN. This motivates the PN removal technique to be introduced in the next section.
5.3 PN Removal Technique

PN injection and its circulation in the modulator signal path bring about a dynamic range reduction problem. In memoryless Nyquist-rate converters such as a pipelined ADC, the input/output swing and stage gain determine the PN amplitude to avoid an overflow/underflow problem. Selective PN injection has been reported for pipelined ADCs to avoid dynamic range degradation [36]. In ∑Δ architecture, the injected PN potentially occupies a larger signal swing than the PN amplitude itself due to the feedback loops inherent to the modulator. From Equation (5.14), the DAC input $D_i$ is expressed as

$$D_i(z) = z^{-1}[\alpha X(z) - T_e(z) - E_i(z)] + (1-z^{-1})E_i(z) + T_e(z),$$  \hspace{1cm} (5.18)$$

in which the maximum or minimum input is generated when $t_e(n-1)$ and $t_e(n)$ exhibit opposite signs. Since $T_e$ with an amplitude of two levels (±1 LSB) is injected, a total of four DAC levels will be occupied by $T_e$. Fortunately, the
proposed calibration does not inject PN for the maximum and minimum DAC inputs (in these cases, \( D_i \neq 0 \) for all \( i \)). As a result, \( T_c \) occupies two quantization levels in the first-order modulator.

Figure 5.6. Analog-domain PN removal scheme for (a) first- and (b) second-order \( \Sigma \Delta \) modulator calibration.

The problem can be mitigated by an analog-domain PN removal scheme as shown in Figure 5.6(a), where \( T_c \) is subtracted out in the analog domain by an additional capacitor at the summing node. Ideally, if all DAC capacitors are matched to the additional PN removal capacitor, no residual injection will end up
circulating the loop at all. In other words, any residual circulation has to derive from capacitor mismatch, which is expected to be at least an order of magnitude smaller than the injection itself. Therefore, the net outcome of the calibration with PN removal would be that every DAC capacitor is eventually calibrated against the PN removal capacitor, which effectively makes it a reference. Since the PN is cancelled out by an additional unit capacitor, it may reduce the feedback factor of the integrator and increase power consumption for low-resolution DACs. However, this effect is not significant for higher-resolution DACs.

Although PN removal minimizes the dynamic range loss, it brings about another mismatch issue of the PN removal capacitor (relative to $C_r$), which is shown as $\Delta_c$ in Figure 5.6(a). Neglecting quantization noise, the output is

$$d(n) = \alpha x(n-1) + \Delta_c t_c(n-1) - e_c(n-1). \quad (5.19)$$

From Equations (5.15), (5.17), and (5.19), the calibrated output $D_c$ is given by

$$d_c(n) = \alpha x(n-1) + \Delta_c t_c(n-1) - \sum_{i=1}^{M} (\hat{\delta}_i - \hat{\delta}_i) d_{i(n-1)}, \quad (5.20)$$

where $\hat{\delta}_i$ is the estimated capacitor mismatch, which is determined by the following equation derived from Equation (5.16):

$$E[d_c(n)\overline{d_i(n-1)t_c(n-1)}] = 0, \quad (5.21)$$

where $D_c$ is used in these equations instead of $E_c$ in Equation (5.16) for clarity. Using Equation (5.20), (5.21) yields
Substituting Equation (5.22) in Equation (5.20), we have
\[
d_c(n) = \alpha x(n-1) + \Delta c \cdot t_c(n-1) - \Delta c \cdot d(n-1)
\]
\[
\frac{\Delta c}{M}
\]
(5.23)

Note that Equation (5.23) is identical to Equation (5.10) by substituting \(\gamma - 1\) with \(\Delta c\) and \(d(n) + e_c(n)\) with \(d_c(n)\), which implies that the effects of PN mismatch and DAC gain error are not distinguishable. As a result, \(\Delta c\) can be considered as a linear gain error of the DAC feedback path.

To calibrate the effect of \(\Delta c\), an additional PN signal \(T_\alpha\) is introduced at the summing node as shown in Figure 5.6(a). The circulation of \(T_\alpha\) in the modulator loop will help to identify \(\Delta c\). In this approach, even though the incomplete PN removal leads to the introduction of another PN, the amplitude of \(T_\alpha\) is much more adjustable than that of \(T_c\) (recall that \(T_c\) is always \(\pm 1\) LSB). In addition, \(T_\alpha\) can be exploited to calibrate integrator nonlinearities, and the scheme can be generalized to higher-order modulators.

With \(T_\alpha\) injected, the modulator output in Equation (5.19) becomes
\[
d(n) = \alpha x(n-1) + (1 + \Delta_c) t_{\alpha}(n-1) + \Delta c \cdot t_c(n-1) - e_c(n-1),
\]
(5.24)
where \(\Delta_{\alpha}\) captures the mismatch of \(T_{\alpha}\). With calibration, Equation (5.23) is revised to
where \( \Delta_c \) and \( \Delta_{a} \) are the estimated mismatch errors. Using Equation (5.24), Equation (5.25) can be expanded as

\[
d_c(n) = \alpha x(n-1) + \left( \Delta_a - \hat{\Delta}_a \right) t_a(n-1) - \left( \Delta_c - \hat{\Delta}_c \right) d(n-1),
\]

If \( \Delta_c \neq \Delta_a \) and/or \( \Delta_{a} \neq \Delta_{a} \), Equation (5.26) indicates that \( d_c(n) \) will contain memory error terms of \( T_a \) and \( T_c \). Out of these terms, \( t_a(n-1) \) is related to \( \Delta_a \) and \( t_a(n-2) \) is related to \( \Delta_c \). Note that the term \( e_c(n-2) \) in Equation (5.26), determined by \( d_c(n-2) \) in Equation (5.15), contains \( t_a(n-k) \) and \( t_c(n-k) \) only for \( k \geq 3 \) (\( T_a \) appears in \( D_t \) with one clock cycle delay). Equation (5.26) leads to the following LMS update equations for \( \Delta_c \) and \( \Delta_{a} \):

\[
\hat{\Delta}_c^{m+1} = \hat{\Delta}_c^m + \mu e_c(n) t_a(n-2),
\]

\[
\hat{\Delta}_{a}^{m+1} = \hat{\Delta}_{a}^m + \mu e_c(n) t_a(n-1).
\]

Note that in Equation (5.26), the term \( t_c(n-2) \) can also be used to derive \( \Delta_c \); however, since this term is proportional to \( \Delta_c^2 \), it requires a long convergence time as \( \Delta_c \) is a small quantity, which compares to the adaptation using the first-
order \( t_\alpha(n - 2) \) term. In addition, as explained in Equation (5.23), the effect of PN mismatch \( \Delta_c \) is equivalent to a feedback gain error. Thus, when the gain mismatch between the DAC and integrator feedback paths is calibrated, all memory errors through these two paths will disappear. This fact holds despite the fact that we always assume an ideal integrator path with a gain of unity. The calibration scheme will ensure that by adapting Equations (5.27) and (5.28), all memory terms in Equation (5.26) disappear eventually. Based on Equations (5.20), (5.22), and (5.25), the final calibrated output with PN removal is given by

\[
d_t(n) = d(n) - (1 + \hat{\Delta}_e) t_e(n - 1) + \hat{\alpha}_t(n - 1) + \hat{\Delta}_c d(n - 1),
\]

\[
\hat{\alpha}_t(n - 1) = \sum_{i=1}^{M} \hat{\delta}_i d_{t(i)}(n - 1) = \sum_{i=1}^{M} \left[ \delta_i d_{t(i)}(n - 1) \right] - \Delta_c d_t(n - 1).
\]  

(5.29)

Ignoring quantization noise, the output of a second-order modulator with PN injection illustrated in Figure 5.5 is obtained from Equations (5.12) and (5.13):

\[
D(z) = z^{-2}[\alpha_1 \alpha_2 X(z) - T_{c1}(z) - E_{c1}^t(z)]
- 2z^{-1}(1 - z^{-1})[T_{c2}(z) + E_{c2}^t(z)],
\]

\[
\hat{\alpha}_t^t(n) = \sum_{i=1}^{M} \Delta_z^t \left[ (1 + \Delta_2) \delta_i \right] d_{t(i)}(n) = \sum_{i=1}^{M} \tilde{\delta}_i d_{t(i)}(n)
\]

(5.30)

Thus, the DAC inputs are written as

\[
D_{\alpha(i)}(z) = z^{-2}[\alpha_1 \alpha_2 X(z) - T_{c1}(z) - E_{c1}(z)]
- 2z^{-1}(1 - z^{-1})[T_{c2}(z) + E_{c2}(z)] + T_{\alpha(i)}(z),
\]  

(5.31)

- 76 -
where \( l = 1 \) and 2. Since \( T_{c1} \) and \( T_{c2} \) each have an amplitude of two DAC levels (±1 LSB), a total of 12 levels will be occupied by \( T_{c1} \) and \( T_{c2} \) in the worst case (when \( t_{c1}(n-2) = t_{c2}(n-1) = -t_{c2}(n-2) = -t_{c1}(n) \)). This calibration scheme does not inject \( T_{c1} \) for the maximum and minimum DAC inputs (in these cases, \( D_i \neq 0 \) for all \( i \)). As a result, \( T_{c1} \) and \( T_{c2} \) occupy 10 quantization levels in a second-order modulator. With PN removal, as illustrated in Figure 5.6(b), both test signals, \( T_{c1} \) and \( T_{c2} \), are eliminated in the analog domain. In this case, Equation (5.31) is rewritten as

\[
D(z) = z^{-2} [\alpha_1, \alpha_2 X(z) + \Delta_{11} T_{c1}(z) - E_{c1}'(z)] \\
+ 2 z^{-1} \left( 1 - z^{-1} \right) [\Delta_{22} T_{c2}(z) - E_{c2}(z)],
\]

and the calibrated output is given by

\[
\hat{D}_c(n) = D(z) + z^{-2} \hat{E}_{c1}(z) + 2 z^{-1} \left( 1 - z^{-1} \right) \hat{E}_{c2},
\]

\[
\hat{e}_{c1}(n) = \sum_{i=1}^{M} \hat{d}_{i j}(n) d_{i j}(n),
\]

where \( \hat{E}_{c1} \) and \( \hat{E}_{c2} \) are the estimated DAC errors. In Equation (5.32), \( E_{c1}' \) and \( E_{c2} \) are correlated with \( T_{c1} \) and \( T_{c2} \), respectively, and we obtain the following LMS update equations:

\[
\hat{\delta}_{l 1}^{m+1} = \hat{\delta}_{l 1}^{m} + \mu d_{l 1}(n) \tilde{d}_{l 1}(n-2) t_{c1}(n-2),
\]

\[
\hat{\delta}_{2 1}^{m+1} = \hat{\delta}_{2 1}^{m} + \mu d_{2 1}(n) \tilde{d}_{2 1}(n-1) t_{c2}(n-1).
\]

- 77 -
Similar to the first-order case, $D_c$ can be filtered before being fed to the LMS update block. From Equation (5.34), $\hat{\delta}_i$ and $\hat{\delta}_{2i}$ are determined by solving the following equations:

$$
E[d_c(n)\bar{d}_i(n-2)t_i(n-2)] = 0,
E[d_c(n)\bar{d}_i(n-1)t_{i2}(n-1)] = 0.
$$

Using Equations (5.32) and (5.33), Equation (5.35) yields

$$
\hat{\delta}_i = \delta_i^i - \frac{\Delta_{ci}}{M}, \quad \hat{\delta}_{2i} = \delta_{2i} - \frac{\Delta_{c2i}}{M}.
$$

Substituting Equations (5.32) and (5.36) in Equation (5.33), $D_c$ with incomplete PN re removal becomes

$$
d_c(n) = \alpha_i\alpha_x(n-2) - 2\Delta_{ci}d(n-1) - (\Delta_{c1} - 2\Delta_{c2})d(n-2)
$$

where $\Delta_{c1}' = \Delta_{c1} - 2\Delta_{c2}$. Equation (5.37) shows that a proper gain correction is needed to complete the calibration. In Figure 5.6(b), the additional PN signal $T_\alpha$ is introduced for the gain calibration and Equation (5.32) is modified to

$$
D(z) = z^{-2}\left[\alpha_i\alpha_x X(z) + (1 + \Delta_{c1})T_\alpha(z) + \Delta_cT_{c1}(z) - E_{c1}'(z)\right]
+ 2z^{-1}(1 - z^{-1})\left[\Delta_{c2}T_{c2}(z) - E_{c2}'(z)\right].
$$

Based on the result of Equation (5.37), the $D_c$ after $T_\alpha$ and gain error ($\Delta_{c1}'$ and $\Delta_{c2}'$) compensation is rewritten as

$$
D_c(z) = \left[1 + 2\hat{\Delta}_{c2}z^{-1} + \hat{\Delta}_{c2}'z^{-2}\right]D(z) - (1 + \hat{\Delta}_a)z^{-2}T_\alpha
+ z^{-2}\hat{E}_{c1}(z) + 2z^{-1}(1 - z^{-1})\hat{E}_{c2}.
$$
In Equation (5.39), since \( T_\alpha \) appears at the modulator output \( D \) with a two clock-cycle delay, \( \hat{\Delta}_\alpha \) is identified by correlating \( D_c \) to \( t_\alpha(n-2) \):

\[
\hat{\Delta}_{\alpha}^{n+1} = \hat{\Delta}_\alpha^n + \mu e_c(n)t_\alpha(n-2).
\] (5.40)

On the other hand, when the first- and second-stage feedback gain errors exist, \( D \) will include memory terms of \( T_\alpha \), i.e., \( t_\alpha(n-3) \) and \( t_\alpha(n-4) \), respectively; and the following update equations are obtained:

\[
\begin{align*}
\hat{\Delta}_{\alpha}^{n+1} &= \hat{\Delta}_\alpha^n + \mu e_c(n)t_\alpha(n-3), \\
\hat{\Delta}_{c1}^{n+1} &= \hat{\Delta}_{c1}^n + \mu e_c(n)t_\alpha(n-4).
\end{align*}
\] (5.41)

This gain error calibration can be explained more intuitively using the generalized transfer function of the second-order modulator of Figure 5.6(b):

\[
D(z) = \frac{\alpha z^{-2}X(z) + z^{-2}(1 + \Delta_\alpha)\frac{T_\alpha(z)}{1 + k_1z^{-1} + k_2z^{-2}} + E_d(z)}{1 + k_1z^{-1} + k_2z^{-2}},
\] (5.42)

where quantization noise is ignored, \( k_1 \) and \( k_2 \) can include any gain error in the signal paths, and \( E_d \) is the total DAC error including PN mismatch \( \Delta'_{c1} \) and \( \Delta'_{c2} \), which is given by

\[
E_d(z) = z^{-2}\left[\Delta_{c1}T_{c1}(z) - E'_{c1}(z)\right] + 2z^{-1}\left(1 - z^{-1}\right)\left[\Delta_{c2}T_{c2}(z) - E'_{c2}(z)\right].
\] (5.43)

By multiplying the denominator on both sides, the time-domain expression of Equation (5.42) is obtained as

\[
d(n) + k_1d(n-1) + k_2d(n-2) = \alpha x(n-2) + (1 + \Delta_\alpha)t_\alpha(n-2) + e_d(n).
\] (5.44)
Comparing Equations (5.39) and (5.44), it is evident that $d_{c}(n)$ will not contain other memory terms of $T_{a}$ except $t_{a}(n–2)$ when properly compensated by $d(n–1)$ and $d(n–2)$; i.e., the denominator of Equation (5.42) is treated, and the total error $E_{d}$ can be formulated using the FIR form of $\hat{E}_{c1}$ and $\hat{E}_{c2}$ in Equations (5.33) and (5.36). This path gain error identification method using a single PN $T_{a}$ can be generalized to any high-order, single-loop modulators.

5.4 Behavioral Simulation

Behavioral simulations of first- and second-order modulators are performed to verify the proposed calibration algorithm. A 2 % standard deviation of capacitor mismatch is assumed for the DAC unit capacitors. The quantizer resolution is approximately 4-bit (17-level), and the amplifiers are assumed ideal. The mismatch between the integrator feedback capacitor and the PN removal capacitor is properly included, and the integrator gain error calibration using an additional PN signal ($T_{a}$) is applied, with its own mismatch included as well. The sinusoidal input frequency is set to around 1 % of the clock frequency and the modulator dynamic performance is calculated for a 16 OSR. An 8th-order FIR HPF is also employed to accelerate the convergence of the calibration. The in-band rejection of the HPF is around 40 dB.
Comparing the output spectra of an ideal modulator in Figure 5.7(a) and the one with multi-bit DAC mismatch in Figure 5.7(b) shows that capacitor mismatch clearly causes large harmonic distortions. Figure 5.7(c) reveals that the proposed calibration effectively eliminates all harmonic tones. Comparing Figures 5.7(a) and (c), one sees that Figure 5.7(c) does not show any pattern noise and achieves an even better SFDR, thanks to the randomization effect of the PN signal.
The input dynamic range reduction due to $T_\alpha$ results in a 0.6-dB SNDR difference between Figures 5.7(a) and (c). The amplitude of $T_\alpha$ determines the calibration speed and accuracy. In contrast, an LPF instead of HPF can potentially make the calibration less sensitive to quantization noise and improve the calibration accuracy for a smaller $T_\alpha$. However, a smaller step size of LMS update is necessary with an LPF to mitigate the large interference from the (unfiltered) in-band signal, which will slow down the training process significantly.

In Figure 5.7(d), the differential nonlinearity (DNL) and integral nonlinearity (INL) are normalized, i.e., $DNL = (C_{S(i) - C_S}/M)/C_S$. Figures 5.7(e) and (f) show the coefficient learning curves and the modulator dynamic performance, respectively. Since $C_F/M$ is set to 2% larger than the reference capacitor, which is used for the PN removal ($\Delta_c = 0.02$), the estimated capacitor mismatch $\hat{\delta_i}$ is biased towards negative values, as revealed by Figure 5.7(e). By excluding this bias from each $\hat{\delta_i}$, the settled values in Figure 5.7(e) correspond to the DNL profile in Figure 5.7(d). The SNDR and SFDR fluctuations in Figure 5.7(f) are mainly caused by the periodicity of the sinusoidal input and PN sequence.

One important issue related to this modulator calibration is that the calibration will increase the final word length, which significantly complicates the following decimation filter design. To clarify this effect, the simulation in Figure 5.7(c) is performed with different numbers of bits for $D_c$, and the results are
summarized in Table 5.3. When $D_c$ is more than 11 b, the calibration performance is nearly identical to that of the floating-point result shown in Figure 5.7(c). When $D_c$ is less than 7 b, the calibration performance drops abruptly (even worse than that of without calibration in Figure 5.7(b)) because the injected PN signal is not properly removed from $D_c$.

Table 5.3 Calibration results with finite word length.

<table>
<thead>
<tr>
<th></th>
<th>4 b</th>
<th>5 b</th>
<th>6 b</th>
<th>7 b</th>
<th>8 b</th>
<th>9 b</th>
<th>10 b</th>
<th>11 b</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNDR [dB]</td>
<td>32.6</td>
<td>47.7</td>
<td>44.2</td>
<td>53.2</td>
<td>54.7</td>
<td>56.0</td>
<td>56.2</td>
<td>56.4</td>
</tr>
<tr>
<td>SFDR [dB]</td>
<td>34.6</td>
<td>50.2</td>
<td>47.6</td>
<td>68.1</td>
<td>65.6</td>
<td>73.1</td>
<td>77.2</td>
<td>79.8</td>
</tr>
</tbody>
</table>

In a second-order modulator, the feedback capacitors are again assumed to be 2% larger than their ideal values, which generate a negative gain error in each signal path. Figures 5.8(a), (b), and (c) indicate that every harmonic tone caused by the capacitor mismatch is eliminated effectively by the proposed calibration. Comparing Figures 5.8(a) and (c) shows that the signal dynamic range occupied by $T_{\alpha}$ and the coefficient fluctuation in steady state degrade the SNDR by approximately 1.7 dB. Simulation reveals that the low-frequency noise floor of the modulator is very sensitive to the values of the model coefficients, which dictates a small step size and thus longer calibration time compared to the first-order case. In this respect, the proposed scheme is more appropriate for wide-band, high-speed applications. Comparing Figures 5.8(d) and (e) shows that the steady-state dynamic performance is less sensitive to the fluctuation of the second-stage
coefficients than that of the first-stage ones in Equation (5.39). Therefore, a larger step size is applied to the second-stage coefficients update. Due to the positive mismatch errors of the feedback capacitors, $\hat{\delta}_1$ and $\hat{\delta}_2$ shown in Figures 5.8(d) and (e), respectively, are all biased to negative values.

Figure 5.8 MATLAB simulation results of a 4-bit second-order $\Sigma\Delta$ modulator (OSR = 16) before calibration (a) without capacitor mismatch, (b) with mismatch of a 2% standard deviation, (c) after calibration, (d) $\hat{\delta}_1$, (e) $\hat{\delta}_2$, and (f) SNDR and SFDR learning curves.
5.5 Extension of Selective PN Injection and Cancellation

The major advantages of the proposed selective PN injection/cancellation over the conventional DEM based PN correlation are the simplicity of its circuit implementation and the ease of its error analysis exploiting the differential DAC structure. This technique can be applicable to the DAC calibration in pipeline ADCs. An N-bit stage of pipeline ADCs with the proposed calibration technique is illustrated in Figure 5.9 where the gain error in residue amplification is separately represented as $\alpha$. The stage residue output $R$ after subtracting the DAC output $W$ from the input $X$ is given by

$$
R = \alpha(X - W) = \alpha \left[ X + T_a - \left[ D + \Delta_{T_c} T_c + \sum \delta_i D_{T_c(i)} \right] \right]
$$

$$
= \alpha \left[ X + T_a - \left[ (1 - \Delta_{T_c}) D + \sum (\delta_i + \Delta_{T_c}) D_{T_c(i)} \right] \right],
$$  \hspace{1cm} (5.45)

where $D_{T_c(i)}$ is the differential thermometer code of $D_{T_c}$, and $\Delta_{T_c}$ is the mismatch of $T_c$ in analog domain PN cancellation. Thus, the input is recovered by adding sub-ADC output $D$ to $R$:

$$
\hat{X} = R + D = \alpha \left[ X + T_a - \left[ \left(1 - \frac{1}{\alpha} \right) \Delta_{T_c} D + \sum (\delta_i + \Delta_{T_c}) D_{T_c(i)} \right] \right],
$$  \hspace{1cm} (5.46)

As explained in this chapter, $\alpha$ and $\Delta_{T_c}$ need to be figured out separately by using an additional PN $T_a$ at the input stage. In the pipeline ADC, $\Delta_{T_c}$ is eventually the mismatch between the PN cancellation and feedback capacitors. Thus, by reusing the PN cancellation capacitor as the feedback capacitor, we can eliminate $\Delta_{T_c}$. $\alpha$ can be easily figured out by $T_a$. It is noticeable that the residue error $R$ in Equation
(5.45) is utilized instead of $\hat{X}$ in correlation to train the capacitor mismatch $\delta_i$ ($\delta_i$ is updated only when $D_i = 0$). The variance of $R$ is much smaller than that in $\hat{X}$, i.e., $D$ includes most of the input signal component $X$, so the adaptation with the selective PN injection can be much faster than the calibration based on DEM circuits which updates $\delta_i$ every sample using $\hat{X}$ [62].

Figure 5.9. DAC calibration in a pipelined ADC structure.
CHAPTER 6

PROTOTYPE 1-0 MASH ADC DESIGN

Although it is presented that the proposed techniques can be theoretically extended to higher-order $\Sigma\Delta$ modulators, many practical issues need to be further investigated at the architectural and circuit levels to minimize the overhead in digital calibration and optimize the ADC performance. This chapter describes design issues in applying the proposed schemes to a prototype ADC.

6.1 Nonlinear Calibration of MASH Structure

Circuit non-idealities in $\Sigma\Delta$ modulators conventionally have been regarded as linear effects and represented as IIR-form signal and noise transfer functions (STF and NTF) shown in Figure 6.1(a) [21]. A direct extension of the model to account for the signal-path nonlinearities, i.e., signal-dependent coefficients $\alpha$, $\beta$, and $\gamma$, yields complicated analysis and no readily useful results. In analog-to-digital conversion, the ultimate interest is not the modulator output $D$ but the input signal $X$ eventually digitized, so the closed form of the modulator error model needs to be rederived as a form of $D$ for the purpose of digital compensation.
A unique finding of this work is that an FIR form, e.g., the two-tap model shown in Figure 6.1(b) for a first-order \( \Sigma \Delta \) modulator, can be sufficient and accurate in representing the long-term nonlinear memory errors in the modulator. In the output referred error analysis, the signal-path distortion is comprehended as three additive error terms \( E_a \), \( E_b \), and \( E_c \) which are all dependent on the modulator output \( D \). Among the three terms, \( E_a \) and \( z^{-1}E_b \) (dependent on \( d(n) \) and \( d(n-1) \), respectively) capture the integrator nonlinearity while \( z^{-1}E_c \) (dependent on the
DAC thermometer code $d(n-1)$ expresses the component mismatch error of the feedback DAC. Compensating for the error sources $E_a$, $E_b$, and $E_c$ is equivalent to multiplying the inverse form of the nonlinear STF, so the non-ideality of any all-pole STF can be represented as an FIR form of the output; that is, when the modulator has zeros, the error model becomes an IIR form of the output. Since the error is derived from the modulator output containing the effect of the other error sources, the interaction from the other errors is naturally reflected in the model of each error source.

The model embodying a direct digital treatment of the nonlinear STF without any approximation is clearly advantageous over the conventional correction method for MASH. Figure 6.2 instantiates the difference with a two-stage 1-0 MASH structure. The conventional approach in Figure 6.2(a) [21]–[24] leaves the non-ideality of the first stage output $D_1$ as it is and matches the STF of the second stage to the NTF of the first stage in digital domain by using an IIR model $NTF_1$ (in practice, $NTF_1$ is approximated to an FIR form). However, the proposed approach in Figure 6.2(b) directly treats the nonlinear STF of the first stage by an FIR model $(1/STF_1)$, such that another FIR model $(NTF_1/STF_1)$ compensates for the gain mismatch between the stages. Note that the NTF and STF generally have the same poles and the STF is assumed to have only poles without zeros.
Figure 6.2. (a) Conventional and (b) proposed 1-0 MASH calibration.

For the identification of the given model, the learning algorithm proposed here, based on the principle of ICA, trains multiple model parameters using a single one-bit PN injected into the input [58]. Figure 6.3 illustrates the general setup of the proposed nonlinear modulator calibration. Ignoring the quantization noise (the effect of the quantization noise in calibration is analyzed in Chapter 3), the first-order modulator with a non-ideal integrator is modeled as two error sources $E_a$ and $E_b$ from Figure 6.1. $e_a(n)$ is dependent on $d(n)$ which mainly consist of the moments of $x(n-1)$ and $pn(n-1)$. If $E_a$ is successfully removed in
digital domain, the calibrated output $\hat{X}$, which is also a digitized version of the input $X$, will not contain any inter-modulation products between $x(n-1)$ and $pn(n-1)$. In other words, $\hat{X}$ is independent to any cross moments of $x(n-1)$ and $pn(n-1)$. Since $x(n-1)$ is unknown, it is recursively estimated from $\hat{x}(n-1)$. In turn, the involved model parameters $a_i$'s are trained by the LMS adaptation until the correlation with all cross terms $[\hat{x}^{i-1}(n-1)pn(n-1)]$ is minimized.

$$
\hat{E}_d = \sum a_i(n)d^i(n) + \sum b_i(n)d^i(n-1)
$$

Figure 6.3. Multiple error parameter identification by using a one-bit PN.

Similarly, the coefficients $b_i$'s are updated using the inter-modulation products between the delayed moment of $\hat{x}(n-2)$ and $pn(n-2)$. Before completing calibration, $\hat{X}$ includes other memory moments of $X$ and $PN$ such as $x(n-3)$,
$pn(n-3)$, $x(n-4)$, and so on. However, under the condition that the first two memory terms are minimized from $\hat{X}$ with a given FIR model, all the other memory terms are also minimized. The DAC distortion term $E_c$, which is not shown in Figure 6.3, is identified by a second PN injected into the DAC code to reduce the order of the polynomial model of the nonlinear integrator. The HPF removes most of the in-band content in $\hat{X}$ before feeding it to the LMS coefficient update unit to accelerate the parameter learning. To better preserve error information and further accelerate the calibration speed, sophisticated filtering techniques can be considered [61].

6.2 Prototype 1-0 MASH Architecture

With the digital calibration, the integrator nonlinearity needs to be separately identified by injecting a wideband PN signal to each stage, so with multiple integrators in a loop, PN signals circulating in a high-order loop occupy a large portion of the input signal dynamic range and slow down the integrator output settling. The proposed approach compensates for the static circuit nonlinearity, whereas the dynamic error involved with signal settling and clock jitter is still difficult to accurately model and calibrate. In this aspect, a low-order noise shaping structure with a low OSR is preferred to reduce the speed overhead in analog circuit design.
The other architectural considerations are calibration speed and accuracy. In LMS adaptation, the identified error parameters continuously fluctuate within a certain range depending on the variance of the residue error signal, which locally correlates to the PN, and the update step size [62]. Among the residue error, the in-band input signal is removed by HPF, so the out-of-band quantization noise becomes a major source perturbing the settled parameter values. In turn, the high-order noise shaping usually accompanying a low resolution quantizer and large out-of-band quantization noise necessitates a smaller step size, significantly elongating the calibration time. Lastly, as the complexity of the digital circuit greatly increases in high-order modulators, the area and power consumption of the digital part can outweigh the benefit in analog design. Considering all these practical issues, we adopt the 1-0 MASH structure as a prototype ADC.

Figure 6.4 depicts the overall architecture of the proposed $\Sigma\Delta$ ADC. This first-order structure looks similar to the 11b Nyquist rate pipeline structure except for the first stage integrator. The frontend consists of a first-order $\Sigma\Delta$ modulator employing a low-gain amplifier and a 9-level feedback DAC, with both integrator distortion and DAC capacitor mismatch calibrated. The integrator obviates the need for a front-end sample-and-hold circuit and gains additional SNR from the noise shaping. Theoretically, this first-order structure can achieve more than 14 effective number of bits (ENOB) with an OSR of 8 (the ENOB of the ADC is eventually limited by the kT/C noise). The backend is a four-stage 2.5b/s
pipelined ADC. The mid-tread structure is used for the sub-ADCs, so 8 comparators are used in the first stage sub-ADC and 6 comparators are used for the other sub-ADCs. Memoryless inter-stage gain calibration is applied to the backend multiplying DACs (MDACs), greatly relaxing the gain requirement of the residue amplifiers. All analog functions are integrated on the prototype chip. The digital calibration logic is synthesized, but not implemented on chip. In testing, error correction is processed in a workstation due to its total feed-forward architecture.

Figure 6.4. Prototype 1-0 MASH architecture.
To calibrate the integrator, a one-bit PN signal $T_a$ is injected at the summing node of the integrator. For the first stage DAC calibration, another test signal $T_c$ is added in digital domain, which is eliminated in analog domain to minimize the conversion dynamic range loss. $T_a$ and $T_c$ are internally generated from 31 linear feedback shift registers with the maximal-length implementation, i.e., the length of the PN sequence is $2^{31} - 1$ [68],[69].

From the output-referred error analysis, the first stage digital output $D_1$ is written by

$$D_1 = z^{-1} \left( \frac{X}{1+\Delta_a} + T_a \right) + E_d - z^{-1} E_c + (1 - \beta z^{-1}) E_{q1},$$

(6.1)

where the net effect of the PN mismatches $\Delta_a$ and $\Delta_c$, after calibration, is conceived as the residue gain error of the input and feedback paths of the modulator (Appendix B.1), so $\beta$ is inserted to represent the scaling of the quantization noise $z^{-1} E_{q1}$. $E_d$ and $E_c$ are the integrator nonlinearity and capacitor mismatch, respectively, given by

$$e_d(n) = e_a(n) + e_b(n-1) = \sum_{i=1}^{N} a_{ji} y'(n) + \sum_{i=4}^{N} b_{ji} y'(n-1),$$

$$e_c(n) = \sum_{i=1}^{M} \delta_i d_{(i)} (n-1),$$

(6.2)

where $D_{1(i)}$ is the thermometer code of $D_1$ ($D_1 = D_{1(M)}...D_{12}D_{11}$). The first stage quantization noise $E_{q1}$ is amplified and further quantized by the following MDAC
stages. Thus, the integrator output $Y$ is derived from $D_1$ and the calibrated pipeline stage output $D_{c2}$:

$$Y = D_1 - E_{q1} \approx D_1 + D_{c2}.$$  \hfill (6.3)

The identification of the first stage integrator error $E_d$ and $E_c$ has been already detailed in Chapters 3, 4, and 5, so, in this chapter, we will focus on the compensation of the inter-stage gain error assuming that the first stage non-ideality is properly addressed.

In Figure 6.4, the digital output of the second to fifth pipeline stage is expressed by the $i$-th stage gain $\alpha_i$ and quantization noise $E_{q(i)}$.

$$D_i = -\alpha_iE_{q(i-1)} + E_{q(i)} (i = 2, 3, 4, 5).$$  \hfill (6.4)

On the other hand, it is supposed that the MDAC performs memoryless functions and its nonlinearity is sufficiently suppressed by the former stages. Hence, the memoryless linear gain-error model is applied for the MDAC calibration, and the circuit bandwidth and linearity specifications are determined accordingly. After the calibration, the final output $D_c$ is derived as

$$D_c = D_1 - T_a - \hat{E}_d + z^{-1}\hat{E}_c + \left(1 - \hat{\beta} z^{-1}\right)\left(\frac{D_2}{\hat{\alpha}_2} + \frac{D_3}{\hat{\alpha}_3} + \frac{D_4}{\hat{\alpha}_4} + \frac{D_5}{\hat{\alpha}_5}\right),$$  \hfill (6.5)

where $\hat{E}_d$ and $\hat{E}_c$ are the estimated error of $E_d$ and $E_c$, respectively. $\hat{\alpha}_i$ is the estimated MDAC gain and $\hat{\beta}$ compensates for the PN mismatch induced term $\beta$.

Using Equations (6.1) and (6.4), Equation (6.5) is rewritten by
\[
D_c = z^{-1} \frac{X}{1 + \Delta_n} + (E_d - \hat{E}_d) - z^{-1} (E_c - \hat{E}_c) + \left[ \frac{(\alpha_2 - \alpha_3) - (\alpha_2 \beta - \alpha_3 \hat{\beta}) z^{-1}}{\hat{\alpha}_2} \right] E_q^{(1)} + \left(1 - \hat{\beta} z^{-1}\right) \left[ \frac{(\alpha_3 - \alpha_3) E_q^{(2)}}{\hat{\alpha}_2 \hat{\alpha}_3} + \frac{(\alpha_4 - \alpha_4) E_q^{(3)}}{\hat{\alpha}_2 \hat{\alpha}_3 \hat{\alpha}_4} + \frac{(\alpha_5 - \alpha_5) E_q^{(4)}}{\hat{\alpha}_2 \hat{\alpha}_3 \hat{\alpha}_4 \hat{\alpha}_5} + \frac{E_q^{(5)}}{\hat{\alpha}_2 \hat{\alpha}_3 \hat{\alpha}_4 \hat{\alpha}_5 \hat{\alpha}_5} \right].
\]

(6.6)

In Equation (6.6), when \( \hat{\alpha}_i = \alpha_i \) and \( \hat{\beta} = \beta \), each stage quantization noise \( E_{q(i)} \) is cancelled out with the next stage output \( D_{(i+1)} \) and only the last stage quantization noise \( E_{q5} \) remains in \( D_c \). Otherwise, \( D_c \) includes the leakage of the \( i \)-th stage quantization noise \( E_{q(i)} \). In correlation, the in-band signal \( X \) is filtered out from \( D_c \) by a HPF. Thus, the correlation between \( D_c \) and \( E_{q(i-1)} \) gives the clue of \( \hat{\alpha}_i \). From Equation (6.4), \( E_{q(i-1)} \) is approximated to the calibrated \( i \)-th stage output \( D_{c(i)} \):

\[
E_{q(i-1)} = E_{q(i)} - \frac{D_i}{\alpha_i} \approx \frac{E_{q(i)} - D_i}{\hat{\alpha}_i} = \frac{E_{q5}}{\hat{\alpha}_i \hat{\alpha}_{i+1} \cdots \hat{\alpha}_5} - \left( \frac{D_i}{\hat{\alpha}_i} + \frac{D_{i+1}}{\hat{\alpha}_i \hat{\alpha}_{i+1}} + \cdots + \frac{D_5}{\hat{\alpha}_i \hat{\alpha}_{i+1} \cdots \hat{\alpha}_5} \right) \approx -D_{c(i)}.
\]

(6.7)

As a result, the LMS update equation is obtained by

\[
\hat{\alpha}_i^{m+1} = \hat{\alpha}_i^m - \mu_i d_{c(i)}(n) e(n), (i = 2, 3, 4, 5).
\]

(6.8)

where \( E \) is the filtered version of \( D_c \). In Equations (6.6) and (6.7), both \( E \) (or \( D_c \)) and \( D_{c(i)} \) include \( E_{q5} \), so the final value of \( \hat{\alpha}_i \) deviates from \( \alpha_i \) (estimation error).

However, for the former stages, \( E_{q5} \) in \( D_{c(i)} \) is attenuated by the later stage gain in Equation (6.7), which makes this correlation weak enough to satisfy the accuracy requirement. The estimation error \( (\hat{\alpha}_i - \alpha_i) \) of the later stage does not affect the
calibration performance because the error term is divided by the overall former stage gain in Equation (6.6). From the similar understanding, \( \hat{\beta} \) is trained by

\[
\hat{\beta}^{n+1} = \hat{\beta}^n + \mu d_{c2}(n-1)e(n).
\]

The consecutive quantization noise of the first stage 3-b quantizer, \( d_{c2}(n) \) and \( d_{c2}(n-1) \), contains mutually exclusive moment, which permits the coefficients \( \hat{\alpha}_2 \) and \( \hat{\beta} \) to be trained in Equations (6.8) and (6.9), respectively. As a result, in MASH structures, the mismatch between stages can be figured out without PN signals by correlating the residue error to the calibrated data of each stage. It should be recognized that this gain calibration approach is possible since the input signal is band-limited and efficiently removed in oversampling structure. Each stage quantization noise leakage, especially the front stage, is strongly correlated to the input signal, so the gain error cannot be identified in the presence of the input [70],[71]. Although only the gain calibration is applied in the MDAC stages, this approach is extensible to nonlinear calibration by adopting ICA approach [58].

### 6.3 Circuit Design

One of the challenges of analog circuit design using digital low-leakage process is the high device threshold voltage in low supply voltage. The high threshold voltage restricts the choice of circuit topologies due to the following issues. First, the switch on-resistance is an order of magnitude larger than that of
the low-threshold devices generally supported in analog processes, which restricts the bias voltage range of analog switches to lower the on-resistance. Second, the input common-mode (CM) of the differential amplifier should be properly chosen allowing the saturation voltage of the current source. Lastly, the logic gate delay is roughly double that in the analog process, so the delay of the critical path, e.g., the path from the sub-ADC output to the DAC input, needs to be carefully optimized. This section describes the implementation of the prototype ADC addressing these design obstacles and presents the feasibility of the performance enhancement of ΣΔ ADCs with high-speed but less accurate analog circuits.

6.3.1 DAC and PN Injection

The circuit diagram of the first stage integrator and DAC is shown in Figure 6.5. For a wide input bandwidth, the input sampling capacitor \(C_{IN}\) is separated from the DAC capacitors \(C_u\). The sampling switch nonlinearity is sufficiently attenuated by the clock bootstrap technique [44]. To calibrate the integrator, a one-bit PN \(T_a\) is injected through the capacitor \(C_{Ta}\) whose size determines the calibration speed and accuracy. \(C_{PNa}\) is set to half of the unit capacitor \(C_u\), such that the test signal occupies around 1 dB of the input signal dynamic range with a moderate calibration speed. The DAC calibration PN \(T_c\) is selectively injected through one of \(C_u\) and immediately removed in analog domain by an additional capacitor \(C_{Tc}\) to minimize the dynamic range loss due to the injection.
Figure 6.5. First stage integrator with a 3-b DAC.

Figure 6.6 compares the conventional differential DAC encoder and the proposed one with PN injection. In the conventional differential structure of Figure 6.6(a), two adjacent comparator outputs $Q_2Q_{2i-1}$ are converted into the corresponding differential code $D_i$, −1, 0, or 1, which build 9 encoder output levels from the 8 comparators. The proposed encoding scheme in Figure 6.6(b) is identical to the conventional one when $Q_2Q_{2i-1}$ is either 00 or 11. However, when $Q_2Q_{2i-1}$ is 01, this code is converted to either −1 or 1 instead of 0 depending on $T_c$. In other words, $T_c$ taking on the value of −1 or 1 is injected whenever one of the
$Q_2Q_{2i-1}$ sets is 01. In turn, the five level encoder output and one-bit PN represents 9 levels, eliminating the middle switch and thus simplifying the circuit.

![Diagram](image)

Figure 6.6. DAC encoder circuit: (a) conventional and (b) proposed with PN injection.

For the PN injection, an addition logic gate is inserted at the data path of the encoder, which possibly delays the connection of $C_{u1}$ to $C_{u4}$, but the overhead is
minimal because the number of switches driven by the encoder is much reduced with $PN_c$ generated in advance. The delay of the data path with and without PN injection is 29 ps and 25 ps, respectively, in this process. The critical path of the encoder has five logic gate delay (75 ps) determining the cancellation of the injected PN (detection of 0 in four $D_i$’s). This delay, however, affects only the connection of $C_{Te}$ and the operation still can be accomplished between the fast falling and rising edges of non-overlap clocks, i.e., the falling edge of $\phi_{1P}$ and the rising edge of $\phi_2$ in Figure 6.7. This is the one of the major benefits compared with the conventional DEM circuit.

![Figure 6.7. Clock timing.](image)

6.3.2 Integrator

Using a non-overlapping two-phase clock, the integrator in Figure 6.5 dumps the previously sampled charge in $C_{IN}$ on the feedback capacitor $C_F$ at $\phi_2$, and holds the output level until the end of $\phi_1$. Thus, the integrator output can be
sampled at the end of either the $\phi_1$ or $\phi_2$ phase by the following blocks. At $\phi_2$, the summing node switch $SW_1$ is turned on, so the amplifier is more stable with a low feedback factor (a low loop gain). In contrast, at the next clock phase $\phi_1$, $SW_1$ is turned off, and the increased feedback factor makes the loop less stable. In the typical high-gain two-stage amplifier design, the second-stage load capacitance does not have a strong influence on the operating speed, whereas it is one of the dominant factors degrading the phase margin. In turn, the $\phi_2$ phase sampling is usable to ensure a sufficient phase margin at $\phi_1$ with low output capacitance. This sampling timing is also suitable for the conventional analog domain linearization techniques [11],[12],[16]–[19], which usually utilize the $\phi_1$ phase for the additional analog signal processing.

The low-gain two-stage amplifier has a different design perspective from the conventional high-gain one. Unlike to the conventional design, the operating speed of the low-gain two-stage amplifier is affected by both first and second stage output capacitance. These features will be detailed in the amplifier design part. From these observations, in this work, the integrator output is sampled at $\phi_1$, such that the amplifier has a fast operating speed at $\phi_2$ with load-less configuration. The operating speed is still fast even with the load capacitance at $\phi_1$ because the amplifier has a higher feedback factor. The phase margin at $\phi_1$ is ensured by a compensation capacitor which is necessary to satisfy the noise specification.
Owing to the digital linearization techniques, neither the stage gain error nor nonlinearity is a major concern dictating the amplifier gain. Instead, the lower limit of the dc gain is given by the quantization noise leakage in a single stage structure, which is equivalent to the minimum fixed-point error in calibration. For the design in Figure 6.5, the fixed-point error in Equation (3.16) is modified to

\[
\bar{E}_e = \frac{1}{12 \text{OSR} A_{\text{max}}^2 4^N} \left[ \left( \frac{C_{\text{St}} + C_{P1}}{4C_u} \right)^2 + \gamma (\gamma - 1) \frac{\pi^2}{3 \text{OSR}^2} \right],
\]

(6.10)

This equation indicates that \( C_{P1} \), which is reset every cycle, has a larger effect on \( \bar{E}_e \) than \( C_{P2} \). \( C_{IN} \) is set to smaller than \( 4C_u \) (\( C_{IN} = 512 \text{ fF}, 4C_u = C_F = 640 \text{ fF} \) (single ended)) to reduce \( \bar{E}_e \). The estimated \( kT/C \) noise of this sampling capacitance is around 12 b level for a 1.2 VPP differential input and an OSR of 8.

Considering the signal dynamic range of the PN and quantization noise, the top and bottom reference voltage are set to 0.9 V and 0.3 V, respectively. The amplifier gain requirement needs to be larger than 26 dB (20 V/V) to guarantee that the SNDR loss is less than 2 dB. Without digital calibration, the amplifier gain should be far larger than 60 dB (1000 V/V) to accomplish the similar linearity in the MASH structure.

The first stage integrator simulation results are summarized in Table 6.1. The amplifier gain in the open-loop configuration reduces around 7% for the
maximum output of ±0.6 V PP, which mostly stems from the signal dependent output impedance of the second stage.

Table 6.1. Simulation summary of the integrator (1.2 V supply, 27 °C, Typical)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplifying phase settling time</td>
<td>0.96 ns (10 b level)</td>
</tr>
<tr>
<td>Open-loop Amplifier Gain</td>
<td></td>
</tr>
<tr>
<td>Maximum (AC gain)</td>
<td>29 V/V</td>
</tr>
<tr>
<td>Minimum (maximum output)</td>
<td>27 V/V</td>
</tr>
<tr>
<td>CM gain</td>
<td>~ 0.2 V/V</td>
</tr>
<tr>
<td>Input-referred noise (1 kHz ~ 1 THz) / (OSR)(^{0.5})</td>
<td>~ 200 (\mu)V(_{rms})</td>
</tr>
<tr>
<td>Compensation capacitor</td>
<td>320 fF</td>
</tr>
<tr>
<td>Load capacitance (amplifying / holding phases)</td>
<td>~ 600 fF / ~ 280 fF</td>
</tr>
<tr>
<td>Transconductance ((G_{m1} / G_{m2}))</td>
<td>22.4 mA/V / 35.9 mA/V</td>
</tr>
<tr>
<td>(f_{-3\text{dB}}) (amplifying phase)</td>
<td>1.8 GHz</td>
</tr>
<tr>
<td>(1 / \beta) ((\text{@ 1 kHz / @ 100 kHz / @} f_{-3\text{dB}}))</td>
<td>15.2 / 4.6 / 3.2 V/V</td>
</tr>
<tr>
<td>Phase margin (amplifying / holding phases)</td>
<td>86 ° / 72 °</td>
</tr>
<tr>
<td>Power consumption</td>
<td>11.1 mW</td>
</tr>
</tbody>
</table>

The integrator noise is inversely proportional to the feedback factor \(\beta\). In deep sub-micron technologies, \(\beta\) much increases at low frequency due to the gate leakage current, so the integrated in-band noise power from 10 Hz to 1 kHz is
even larger than the noise from 1 kHz to 1 THz reflecting the OSR. In wideband applications, the input information is not carried at lower than a few tens of kHz and the noise at this band is filtered out in digital domain, so the design is focused on the noise larger than 1 kHz frequencies.

6.3.3 Low Gain Two-Stage Amplifier

In this process technology, the intrinsic gain of a single stage amplifier is less than 20 dB, so the two-stage amplifier in Figure 6.8(a) is selected in the integrator design to realize higher than 26 dB gain. All MDAC stages also have the same structure supporting the similar level of the dc gain to sufficiently attenuate the amplifier nonlinearity in analog domain (the nonlinearity of the MDACs is not calibrated).

The input CM voltage is set to a quarter of the supply voltage $V_{DD}$, which allows a smaller size of NMOS switches at the summing node and ensures flexible input CM voltage, thereby allowing high threshold voltage (around 0.5 V) of the PMOS input pair. The output CM voltage is around half $V_{DD}$ for the single-ended output swing of 0.6 $V_{PP}$ (differentially 1.2 $V_{PP}$).

One drawback of this simple resistive CM feedback output stage is the dependency of the output CM voltage on $|V_{GS}|$ of the PMOS load transistors. With a high threshold voltage, a large PMOS device takes a large portion of the load capacitance at the amplification phase $\phi_l$ to ensure the output swing. Particularly,
in a pipeline structure, the CM voltage deviation needs to be cautiously investigated because the CM error propagated from the former stage possibly reduces the signal dynamic range of the later stages.

Figure 6.8. (a) Two-stage amplifier and (b) simplified model of (a).

The CM propagation can be calculated from the CM gain of each stage.

The two-stage amplifier with the diode connected configuration has an open-loop
CM gain $A_{CM(op)}$ of around one. Using $A_{CM(op)}$, the closed-loop CM gain $A_{CM(cl)}$ is written as

$$A_{CM(cl)} = \frac{C_S}{C_F - \frac{C_{TOT}}{A_{CM(op)}}} \approx -\frac{C_S}{C_{TOT} - C_F},$$  \hspace{1cm} (6.11)

where $C_{TOT}$ is the total capacitance connected to the summing node including the parasitic capacitance ($C_{TOT} > C_F$), and $C_S$ and $C_F$ is the sampling and feedback capacitor, respectively. Note that the closed-loop has a positive CM feedback ($A_{CM(op)} > 0$) in the two-stage amplifier. In spite of the positive feedback, the loop is still stable with the CM loop gain smaller than one. Since $A_{CM(op)}$ is around one, $A_{CM(cl)}$ becomes a negative value larger than $-1$ ($-1 < A_{CM(cl)} < 0$). Using Equation (6.11), the total CM error at the last (5th) stage output becomes

$$\Delta V_{CM,\text{tot}} = A_{CM(cl)}^2 \Delta V_1 + A_{CM(cl)}^3 \Delta V_2 + A_{CM(cl)}^2 \Delta V_3 + A_{CM(cl)} \Delta V_4 + \Delta V_5 \approx \left(A_{CM(cl)}^4 + A_{CM(cl)}^3 + A_{CM(cl)}^2 + A_{CM(cl)} + 1\right)\Delta V_{CM},$$  \hspace{1cm} (6.12)

where $\Delta V_i$ is the output referred CM error of the individual $i$-th stage. Since every stage has the same structure, it can be safely assumed that all stages have similar amounts of CM error $\Delta V_{CM}$ or, at least, the same polarity of $\Delta V_{CM}$. Based on this assumption, Equation (6.12) manifests that each stage CM error is not accumulated in the pipeline stages but first-order cancelled out between two consecutive stages due to the negative $A_{CM(cl)}$. It should be noticed that this cancellation arises from the positive open-loop CM gain of the two-stage structure.
With a single stage amplifier, the closed-loop CM gain is negative, so the CM error is accumulated through the pipeline stages.

Now, two different design perspectives in low-gain two-stage amplifier are detailed. First, as the amplifier gain approaches the inverse of the feedback factor, i.e., the loop gain approaches one, the phase margin greatly increases (when the loop gain is less than one, the loop is unconditionally stable). Second, the second stage load capacitance as well as the first stage load is an important factor determining the amplifier bandwidth.

In a simplified two-stage amplifier model shown in Figure 6.8(b), the first and second poles, \( p_1 \) and \( p_2 \), and phase margin \( PM \) are given by [10]

\[
p_1 \approx \frac{1}{g_{m2} R_2 C C_1 R_1 + C_2 R_2}, \quad p_2 \approx \frac{g_{m2} C}{C C_2 + C_2 C_1},
\]

\[
PM : 180 - \tan^{-1} \left( \frac{\omega_0}{p_1} \right) - \tan^{-1} \left( \frac{\omega_0}{p_2} \right).
\]

In Equation (6.14), \( \omega_0 \) is the frequency where the loop gain is one:

\[
A_{sp}(j\omega_0) \beta = \frac{A_{DC} \beta}{1 + j\omega_0 / p_1} = 1,
\]

where a single-pole system with a feedback factor \( \beta \) and a dc gain \( A_{DC}(= A_{sp}(0)) \) is assumed considering \( \omega_0 < p_2 \). From Equation (6.15), \( \omega_0 \) is derived as

\[
\omega_0 = p_1 (A_{DC} \beta - 1).
\]
With a high loop gain ($A_{DC\beta} > 10$), $\omega_0$ resides at a much higher frequency than first pole $p_1$, so the second pole $p_2$ should be placed far higher than $\omega_0$ to ensure enough phase margin from Equation (6.14). As the loop gain is lowered, $p_1$ approaches $\omega_0$, reducing the phase shift at $\omega_0$. When the loop gain becomes two, $p_1$ is the same as $\omega_0$. Under this condition, the phase shift originating from $p_1$ becomes only $45^\circ (= \tan^{-1}(\omega_0/p_1))$, so more than $90^\circ$ phase margin can be offered even with $p_2$ placed close to $\omega_0$. In this work, the loop gain of the MDACs is around 3.5 with a feedback factor of around 0.11, which causes the phase shift of only $68^\circ$ from the first pole. This analysis reveals that a low gain amplifier with a low loop gain (a large number of bit per stage) allows smaller $g_{m2}$ and $C_c$ in the phase margin consideration from Equations (6.13) and (6.14). It is important to note that, for a low-loop gain, $\omega_0$ is not same as the closed-loop $-3$dB frequency ($\omega_{-3\text{dB}}$), which dictates the operating speed of the amplifier. From the definition of $-3$dB frequency,

$$|A_d(j\omega_{-3\text{dB}})| = \frac{1}{2}|A_d(0)|,$$  \hspace{1cm} (6.17)

and the closed-loop gain $A_c(j\omega)$ is derived from the charge conservation law:

$$A_c(j\omega) = \frac{C_s}{C_F + \frac{C_{TOT}}{A_{DC}} \left[ 1 + \frac{j\omega}{p_1} \right]}.$$ \hspace{1cm} (6.18)

From Equations (6.17) and (6.18), $\omega_{-3\text{dB}}$ is obtained as

- 110 -
\[ \omega_{3\text{dB}} = p_1(A_{\text{DC}} \beta + 1). \] 

Equation (6.16) implies that the amplifier is always stable with the loop gain close to one, whereas Equation (6.19) shows that \( \omega_{3\text{dB}} \) is same as \( p_1 \) when the loop gain is zero (open-loop configuration). Using Equation (6.13), Equation (6.19) is rewritten by

\[ \omega_{3\text{dB}} = \beta \left( \frac{g_{m1} + \frac{1}{R_1g_{m2}R_2\beta}}{C_c + C_2 / R_1g_{m2}} \right). \] 

It becomes rather interesting to investigate \( \omega_{3\text{dB}} \) in Equation (6.20). Assuming large \( g_{m2}R_1 (> 10) \), \( \omega_{3\text{dB}} \) is typically approximated to \( (\beta g_{m1}) / C_c \) (the effect of \( C_2 \) can be ignored). In contrast, when the amplifier intrinsic gain is small, \( C_2 \) becomes an important factor in \( \omega_{3\text{dB}} \). For the first stage integrator, a large \( C_c \) (around 60% of \( C_2 \)) is used to ensure the noise and phase margin. For the MDAC stages, enough phase and noise margin can be still achieved with \( C_c \) far smaller than \( C_2 \), so \( \omega_{3\text{dB}} \) is even more sensitive to \( C_2 \). In this case, the second stage transconductance \( g_{m2} \) is optimized for a better \( \omega_{3\text{dB}} \) and the phase margin becomes the second consideration.

The simulation results of the first MDAC (the second pipeline stage) are summarized in Table 6.2. The settling time of the output is around 1 ns at 7 b level and amplifier dc-gain variation is around 7%. Owing to the low feedback factor and loop gain, the phase margin of 90° is achieved with only 50 fF.
compensation capacitor and the small transconductance ratio between the first and second stage amplifiers. The first and second stage transconductance is optimized for the maximum $f_{-3\text{dB}}$.

### Table 6.2. Simulation summary of the first stage MDAC (1.2 V supply, 27 °C, Typical)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplifying phase settling time</td>
<td>0.95 ns (7 b level)</td>
</tr>
<tr>
<td>Gain</td>
<td></td>
</tr>
<tr>
<td>Maximum (minimum output)</td>
<td>34 V/V</td>
</tr>
<tr>
<td>Minimum (maximum output)</td>
<td>32 V/V</td>
</tr>
<tr>
<td>Stage gain</td>
<td>3.47 V/V</td>
</tr>
<tr>
<td>Compensation capacitor</td>
<td>50 fF</td>
</tr>
<tr>
<td>Load capacitance</td>
<td>~ 450 fF</td>
</tr>
<tr>
<td>Transconductance ($G_{m1} / G_{m2}$)</td>
<td>10.2 mA/V / 13.2 mA/V</td>
</tr>
<tr>
<td>$f_{-3\text{dB}}$</td>
<td>0.8 GHz</td>
</tr>
<tr>
<td>$1 / \beta$ (@ 100 kHz / @ $f_{-3\text{dB}}$)</td>
<td>8.8 / 8.1</td>
</tr>
<tr>
<td>Phase margin (amplifying / holding phases)</td>
<td>90 °</td>
</tr>
<tr>
<td>Power consumption</td>
<td>6.7 mW</td>
</tr>
</tbody>
</table>

#### 6.3.4 Comparator

The two-stage preamplifier in Figure 6.9 is designed with a dc-gain of 30 V/V and $-3$ dB frequency $f_{-3\text{dB}}$ of 610 MHz. The gain of the input network is 0.72,
and the first and second stage gains are 6.3 and 6.7, respectively. The effective gain at $f_{-3\text{dB}}$ is 22 V/V. To simplify the current bias, the first stage bias current is replicated to the second stage from the device ratio between first stage load and second stage input pair. The NMOS load of the second stage preamplifier supports high CM voltage to the following latch block, such that the dynamic latch with NMOS input pair has high operating speed. The input CM voltage $V_{CM}$ of the first stage preamplifier is set to $0.75 \times V_{DD}$ to ensure the bias voltage margin of the current source. In Monte Carlo simulation, the input referred offset (standard deviation) is 26 mV, which is sufficiently smaller than the 1 LSB of the sub-ADC (150 mV).

Figure 6.9. Comparator schematic.
The latch delay including two output inverters in Figure 6.10 is less than 70 ps for larger than 4 mV input and it increases up to 110 ps when the latch input is small (meta-stable).

Figure 6.10. Latch schematic.

6.4 Measurement

The ADC of Figure 6.11 is implemented in a 65-nm CMOS low-leakage digital process and its active area is 0.28 mm$^2$ (=0.73 mm × 0.38 mm). The analog input applied to the left-hand side is processed by an integrator and four MDAC stages and converted into 16 b digital output from the five sub-ADCs. Analog (integrator and MDACs) and digital (sub-ADCs) blocks are located at bottom and
top sides, respectively, and separate power supply lines are applied to minimize the coupling of the digital noise to the analog blocks. The reference voltage lines are placed between the analog and digital blocks and shared. The differential clock line is supported to improve the CM noise immunity and to generate 50 % duty cycle in multi-phase clock generation circuit.

![Die photo of the prototype ADC.](image)

Figure 6.11. Die photo of the prototype ADC.

The chip shown in Figure 6.12 has a die area of 2 mm × 2 mm, which is determined by the pad size. The differential output driver is placed near the pads, and the power is supported from the dedicated supply and ground lines. The spare chip area excluding ADC core is filled with decoupling capacitors using PMOS transistors for the power supply and reference voltage.
The chip is directly attached on the test board (chip on board package).

The body size of the 64 pads in Figure 6.12 is 5 mm \( \times \) 5 mm on the board, so the expected length of the bonding wire is around 1.5 ~ 2.0 mm. The pad description is summarized in Table 6.3.

![Figure 6.12. Bonding diagram.](image)

Table 6.3. Pad description.

<table>
<thead>
<tr>
<th>PIN #</th>
<th>Pin Name</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1     | EXPN     | External PN enable (default High)  
High: external PN, Low: internal | Input |
Table 6.3 Continued

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>PNI</td>
<td>External PN input</td>
<td>Input</td>
</tr>
<tr>
<td>3</td>
<td>SET</td>
<td>PN generator set (default Low)</td>
<td>Input</td>
</tr>
<tr>
<td>4</td>
<td>CMLD</td>
<td>Digital common mode (0.9 V)</td>
<td>I/O</td>
</tr>
<tr>
<td>5</td>
<td>REFTOP</td>
<td>Positive reference voltage (0.9 V)</td>
<td>Input</td>
</tr>
<tr>
<td>6</td>
<td>REFBOT</td>
<td>Negative reference voltage (0.3 V)</td>
<td>Input</td>
</tr>
<tr>
<td>7</td>
<td>CMLA</td>
<td>Analog common mode (0.6 V)</td>
<td>I/O</td>
</tr>
<tr>
<td>8, 18, 64</td>
<td>SUBST</td>
<td>Substrate supply</td>
<td>Power</td>
</tr>
<tr>
<td>9 / 10</td>
<td>INP</td>
<td>Positive analog input (CM: 0.3 V)</td>
<td>Input</td>
</tr>
<tr>
<td>11 / 12</td>
<td>INN</td>
<td>Negative analog input (CM: 0.3 V)</td>
<td>Input</td>
</tr>
<tr>
<td>13</td>
<td>PNEN</td>
<td>PN enable (default high)</td>
<td>Input</td>
</tr>
<tr>
<td>14</td>
<td>AVSS</td>
<td>Analog ground (0 V)</td>
<td>Power</td>
</tr>
<tr>
<td>15</td>
<td>AVDD</td>
<td>Analog supply (1.2 V)</td>
<td>Power</td>
</tr>
<tr>
<td>16</td>
<td>BSIN</td>
<td>Bias current (50μA sinking from AVDD) (1.37 kΩ connected to AVDD)</td>
<td>I/O</td>
</tr>
<tr>
<td>17</td>
<td>BSTS</td>
<td>Bias test (25μA sinking from AVDD)</td>
<td>I/O</td>
</tr>
<tr>
<td>19</td>
<td>BSO</td>
<td>1mA sinking from OVDD</td>
<td>I/O</td>
</tr>
<tr>
<td>20, 32, 48</td>
<td>OVSS</td>
<td>Output driver ground (0 V)</td>
<td>Power</td>
</tr>
<tr>
<td>21, 33, 49</td>
<td>OVDD</td>
<td>Output driver supply (1.2 V)</td>
<td>Power</td>
</tr>
<tr>
<td>22</td>
<td>PNO</td>
<td>Positive PN output 0 ~0.5V (100Ω)</td>
<td>Output</td>
</tr>
<tr>
<td>23</td>
<td>PNOB</td>
<td>Negative PN output 0 ~0.5V (100Ω)</td>
<td>Output</td>
</tr>
<tr>
<td>24 ~ 58</td>
<td>DOx</td>
<td>Positive digital output (24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 50, 52, 54, 56, 58) MSB (PAD24) ~ LSB (PAD 58) 0 ~0.5V (100Ω)</td>
<td>Output</td>
</tr>
<tr>
<td>25 ~ 59</td>
<td>DOBx</td>
<td>Negative digital output (25, 27, 29, 31, 33, 35, 39, 41, 43, 45, 47, 51, 53, 55, 57, 59) MSB (PAD25) ~ LSB (PAD 59) 0 ~0.5V (100Ω)</td>
<td>Output</td>
</tr>
<tr>
<td>60</td>
<td>DVDD</td>
<td>Digital supply (1.2 V)</td>
<td>Power</td>
</tr>
<tr>
<td>61</td>
<td>DVSS</td>
<td>Digital ground (1.2 V)</td>
<td>Power</td>
</tr>
<tr>
<td>62</td>
<td>CK</td>
<td>Positive clock input (0.6 V ± 0.6 V)</td>
<td>Input</td>
</tr>
<tr>
<td>63</td>
<td>CKB</td>
<td>Negative clock input (0.6 V ± 0.6 V)</td>
<td>Input</td>
</tr>
</tbody>
</table>
A four-layer printed circuit board (PCB) is manufactured using FR-4. The layout of the top layer shown in Figure 6.13 includes most of the signal lines. The second and third layers are assigned for supply and ground planes, respectively. The bottom layer is used for some of the digital signal routing. The board has two power domains to isolate the digital noise of the output driver. The analog and digital power of the core ADC share one power domain and the other power domain is assigned to the output driver circuits.

Figure 6.13. PCB layout of the top layer.
The measured FFT results for $-1\ dBFS$ 5 MHz input at 240 MS/s and $\times 8$ OSR are shown in Figure 6.14. Figures 6.14(a) and (b) are the reconstructed spectra of the raw modulator output without and with PN injection, respectively.

![FFT Spectra](image)

Figure 6.14. Normalized FFT spectra of raw ADC output (a) without PN and (b) with PN injection, and after calibration using (c) integrator nonlinear calibration disabled, (d) DAC calibration disabled, (e) memoryless, and (f) memory models.
The PN dithers the harmonic tones and pattern noise, which improves the SFDR by 6 dB, but the SNDR is degraded by 4 dB due to the incomplete PN removal in Figure 6.14(b). Figure 6.14(f) illustrates that the calibration improves the SNDR and SFDR by around 23 dB and 29 dB, respectively, from Figure 6.14(b). The third-order two-tap polynomial model is employed to the integrator calibration, and capacitor mismatch and gain error calibrations are also performed.

With the integrator nonlinear calibration disabled, in Figure 6.14(c), the gain and DAC calibration attenuates the noise level, whereas the performance is limited by the third-order harmonic of the integrator. On the other hand, without the DAC calibration in Figure 6.14(d) (with the nonlinear integrator and gain calibration enabled), the mismatch between the first integrator and the following pipeline stages causes the large second-order harmonic tone. Comparing Figures 6.14(e) and (f), the memory nonlinear calibration exhibits around 20 dB gain in SFDR relative to a memoryless nonlinear calibration, which clearly demonstrates the necessity of the memory error compensation.

Figure 6.15 plots the performance versus input magnitude. The calibration improves the SFDR by more than 25 dB for all input levels compared with that before calibration. The maximum input signal (0 dBFS) just before the saturation of the output is 1.2 Vpp. When the input is larger than 0 dBFS, the injected PN cannot be removed in digital domain, thereby seriously degrading the calibration performance. The SFDR for the input larger than around –2 dB is limited by the
third-order distortion of the first-stage amplifier, while, for the small input, the
SFDR is limited by the even-order distortion, which is expected mainly from the
gain mismatch between stages. The SNDR before calibration starts to saturate due
to the increased amplifier distortion when the input signal is larger than around −5
dB. After calibration, the SNDR slope is constant up to the point where the output
is saturated. When the input signal is less than −15 dBFS, the SFDR both before
and after calibration increases again.

Figure 6.15. Performance versus input signal power.

One concern of nonlinear parameter learning is its dependency on the
input signal amplitude. The training curves for −1 dBFS and −7 dBFS 5 MHz
input are compared in Figure 6.16. For a smaller input, the settling of nonlinear
coefficients $a_{ii}$’s in Figure 6.16(d) is far slower than that in Figure 6.16(a)
because the identification of the nonlinear parameter relies on the intermodulation product between the input and PN.

Figure 6.16. Training curves (240MS/s update rate) for −1 dBFS input (a) memoryless nonlinear coefficient \( a_i \)'s, (b) capacitor mismatch coefficient \( c_i \)'s, (c) performance, and for −7 dBFS input (d) \( a_i \)'s, (e) \( c_i \)'s, and (f) performance.

In DAC calibration, due to the signal dependent PN injection, \( C_2 \) and \( C_3 \) in Figure 6.16(e) are trained faster than \( c_1 \) and \( c_4 \) for a small input signal, while, in
Figure 6.16(b), the capacitor mismatch is identified with a similar speed for a large input. However, as described in Figure 6.16(f), the amplifier nonlinearity for a small input signal is not a critical limiting factor of the performance, and calibration speed is still not much different between Figures 6.16(c) and (f). The LMS adaptation achieves larger than 83 dBc within 1.8 Msamples (7.5 msec at 240 MS/s) for −1 dBFS input. The steady-state SFDR fluctuation is within ±1dB centered at 84.5dBc in Figure 6.16(c).

Figure 6.17 is the SNDR and SFDR for different input frequency at 240 MS/s clock frequency. The performance is measured at 1 MHz, 3 MHz, 5 MHz, and 10 MHz where the input signal band-pass filters removing the harmonic tone of the input source are available in the experiment.

![Figure 6.17](image)

Figure 6.17. Measured performance versus input frequency before and after calibration.
For a high frequency single-tone input, the harmonic tones appear at the out-of-band on top of the high quantization noise, so the third-order intermodulation (IM3) using two-tone signal is measured to demonstrate the modulator linearity for high frequency input.

![Normalized FFT spectra](image)

Figure 6.18. Normalized FFT spectra of raw ADC output (a) without PN and (b) with PN injection, and after calibration using (c) memoryless and (d) memory models.
The IM3 for −6.5 dBFS two-tone signals at 14.9 MHz and 15.1 MHz is shown in Figure 6.18. The calibration in Figure 6.18(d) improves the IM3 more than 30 dB from that without calibration of Figure 6.18(a). The average IM3 after calibration is 87.1dBc and the steady-state fluctuation is within ±2 dB. Again, Figures 6.18(c) and (d) illustrate that the memory model more accurately represents the modulator non-ideality. The IM3 with different input frequency is measured in Figure 6.19. Two-tone signal with 200 kHz space is applied as an input signal.

![Figure 6.19. Measured IM3 versus input frequency.](image_url)
Figure 6.20 is the SNDR and SFDR for different sampling frequencies. Up to 240 Ms/s, the SFDR after calibration is higher than 80 dBC. At low sampling rate, the SNDR is around 67.8 dB. When the sampling frequency is higher than 240 MS/s, the third order harmonic tone starts to increase and degrades the ADC performance. All measured data are obtained for –1 dBFS 5 MHz input. Due to the unbalanced input network, the second-order harmonic at the output FFT spectra changes more than 10 dB according to the sampling frequency. This second-order harmonic tone is attenuated by changing the coupling capacitor between the positive and negative analog input. As the sampling frequency increases, a small capacitor is applied to balance the differential input network. In this measurement, 56 pF and 11 pF are used for a 110 MS/s and 240 MS/s sampling rates, respectively.

Figure 6.20. Measured SNDR and SFDR versus sampling frequency.
The chip performance is summarized in Table 6.4. The power consumption of the core ADC is around 37 mW (analog 28 mW / digital 9 mW) from 1.25 V supply. The synthesized calibration circuit consumes around 12 mW and occupies 0.08 mm², which are mainly determined by the number of multipliers in the adaptive filter (APF) and LMS weight update block (WUB). The two-tap third-order polynomial model for the integrator calibration requires 10 and 6 multipliers in APF and WUB, respectively (the DAC calibration performing only bitwise operation consists of adders and multiplexers). Each multiplier with a 28-b (14-b × 14-b) word length consumes around 0.4 mW at 240 MS/s clock speed. When the calibration operates in foreground, i.e., WUB is disabled, the power consumption becomes around 9 mW.

<table>
<thead>
<tr>
<th>Sample Rate</th>
<th>240 MS/s</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before</td>
</tr>
<tr>
<td>SNDR (@ 5 MHz input)</td>
<td>47.8 dB</td>
</tr>
<tr>
<td>SFDR (@ 5 MHz input)</td>
<td>50.2 dBc</td>
</tr>
<tr>
<td>IM3 (−6.5 dBFS two-tone @15.1 MHz and 14.9 MHz)</td>
<td>55.8 dBc</td>
</tr>
<tr>
<td>Calibration time</td>
<td>~ 8 msec</td>
</tr>
<tr>
<td>Power consumption</td>
<td>ADC core</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology</td>
<td>65-nm CMOS digital low-leakage process</td>
</tr>
<tr>
<td>Area</td>
<td>ADC core</td>
</tr>
<tr>
<td></td>
<td>Calibration</td>
</tr>
</tbody>
</table>
CHAPTER 7

PIECEWISE LINEAR MODEL

The polynomial model is mainly used for nonlinear calibration in this work, but the polynomial model has two critical drawbacks. First, the error parameter identification reliant on the input signal amplitude limits the ADC performance when the input statics change faster than the calibration speed. Second, to implement the polynomial model, the calibration circuit includes many multipliers, which increases the overhead of the chip area and power consumption. The possible solution for handling these issues can be a piecewise linear model [72],[73]. Using this linear approximation, we can significantly reduce the number of multipliers and simplify the digital circuits. In addition, since the coefficients for each segment are updated by the input corresponding to the segment, it is less sensitive to the input statistics. To improve the accuracy of the model with a moderate digital complexity, this segmented linear model can even be extended to a segmented low-order polynomial model. This chapter compares the piecewise and polynomial models to understand the advantages and disadvantages of each model. Although the detailed analysis here is mostly about the accuracy and complexity in terms of the error modeling, the comparison in terms of the identification is also briefly mentioned.
7.1 Polynomial Model

In weakly nonlinear circuits such as amplifiers and switches, the polynomial function in Equation (7.1) can faithfully represent the nonlinear error with a reasonable number of error parameters.

\[ \hat{X} = \alpha_1 D_{out} + \alpha_2 D_{out}^2 + \alpha_3 D_{out}^3 + \ldots + \sum_{i=1}^{N} \alpha_i D_{out}^i \]  

(7.1)

where \( \hat{X} \) is the recovered input from the distorted output \( D_{out} \) of the circuit. The practical issue in adapting this model to digital calibration rises from the fact that polynomial function is based on the multiplication. Whenever one order of polynomial function is increased, two additional multipliers are required. Note that \( D_{out}^n \) is obtained by multiplying \( D_{out} \) and \( D_{out}^{n-1} \), which can be obtained from the lower order term. Thus, the total number of multiplications required to implement an \( N \)-th order polynomial model is \( 1+2(N-1) \). On the other hand, the word length of the multiplier is calculated from the required accuracy of the model. For the purpose of the comparison among different nonlinear models, the integral nonlinearity (INL) can be used as an indicator of nonlinearity. Although it is arguable that the INL does not necessarily represent the overall ADC nonlinearity, the linearity specification can be satisfied by attenuating the INL within a certain range.
The INL from the fixed-point multiplication, i.e., the maximum error between floating-point and fixed-point models, is derived by

\[ e = X - \hat{X} = \sum_{i=0}^{N} a_i Y^i - \sum_{i=0}^{N} a_i D_{out}^i = \sum_{i=0}^{N} a_i Y^i - \sum_{i=0}^{N} a_i \left( Y + e_q \right) \]

\[ \approx -i \sum_{i=1}^{N} a_i Y^{i-1} e_q, \]  

(7.2)

where \( |Y| >> |E_q| \) is assumed. From Equation (7.2), the normalized maximum error with \( K_i \)-bit word length is caused by the maximum output signal \( Y (Y_{\text{max}} = 1) \) and given by

\[ e_{\text{max}} \approx \left| \sum_{i=1}^{N} \frac{i \cdot a_i}{2^{K_i}} \right| \]  

(7.3)

In Equation (7.3), the linear error coefficient \( a_1 \) of circuits is generally much larger than the other higher-order coefficients \( a_i (i \neq 1) \), so using different word-length \( K_i \) for each error term realizes efficient implementation.

Figure 7.1 illustrates the simulated INL using ‘fminsearch’ function in MATLAB minimizing the MSE between the floating-point and fixed-point models. To clarify the relation between the nonlinear error and required word length, it is assumed that the offset and gain error terms are completely removed (\( K_0 = K_1 = \infty \)) and only third-order distortion is considered (\( a_3 = 0.05, a_i = 0 \) if \( i = 2 \) or \( i > 3 \)). When the bit used for the third-order model is larger than three (\( K_3 > 3 \)), the simulation results are well matched to Equation (7.3). For this simulation
condition, the word length for the third-order model should be larger than 8 bit to achieve 10 bit level INL.

![INL Graph](image)

Figure 7.1. Simulated INL for a fixed-point polynomial model ($a_3 = 0.05$).

### 7.2 Piecewise Linear Model

In the piecewise linear model, the output is divided into multiple segments $M$ and the output corresponding to each segment $m$ is approximated to a linear model with a gain $\alpha_m$ and a coordinate $(D_m, X_m)$ in the segment as shown in Figure 7.2, and the recovered input is given by

$$\hat{X} = \alpha_m (D_{\text{out}} - D_m) + X_m = \alpha_m D_{\text{out}} + \Delta_m.$$  \hfill (7.4)

In this model, only one corresponding linear model is applied to each output, so it can be implemented with one multiplier and $M$ multiplexers. The number of the
error parameters is $2M$, which is much larger than that in the polynomial model. Theoretically, this segmented model reduces the input-dependent variance of the output and expedites the calibration [62]. However, in the calibration of the oversampling structure, the in-band input signal can be eliminated before the correlation, and the variance of the output is mainly dictated by the quantization noise. Considering that the variances of the quantization noise in each segment are similar, the calibration speed will increase proportionally to the number of segments $M$. On the other hand, unlike the polynomial model where the coefficient training is all correlated, the error parameters in each segment are identified independently from the other segments.

$$
\hat{X} = \alpha_m (D_{\text{out}} - D_m) + X_m
$$

Figure 7.2. Approximation of nonlinear function to piecewise linear model.
Again, assuming that the inverse of the nonlinear circuit is well represented by a polynomial function, the estimation error is calculated from the difference between the floating-point polynomial function and fixed-point piecewise model.

$$e = \sum_{i=0}^{N} a_i Y^i - [\alpha_m D_{out} + \Delta_m]$$  \hspace{1cm} (7.5)

One of the popular ways to determine the optimum error coefficients is by using the minimization of the MSE. From Equation (7.5), MSE of each segment \( m \) is given by

$$\bar{e}^2 = \frac{1}{e_q} \sum_{m-1}^{N} \left\{ \sum_{i=0}^{N} a_i Y^i - [\alpha_m D_{out} + \Delta_m] \right\}^2.$$  \hspace{1cm} (7.6)

Solving Equation (7.6) for arbitrary nonlinear function \( \sum_{i=0}^{N} a_i Y^i \) and \( m \) in the presence of the quantization noise is complicated. For simplicity, the third-order nonlinear model without second order \( (N = 3, a_2 = 0) \) is assumed. In piecewise linear model, only one multiplier is needed, so a sufficiently large word length can be used for representing \( Y \). Thus, \( D_{out} \) is replaced by \( Y \) in Equation (7.7).

$$\bar{e}^2 = \frac{1}{e_q} \int_{(M-1)e_q}^{Me_q} \left\{ a_0 + a_1 Y + a_3 Y^3 - [\alpha_M Y + \Delta_M] \right\}^2,$$  \hspace{1cm} (7.7)

where the number of segment \( M \) is used instead of \( m \) considering that the maximum error is generated when \( m = M \). To obtain optimum \( \alpha_M \) and \( \Delta_M \),
From Equation (7.8), we get

$$
\frac{d \bar{e}^2}{d \Delta_M} = \frac{1}{e_q} \int_{(M-1)e_q}^{Me_q} \left[ a_0 + a_1 Y + a_2 Y^2 - [a_M Y + \Delta_M] \right]^2 = 0.
$$

(7.8)

Using Equations (7.10), Equation (7.9) is rewritten by

$$
\Delta_{M, opt} = a_0 + a_1 (M - 0.5) e_q + a_3 \left( \frac{3}{2} M^2 + M - \frac{1}{4} e_q^2 \right) + a_M \left( Y_0 - (M - 0.5) e_q \right),
$$

(7.9)

where $M >> 1$ ($e_q << 1$) is assumed. Substituting $\Delta_M$ in Equation (7.7) by Equation (7.9) and taking the derivative with respect to $\alpha_M$, we get

$$
\alpha_{M, opt} = a_i + 3a_i \left( M^2 - \frac{3}{4} M + \frac{1}{5} \right) e_q^2 \approx a_i + 3a_i M^2 e_q^2.
$$

(7.10)

From Equations (7.5), (7.10), and (7.11), the error between the piecewise and the third-order polynomial model is derived as

$$
e = a_0 + a_1 Y + 3a_2 M^2 e_q^2 Y - 2a_3 M^3 e_q^3.
$$

(7.11)

Since $Me_q = Y_{max}$, the normalized maximum error ($Y_{max} = 1$) with $K_j$-bit segments ($2^{K_j} = M$) is

$$
e_{max} \approx \frac{3a_3}{2^{2K_j}}.
$$

(7.12)

Equation (7.13) shows that the maximum fixed-point error decreases by four times when the number of segments is doubled.
Figure 7.3 illustrates the INL between the floating-point polynomial function and piecewise linear model with a fixed number of segments under the same method and simulation condition as Figure 7.1 ($K_0$ and $K_1 = \infty$, $a_3 = 0.05$, $a_i = 0$ (if $i = 2$ or $i > 3$)). Similar to the first-order analysis in Equation (7.13), the INL in Figure 7.3 reduces as a function of $2^{K_s}$. To achieve 10 b level INL, the number of segment ($2^{K_s}$) larger than 8 (3 bit) is required.

![Figure 7.3. Simulated INL for a piecewise linear model ($a_3 = 0.05$).](image)

In the piecewise linear model in Equation (7.4), if we set gain error $\alpha$ term to be 0, it becomes a look-up table. Assuming that the word length is sufficiently large, the number of columns needed in a look-up can be derived in a similar way, and the maximum fixed-point error is given by
In the previous sections, the relation between the maximum fixed-point error and the hardware cost is analyzed. Although the polynomial model efficiently expresses the weakly nonlinear circuit, the number of the multiplier and its word length increase when the order of polynomial or amount of nonlinearity increase, thereby requiring high hardware cost in modeling nonlinear error in Equation (7.3). The maximum error in the piecewise model can be greatly attenuated according to the number of segments which can be implemented by simple multiplexers, so that it can be more efficient to represent highly nonlinear circuits.

The number of error parameters in the piecewise linear model is much larger than that in the polynomial model, which is expected to increase the effort in the identification of the model. In this aspect, the proper trade-off between these two models, e.g., a segmented polynomial model, can be a good future research topic.

Figure 7.4 compares the calibration performance between the third-order two-tap polynomial and 8-segment two-tap piecewise linear model using the measured ADC data from Figure 6.14. In the PN-based calibration, the minimum
segment larger than the PN signal amplitude is necessary for the identification, so the maximum number of segments is 8 in this work (the PN amplitude is around 1/8 of the input dynamic range). Although the adaptation method and mechanism need to be further investigated in the future, the measured results illustrate that this simple piecewise linear model can replace the polynomial model. The estimated total power of the 8-segment piecewise calibration is around 50 % of that using the third-order polynomial model.

Figure 7.4. Measured FFT results after calibration using (a) third-order polynomial and (b) 8-segment piecewise linear models.
CHAPTER 8

CONCLUSION

This research presents the first attempt to digitally compensate the nonlinear memory errors in $\Sigma\Delta$ modulators. The output-referred distortion analysis helps to analyze the nonlinear memory error in switched-capacitor circuits and treat multiple error sources in one frame. The analysis shows that the nonlinear errors in a DTI can be represented by an APTF with one-tap memory. To efficiently address the capacitor mismatch, the selective PN injection utilizing the differential DAC structure reduces the circuit overhead involved with the calibration. The injected PN is cancelled out in the analog domain, which resolves the input signal dynamic range reduction. The error parameters identification based on the independence between the input signal and a one-bit PN significantly simplifies the complexity of the calibration circuit.

The proposed techniques are applied to the prototype 1-0 MASH structure to show the effectiveness and limitations of the proposed schemes. The prototype ADC successfully demonstrates that the static nonlinear memory error can be compensated in the digital domain allowing simple analog circuit topologies. Unlike to the computer simulation, which generates only the deterministic error, the dynamic error of the real chip is difficult to model and calibrate, so the
performance enhancement is limitedly achieved when comparing the measurement results to simulated ones. Thus, either a more sophisticated model reflecting the dynamic errors or careful analog design minimizing the error will be required to further improve the efficiency of the proposed approaches.

The complexity of the digital calibration is dependent on the nonlinearity of the analog circuits. For weakly nonlinear amplifiers, a simple third-order model sufficiently represents the nonlinear error. With a moderate OSR and gain, this nonlinear calibration is required for only the first stage. Thus, finding the optimal trade-off between analog and digital parts will be a good topic for future research.

The comparison between the polynomial and piecewise models has been described briefly in Chapter 7. The identification of the piecewise model needs further investigation for the complete comparison. Although the memory model is mainly applied in the context of the $\Sigma\Delta$ conversion, the concept of the nonlinear memory error analysis is extensible to the other types of mixed-signal processing. With technology downscaling, the digital technique can potentially lead to a viable approach to achieve wideband, highly linear $\Sigma\Delta$ ADCs.
APPENDIX A

ANALYSIS OF FIXED-POINT ERROR IN CALIBRATION

A.1 First-Order Modulator

Proof of Equation (3.16):

Consider only the linear error \( i = 1 \) in Equation (3.15)

\[
e_i(n) \approx -a_ie_i(n) + b_ie_i(n-1).
\]  

(A.1)

Taking the \( z \)-transform of both sides and setting \( z = e^{j2\pi f} \), the power spectral density (PSD) of \( E_e \) is given by

\[
|E_e(e^{j2\pi f})|^2 \approx |a_i - b_ie^{j2\pi f}|^2 |E_q(e^{j2\pi f})|^2
\]

\[
= \left[ (a_i - b_i)^2 + 4a_i b_i \sin^2 \pi f \right] |E_q(e^{j2\pi f})|^2. 
\]  

(A.2)

Therefore, the in-band noise power for a given OSR is

\[
\overline{E_e^2} \approx 2\sigma_e^2 \int_0^{\frac{1}{2OSR}} \left[ (a_i - b_i)^2 + 4a_i b_i \sin^2 \pi f \right] df
\]

\[
= \frac{1}{12\text{OSR} 4^N} \left( \frac{1}{A_{\max}^2} + \frac{\gamma (\gamma - 1)}{A_{\max}^2} \pi^2 \right). 
\]  

(A.3)

where \( \sin(\pi f) \approx \pi f \) for \( f << 1 \) and \( \gamma = \frac{C_S + C_F + C_P}{C_S} \).
Relationship between fixed-point error and quantization noise leakage:

Assuming a linear model of Equation (3.1) for the modulator, the input-referred quantization noise leakage is

\[
E_q(z) = \left[ \frac{NTF_{ideal}}{STF_{ideal}} \frac{NTF_{real}}{STF_{real}} \right] E_q(z) \\
= \left[ \frac{1-z^{-1}}{z^{-1}} - \frac{1-\beta z^{-1}}{\alpha z^{-1}} \right] E_q(z) = \frac{-a_i + b_i z^{-1}}{z^{-1}} E_q(z),
\]  
(A.4)

which essentially reduces to Equation (A.1).

A.2 Second-Order Modulator

Proof of Equation (3.24):

Consider only the first-order \((i = 1)\) terms in Equation (3.22)

\[
e_i(n) \approx -a_{i1}y_1(n-1) + b_{i1}y_1(n-2) \\
- a_{i2}y_2(n) + \left( b_{i2} + a_{i2} \right) y_2(n-1) - b_{i2}y_2(n-2),
\]  
(A.5)

where

\[
a_{i1} = \frac{C_{F1} - C_{S1}}{C_{S1}} + \frac{C_{S1} + C_{F1} + C_{P1}}{A_i C_{S1}}, \quad b_{i1} = \frac{C_{F1} - C_{S1}}{C_{S1}} + \frac{C_{F1} + C_{P1}}{A_i C_{S1}},
\]

\[
a_{i2} = \frac{C_{F2} - C_{S2}}{C_{S2}} + \frac{C_{S2} + C_{F2} + C_{P2}}{A_i C_{S2}}, \quad b_{i2} = \frac{C_{F2} - C_{S2}}{C_{S2}} + \frac{C_{F2} + C_{P2}}{A_i C_{S2}},
\]

Using Equation (3.23), Equation (A.5) is rewritten as
Thus, the fixed-point error is

\[ e_x(n) = e_q(n) - e_q(n) \approx c_1 e_q(n) + c_2 e_q(n-1) + c_3 e_q(n-2), \quad (A.7) \]

where

\[ c_1 = -a_{11} - a_{21} - a_{1} a_{21}, \quad c_2 = a_{11} + a_{21} + b_{11} + a_{1} b_{21} + a_{2} b_{11}, \]

\[ c_3 = -b_{11} - b_{21} - b_{1} b_{21}. \]

The PSD of \( E_e \) is

\[
\left| E_e(e^{i2\pi f}) \right|^2 \approx \left| c_1 + c_2 e^{-j2\pi f} + c_3 e^{-j4\pi f} \right|^2 \left| E_q(e^{i2\pi f}) \right|^2
\]

\[ = \left( c_1 + c_2 + c_3 \right)^2 - 4 \left[ (c_1 c_2 + c_2 c_3) \sin^2 \pi f + c_1 c_3 \sin^2 2\pi f \right] \left| E_q(e^{i2\pi f}) \right|^2. \quad (A.8) \]

Therefore, the in-band noise power is

\[
\overline{E_e^2} \approx 2\sigma_e^2 \int_{-\frac{1}{2}}^{1} \left| E_e(e^{i2\pi f}) \right|^2 df \approx \frac{1}{12OSR^4} \left[ \frac{1}{A_1^2 A_2^2} + \left( \frac{1}{A_1} + \frac{1}{A_2} \right)^2 \frac{\pi^2}{3OSR^2} \right]. \quad (A.9)
\]

**Relationship between fixed-point error and quantization noise leakage:**

From Equation (3.22), the total modulator error is

\[
E_a(z) = z^{-1} E_{a1}(z) + \left( 1 - z^{-1} \right) E_{a2}(z)
\]

\[ = \left( \frac{\alpha_1 - 1}{\alpha_2} \right) z^{1} Y_1(z) + \left( \frac{\alpha_2 - 1}{\alpha_2} \right) z^{-1} Y_2(z). \quad (A.10) \]
Since

\[ Y_1(z) = \frac{1 + (2\alpha_2 - \beta_2)z^{-1}}{\alpha_2 z^{-1}} Y_2(z) + 2E_q(z), \]  
(A.11)

Equation (A.10) can be rewritten as

\[ E_d(z) = \frac{\alpha_1\alpha_2 - 1 + (\beta_1 + \beta_2 - 2\alpha_2)z^{-1} - (\alpha_1\alpha_2 + \beta_1\beta_2 - 2\alpha_2\beta_2)z^{-2}}{\alpha_1\alpha_2} Y_2(z) \]
\[ + 2 \left( \frac{\alpha_1 - 1 + (\beta_1 - \alpha_1)z^{-1}}{\alpha_1} \right) z^{-1} E_q(z). \]  
(A.12)

Therefore, the fixed-point error is given by

\[ \frac{E_q(z)}{E_q(z)} = \left[ \frac{\alpha_1 \alpha_2 - 1 + (\beta_1 + \beta_2 - 2\alpha_1\alpha_2)z^{-1} + (\alpha_1 \alpha_2 - \beta_1 \beta_2)z^{-2}}{\alpha_1 \alpha_2} \right]. \]  
(A.13)

In contrast, in Figure 3.3(a), the modulator output is

\[ D(z) = \frac{\alpha_1 \alpha_2 z^{-2} X(z) + (1 - \beta_1 z^{-1})(1 - \beta_2 z^{-1}) E_q(z)}{(\beta_1 \beta_2 - 2\alpha_1 \beta_1 + \alpha_1 \alpha_2)z^{-2} + (2\alpha_2 - \beta_2)z^{-1} + 1}. \]  
(A.14)

Thus, the input-referred quantization noise leakage is

\[ \frac{E_q(z)}{E_q(z)} = \frac{\text{NTF}_{\text{ideal}} - \text{NTF}_{\text{real}}}{\text{SF}_{\text{ideal}} - \text{SF}_{\text{real}}} \]
\[ = \frac{\alpha_1 \alpha_2 - 1 + (\beta_1 + \beta_2 - 2\alpha_1 \alpha_2)z^{-1} + (\alpha_1 \alpha_2 - \beta_1 \beta_2)z^{-2}}{\alpha_1 \alpha_2 z^{-2}}, \]  
(A.15)

which is equivalent to Equation (A.13).
ANALYSIS OF CAPACITOR MISMATCH IN PN INJECTION

B.1 First-Order Modulator

In Figure B.1(a), $\Delta_T$ and $\Delta_D$ represent the capacitor mismatch errors in the PN injection and DAC feedback paths, respectively. Neglecting quantization noise, i.e., $Y = D$, the modulator output with mismatch can be expressed as

$$d^n = x^{n-1} + (1 + \Delta_T) \cdot t^{n-1} + e_d^n - \Delta_D \cdot d^{n-1}, \quad (B.1)$$

where $e_d^n \approx -\sum_{i=1}^N a_i \cdot (d^n)^i + \sum_{i=1}^N b_i \cdot (d^{n-1})^i$ and the superscripts $n$ and $n-1$ are shorthand notations of sample index. The error-corrected output $D_c$ is then given by

$$d_c^n = d^n - t^{n-1} - \hat{e}_d^n, \quad (B.2)$$

where $\hat{e}_d^n \approx -\sum_{i=1}^N \hat{a}_i \cdot (d^n)^i + \sum_{i=1}^N \hat{b}_i \cdot (d^{n-1})^i$ is the estimated modulator distortion.
Figure B.1 (a) Error model of first-order \( \Sigma \Delta \) modulator with capacitor mismatches in the test signal and DAC feedback paths with quantization noise neglected, (b) and (c) the transformation leading to (B3), and (d) the effect of the PN mismatch on \( X, a_1 \) and \( b_1 \).

In the following analysis, the linear (first-order) effect of \( \Delta_T \) and \( \Delta_D \) is identified first, and the result will subsequently be generalized to the nonlinear case by induction. Neglecting the terms of \( i > 1 \) in Equations (B.1) and (B.2), we have

\[
d^n = x^{n-1} + (1 + \Delta_T)t^{n-1} - a_1d^{n-1} + b_1d^{n-1} - \Delta_d d^{n-1}
\]

\[
dx^n = x^{n-1} + (1 + \Delta_T)t^{n-1} - a_1d^{n-1} + b_1d^{n-1}, \tag{B.3}
\]

\[
d_c^n = d^n - t^{n-1} + \hat{a}_1d^n - \hat{b}_1d^{n-1}
\]

\[
dx_c^n = x^{n-1} + \Delta_T t^{n-1} + (\hat{a}_1 - a_1)d^n - (\hat{b}_1 - b_1)d^{n-1}, \tag{B.4}
\]
where \( b'_i = b_i - \Delta_d \). From Equations (3.3) and (3.4), \( \hat{a}_i \) and \( \hat{b}_i \) can be determined by jointly solving the following equations:

\[
E[d_c(n) \cdot t(n-1)] = 0, \\
E[d_c(n) \cdot t(n-2)] = 0.
\]  

(B.5)

Using Equations (B.4), (B.5) is rewritten as

\[
\begin{align*}
\Delta_T + (\hat{a}_i - a_i)E_{d'}(1) - (\hat{b}_i - b'_i)E_{d'}(0) &= 0, \\
(\hat{a}_i - a_i)E_{d'}(2) - (\hat{b}_i - b'_i)E_{d'}(1) &= 0,
\end{align*}
\]

(B.6)

where \( E_{d'}(j) = E[d(n) \cdot t(n - j)] \) and \( d(n) \) and \( t(n) \) are assumed stationary. Using Equation (B.3), we have

\[
\begin{align*}
E_{d'}(0) &= -a_iE_{d'}(0) + b'_iE_{d'}(-1), \\
E_{d'}(1) &= 1 + \Delta_T - a_iE_{d'}(1) + b'_iE_{d'}(0), \\
E_{d'}(2) &= -a_iE_{d'}(2) + b'_iE_{d'}(1).
\end{align*}
\]

(B.7)

Since \( E_{d'}(-1) = 0 \), Equation (B.7) yields

\[
\begin{align*}
E_{d'}(0) &= 0, \\
E_{d'}(1) &= \frac{1 + \Delta_T}{1 + a_i}, \\
E_{d'}(2) &= \frac{1 + \Delta_T}{1 + a_i} \cdot \frac{b'_i}{1 + a_i}.
\end{align*}
\]

(B.8)

Therefore, Equation (B.6) solves to

\[
\begin{align*}
\hat{a}_i &= -\frac{\Delta_T}{1 + \Delta_T} + \frac{a_i}{1 + \Delta_T}, \\
\hat{b}_i &= \frac{b'_i}{1 + \Delta_T} = -\frac{\Delta_T}{1 + \Delta_T} + \frac{b'_i}{1 + \Delta_T}.
\end{align*}
\]

(B.9)

Substituting Equation (B.9) in Equations (B.4) and using Equation (B.3), we have

\[
d_c(n) = \frac{1}{1 + \Delta_T} x(n-1),
\]

(B.10)
i.e., the net effect of PN mismatch is a scaling of $X$ inversely proportional to $1 + \Delta_T$. In general, when nonlinear terms are included, we can induce from the above calculation that

$$\hat{a}_i = -\frac{\Delta_T}{1 + \Delta_T} + \frac{a_i}{1 + \Delta_T}, \quad \hat{b}_i = -\frac{a_i}{1 + \Delta_T} (i = 2, 3, \ldots),$$

$$\hat{b}_i = -\frac{\Delta_D}{1 + \Delta_T} + \frac{b_i}{1 + \Delta_T}, \quad \hat{b}_i = -\frac{b_i}{1 + \Delta_T} (i = 2, 3, \ldots).$$

(B.11)

Results in Equation (B.11) are confirmed by behavioral simulations.

**B.2 Second-Order Modulator**

Figure B.2, $\Delta_{T1}$, $\Delta_{T2}$ $\Delta_{D1}$, and $\Delta_{D2}$ represent the capacitor mismatch errors in the PN injection and DAC feedback paths of the two integrator loops in a second-order modulator.

![Figure B.2](image)

Figure B.2 Error model of second-order sigma-delta modulator with capacitor mismatches in the test signal and DAC feedback paths in both integrator loops (quantization noise neglected).
Again, neglecting quantization noise, i.e., $Y_1 = D_1$ and $Y_2 = D_2$, the integrator outputs with mismatch can be expressed as

$$d_2^n = x^{n-2} + (1 + \Delta_1) t_1^{n-2} + (1 + \Delta_2) t_2^{n-1} - (1 + \Delta_2) t_2^{n-2} + e_{a_1}^{n-1} + e_{a_2}^n - e_{a_2}^{n-1} - \Delta_4 d_2^{n-1} + (\Delta_4 - \Delta_3) d_2^{n-2},$$  \hspace{1cm} \text{(B.12)}$$

$$d_1^{n-1} = y_1^{n-1} = d_2^n + (1 + \Delta_3) d_2^n - (1 + \Delta_2) t_2^{n-1} - e_{a_2}^n,$$ \hspace{1cm} \text{(B.13)}

where

$$e_{a_{ij}}^n \approx -\sum_{i=1}^{N} a_{(ij)} (d_i^n)^{'} + \sum_{i=1}^{N} h_{(ij)} (d_i^{n-1})^{'}.$$ \hspace{1cm} \text{(B.14)}$$

According to Figure 3.2(b), the calibrated output $D_c$ of the second stage is given by

$$d_c^n = d_2^n - (1 + c) t_1^{n-2} - t_2^{n-1} + t_2^{n-2} - \hat{e}_{a_1}^{n-1} - \hat{e}_{a_2}^n + \hat{e}_{a_2}^{n-1},$$ \hspace{1cm} \text{(B.15)}

where

$$\hat{e}_{a_2}^n \approx -\sum_{i=1}^{N} \hat{a}_{a_i} (d_i^n)^{'} + \sum_{i=1}^{N} \hat{h}_{a_i} (d_i^{n-1})^{'}.$$ \hspace{1cm} \text{(B.16)}$$

and the estimated first-stage output is

$$\hat{d}_{1}^{n-1} = d_2^n + d_2^{n-1} - t_2^{n-1} - \hat{e}_{a_2}^n.$$ \hspace{1cm} \text{(B.17)}$$

The coefficient $c$ is to precisely eliminate $T_1$.

We will examine the first-order behavior of the mismatch by neglecting all nonlinear terms in (B14) and (B16), which yields

$$e_{a_2}^n = -a_{21} \cdot d_1^n + b_{21} \cdot d_1^{n-1},$$ \hspace{1cm} \text{(B.18)}$$

$$e_{a_1}^{n-1} = -a_{11} \cdot d_1^{n-1} + b_{11} \cdot d_1^{n-2}$$

$$= a_{11} (1 + \Delta_2) t_2^{n-1} - b_{11} (1 + \Delta_3) t_2^{n-2} + a_{11} e_{a_2}^n - b_{11} e_{a_2}^{n-1}$$

$$- a_{11} d_2^n - [a_{11} (1 + \Delta_4) - b_{11}] d_2^{n-1} + b_{11} (1 + \Delta_4) d_2^{n-2}. $$ \hspace{1cm} \text{(B.19)}$$
Combining Equations (B.18) and (B.19), we have

\[
\begin{align*}
\frac{e_{a_1}^n + e_{a_2}^n - e_{a_2}^{n-1}}{2} &= a_{11} (1 + \Delta_2) t_{21}^{n-1} - b_{11} (1 + \Delta_2) t_{22}^{n-2} \\
- (a_{11} + a_{1}, a_{21} + a_{2}) d_{22}^{n} &
- \left[ a_{11} (1 + \Delta_2) - a_{11} b_{21} - b_{21} - b_{11} a_{21} - a_{21} \right] d_{22}^{n-1} \\
- \left[ -b_{11} (1 + \Delta_4) + b_{11} b_{21} + b_{21} \right] d_{22}^{n-2}.
\end{align*}
\]  

(B.20)

Thus, Equation (B.12) yields

\[
\begin{align*}
d_{22}^n &= x^{n-2} + (1 + \Delta_4) t_{11}^{n-2} \\
&+ (1 + \Delta_2) (1 + a_{11}) t_{22}^{n-1} - (1 + \Delta_2) (1 + b_{11}) t_{22}^{n-2} \\
&- k_{0} \cdot d_{22}^{n} - k_{1} \cdot d_{22}^{n-1} - k_{2} \cdot d_{22}^{n-2},
\end{align*}
\]

(B.21)

where

\[
k_{0} = a_{11} + a_{1}, a_{21} + a_{2},
\]

\[
k_{1} = a_{11} (1 + \Delta_4) - a_{11} b_{21} - b_{21} - b_{11} a_{21} - a_{21} + \Delta_4,
\]

\[
k_{2} = -b_{11} (1 + \Delta_4) + b_{11} b_{21} + b_{21} + \Delta_4 - \Delta_4.
\]

A similar procedure performed on Equation (B.15) results

\[
\begin{align*}
d_{c}^n &= x^{n-2} + (\Delta_4 - c) t_{11}^{n-2} \\
&+ \left[ a_{11} - \hat{a}_{11} + \Delta_2 (1 + a_{11}) \right] t_{22}^{n-1} \\
&- \left[ b_{11} - \hat{b}_{11} + \Delta_2 (1 + b_{11}) \right] t_{22}^{n-2} \\
&+ \left( \hat{k}_{0} - k_{0} \right) d_{22}^{n} + \left( \hat{k}_{1} - k_{1} \right) d_{22}^{n-1} + \left( \hat{k}_{2} - k_{2} \right) d_{22}^{n-2},
\end{align*}
\]

(B.22)

where

\[
\hat{k}_{0} = \hat{a}_{11} + \hat{a}_{1}, \hat{a}_{21} + \hat{a}_{2},
\]

\[
\hat{k}_{1} = \hat{a}_{11} - \hat{a}_{11} \hat{b}_{21} - \hat{b}_{21} - \hat{b}_{11} \hat{a}_{21} - \hat{a}_{21},
\]

\[
\hat{k}_{2} = -\hat{b}_{11} + \hat{b}_{11} \hat{b}_{21} + \hat{b}_{21}.
\]

In Equation (B.22), the known variables are \(a_{11}, b_{11}, a_{21}, b_{21}, \Delta_{1}, \Delta_{2}, \Delta_{3}, \Delta_{4}\), and \(\Delta_{4}\), and the unknowns are \(\hat{a}_{11}, \hat{b}_{11}, \hat{a}_{21}, \hat{b}_{21}\), and \(c\), which can be determined by jointly solving the following five (independent) equations:
\[ E[d_t(n) \cdot t_{t(n-j)}] = 0, \quad \text{for } j = 2 \text{ and } 3, \]  
\[ E[d_t(n) \cdot t_{t(n-j)}] = 0, \quad \text{for } j = 1, 2, \text{ and } 3. \]  
(B.23)

Using Equation (B.22), we have

\[ E[d_t(n) \cdot t_{(n-2)}] = 0 \]
\[ = (\Delta_1 - c) + (\hat{k}_0 - k_0) E_{d_{2t}}(2) \]
\[ + (\hat{k}_1 - k_1) E_{d_{2t}}(1) + (\hat{k}_2 - k_2) E_{d_{2t}}(0), \]  
(B.24)

\[ E[d_t(n) \cdot t_{(n-3)}] = 0 \]
\[ = (\hat{k}_0 - k_0) E_{d_{2t}}(3) \]
\[ + (\hat{k}_1 - k_1) E_{d_{2t}}(2) + (\hat{k}_2 - k_2) E_{d_{2t}}(1), \]  
(B.25)

\[ E[d_t(n) \cdot t_{(n-1)}] = 0 \]
\[ = (a_{11} - a_{12}) + \Delta_2 (1 + a_{12}) + (\hat{k}_0 - k_0) E_{d_{2t}}(1) \]
\[ + (\hat{k}_1 - k_1) E_{d_{2t}}(0) + (\hat{k}_2 - k_2) E_{d_{2t}}(-1), \]  
(B.26)

\[ E[d_t(n) \cdot t_{z(n-2)}] = 0 \]
\[ = - (b_{11} - b_{12}) - \Delta_2 (1 + b_{12}) + (\hat{k}_0 - k_0) E_{d_{2z}}(2) \]
\[ + (\hat{k}_1 - k_1) E_{d_{2z}}(1) + (\hat{k}_2 - k_2) E_{d_{2z}}(0), \]  
(B.27)

\[ E[d_t(n) \cdot t_{z(n-1)}] = 0 \]
\[ = (\hat{k}_0 - k_0) E_{d_{2z}}(3) \]
\[ + (\hat{k}_1 - k_1) E_{d_{2z}}(2) + (\hat{k}_2 - k_2) E_{d_{2z}}(1), \]  
(B.28)

where \( E_{d_{2t}}(j) = E[d(z(n) \cdot t_t(n-j)), \quad E_{d_{2z}}(j) = E[d(z(n) \cdot t_z(n-j)), \]

and \( d(n) \) and \( t(n) \) are again assumed stationary. Using Equation (B.21), we obtain
\[ E_{d21}(0) = E_{d21}(1) = 0, \]
\[ E_{d21}(2) = \frac{1 + \Delta_1}{1 + k_0}, \quad E_{d21}(3) = -\frac{k_i (1 + \Delta_i)}{(1 + k_0)^2}, \quad \text{(B.29)} \]
\[ E_{d22}(-1) = E_{d22}(0) = 0, \quad E_{d22}(1) = \frac{(1 + \Delta_2)(1 + a_{11})}{1 + k_0}, \quad \text{(B.30)} \]
\[ E_{d22}(2) = -\frac{k_i (1 + \Delta_2)(1 + a_{11}) - (1 + \Delta_2)(1 + b_{11})}{(1 + k_0)^2}, \quad \text{(B.31)} \]
\[ E_{d22}(3) = \frac{k^2 (1 + \Delta_2)(1 + a_{11}) + k_1 (1 + \Delta_2)(1 + b_{11})}{(1 + k_0)^3} \]
\[ - \frac{k_i (1 + \Delta_2)(1 + a_{11})}{(1 + k_0)^2}. \quad \text{(B.32)} \]

Solving Equations (B.24)–(B.28) with Equations (B.29)–(B.32), we have
\[ \hat{a}_{11} = \frac{\Delta_2 - \Delta_3 + (1 + \Delta_2) a_{11}}{1 + \Delta_3}, \quad \hat{b}_{11} = \frac{\Delta_2 - \Delta_3 + (1 + \Delta_2) b_{11}}{1 + \Delta_3}, \]
\[ \hat{a}_{21} = \frac{-\Delta_2}{1 + \Delta_2} + \frac{a_{21}}{1 + \Delta_2}, \quad \hat{b}_{21} = \frac{\Delta_2 - \Delta_3 + b_{21}}{1 + \Delta_2}, \quad \text{(B.33)} \]
\[ c = \frac{\Delta_3}{1 + \Delta_3}. \]

In addition, we also obtain from Equation (B.33)
\[ \hat{k}_0 = \frac{-\Delta_3 + k_0}{1 + \Delta_3}, \quad \hat{k}_1 = \frac{k_i}{1 + \Delta_3}, \quad \hat{k}_2 = \frac{k_2}{1 + \Delta_3}. \quad \text{(B.34)} \]

Substituting Equations (B.32) and (B.31) in Equation (B.22), using Equation (B.21), we have
\[ d_c(n) = \frac{1}{1 + \Delta_3} x(n-2), \quad \text{(B.35)} \]
i.e., the net effect of mismatch is a scaling of $X$ inversely proportional to $1 + \Delta_3$.

In general, when nonlinear terms are included, we can induce from the above calculation that

$$
\hat{a}_{11} = \frac{\Delta_2 - \Delta_3}{1+\Delta_3} + \frac{(1+\Delta_2) a_{11}}{1+\Delta_3}, \quad \hat{a}_{ii} = \frac{(1+\Delta_2)^i a_{ii}}{1+\Delta_3} (i = 2, 3, \ldots),
$$

$$
\hat{b}_{11} = \frac{\Delta_2 - \Delta_3}{1+\Delta_3} + \frac{(1+\Delta_2) b_{11}}{1+\Delta_3}, \quad \hat{b}_{ii} = \frac{(1+\Delta_2)^i b_{ii}}{1+\Delta_3} (i = 2, 3, \ldots), \quad (B.36)
$$

$$
\hat{a}_{21} = \frac{-\Delta_4}{1+\Delta_2} + \frac{a_{21}}{1+\Delta_2}, \quad \hat{a}_{2i} = \frac{a_{2i}}{1+\Delta_2} (i = 2, 3, \ldots),
$$

$$
\hat{b}_{21} = \frac{\Delta_4}{1+\Delta_2} + \frac{b_{21}}{1+\Delta_2}, \quad \hat{b}_{2i} = \frac{b_{2i}}{1+\Delta_2} (i = 2, 3, \ldots).
$$

Results in Equation (B.36) are confirmed by behavioral simulations.
REFERENCES


