2020

Design techniques to improve analog and mixed-signal circuits performance

Nanqi Liu
Iowa State University

Design techniques to improve analog and mixed-signal circuits performance

by

Nanqi Liu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical and Computer Engineering (Very Large Scale Integration)

Program of Study Committee:
Degang Chen, Major Professor
Nathan Neihart
Cheng Huang
Meng Lu
Aditya Ramamoorthy

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

TABLE OF CONTENTS

Page

LIST OF FIGURES v

LIST OF TABLES ix

ACKNOWLEDGMENTS x

ABSTRACT xi

CHAPTER 1. GENERAL INTRODUCTION 1
1.1 Background and Motivation 1
1.2 Organization of Dissertation 3

CHAPTER 2. PASSIVE ARRAY MATCHING CHARACTERIZATION WITH GRADIENT ERRORS AND RANDOM MISMATCH DECOMPOSITION 6
2.1 Abstract 6
2.2 Introduction 6
2.3 Model of Device Errors 8
2.4 Extraction and Decomposition 10
2.4.1 Extraction of Mismatch Data 10
2.4.2 Polynomial Fitting with Least Squares Method 11
2.4.3 Random Measurement Noise Robustness 13
2.5 Measurement Results 14
2.5.1 Testbench 14
2.5.2 Results for Systematic Mismatch 15
2.5.3 Results for Random Mismatch 16
2.6 Conclusion 16
2.7 References 17

CHAPTER 3. AN 8-BIT LOW-COST STRING DAC WITH GRADIENT ERRORS SUPPRESSION TO ACHIEVE 16-BIT LINEARITY 18
3.1 Abstract 18
3.2 Introduction 18
3.3 Resistor Array Mismatch 21
3.3.1 Gradient Errors and Their Effects on Linearity 22
3.3.2 Common Layout Patterns 24
3.3.3 Proposed Layout Structure 26
3.4 Outputs Averaging and Interpolation 28
3.4.1 Averaging of the Substrings' Outputs 28
3.4.2 Interpolation with Matched Resistors 31
3.4.3 Interpolation with a Differential Difference Amplifier 36
3.4.4 Discussions 37

3.5 Implementation 39
3.5.1 DAC Design 39
3.5.2 Interpolation on Board 42
3.6 Measurement Results 43
3.7 Conclusion 46
3.8 References 46

CHAPTER 4. A LOW-POWER AND AREA-EFFICIENT ANALOG DUTY CYCLE CORRECTOR FOR ADC'S EXTERNAL CLOCKS 49
4.1 Abstract 49
4.2 Introduction 49
4.3 Proposed Duty Cycle Corrector 52
4.3.1 Differential Charge Pump with Fast Startup 52
4.3.2 Frequency to Current Converter and Adaptive Biasing 54
4.3.3 Current-starved Pulse Width Control Stage 56
4.4 Simulation Results 57
4.5 Conclusion 58
4.6 References 59

CHAPTER 5. SUB-PPM/°C VOLTAGE REFERENCE WITH NATURAL BASE EXPANSION FOR CURVATURE CANCELLATION 61
5.1 Abstract 61
5.2 Introduction 61
5.3 Natural Base Expansion and V_{GOr} Extraction 64
5.3.1 Temperature Characteristic of V_{BE} 64
5.3.2 V_{GOr} Extraction 67
5.4 Bandgap References Design 69
5.4.1 Constant Current Reference 70
5.4.2 Constant Voltage Reference 71
5.4.3 Error Sources and Trimming Approach 73
5.4.4 Simulation Results 76
5.5 Implementation and Measurements 77
5.5.1 Test Setup 77
5.5.2 Integrated Constant Current Reference 79
5.5.3 Partial Chip and Partial Board Constant Voltage Reference 81
5.6 Conclusion 84
5.7 References 85

CHAPTER 6. A TRANSIENT-ENHANCED OUTPUT-CAPACITORLESS LDO WITH LOCAL FAST LOOP AND OVERSHOOT DETECTION 88
6.1 Abstract 88
6.2 Introduction 88
6.3 Proposed LDO Architecture 91
6.3.1 Fast Local Loop 92
6.3.2 Overshoot Detection Circuit 98
6.3.3 Global Loop for Regulation 100
6.4 Simulated Load Transient Responses 103

6.5 Measurement Results and Discussion 105
6.5.1 Load Transient Measurement 107
6.5.2 V_{REF} and V_{IN} Transient Measurement 109
6.5.3 PSR Measurement 110
6.5.4 Comparison with State-of-the-art 111
6.6 Conclusion 112
6.7 References 113

CHAPTER 7. GENERAL CONCLUSIONS 116

LIST OF FIGURES

Page

Figure 2.1 Basic device array layout..... 9

Figure 2.2 Simulated estimation error in the decomposition..... 12

Figure 2.3 Chip photo. 14

Figure 2.4 (a) Measured 32 sweeps of DAC1, (b) averaged DNL and INL..... 15

Figure 2.5 (a) Extracted mismatch data, (b) decomposed systematic mismatch, (c) residue part, (d), (e) and (f) are corresponding INL..... 15

Figure 3.1 (a) Linear up and down pattern, (b) serpentine pattern. 23

Figure 3.2 (a) x gradient error, (b) resulted INL of an 8-bit string DAC..... 24

Figure 3.3 (a) Two substrings shifted in x-direction, (b) x^2 gradient error..... 25

Figure 3.4 INL of two substrings under x^2 gradient error..... 26

Figure 3.5 Proposed layout with shift INL methodology for suppressing up to 2^{nd} order gradients. 28

Figure 3.6 (a) Proposed averaging technique, (b) small error model. 29

Figure 3.7 (a) Interpolation between code k and k+1, (b) interpolation with matched resistors..... 31

Figure 3.8 String DAC's output resistance..... 32

Figure 3.9 Interpolating between code 0 and 1..... 32

Figure 3.10 (a) Two substrings under linear gradient in the x-direction, (b) interpolation without Rm and with Rm. 33

Figure 3.11 Simulated INL under randomly generated gradient patterns. 34

Figure 3.12 Possible digital interface..... 35

Figure 3.13 Interpolation with a differential difference amplifier..... 36

Figure 3.14 A differential difference amplifier with four input pairs..... 37

Figure 3.15 Possible two-stage DAC with high linearity. 38

Figure 3.16 Chip micrograph. 39

Figure 3.17 8-bit resistor array (configured into four 6-bit DACs). 40

Figure 3.18 Block diagram of a 6-bit DAC. 41

Figure 3.19 Evaluation board. 42

Figure 3.20 Measured INL of four DACs. 43

Figure 3.21 INLs at the averaged and interpolated outputs of DAC1&2. 44

Figure 3.22 INL at the 8-bit interpolated output of four DACs. 45

Figure 4.1 Proposed duty cycle corrector topology 50

Figure 4.2 Differential charge pump. 52

Figure 4.3 Fast startup circuit. 53

Figure 4.4 (a) Regulated bias, (b) proposed simple adaptive bias. 54

Figure 4.5 Current-starved pulse width control stage. 55

Figure 4.6 I_{CP} adaptive biasing at 4MHz, 25MHz and 100MHz. 56

Figure 4.7 Duty cycle corrector timing with 4 MHz input. 57

Figure 4.8 Duty cycle corrector timing with 100 MHz input. 58

Figure 4.9 100 Monte Carlo runs with 100 MHz input clock. 59

Figure 5.1 V_{GOr} extrapolation. 65

Figure 5.2 $V_G(T)$ fitting error in the interested temperature range. 66

Figure 5.3 Illustration of using three BJT connected as diodes to extract the V_{GOr}. 67

Figure 5.4 Weighted summation to extract the V_{GOr}. 68

Figure 5.5 Constant current reference based on Banba structure. 69

Figure 5.6 Constant voltage reference based on Kuijk structure. 71

Figure 5.7 V_{REF} drift at typical corner. 72

Figure 5.8 100 Monte Carlo runs of V_{REF}. 74

Figure 5.9 Randomly selected 5 runs: (a) V_{REF} before trimming, (b) V_{REF} after trimming..... 75

Figure 5.10 Lab setup. 76

Figure 5.11 TC characterization data of LTC6655 EVM with the lab setup..... 77

Figure 5.12 Die micrograph overlaid by the layout of the current mode BGR. 78

Figure 5.13 Measured drift curves in multiple fast temperature ramps. 79

Figure 5.14 Averaged V_{REF} of 6 ramps and its fitted curve. 80

Figure 5.15 Schematic of the voltage mode BGR prototype. 81

Figure 5.16 Die micrograph overlaid by the layout. 82

Figure 5.17 Noise (V_{REF}-mean(V_{REF})) at T_r in 30 minutes. 83

Figure 5.18 V_{REF} drift after the proposed two-temperature trimming. 84

Figure 5.19 V_{REF} drift during temperature ramps after optimal trimming. 85

Figure 6.1 Output-capacitorless low-dropout regulator. 89

Figure 6.2 Proposed LDO architecture. 91

Figure 6.3 Buffer structures: (a) super source follower (SSF), (b) flipped voltage follower (FVF). 92

Figure 6.4 Fast local loop including load. 93

Figure 6.5 Small signal model of the local loop approximated as a two-stage amplifier. 96

Figure 6.6 Simulated Bode plots of the local loop with C_L = 10 pF, V_{IN} = 1.1 V and V_{OUT} = 0.9 V under different I_L. 97

Figure 6.7 Inverter-based overshoot detection circuits. 98

Figure 6.8 Robust overshoot detection circuit. 99

Figure 6.9 Error amplifier with improved positive slewing. 100

Figure 6.10 Small-signal model of the whole system. 101

Figure 6.11 Simulated Bode plots of the global loop with C_L = 100 pF, V_{IN} = 1.1 V and V_{OUT} = 0.9 V under different I_L. 102

Figure 6.12 Transient responses (undershoot only) at the LDO's output with (a) $T_{edge} << T_R$ and (b) $T_{edge} > T_R$. (Redrawn from [4]). 103

Figure 6.13 Simulated transient responses under load steps with different T_{edge} with C_L = 10 pF. 104

Figure 6.14 Simulated transient responses with different C_L under a load step with T_{edge} = 500 ps. 105

Figure 6.15 Chip micrograph. 106

Figure 6.16 Testing setup for investigating the transient responses of the proposed OCL-LDO (architecture not redrawn here). 106

Figure 6.17 Measured load transient responses at V_{IN} = 1.1 V, V_{OUT} = 0.9 V with C_L = 10 pF under (a) $T_{edge1} \approx$ 500 ps and (b) T_{edge2} = 5 ns. 107

Figure 6.18 Measured load transient responses at V_{IN} = 1.1 V, V_{OUT} = 0.9 V with C_L = 100 pF under (a) $T_{edge1} \approx$ 500 ps and (b) T_{edge2} = 5 ns. 108

Figure 6.19 Measured V_{REF} transient responses at V_{IN} = 1.1 V, I_L = 0.1 mA with C_L = 10 pF. 109

Figure 6.20 Measured line transient responses at V_{OUT} = 0.9 V, C_L = 10 pF with (a) I_L = 20 mA and (b) I_L = 0.1 mA. 109

Figure 6.21 Measured PSR versus frequency at V_{OUT} = 0.9 V, I_L = 20 mA with C_L = 10 pF. .. 110

Figure 6.22 Measured transient waveforms at full load for PSR calculation at 10 MHz. 111

LIST OF TABLES

Page

Table 3.1 Measured INL results of four chips 45

Table 3.2 Comparison of state-of-art 46

Table 6.1 Performance Summary and Comparison with State-of-the-art OCL-LDOs 112

ACKNOWLEDGMENTS

First of all, I would like to express my deepest gratitude to my major professor Dr. Degang Chen for his guidance and support throughout my PhD study and research.

I would also like to thank my committee members Dr. Nathan Neihart, Dr. Huang Chen, Dr. Meng Lu and Dr. Aditya Ramamoorthy for their contributions to this work.

I would like to thank Dr. Randall Geiger for his support, advice and encouragement.

I would like to thank Jim Todsen and Chris Lash from Texas Instrument for their collaboration, support and inspiration.

I would like to thank my friends and colleagues Tao Chen, Zhiqiang Liu, You Li, Hao Meng, Shravan Chaganti, Pangzhou Li, and anyone who helped and supported me in various aspects.

I would like to thank Semiconductor Research Corporation (SRC), Texas Instruments, and Iowa State University for the financial and technical support of my research work.

Finally, I would like to thank my parents Suolin Liu and Xiaofang Zhang for their support and sacrifices.

ABSTRACT

System-on-chips (SoCs) with analog and mixed-signal (AMS) blocks are used everywhere in real-life applications, e.g., smart sensors, medical monitoring devices, automotive components, etc. Such applications typically require some sort of power management, sensor interface, signal conditioning, and processing. While modern deep sub-micron technologies benefit digital subsystems with low cost and high data rates, the performance of analog subsystems becomes more difficult to manage because of factors such as lower channel resistance, poorer matching, inferior linearity, and smaller operating voltage ranges. In the coming Internet of Things (IoT) era with more embedded analog functions, the design of the analog portion of a system becomes a bottleneck with respect to performance, time to market and costs.

This dissertation introduces several design techniques for helping recover analog performance for common AMS blocks in SoCs. The main bottleneck of most signal processing SoCs is analog signal processing whose accuracy relies mostly on the matching property of a device array configured into a ratio matrix for signal processing. A general methodology for characterizing device array mismatch, both systematic gradient and random mismatch, is introduced. It can provide information for process engineers to use in optimizing fabrication processes and for circuit designers to use in applying error reduction techniques for achieving acceptable yield. To address gradient errors induced nonlinearity, a practical layout structure and an interpolation method are proposed for string digital-to-analog converter (DAC). It divides a string into multiple smaller substrings and places them in a pattern where both linear and 2^{nd} order gradients are suppressed, with averaging and interpolation implemented at the substrings' outputs to recover bits of resolution without sacrificing linearity improvement. The proposed structure is validated by both simulation and measurement results. An 8-bit DAC prototype was designed in

GlobalFoundries 130nm process and achieves 16-bit linearity performance without trimming and calibration.

Good clocks are also important for mixed-signal systems, because without uniform sampling all standard signal processing theory can fall apart. While smaller geometry size and less parasitic can lead to increased clock speed, with process variations, clocks can suffer from unpredictable duty cycle distortions during generation and distribution. To address this issue, a low-cost and area-efficient analog duty cycle corrector with feedback is proposed.

Good linearity and clean clocks contribute to the relative accuracy of a system, and to achieve high absolute accuracy, a precise and constant reference is needed. A general approach based on natural base expansion is proposed for canceling the V_{BE} curvature of a bipolar transistor and creating sub-ppm/°C voltage references. A prototype realized with both on-chip devices and off-chip components was designed, fabricated, and measured with results demonstrating the first voltage reference with sub-ppm/°C, or 0.8 ppm/°C to be exact, temperature coefficient over a temperature range of 0 °C to 80 °C after optimal trimming.

Good power supplies are also essential for high precision systems, and output-capacitorless low-dropout regulators (OCL-LDOs) have been widely used in SoCs for the purpose of generating clean supplies and providing isolation from noisy blocks. Stability is a basic requirement and achieving satisfactory performance in terms of accuracy, transient response, and power-supply rejection (PSR) is usually a more difficult task. To achieve fast transient response with reasonable current, a dual loop OCL-LDO structure utilizing a super source follower and a flipped voltage follower is proposed. A prototype that can deliver a 20 mA current with only 0.01 mm^2 active area was implemented in the UMC 65nm process. While the overall structure is simple and straightforward, it achieves a FOM comparable to the state-of-the-art.

CHAPTER 1. GENERAL INTRODUCTION

Analog and mixed-signal (AMS) circuits play a significant role in modern electronic systems. A system on a chip (SoC) is an integrated circuit that combines all or most components and typically contains digital, analog, mixed-signal, and often radio frequency signal processing functions. Because SoCs consume much less power and occupy much less area than multi-chip designs with equivalent functionality, they are very commonly found in smart phones, edge computing, and the Internet of Things (IoT) markets. In the coming IoT era, more analog functions are embedded for interconnecting physical entities such as smart machines, wearables, autonomous driving systems, and technology in virtually every environment, ranging from streetlights to parking and traffic sensors. Over the past decades, process feature size has continued shrinking with associated digital performance improvement and cost reduction, so that analog performance primarily determines overall system performance. As the system integration level keeps growing, AMS parts occupy an increased share of overall chip design efforts.

1.1 Background and Motivation

It is a fact that technology advancement has so far not greatly benefitted many of the important parameters of AMS circuits such as gain, matching, linearity, and noise. At submicron technology nodes and beyond, integrated device matching is poor and usually imposes difficult challenges to design of circuits based on device arrays providing ratio-metric high linearity signal processing functions, e.g., digital-to-analog converters, etc. In many systems involving both digital and real world, data converter performance determines overall system performance. While increasing chip area is one of the most effective techniques commonly used by IC designers for reducing random mismatch, it results in another problem: more gradient errors in the device array that can accumulate across the array and dramatically degrade linearity performance. To address

this problem, a low-cost practical string DAC structure with gradient error suppression that achieves good linearity without trimming and calibration will be proposed.

Since standard signal-processing theory can fail badly without uniform sampling, clean clocks with low jitter and low duty cycle distortion are critically important for mixed-signal systems. In deep submicron technologies, process variation and asymmetric bias temperature instability (BTI) due to aging can cause duty cycle errors. Therefore, a low-power and area-efficient analog duty cycle corrector with feedback is introduced to address the duty cycle distortion problem.

Most analog and mixed-signal circuits require a good voltage reference, either as part of the internal biasing configuration or as a target of operation. While good matching and clean clocks can help a system achieve high relative accuracy, it can only be good up to the accuracy of the reference. Achieving absolutely accurate results requires a precise and constant reference that doesn't vary with temperature or environment. Performance at the level of a few ppm/°C is desired in many standard applications, and some systems target even better precision performance. For example, if the temperature dependence of the reference voltage for a 16-bit analog-to-digital converter (ADC) with a full-scale input of 1V is symmetric around the mid-range temperature, with a state-of-the-art temperature coefficient of 1 ppm/°C over a 100 °C operating range, the temperature dependence of the reference alone will contribute to about a 6 LSB error in the ADC output at the specified temperature extremes. This dissertation proposes a general approach utilizing natural base expansion to cancel the curvature in the base emitter voltage of a bipolar transistor that can be implemented on any of the widely used Banba, Kuijk, or Brokaw structures to build references with sub-ppm/°C temperature coefficient.

Variability in supply voltage can also limit performance and reliability of circuits and systems designed in advanced technologies. Output-capacitorless low-dropout regulators (OCL-LDOs) have been widely used in SoCs aimed at generating a clean supply, providing isolation, and avoiding coupling between blocks. Stability is a basic requirement for an LDO, and achieving desired performance in terms of accuracy, transient response, and power-supply rejection (PSR) is often a more difficult task. In nano-scale technology, current consumption of digital cells can vary between nearly zero and some maximum value over a very short time. While removal of the bulky capacitor in OCL-LDOs is beneficial in terms of integration, it degrades load transient responses. To address this issue, a dual loop OCL-LDO is proposed. A fast local loop based on a flipped voltage follower and a super source follower can quickly respond to load steps and the global loop, including feedback and an error amplifier, handles the regulation. While the overall structure is very simple and straightforward, implying robustness and reliability for analog circuits, it achieves a FOM comparable to the state-of-the-art.

1.2 Organization of Dissertation

The remainder of this dissertation is arranged in the following order:

Chapter 2 is a modified version of a paper published in the IEEE International Midwest Symposium on Circuits and Systems (MWSCAS). It introduces a general methodology for characterizing device array matching performance. Systematic and random mismatch are separately characterized using a proposed decomposition method that provides robustness with respect to additive measurement noise. An 8-bit high precision TaN thin film resistor array has been fabricated, with measurement results indicating that gradient error induced systematic mismatch normally contributes to a large portion of the total mismatch when the matching accuracy diminishes to a 0.01% level.

Chapter 3 is modified from a manuscript published in the IEEE Transactions on Circuits and Systems I. It introduces a low-cost practical string DAC structure with gradient error suppression that achieves high linearity by dividing a string into multiple smaller substrings and placing them into a pattern where both linear and 2^{nd} order gradients are suppressed. A derivation explaining how the 2^{nd} order gradient is mitigated is provided. Averaging and interpolation are implemented at substrings' outputs to recover bits of resolution without sacrificing linearity improvement. An 8-bit prototype was designed, fabricated, and tested, with measurement results demonstrating achievement of 16-bit linearity performance without trimming and calibration.

Chapter 4 is a modified version of a paper to be published in the IEEE International Symposium on Circuits and Systems (ISCAS). It presents a low power and area-efficient analog duty cycle corrector with feedback. A differential charge pump with fast startup is used to detect duty cycle error and outputs a control voltage for current-starved inverters to adjust pulse width. Since the pump current is designed to be adaptive to the input frequency, a wide frequency range is achieved without requiring a large load capacitor.

Chapter 5 is a modified version of a manuscript to be submitted to the IEEE Transactions on Circuits and Systems I. It presents a general approach to cancel the curvature in the base emitter voltage of a bipolar transistor and build references with sub-ppm/°C temperature coefficients. One embodiment based on a Kuijk structure is introduced and a two-temperature trimming method is proposed to reduce error sources' effects on the temperature drift. Transistor level simulations demonstrate that the design can achieve sub-ppm/°C temperature coefficient over the temperature range of -40 °C to 125 °C. A prototype realized with both on-chip devices and off-chip components was designed, fabricated, and measured.

Chapter 6 is modified from a manuscript published in the IEEE Transactions on Circuits and Systems I. It proposes an OCL-LDO with a very simple structure and fast transient responses. A super source follower is inserted into a cascoded flipped voltage follower to drive the power transistor, forming a fast local loop for providing quick turn-on. A robust overshoot detection circuit consuming only leakage current is proposed for fast turn-off. A simple yet effective additional turn-around stage is added to the error amplifier to improve its positive phase slew rate for potential dynamic voltage scaling (DVS) function in battery-operated systems. The LDO, implemented in the UMC 65nm process, can deliver a 20 mA load current with a 150 mV dropout voltage. It occupies an active area of 0.01 mm^2 and can work stably over the entire load range with 65 µA quiescent current. Measured results indicate a settling time of about 100 ns for both load steps from 100 µA to 20 mA and V_{REF} and V_{IN} steps.

Chapter 7 concludes the dissertation.

CHAPTER 2. PASSIVE ARRAY MATCHING CHARACTERIZATION WITH GRADIENT ERRORS AND RANDOM MISMATCH DECOMPOSITION

Nanqi Liu*, Jim Todsen, and Degang Chen*

*Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010

Texas Instruments, Tucson, AZ 85711

Modified from a manuscript to be submitted to IEEE Transactions on Semiconductor Manufacturing

2.1 Abstract

A general methodology of device array matching characterization is introduced, analyzed, and verified. Rather than individually measuring each device's parameters, the device array is configured as a data converter and mismatch information is extracted from the DNL of the converter. Systematic and random mismatch are separately characterized using the proposed decomposition method with robustness to additive measurement noise. Measurement results for an 8-bit high precision resistor array fabricated using the Globalfoundries 0.13μm process show that systematic gradient errors contribute a large portion of the total mismatch when the matching requirement diminishes to a 0.01% level.

2.2 Introduction

Precise characterization of device array matching is increasingly important. Matched device arrays are widely used in high performance data converters, such as transistor arrays in high speed current-steering DACs, string or R-2R ladders in resistor DACs and binary weighted capacitor arrays in successive approximation ADCs. Since static performance of such data converters is highly dependent on device matching accuracy, knowledge of matching information at an early design-cycle stage is important.

Mismatch is generally divided into two categories: random and systematic [1]. Random mismatch of integrated devices has been accurately characterized and modeled in [2]-[4]. In such characterizations, pairs of neighboring components with identical layouts are placed across dies, batches, and wafers, and by measuring a large population of such matched pairs, random fluctuations can be properly characterized through statistical analysis. Design engineers can use both analytic equations and EDA tools (e.g., Monte Carlo simulation) to predict the circuit yield and make design trade-offs, such as area vs. mismatch. Systematic mismatch resulting from process and environmental gradients are mitigated in matched pairs because of the small distance between components. However, in high resolution arrays, devices can be distanced at the millimeter level. As reported in [5], the contribution of systematic mismatch is as significant as that of random mismatch. Therefore, precise characterization of systematic mismatch has become an important task in recent years. Accurate characterizations of doping levels or device dimension offsets across device arrays can provide information for process engineers to optimize fabrication processes, and once matching performance is known, circuit designers can apply layout, trimming or calibration techniques to reduce the effect of such gradient errors and achieve acceptable yield.

Measuring device parameters based on Kelvin force-sense approaches through probing pads is a traditional method for characterize mismatch [6], but the large pad area required limits the number of devices in a chip. Terminal multiplexing [7] or switching matrix [8] has been developed as an alternative solution, but array sample size and characterization accuracy are limited due to parasitic and leakage currents. Device matrix array measurement was implemented in [9] to overcome these drawbacks, but this requires a reference component (0 Ω or 0 F) to compensate the parasitic resistances and bus leakage currents, and because each component needs to be measured sequentially, the test is time-consuming. A methodology that extracts the mismatch

information from a fully functional circuit has been proposed in [5]. MOS transistor current mismatch is characterized by measuring a current steering DACs' linearity. When characterization accuracy diminishes to the 0.01% level, the systematic error model used in [5] is no longer accurate and measurement noise also must be taken into consideration.

In this chapter, a general methodology for characterizing array mismatch is developed. After configuring device arrays into data converters, mismatch information is extracted from DNL measurements of the circuits. The total mismatch is then decomposed into systematic mismatch and a residue part using polynomial fitting with the least squares method. The random mismatch is further characterized from the residue part in the existence of random noise from measurements. This chapter is organized as follows. In Section 2.3, systematic mismatch in a device array is modeled and each device's mismatch is defined. Section 2.4 introduces how mismatch information can be extracted from the measured DNL and how the systematic and random parts are decomposed. Measurement results are provided in Section 2.5. Finally, conclusions are drawn in Section 2.6.

2.3 Model of Device Errors

Random variations in physical quantities of devices result from several random processes occurring during fabrication phases. A random mismatch model has been derived and verified in [2]. The standard deviation $\sigma_{\Delta P}$ of random mismatch for electrical parameter P is inversely proportional to the square root of the device area:

$$\sigma_{\Delta P} = \frac{A_{\Delta P}}{\sqrt{W * L}}, \tag{2.1}$$

in which W and L are width and length of the active part of the device and $A_{\Delta P}$ is a process parameter. In the array matching case, since the maximum distance between devices can increase to the millimeter level, systematic mismatch can no longer be ignored.

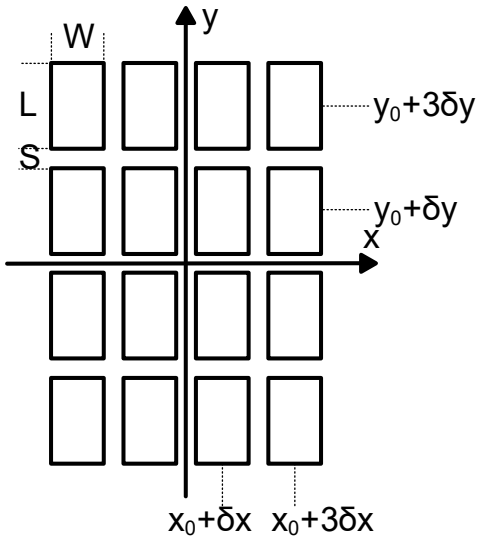

Figure 2.1 Basic device array layout

To analyze and model gradient errors, a basic device array layout in which identical rectangular cells are placed as in Figure 2.1 is adopted here. (x_0, y_0) is the centroid of the array, and pitches in the x- and y- direction are defined as

$$\delta_x = W + S, \delta_y = L + S. \tag{2.2}$$

S is usually the minimum space defined in design rules to make the layout compact. Each unit cell's coordinate can then be calculated from Equation (2.2).

In previous work [2], systematic mismatch was modeled as an additional stochastic process with a long correlation distance. The mismatch equation is for two components and it's assumed there is only linear gradient error. Therefore, it is not appropriate to model array mismatch. Quadratic gradient errors have also been reported in [9] and [10], with systematic mismatch modeled by a Taylor series expansion around the center of the device array:

$$e_{sys}(x_k, y_k) = a_0 + a_1 x_k + a_2 y_k + a_3 x_k^2 + a_4 y_k^2 + a_5 x_k y_k + \ldots, \tag{2.3}$$

where (x_k, y_k) is the coordinate of unit cell k's centroid. The coefficients a_0, a_1, … are dependent on the manufacturing process and the die location on the wafer.

In a device array, each device's mismatch can be calculated by comparing its parameter value with the mean value of the total array. The mismatch of the k-th device is therefore:

$$e_k = \frac{P_k - \overline{P}}{\overline{P}} = \frac{P_k}{\overline{P}} - 1, \tag{2.4}$$

where $\overline{P}$ is the average P value of all devices in the array and e_k is a superposition of random and systematic mismatch:

$$\begin{aligned} e_k &= e_{rndk} + e_{sysk} \\ &= e_{rndk} + \begin{bmatrix} 1 & x_k & y_k & x_k^{\,2} & y_k^{\,2} & \dots \end{bmatrix} * \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & \dots \end{bmatrix}^T, \end{aligned} \tag{2.5}$$

where $e_{rndk} \sim N(0, \sigma_{rm})$ and σ_{rm} is the standard deviation of the random mismatch.

2.4 Extraction and Decomposition

A general methodology of array mismatch characterization will be introduced in this section.

2.4.1 Extraction of Mismatch Data

Unary data converters are more straightforward to consider when explaining extraction of mismatch information. An N-bit unary DAC consists of 2^N elements. Assuming that the actual analog output for a digital code k is denoted as A(k), the least significant bit (LSB) is defined as

$$LSB = \frac{A(2^N - 1) - A(0)}{2^N - 1}. \tag{2.6}$$

DNL is the deviation of the analog step size from 1 LSB. The DNL (in units of LSB) at code k is therefore

$$DNL(k) = \frac{A(k) - A(k-1)}{LSB} - 1, \; k = 1, 2, ..., 2^N - 1. \tag{2.7}$$

Taking into account each component's electrical parameter P, we have

$$A(k) = \frac{\sum_{i=1}^{k} P_i}{\sum_{i=1}^{2^N-1} P_i} * [A(2^N - 1) - A(0)] + A(0) \cdot \tag{2.8}$$

Substituting equation (2.8) into (2.7), the DNL(k) can be rewritten as

$$DNL(k) = \frac{P_k}{\overline{P}} - 1 = e_k, \tag{2.9}$$

with the mismatch of device k thereby extracted from the DNL of the converter.

In high precision (14-bit or higher) mixed-signal circuits, mismatch characterization accuracy has been reduced to the 0.01% level [1], so that measurement noise is usually large and requires much effort to be averaged out with replications over time. Careful design and setup of test benches are required to avoid systematic measurement errors such as code dependent settling error, buffer nonlinearity, and long-term drift. With additive random noise, the measured mismatch of device k is

$$e_{measurek} = e_{sysk} + e_{rndk} + e_{noisek} \cdot \tag{2.10}$$

Since random mismatch and measurement noise are both random fluctuations, they will be lumped together when we separate the systematic and random part.

2.4.2 Polynomial Fitting with Least Squares Method

After extracting the mismatch data from the measured DNL, we can decompose the mismatch of the whole array into a smooth systematic mismatch and a random mismatch:

$$\begin{bmatrix} e_{measure1} \\ \vdots \\ e_{measure2^N} \end{bmatrix} = \begin{bmatrix} 1 & x_1 & y_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{2^N} & y_{2^N} & \cdots \end{bmatrix} * \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{bmatrix} + \begin{bmatrix} e_{r1} \\ \vdots \\ e_{r2^N} \end{bmatrix} \tag{2.11}$$

where $e_{rk} = e_{rndk} + e_{noisek}$. Each device's coordinates can be calculated based on the array's layout. For simplicity, Equation (2.11) can be rewritten into a matrix form:

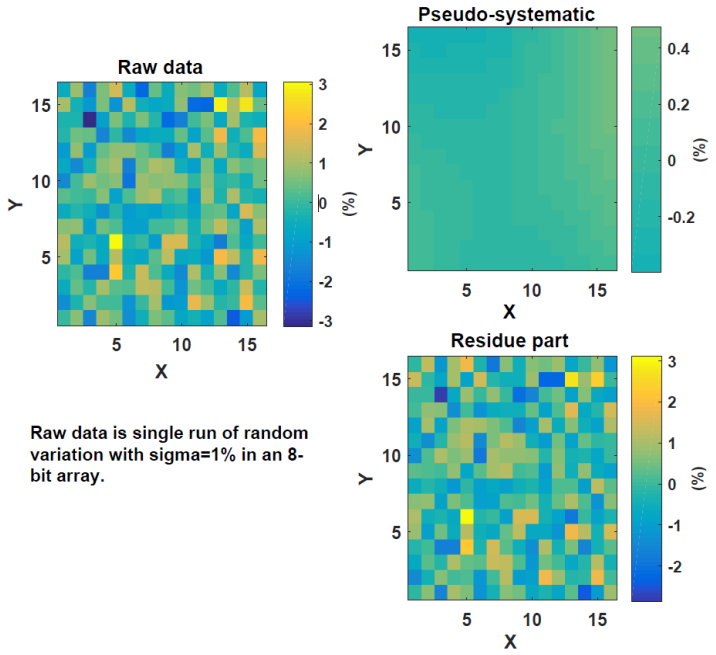

Figure 2.2 Simulated estimation error in the decomposition.

$$e_m = h * a + e_r . \tag{2.12}$$

Since e_m and h are known matrices, we can extract unknown gradient coefficient a using the least squares method. The estimation of a is

$$\hat{a} = [h^T h]^{-1} h^T e_m = a + [h^T h]^{-1} h^T e_r . \tag{2.13}$$

The extracted systematic mismatch of the array is

$$\hat{e}_{sys} = h * \hat{a} . \tag{2.14}$$

Compared to the true systematic mismatch, there exists an estimation error:

$$e_{est\_err} = h\hat{a} - ha = h[h^T h]^{-1} h^T e_r . \tag{2.15}$$

Figure 2.2 shows an 8-bit array with only random mismatch (σ_{rm} = 1%) that was created in MATLAB. Using a 2nd order systematic error model and implementing the decomposition method, an estimation error was observed and designated as a pseudo-systematic error that, according to

Equation (2.15), is dependent both on array size and the polynomial order of the systematic error model:

$$\frac{\mathrm{var}(e_{est\_err})}{\mathrm{var}(e_r)} \sim \frac{1}{2^N - M}, \tag{2.16}$$

where M is the number of gradient coefficients to be fitted. In the measurement results, random mismatch and measurement noise are fixed values and will be projected to smooth polynomial terms with attenuation, resulting in inaccurate characterization of systematic mismatch. Therefore, given the required characterization accuracy and magnitude of random mismatch and measurement noise, minimum sample size based on Equation (2.16) can be calculated.

2.4.3 Random Measurement Noise Robustness

To measure a data converter's linearity, it is better to perform many (16 or more) full code sweeps then average the DNL to reduce noise and drift. To further separate random mismatch from measurement noise, the decomposition step can use two sets of data. We first extract the mismatch from the averaged DNL of all measured sweeps. Implementing the decomposition on this 1st set of data, we have

$$e_{m\_1} = h\hat{a}_{\_1} + \hat{e}_{r\_1}. \tag{2.17}$$

And the standard deviation of $\hat{e}_{r-1}$ can be calculated:

$$\sigma^2_{\hat{e}_{r\_1}} = \sigma^2_{rm} + \sigma^2_{noise\_1}. \tag{2.18}$$

By selecting every second sweep in all the sweeps and averaging, another set of mismatch data can be extracted, decomposed, and analyzed:

$$\sigma^2_{\hat{e}_{r\_2}} = \sigma^2_{rm} + \sigma^2_{noise\_2}. \tag{2.19}$$

Since only half sweeps are used, the magnitude of the additive measurement noise here is $\sqrt{2}$ times that of the 1st set of data:

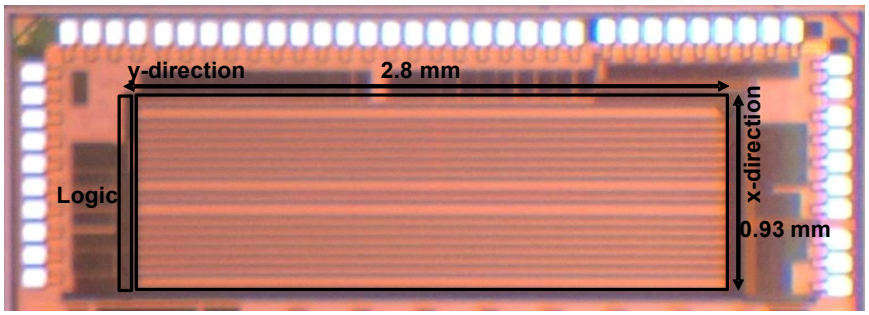

Figure 2.3 Chip photo.

$$\sigma_{noise\_2} = \sqrt{2} \cdot \sigma_{noise\_1}. \tag{2.20}$$

So, the random mismatch of the device array can be recovered as

$$\sigma_{rm} = \sqrt{(2\sigma^2_{\hat{e}_{r\_1}} - \sigma^2_{\hat{e}_{r\_2}})/(2-1)} = \sqrt{2\sigma^2_{\hat{e}_{r\_1}} - \sigma^2_{\hat{e}_{r\_2}}}. \tag{2.21}$$

2.5 Measurement Results

An 8-bit high precision resistor array has been fabricated using the Globalfoundries 0.13μm process. Tantalum Nitride (TaN) thin film-resistors sized to achieve 0.017% random mismatch (1σ) were used, and the layout of the 8-bit resistor array is shown in Figure 2.3. It consists of four 8*8 subarrays with each subarray configured into to a 6-bit string DAC with separate switches and decoders. A simple linear up and down routing pattern was used to minimize metal routing resistances between columns, and dummy resistors were placed around the total array to reduce edge effects.

2.5.1 Testbench

An 18-bit SAR ADC (ADS8881EVM) was used to sequentially measure the output voltages of each 6-bit DAC, and 32 full code sweeps, with each code sampled 20 times, were performed for each DAC. Measurement error was gauged by looking at the standard deviation of integral nonlinearity (INL), with results beyond measurement noise thrown out. The measured

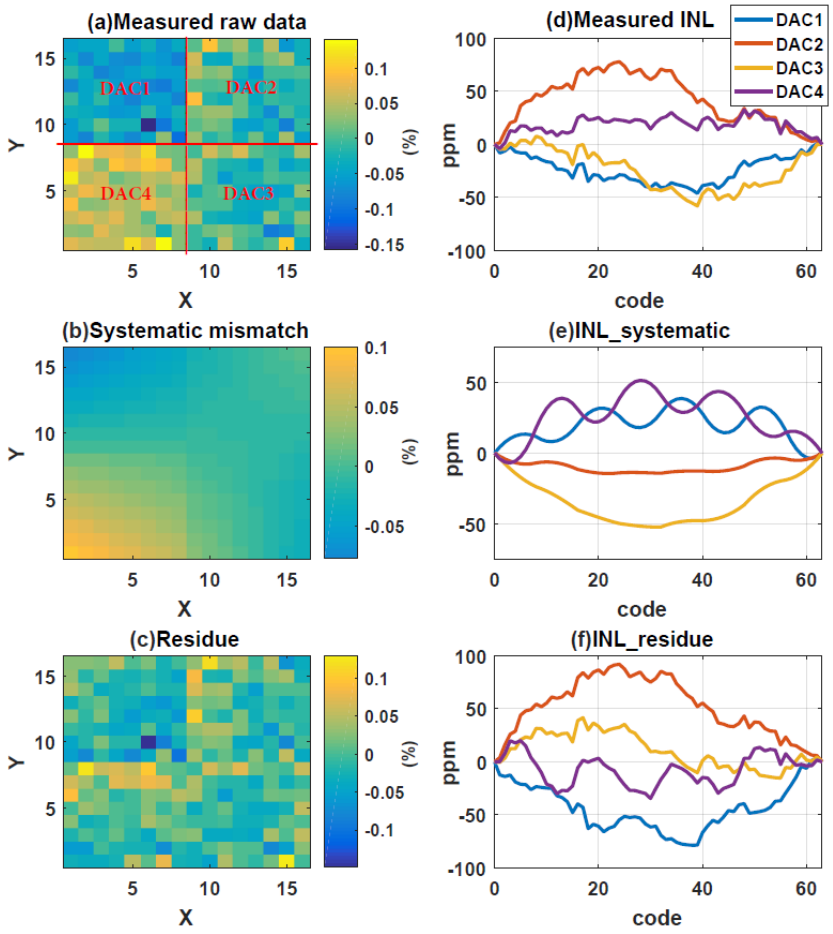

Figure 2.5 (a) Extracted mismatch data, (b) decomposed systematic mismatch, (c) residue part, (d), (e) and (f) are corresponding INL.

results for DAC1 are shown in Fig. 2.4(a). Averaging these 32 sweeps, measurement noise was reduced by $\sqrt{32}$ times. The averaged DNL and INL of DAC1 are plotted in Figure 2.4(b).

2.5.2 Results for Systematic Mismatch

Based on Equation (2.9), mismatch data extracted from DNL measurements are provided in Figure 2.5(a). Using a 2^{nd} order systematic error model and the least squares method, the total mismatch can be decomposed into the systematic and residue parts, respectively shown in Figure 2.5(b) and (c). When matching accuracy diminishes to the 0.01% level, systematic mismatch contributes a large portion of the total mismatch. A more serious problem is that, even

when systematic mismatch is small, it can accumulate quickly to become a large INL. As shown in Figure 2.5(e) and (f), in each 6-bit DAC, the systematic mismatch magnitude is much smaller than the residue part, but it accumulates to INL and is comparable to the INL induced by the residue part.

2.5.3 Results for Random Mismatch

Random mismatch needs to be further extracted from the residue part. For the first set of data, 32 sweeps were averaged and the standard deviation of the residue part $\sigma_{\hat{e}_{r\_1}}$ = 0.039%. By selecting every other sweep and averaging the resulting 16 sweeps, another set of data can be generated. After decomposition, the standard deviation of the residue part $\sigma_{\hat{e}_{r\_2}}$ = 0.044%. Equation (2.21) is used to recover the random mismatch to yield $\sigma_{e_{rm}}$ = 0.033%, larger than the designed value 0.017%. This may be explained by the random variation in routing resistances between unit cells that are not accurately characterized and modeled in this process. Because of the low sheet resistance (about 50m/square) of TaN thin film, routing resistances are not negligible, especially when matching performance is at the 0.01% level.

2.6 Conclusion

An accurate and efficient characterization method of device array mismatch has been presented and verified in this chapter. Mismatch information was extracted from the DNL of the data converter built with the device array, and compared to use of probing pads and switching matrix methods, measurement errors resulted from bus parasitic resistances and switch leakage currents have been removed with little accompanying area overhead. Systematic mismatch and random part have been decomposed using the proposed polynomial fitting along with the least squares method. Measurement data show that systematic gradient errors contribute to a large

portion of the total mismatch. Random mismatch information is further extracted with robustness with respect to additive measurement noise.

2.7 References

[1] H. Tuinhout, N. Wils and P. Andricciola, "Parametric mismatch characterization for mixed-signal technologies", *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 16871696, 2010.

[2] M. J. M. Pelgrom, A. C. J. Duinmaijer and A. P. G. Welbers, "Matching properties of MOS transistors", *IEEE J. Solid-State Circuits*, 1989.

[3] P. G. Drennan and C. C. McAndrew, "Understanding MOSFET mismatch for analog design", *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 450456, 2003.

[4] T. G. Odwyer and M. P. P. Kennedy, "An enhanced model for thin film resistor matching", *Proc. ICMTS*, pp. 4549, 2009.

[5] E. Felt, A. Narayan and A. Sangiovanni-Vlncentelli, "Measurement And Modeling Of MOS Transistor Current Mismatch In Analog ICs", *Proc. ICCAD*, pp. 272-277, 1994.

[6] H. P. Tuinhout, G. Hoogzaad, M. Vertregt, R. L. J. Roovers and C. Erdmann, "Design and characterization of a high-precision resistor ladder test structure", *IEEE Trans. Semicond. Manuf.*, vol. 16, no. 2, pp. 187-193, 2003.

[7] W. Posch and E. Seebacher, "Integrated capacitor array matching characterization", *IEEE Trans. Semicond. Manuf.*, vol. 25, no. 3, pp. 331338, 2012.

[8] U. Schaper, C. G. Linnenbank and R. Thewes, "Precise characterization of long-distance mismatch of CMOS devices", *IEEE Trans. Semicond. Manuf.*, vol. 14, no. 4, pp. 311317, 2001.

[9] S. I. Ohkawa, M. Aoki and H. Masuda, "Analysis and characterization of device variations in an LSI chip using an integrated device matrix array", *IEEE Trans. Semicond. Manuf.*, vol. 17, no. 2, pp. 155165, 2004.

[10] Tian Weidong, P. Steinmann, E. Beach, I. Khan and P. Madhani, "Mis-match characterization of a high precision resistor array test structure", *Proc. ICMTS*, pp. 11-16, 2008.

CHAPTER 3. AN 8-BIT LOW-COST STRING DAC WITH GRADIENT ERRORS SUPPRESSION TO ACHIEVE 16-BIT LINEARITY

Nanqi Liu*, Jim Todsen and Degang Chen*

*Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010

Texas Instruments, Tucson, AZ 85711

Modified from a manuscript published in IEEE Transactions on Circuits and Systems I: Regular Papers

3.1 Abstract

Resistor string is one of the classic digital-to-analog converter (DAC) architectures and has dominant applications where guaranteed monotonicity is important. For high-resolution string DACs, gradient errors limit the linearity performance by causing mismatch in the resistor arrays. In this chapter, a low-cost practical string DAC structure with gradient error suppression is proposed. It achieves high linearity by dividing a string into multiple smaller substrings and placing them in a pattern where up to 2^{nd} order gradients are suppressed. The derivation is provided to explain how the 2^{nd} order gradient is mitigated. Averaging and interpolation are implemented at the substrings' outputs to recover the bits of resolution without sacrificing the linearity improvement. An 8-bit prototype is fabricated in GlobalFoundries 0.13µm technology and occupies 2.6 mm^2. It achieves 16-bit linearity performance without trimming and calibration.

3.2 Introduction

Integrated circuit matching has been one of the major challenges in precision analog and mixed-signal systems [1]. Static linearity performances of most data converters depend upon the matching characteristics of basic circuit components.

String DACs use unary resistor ladders to produce linear output voltages. Because of the inherent monotonicity and low power features, string DACs have been broadly used in many

applications, such as instrumentation, test and measurement, process control, and liquid crystal display (LCD) drivers [2], [3], [4]. Due to area constraints, a pure R-string architecture has not exceeded eight bits of resolution as the number of resistors, switches and decoding circuits grows exponentially with bits of resolution [5]. Moreover, the integral nonlinearity (INL) is generally large and is highly dependent on the resistor array matching. However, IC fabrication technologies cannot produce perfectly matched devices, and with downward process scaling, random mismatch increases significantly [6]. As a result, each resistor requires a large area to meet the stringent matching requirement. However, simply increasing area results in another problem: more gradient errors (e.g., geometry gradients and self-heating temperature gradients [7]). These systematic errors can accumulate across the array and dramatically degrade the linearity performance [8].

To reduce area and achieve high bits of resolution, segmented architectures [9], [10] and interpolation topology [11], [12] have been introduced. These two-stage structures divide the DAC input code into two sets: most significant bits (MSBs) and least significant bits (LSBs). MSBs which play the most critical role in DACs' linearity are commonly implemented with unary R-strings. For resolution above 14-bit, trimming or calibration becomes necessary to overcome the mismatch problem in data converters [13]. However, standard trim methods are rarely used in unary weighted structures because of the large number of cells to be trimmed. Also, trimming is usually expensive due to the requirement of extra layers or more die area for trim pads and the achieved accuracy can shift after packaging processes. Post-assembly calibration with analog redundancy [14], [15] or complicated feedback [7] is another common technique to improve INL. However, the price of implementation is high as it requires accurate measurement and feedback circuits which contradicts the low cost and low power feature of string DACs. Recently a delta-sigma assisted calibration method using minimal circuit overhead was proposed in [16]. Hybrid

current steering structure was used to achieve both high frequency and good linearity with the MSB cells mismatch measured off-chip and calibrated digitally in the LSB delta-sigma path.

Besides trimming and calibration techniques, many layout structures have been proposed to cancel or reduce gradient errors in precision DACs. Multiple switching sequence based placement methods [17], [18] have been proposed for current source arrays. Current cells are additive and thus complex placement and routing are usually allowed. A detailed comparative analysis of known switching sequences has been developed in [19]. As pointed out in the chapter, none of them is suitable for R-strings due to the routing complexity. An anti-parallel connection has been used in [5] to eliminate linear gradient errors in a 4-bit coarse ladder, but routing complexity limits its application in higher resolution strings. An alternative solution has been proposed to average two common centroid strings with two differential trans-conductance stages [8]. Interdigitation has also been used to reduce the accumulation of gradient errors [20]. Although the routing is not complex, gradient errors are only reduced by a factor of five when only the columns of an array are interdigitated.

In this chapter, a practical string DAC structure with high linearity is proposed. A resistor string is first divided into four lower resolution substrings. They are arranged in a common centroid pattern with shift INL methodology [21] implemented to cancel linear gradient and reduce quadratic gradient errors. The reason why the 2^{nd} order gradient can be mitigated is explained with derivations. Then an output averaging technique is used to overcome the complex routing problem.

Small signal error analysis with superposition method proves that gradient-errors-induced nonlinearity can be suppressed significantly at the shorted output of four substrings. Since the resolution at the averaged output is reduced resulting from substrings having fewer resistors, an interpolation concept is further introduced to recover the original bits of resolution by averaging

the substrings' outputs at consecutive codes. The proposed structure method offers a unique advantage to production test and calibration methods as it apparently functions when external stress, temperature change or aging happens.

This chapter is organized as follows. In Section 3.3, gradient errors in resistor arrays and their effects on INL are analyzed. Common layout patterns are compared, and the proposed structure is explained. Then outputs averaging and interpolation methods are analyzed and discussed in Section 3.4. Section 3.5 shows the implementation of the structure and Section 3.6 provides the measurement results. Finally, the conclusion is drawn in Section 3.7.

3.3 Resistor Array Mismatch

Resistor array mismatch directly determines string DACs' linearity. Two major error sources in a device array are local random errors and systematic gradient errors. Local variations of physical quantities including line edge roughness [22] and doping level contribute to differences between devices. For an N-bit string DAC, the following relations hold when only random variations are considered:

$$\sigma_{INLk} = \sqrt{k(1-\frac{k}{2^N})} \cdot \sigma_{rnd}\,, \tag{3.1}$$

where k is the input code and σ_{rnd} is the standard deviation of mismatch between identical, adjacent unit resistors. Increasing area is one of the most effective techniques to reduce random mismatch and is commonly used by IC designers.

Increasing each unit cell's size results in a large array area which can contain significant gradient errors. For linearity above 14-bit, gradient errors were shown to be a bigger problem [17]. Many analyses assume linear (i.e., first order) gradients which can come from imperfect processing during resistor fabrications. There is also evidence from wafer mapping [23] showing radial process gradients and die stress mapping [24] showing high order stress gradients.

3.3.1 Gradient Errors and Their Effects on Linearity

Process and environment gradients including geometry, doping level, and temperature affect resistance differently. However, as they are all small errors, the superposition method can be applied here. Still using the basic array layout shown in Figure 2.1, the overall gradient errors are approximated by a Taylor series expansion around the center of the resistor array [1]:

$$\varepsilon(x_k, y_k) = a_0 + a_1 x_k + a_2 y_k + a_3 x_k^{\,2} + a_4 y_k^{\,2} + a_5 x_k y_k + \ldots, \tag{3.2}$$

where (x_k, y_k) is the coordinate of unit cell k's centroid. We haven't found any experimental data show that gradients higher than second order are important in practice. So up to 2nd order gradients are considered here.

The gradient induced variation of a unit cell's resistance is modeled by:

$$R_k = R_0 \cdot [1 + \varepsilon(x_k, y_k)], \tag{3.3}$$

where R_0 is the nominal resistance. For a unipolar N-bit string DAC, its output voltage at code k equals to

$$V_k = \left(\sum_{i=1}^{k} R_i \Big/ \sum_{i=1}^{2^N} R_i\right) \cdot V_{REF}, \text{ k=1, 2, ..., } 2^N\text{-1,} \tag{3.4}$$

where V_{REF} is the reference voltage of the DAC. The DNL can be calculated as

$$DNL_k = \frac{V_k - V_{k-1} - LSB}{LSB} = \frac{R_k - \overline{R}}{\overline{R}}. \tag{3.5}$$

Here $\overline{R}$ is the mean resistance of the array and it equals to

$$\overline{R} = \frac{R_0 \cdot \sum_{i=1}^{2^N} [1 + \varepsilon(x_k, y_k)]}{2^N}. \tag{3.6}$$

Substituting equation (3.6) into equation (3.5), DNL_k becomes

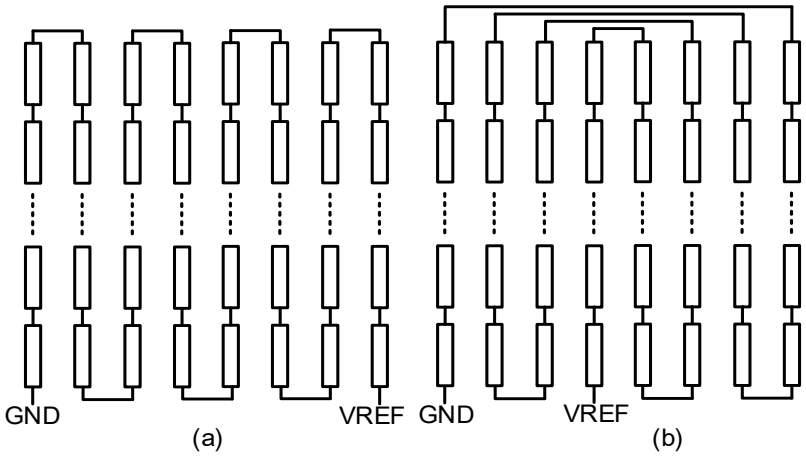

Figure 3.1 (a) Linear up and down pattern, (b) serpentine pattern.

$$DNL_k = \frac{R_k}{\overline{R}} - 1 = \frac{1+\varepsilon(x_k, y_k)}{\sum_{i=1}^{2^N}[1+\varepsilon(x_k, y_k)]/2^N} - 1 = \frac{1+\varepsilon(x_k, y_k)}{1+\sum_{i=1}^{2^N}\varepsilon(x_k, y_k)/2^N} - 1. \quad (3.7)$$

When $|\Delta x| << 1$, we have

$$1/(1+\Delta x) \approx 1 - \Delta x. \quad (3.8)$$

Applying the approximation equation (3.8) into equation (3.7), DNL_k is simplified to

$$DNL_k \approx [1+\varepsilon(x_k, y_k)]\cdot[1-\sum_{i=1}^{2^N}\varepsilon(x_k, y_k)/2^N] - 1 \approx \varepsilon(x_k, y_k) - \sum_{i=1}^{2^N}\varepsilon(x_k, y_k)/2^N. \quad (3.9)$$

Because of the unary weighted feature, gradient errors only result in a small DNL. Then the INL at code k can be calculated as

$$INL_k = \sum_{i=1}^{k} DNL_i \approx \sum_{i=1}^{k}\varepsilon(x_i, y_i) - k/2^N \cdot \sum_{i=1}^{2^N}\varepsilon(x_k, y_k). \quad (3.10)$$

As shown in equation (3.10), although gradient errors only cause a small resistance change for each unit resistor, it can accumulate across the whole array and result in a large INL error which equals to $\max(|INL_k|)$.

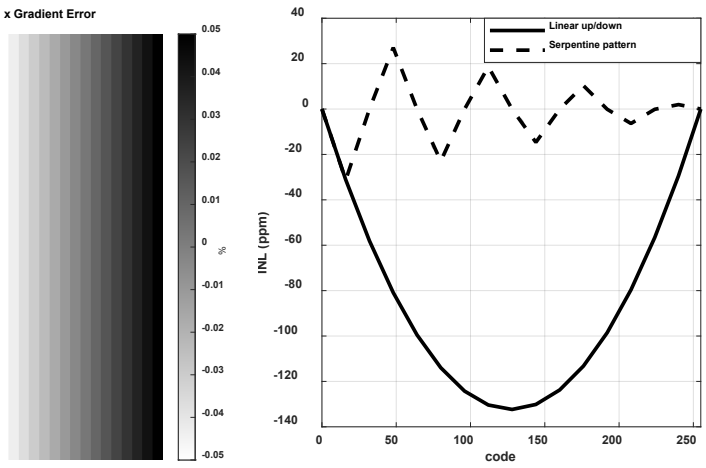

Figure 3.2 (a) x gradient error, (b) resulted INL of an 8-bit string DAC.

3.3.2 Common Layout Patterns

Routing metals and via both add parasitic resistances into the string. If parasitic resistances between adjacent cells are equal, then each parasitic can be considered as an additive to the unit cell and its effect on the DAC's linearity will be eliminated. The switching sequences in the cell level are not suitable in resistor strings because the complex routing induced parasitic is hard to match.

In industry, one of the most common layout patterns for a resistor ladder is the so-called linear up and down pattern as shown in Figure 3.1(a). It has short and systematic routings between columns and the string is periodic every two columns under any gradients in the y-direction. Correspondingly, the INL_k will also be periodic according to equation (3.10) and the resulted INL error is usually not large. However, x-direction gradients can integrate across the whole string and result in a large nonlinearity error in this pattern. To address this issue, dispersion can be added in the column level with careful routing. A serpentine shape pattern provided in Figure 3.1(b) requires no crossed routings and has periodicity every two columns under linear gradients in the x-direction.

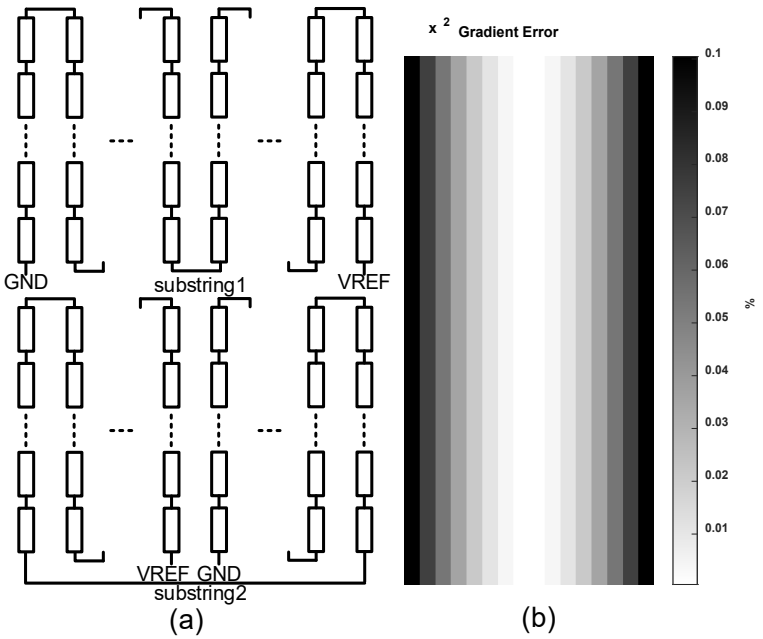

Figure 3.3 (a) Two substrings shifted in x-direction, (b) x^2 gradient error.

An x gradient error with a magnitude of 0.1 % is generated in MATLAB as shown in Figure 3.2(a). Using a behavior model of the string DAC with resistors' values calculated using equation (2.4), the simulated INL of two patterns are given in Figure 3.2(b). The effects of x gradients are mitigated approximately by a magnitude of four in the serpentine pattern.

The common centroid layout is another type of approach to suppress the linear gradient errors. Each device is divided into multiple sections which are placed so that their spatial centroids coincide. However, it has no effects on quadratic gradients which also exist as mentioned before. Neither does the serpentine pattern, it has almost the same INL curve as linear up and down pattern under 2^{nd} order gradient centered in the array. In [21], we have proposed a topology combining common centroid and shift INL methodology to suppress up to 2^{nd} order gradient errors. A short review of this topology is given below with an 8-bit example and the suppression of 2^{nd} order gradient is explained with derivations.

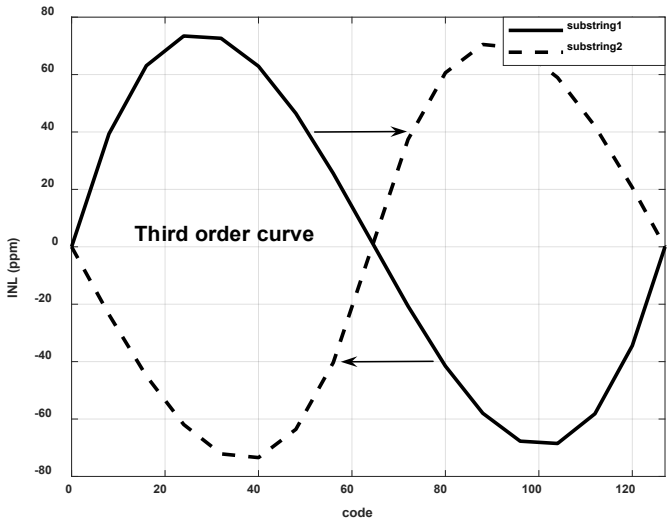

Figure 3.4 INL of two substrings under x^2 gradient error.

3.3.3 Proposed Layout Structure

The best remedy to reduce gradient errors in an array is to reduce the array area. When an 8-bit 16*16 array is divided into two 7-bit 16*8 array as shown in Figure 3.3(a), the y error term can be reduced by two times and the y^2 error term can be reduced by a factor of four. Because of the periodicity mentioned earlier, nonlinearity resulted from y^2 gradient should be 16 times (i.e., number of columns) smaller than nonlinearity resulted from x^2 term when they are in the same magnitude. To suppress x^2 errors, a novel shift INL methodology is introduced. For the linear up and down routing pattern under x^2 gradient, we have

$$\varepsilon(x_k, y_k) = \varepsilon(x_{2^N+1-k}, y_{2^N+1-k}) \approx \varepsilon(x_{2^N-k}, y_{2^N-k}). \tag{3.11}$$

Combining equation (3.9) and equation (3.11), the DNL curve has the following property:

$$DNL_k \approx DNL_{2^N-k}. \tag{3.12}$$

For k=1, 2, …, 2^N-1,

$$INL_k = \sum_{i=1}^{k} DNL_i = \sum_{i=1}^{2^N-1} DNL_i - \sum_{i=k+1}^{2^N-1} DNL_i \approx 0 - \sum_{i=1}^{2^N-1-k} DNL_i = -INL_{2^N-1-k}. \tag{3.13}$$

According to equation (3.13), the x^2 error term results in a 3rd order INL curve with half codes positive and half codes negative. Given in Figure 3.3(a), substring2 is a shifted version by half period in the x-direction (i.e., half columns) compared to substring1. As a result, the two substrings' INL curves will also be shifted by half period and have approximately opposite shapes.

An x^2 gradient error with a magnitude of 0.1 % is generated in MATLAB as shown in Figure 3.3(b). Behavior model of the two substrings arranged like Figure 3.3(a) are created in MATLAB. Figure 3.4 shows the simulated INL curves of these two substrings. If the two curves are averaged, the residue nonlinearity will be seven times smaller. Note that here we have assumed the x^2 gradient is centered in the middle of the two subarrays. For general 2nd order gradients in the x-direction, it can always be decomposed into $(x\text{-}x_0)^2$ and a linear term:

$$x^2 = (x - x_0)^2 + 2x_0 \cdot (x - x_0) + x_0^2 \tag{3.14}$$

To cancel all the linear gradient errors, the interdigitated common centroid concept is further implemented.

In the proposed topology, an 8-bit 16*16 array is broken into four 6-bit 8*8 subarrays as shown in Figure 3.5. Benefited from the reduced area, we can reduce quadratic errors in the y-direction by a factor of four in each subarray. Substring1 and substring2 have shifted starting points for x^2 error suppression. Substring3 and substring4 are interdigitated common centroid copies of 1 and 2. The interdigitation guarantees

$$\begin{aligned} x_{1k} - x_0 &= -(x_{3k} - x_0) \\ x_{2k} - x_0 &= -(x_{4k} - x_0) \end{aligned}, \tag{3.15}$$

where x_{ik} is the coordinate of resistor k in substring_i and x_0 is the centroid of the total layout in the x-direction. Equation (3.15) proves that all linear gradients are canceled out in the proposed layout including the linear term from uncentered x^2 gradient. With the same analog area, binary

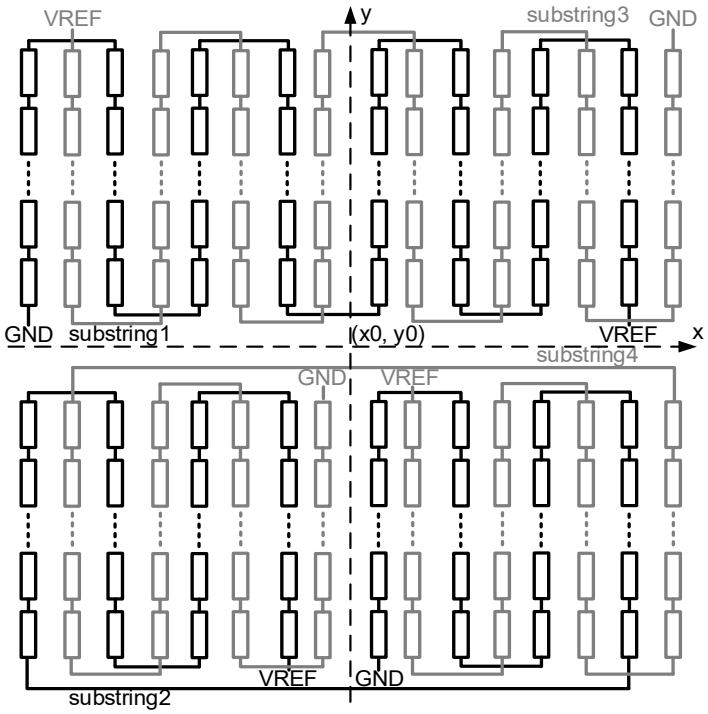

Figure 3.5 Proposed layout with shift INL methodology for suppressing up to 2nd order gradients.

weighted and thermometer coded DACs have the same root mean square (rms) INL considering only random variations [25]. Similarly, since the total area of resistors remains the same in the proposed approach, random mismatch induced nonlinearity magnitude is not changed. In the next section, small error analysis will be used to prove that averaging the substrings' outputs have the same effect at canceling out the errors as wiring substrings in the anti-parallel connection [5].

3.4 Outputs Averaging and Interpolation

This section will derive the INL at the averaged output of the substrings and present how interpolation can be done to recover the bits of resolution.

3.4.1 Averaging of the Substrings' Outputs

In the proposed layout structure, an 8-bit string is divided into four 6-bit substrings. The simplest method to get the averaged output is to tie all the outputs together as shown in Figure

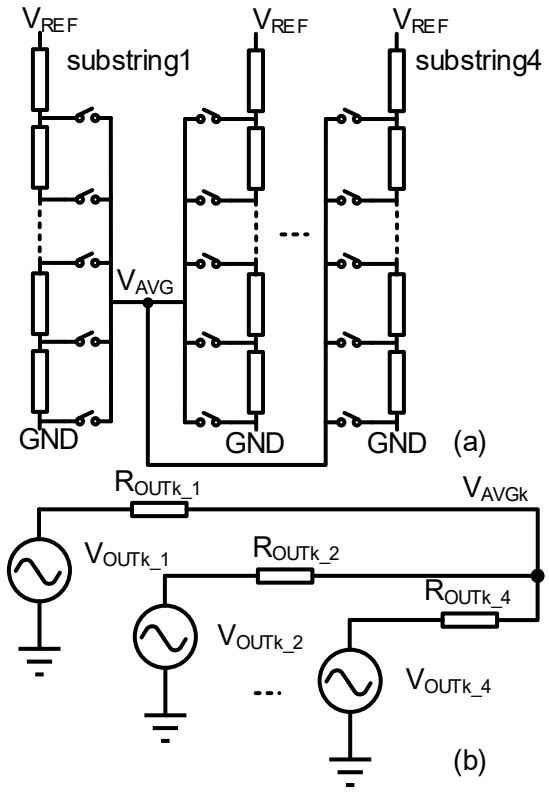

Figure 3.6 (a) Proposed averaging technique, (b) small error model.

3.6(a). A small error model is developed and combined with the superposition method to derive the INL equation.

Given in Figure 3.6(b), each substring's output at code k is modeled as

$$V_{OUTk\_i} = V_{OSi} + k \cdot LSB_i + INL_{k\_i}, \tag{3.16}$$

where V_{OSi} and LSB_i are the offset and the smallest increment voltage of the substring_i, respectively. LSB_i can be modeled as the nominal LSB=$V_{REF}/2^N$ plus a gain error term:

$$LSB_i = LSB + GE_i. \tag{3.17}$$

Following Kirchhoff's current law (KCL), the final output is a weighted average based on each substring's output conductance. The output resistance at code k of substring_i is

$$R_{OUTk\_i} = (\sum_{j=1}^{k} R_{j\_i}) \| (\sum_{j=k+1}^{2^N} R_{j\_i}) + R_{SWk\_i}, \tag{3.18}$$

where R_{j_i} is the jth resistor in substring_i and R_{SWk_i} is the on resistance of the switch at code k in substring_i, respectively. The resistance variations of unit cells and switches can be lumped together and R_{OUTk_i} can be rewritten into:

$$R_{OUTk\_i} = (k \cdot R_0) \| [(2^N - k) \cdot R_0] + R_{SWk} + \Delta R_{OUTk\_i} = R_{OUTk} + \Delta R_{OUTk\_i}. \tag{3.19}$$

Here R_{OUTk} is the nominal output resistance of a substring at code k and ΔR_{OUTk_i} stands for the lumped output resistance variation. The averaged output at code k is calculated below:

$$V_{AVGk} = \sum_{i=1}^{4} (G_{OUTk\_i} / \sum_{j=1}^{4} G_{OUTk\_j} \cdot V_{OUTk\_i}), \tag{3.20}$$

where $G_{OUTk\_i} = 1/R_{OUTk\_i}$. Applying the approximation equation (3.8) in equation (3.19), G_{OUTk_i} can be estimated as

$$G_{OUTk\_i} = 1/[R_{OUTk} \cdot (1 + \frac{\Delta R_{OUTk\_i}}{R_{OUTk}})] \approx \frac{1}{R_{OUTk}} (1 - \frac{\Delta R_{OUTk\_i}}{R_{OUTk}}). \tag{3.21}$$

Substituting equation (3.21) in equation (3.20) and ignoring high order small errors, the averaged output is calculated as follows:

$$V_{AVGk} \approx \frac{1}{4} \sum_{i=1}^{4} V_{OSi} + k \cdot (LSB + \frac{1}{4} \sum_{i=1}^{4} GE_i) + \frac{1}{4} \sum_{i=1}^{4} INL_{k\_i}. \tag{3.22}$$

The first two terms are the offset and gain error of the DAC which are not code dependent, whereas the last term is the nonlinearity term. Therefore, the INL at the averaged output equals to

$$INL_k \approx \frac{1}{4} \sum_{i=1}^{4} INL_{k\_i}. \tag{3.23}$$

Based on equation (3.23), we can conclude that, when substrings have opposite INL curves under gradients, the nonlinearity can be averaged out by simply connecting the substrings' outputs

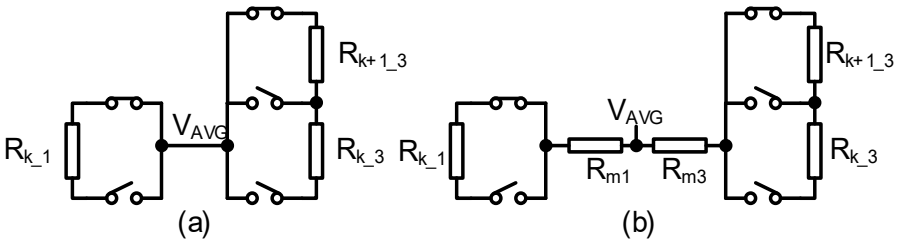

Figure 3.7 (a) Interpolation between code k and k+1, (b) interpolation with matched resistors.

together. When a string is divided, each substring's resolution is reduced, so does the resolution at the averaged output. However, we can encode different digital inputs to the substrings and achieve more bits with further interpolation.

3.4.2 Interpolation with Matched Resistors

Theoretically, with 2^n sub-DACs, n more bits of resolution can be recovered. In the 8-bit example given in Figure 3.5, substring1 and substring3 are two 6-bit strings placed in common centroid locations. Linear gradient errors are expected to be canceled at the averaged output. To recover the bits of resolution to 7-bit, we can interpolate substring1 at code k and substring3 at code k+1 as shown in Figure 3.7(a). One more output is recovered, and it equals to

$$V_{AVG_{k+\frac{1}{2}}} = \frac{G_{OUTk\_1}}{G_{OUTk\_1} + G_{OUTk+1\_3}} \cdot V_{OUTk\_1} + \frac{G_{OUTk+1\_3}}{G_{OUTk\_1} + G_{OUTk+1\_3}} \cdot V_{OUTk+1\_3}. \tag{3.24}$$

As derived in equation (3.10), INL_k results from the integration of resistor variations and is continuous over the entire digital input range. Therefore, the suppression of linear gradient errors is still effective here. Assuming $G_{OUTk\_1} \approx G_{OUTk+1\_3}$, equation (3.24) will be simplified to

$$V_{AVGk+1/2} \approx (V_{OUTk\_1} + V_{OUTk+1\_3})/2. \tag{3.25}$$

Then the INL at the 7-bit interpolated output becomes

$$\begin{aligned} INL_{AVG2k} &\approx (INL_{k\_1} + INL_{k\_3})/2, k = 0,1,...,2^6 - 1 \\ INL_{AVG2k+1} &\approx INL_{AVG2k} + DNL_{k+1\_3}/2 \end{aligned}. \tag{3.26}$$

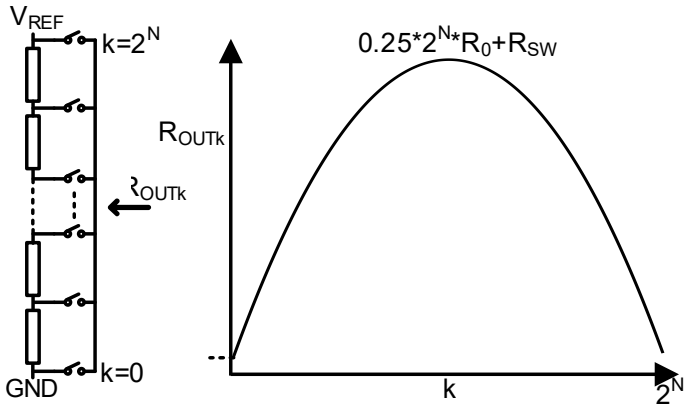

Figure 3.8 String DAC's output resistance.

Since substring 1 and 3 are in common centroid layout, under linear gradient errors, we have $INL_{k\_1} = -INL_{k\_3}$ according to equation (3.10). Given in equation (3.9), the gradient errors induced DNL of a string DAC is small. Therefore, linear gradients caused nonlinearity is canceled even at the interpolated output. However, the assumption $G_{OUTk\_1} \approx G_{OUTk+1\_3}$ is not always true.

The output impedance of a resistor string varies with the position as shown in Figure 3.8 [26]. On both ends, the impedance is small and equal to the impedance of the reference source plus the switch on resistance. Therefore, the G_{OUT} difference between code k and k+1 can be large on both ends. A code dependent error due to the unbalanced interpolation will show up in the INL curve. To solve this problem, a matched resistor network R_{m1} and R_{m3} are added at each substring's output as given in Figure 3.7(b) to compensate for the G_{OUT} difference. An alternative solution is to add top and bottom resistors shared by all substrings with a drawback of a reduced output range. The largest G_{OUT} mismatch between consecutive codes happens at two ends. Given in Figure 3.9, by interpolating substring1 at code 0 and substring3 at code 1, one more output is recovered between 0 V and 1 $LSB_{6\text{-}b} = V_{REF}/2^6$ and it equals to

$$V_{AVG\frac{1}{2}} \approx \frac{R_m + R_{SW}}{2(R_m + R_{SW}) + R_0} \cdot \frac{V_{REF}}{2^6}, \tag{3.27}$$

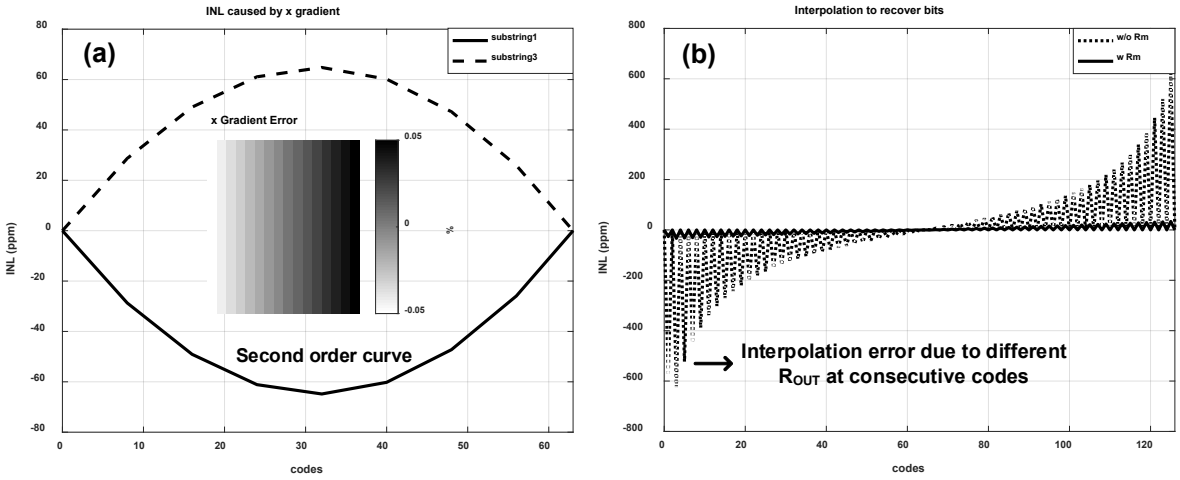

Figure 3.10 (a) Two substrings under linear gradient in the x-direction, (b) interpolation without Rm and with Rm.

where R_m is the nominal value of the matched resistor network. In equation (3.27), the output resistance of substring3 has been approximated by R_0 to simplify the equation. To achieve 16-bit linearity (INL < 1 $LSB_{16\text{-}b}$), the interpolation error needs to satisfy:

$$|V_{AVG\frac{1}{2}} - \frac{1}{2}\frac{V_{REF}}{2^6}| < \frac{V_{REF}}{2^{16}}. \tag{3.28}$$

By substituting equation (3.27) into equation (3.28), we can derive the requirement for R_m:

$$R_m > 2^8 R_0 - R_{SW}. \tag{3.29}$$

The matched resistors require large resistances. However, since they are further interpolating voltages at consecutive codes, the matching performance is not so critical as unit cells in the string. 9-bit matching is sufficient for 16-bit linearity considering only interpolation error resulted from their mismatch. Therefore, area overhead is minimal when high resistivity resistors are used. The extra benefit is that the output resistance of the DAC becomes more constant across the input range, but the settling time does increase which is not stringent for precision string DACs used in pseudo-static applications.

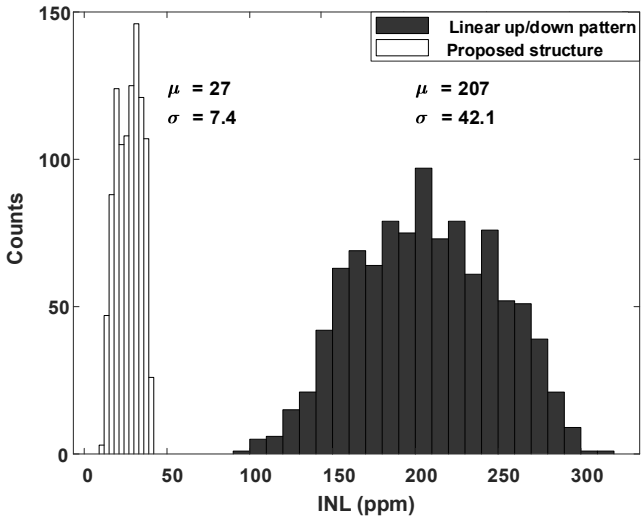

Figure 3.11 Simulated INL under randomly generated gradient patterns.

Using the same MATLAB behavioral model used before, the interpolation processes with and without matched resistors are simulated and the results are shown in Figure 3.10. Since substring1 and 3 are common centroid copies, the resistor errors accumulate to opposite INL shapes as shown in Figure 3.10(a). If we directly interpolate the substrings' outputs, large INL glitches are observed at odd codes in Figure 3.10(b). The problem is most severe at two ends where R_{OUT} difference between consecutive codes are the largest. However, with the matched R_m network, the interpolation error can be mitigated to the targeted 16-bit level.

In the proposed topology, an 8-bit string is broken into four substrings and two more bits of resolution can be recovered through interpolation. For k = 0, 1, …, 2^6 - 1, three more outputs can be recovered between code k and k+1 as shown below:

$$\begin{aligned} V_{AVGk+1/4} &\approx (V_{OUTk\_1} + V_{OUTk+1\_2} + V_{OUTk\_3} + V_{OUTk\_4})/4 \\ V_{AVGk+1/2} &\approx (V_{OUTk\_1} + V_{OUTk+1\_2} + V_{OUTk+1\_3} + V_{OUTk\_4})/4 \\ V_{AVGk+3/4} &\approx (V_{OUTk\_1} + V_{OUTk+1\_2} + V_{OUTk+1\_3} + V_{OUTk+1\_4})/4 \end{aligned} \quad . \tag{3.30}$$

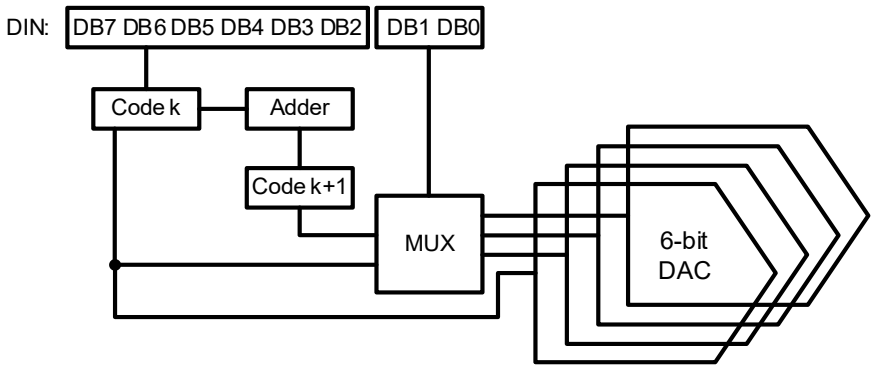

Figure 3.12 Possible digital interface.

Therefore, an 8-bit resolution is achieved at the averaged and interpolated output of four substrings. Note that three output voltages near V_{REF} are lost, but it doesn't hurt most applications of string DACs.

To prove the effectiveness of the proposed averaging and interpolation scheme, randomly generated gradient patterns are assigned on strings in traditional linear up/down pattern and in our proposed structure. Gradient coefficients, i.e., $a_1 \sim a_5$ in (3), are limited in a range of half to two times of what we saw in measured chips. Both DACs are modeled numerically in MATLAB. The random mismatch between unit resistors is assumed to be 0.03% which corresponds to about 16-bit linearity. The simulated INLs are provided in Figure 3.11. The string DAC with linear up/down pattern has mean INL of 207 ppm, while after applying our structure method, the DAC has mean INL of 27 ppm with a standard deviation of 7.4 ppm. The improvement factor really depends on the type of gradients which is unknown before production.

The digital interface for the proposed structure is different from a traditional string DAC where a shift-register chain and row-column decoders will work. A possible solution is provided in Figure 3.12. In the 8-b digital input, the 6-b MSB bits DB7-DB2 corresponding to code k are sent to substring1 and an adder to get the code k+1. Then both codes k and k+1 are sent to a

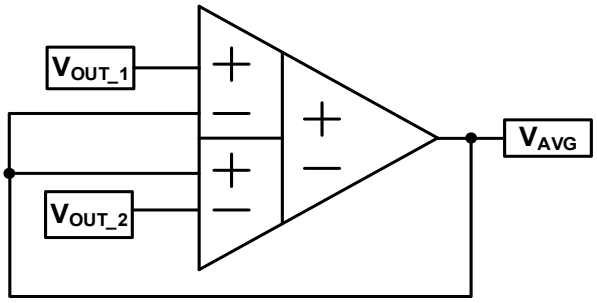

Figure 3.13 Interpolation with a differential difference amplifier.

multiplexer controlled by the last two bits DB1-DB0 which decide what to send to the left three substrings.

3.4.3 Interpolation with a Differential Difference Amplifier

Adding the matched resistor network consumes area and slows down the string DAC's settling performance. Differential difference amplifiers (DDAs) have been configured to perform voltages averaging functions [27], [28]. By sending voltage outputs to transistors' gates, the output resistance difference problem in the interpolation process can be solved. Moreover, the amplifier can simultaneously work as the isolation buffer in a segmented resistor string architecture.

Connecting a differential difference amplifier like Figure 3.13, we have

$$V_{AVG} = A \cdot [(V_{OUT\_1} - V_{AVG}) - (V_{AVG} - V_{OUT\_2}) + \frac{V_{OUT\_1} + V_{AVG} - (V_{AVG} + V_{OUT\_2})}{CMRR}], \quad (3.31)$$

where A is the gain of the amplifier and CMRR is the common-mode rejection ratio. From equation (3.31), the averaged output V_{OUTk} can be derived as

$$V_{AVG} = \frac{A}{1/2 + A} \cdot \frac{V_{OUT\_1} + V_{OUT\_2}}{2} + \frac{1}{1 + 2A} \cdot \frac{V_{OUT\_1} - V_{OUT\_2}}{CMRR}. \quad (3.32)$$

The amplifier should have enough gain and CMRR so that the 2nd term is small enough during interpolation. If these requirements are satisfied, V_{OUTk} can be approximated by

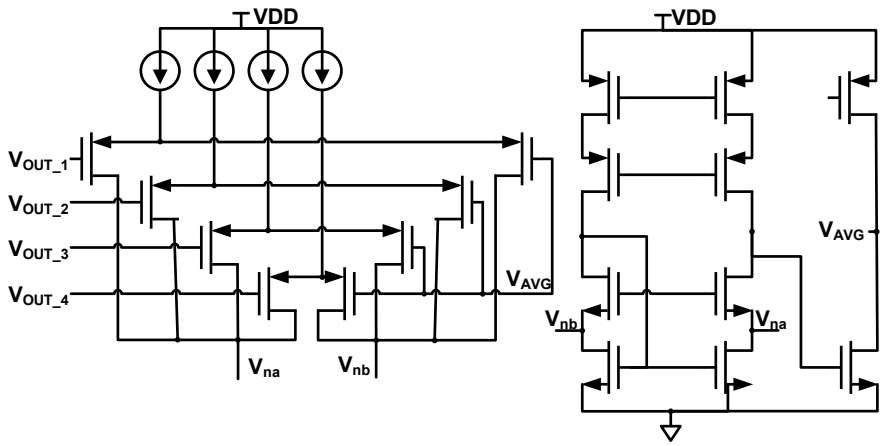

Figure 3.14 A differential difference amplifier with four input pairs.

$$V_{AVG} \approx \frac{A}{1/2+A} \cdot \frac{V_{OUT\_1}+V_{OUT\_2}}{2}. \tag{3.33}$$

The amplifier gain needs to be either constant or large enough so that the coefficient A/(1/2+A) is fixed and doesn't introduce nonlinearity into the averaged output.

In the proposed architecture, there are four substrings to be averaged. A differential difference amplifier with four input pairs as shown in Figure 3.14 can be used to perform the averaging and interpolation function. The circuit is similar to the LSB interpolation circuit in [12] but has a different goal here. The amplifier's input common mode range (ICMR) should cover the substrings' output range. Therefore, the amplifier's supply V_{DD} needs to satisfy:

$$V_{DD} \geq V_{REF} + V_{SGp} + V_{SDp}. \tag{3.34}$$

Good matching between the differential pairs is important to minimize the offsets. The offsets' dependency on common mode input voltage needs to be minimized as they can add nonlinearities to the DAC output.

3.4.4 Discussions

The proposed solution is in topology level and therefore it's robust under PVT (process-voltage-temperature) variations. It's low cost as it doesn't require extra layer for analog trimming

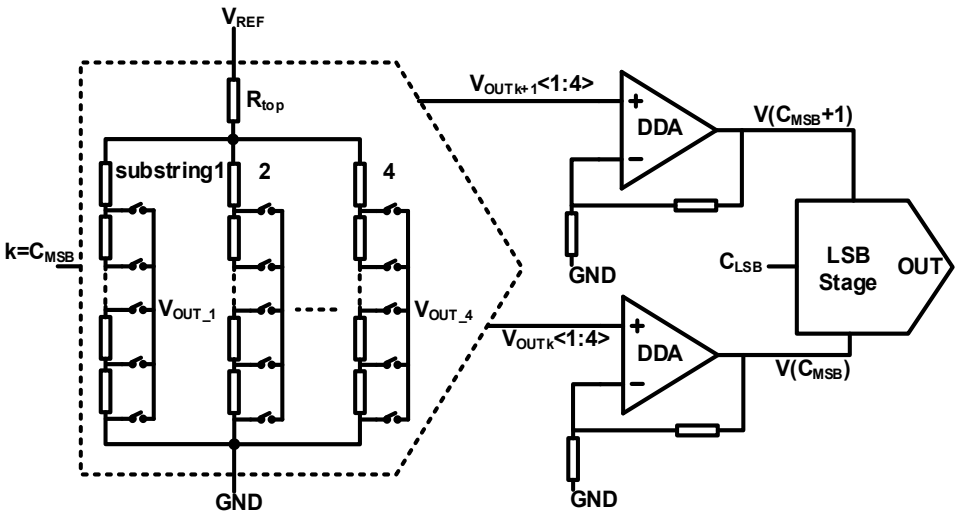

Figure 3.15 Possible two-stage DAC with high linearity.

or on-chip memory for production measurement or calibration. It can be implemented on resistor arrays in cheap processes to suppress the gradient errors' effect and improve linearity. The area overhead in the interpolation scheme using matched resistors is small compared to the string array.

As mentioned before, the nonlinearity resulted from random mismatch is not changed in the proposed solution. However, the current flowing in each substring is four times larger due to the reduced resistance. Since there are four substrings, the total power consumption is sixteen times of a traditional 8-bit string:

$$P_{String} = 4 * \frac{V_{REF}^{\ 2}}{64R_0} = 16 * \frac{V_{REF}^{\ 2}}{256R_0}. \tag{3.35}$$

If the power consumption is a stringent design budget for the DAC, the resistance of each unit cell needs to be increased correspondingly. Moreover, a large resistor R_{top} can be added at the top of the strings to reduce the current as shown in Figure 3.15. If a rail to rail output between V_{REF} and GND is required, a gain stage can be added in the following interpolation amplifier. There is a tradeoff between power consumption and noise performance when deciding the R_{top} value. The DDA also buffers the coarse string's output. Therefore, a fine string or an R-2R network can be

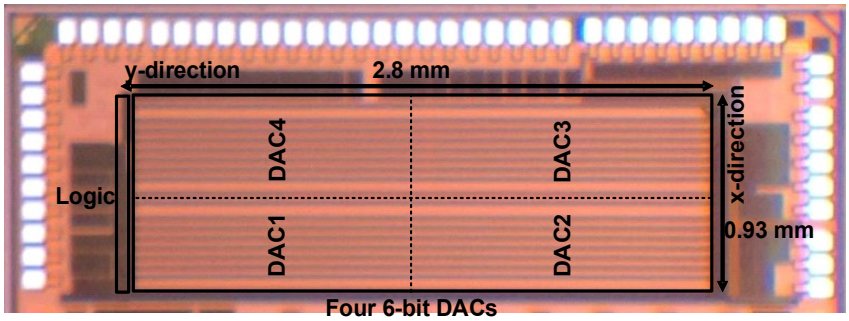

Figure 3.16 Chip micrograph.

loaded between two DDAs' outputs to build a two-stage RDAC as shown in Figure 3.15. The INL is mainly decided by the MSB string stage, so a high-resolution high- linearity DAC can be achieved.

3.5 Implementation

An 8-bit high linearity string DAC is fabricated in a CMOS 130nm process with TaN thin film resistors. This section will explain the design of the DAC and the evaluation of the chip.

3.5.1 DAC Design

Unit resistors are firstly sized so that INL resulted from random mismatch is at 16-bit level (1σ interval used due to limited area). The INL calculation of unary-weighted or binary-weighted DAC was proved in [25]:

$$\sigma_{INL} \approx 2^{N/2-1} \cdot \sigma_{rnd}, \tag{3.36}$$

where σ_{INL} is the standard deviation of the INL error of the DAC. The required area of a unit resistor is then estimated from the standard deviation of its mismatch error σrnd:

$$\sigma_{rnd} = \sqrt{\frac{M_A^2}{W \cdot L} + \frac{M_W^2}{W^2} + \frac{M_L^2}{L^2}}, \tag{3.37}$$

where M_A, M_W and M_L are matching parameters provided in process documents. A relatively large L/W ratio is chosen here to keep the total power consumption of the ladders low.

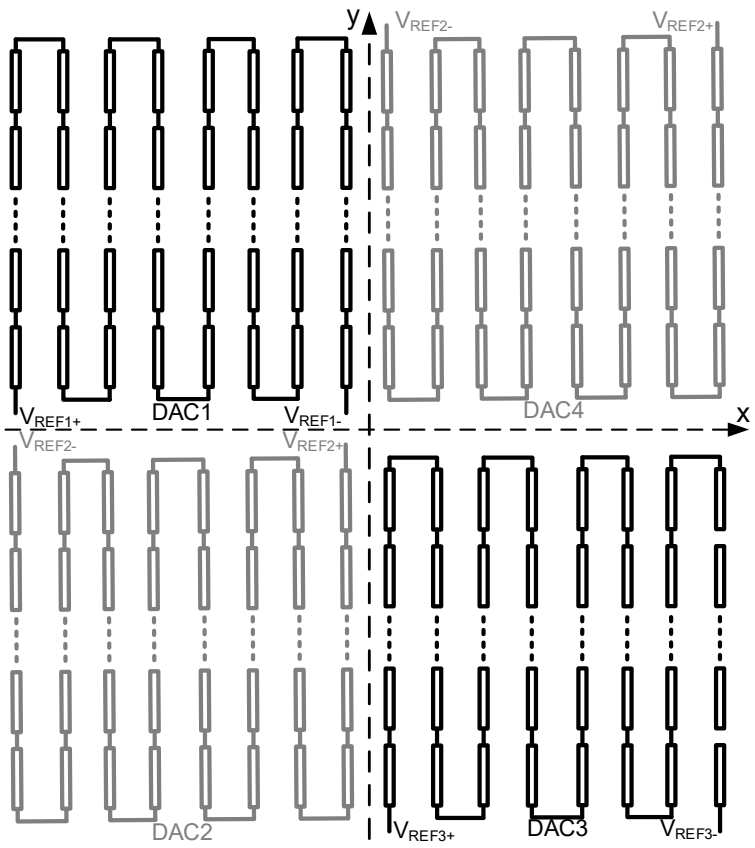

Figure 3.17 8-bit resistor array (configured into four 6-bit DACs).

The chip micrograph of the 8-bit DAC is given in Figure 3.16. It consists of four 6-bit substrings and each substring is configured to be a DAC with separated switches and decoders. The layout and routing of the four DACs are described in Figure 3.17. Common centroid layout is adopted to cancel linear gradient errors. Second order gradient errors in both x and y-direction are suppressed by a factor four because of the reduced area of each substring. The main target of this fabrication is to prove the averaging and interpolation method's effectiveness on canceling out gradient errors. However, by connecting substring1 and 4, substring 3 and 2 in series, two 7-bit DAC can be formed like Figure 3.3(a) to verify the shift INL methodology which can suppress x^2 gradient errors.

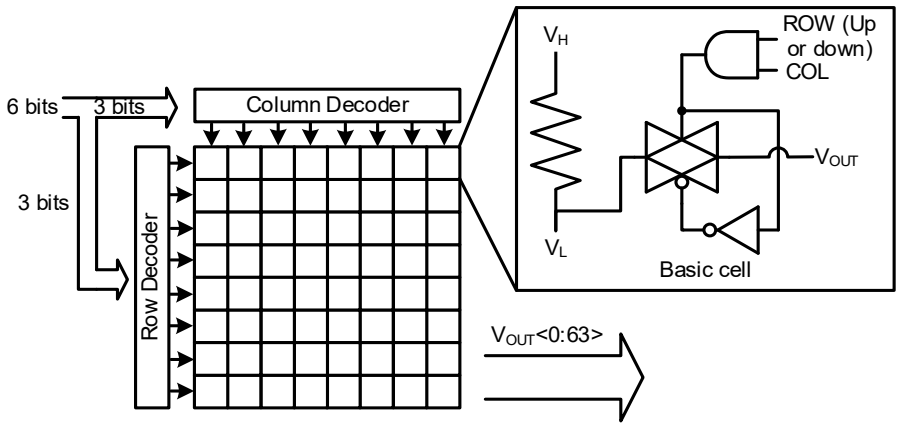

Figure 3.18 Block diagram of a 6-bit DAC.

The block diagram of a 6-bit DAC is described in Figure 3.18. The 6-b digital input word is split into two 3-b words which are decoded by two sets of 3-8 decoders. The row and column decoder lines run horizontally and vertically over the matrix, respectively. The matrix consists of 64 basic cells and each basic cell has a unit resistor, a transmission gate and an AND gate which perform the final decoding. The target of this prototype is to verify the effectiveness of the gradient errors suppression after averaging and interpolation. Therefore, the digital logics are simplified. It includes a 24-b shift register chain which converts a serial 24-b input to parallel outputs and sends four 6-b codes to corresponding DACs where row and column decoders are used to perform the binary to thermometer conversion. For standard interfaces, an 8-b digital input need be processed. The first 6 bits represent the code k whereas the last 2 bits tell which DACs need to output at the code k+1.

Careful floor planning, symmetric and hierarchical layout strategy has been adopted to improve matching. Routing between the ends of columns is done with another metal layer and enough via. Dummy cells are placed around the total array to reduce the edge effects. Switches and AND gates are placed in the space to save area. In this prototype, each substring has separate

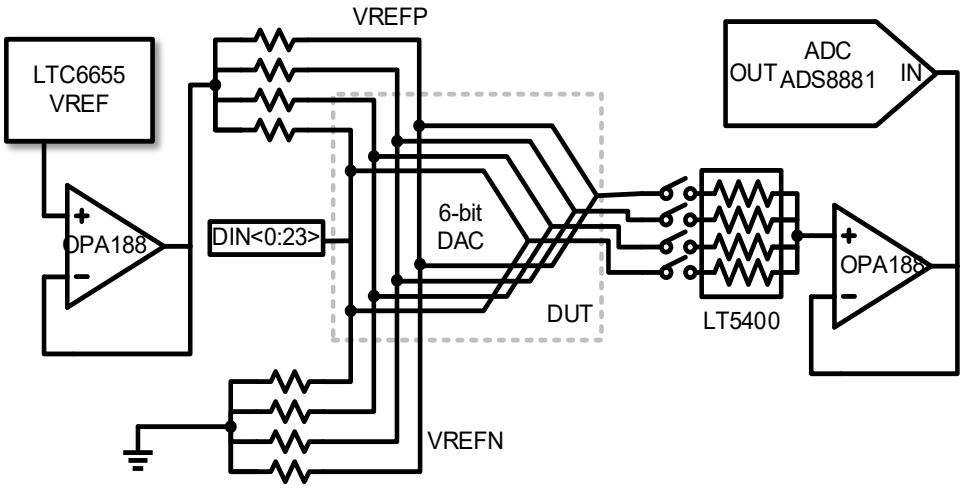

Figure 3.19 Evaluation board.

V_{REFP} and V_{REFN} pins so that they can be connected in series to form 7-bit or 8-bit DACs as references. When building a high-resolution DAC product, V_{REFP} and V_{REFN} of all the substrings should be star connected to reduce the offset and gain error differences.

3.5.2 Interpolation on Board

A resistor based averaging circuit is implemented in the evaluation board shown in Figure 3.19. An analog mux and a matched resistor array LT5400 are used as the averaging and interpolation circuit here. The LT5400's resistance is chosen so that the interpolation error is minimized below 16-bit level. A low drift precision reference LTC6655 is used to provide the reference voltage for the DACs. Top and bottom resistors are added to compensate systematic routing differences between 4 DACs due to separate V_{REF} pins. They are not required when the substrings share the same high and low references on chip. A DE-115 FPGA board is used to generate the 24-b series digital code to control the DACs here.

An 18-bit SAR ADC (ADS8881EVM) with 18-bit linearity (typical) is used to sample the output voltages. To avoid environment drift and reduce measurement noise, 32 full code sweeps with each code sampled 20 times are performed for each INL measurement.

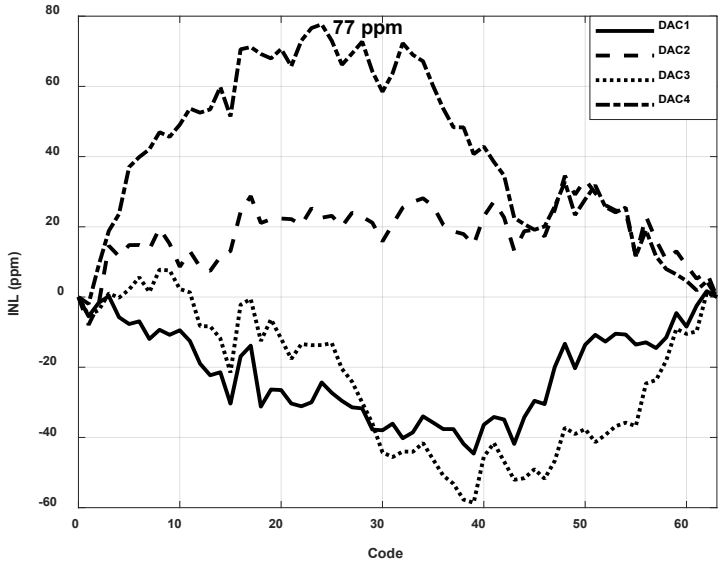

Figure 3.20 Measured INL of four DACs.

3.6 Measurement Results

Firstly, the same digital inputs are sent to four DACs and their output voltages are measured sequentially. The INL of these four 6-bit DACs are given in Figure 3.20. As analyzed before, after the division, quadratic gradient errors in each subarray is reduced by a factor of four and linear gradient errors appear to dominate. DAC1 and DAC2 are common centroid layout with linear up and down patterns and so do DAC3 and DAC4. As shown in Figure 3.20, they have approximately opposite 2nd order INL shapes which mainly comes from the accumulation of linear gradients in the x-direction. The largest INL of the four DACs is 77 ppm which corresponds to about 14-bit linearity. This indicates gradient errors can accumulate to a large INL in an 8-bit string DAC with traditional layout.

By controlling the mux at the DACs' outputs, averaged and interpolated outputs are measured, respectively. For example, when measuring DAC1&2's averaged output, the corresponding switches are turned on while the other two are left off and then the digital input is ramped up. As shown in Figure 3.21, at the averaged output of DAC1 and 2, INL is reduced by 3

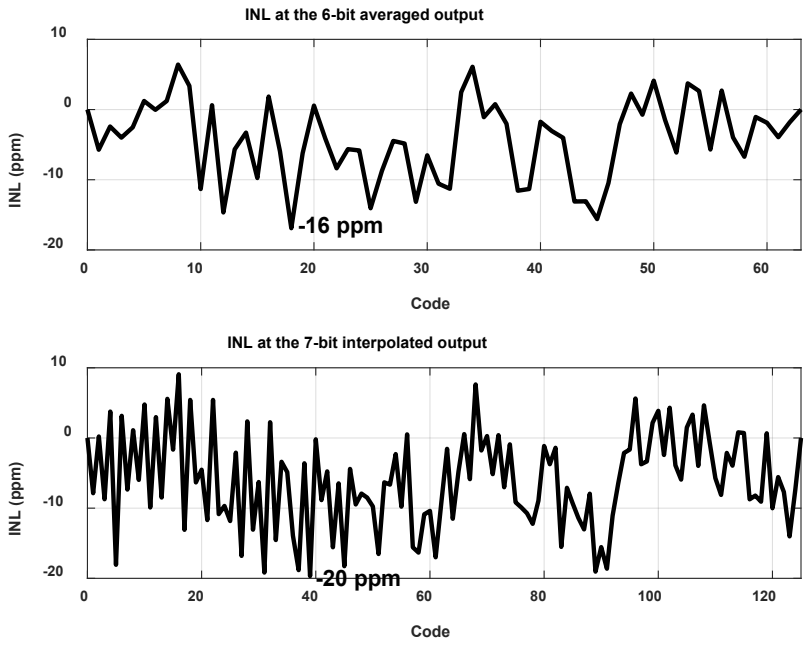

Figure 3.21 INLs at the averaged and interpolated outputs of DAC1&2.

times to 16 ppm, a level limited by the random mismatch between resistors. Consecutive codes are also sent to the DACs to evaluate the interpolation effect on gradient cancellation. The interpolated result of these two DACs to achieve 7-bit is also provided in Figure 3.21. The 7-bit INL curve has a similar pattern as the 6-bit INL curve as derived in equation (3.26). With all four DACs being interpolated, an 8-bit output can be recovered and the measured INL curve is given in Figure 3.22. The suppression of gradient errors still works efficiently and the INL is 21 ppm which corresponds to about 16-bit linearity. When connecting substrings in series, large DNL errors at mid codes are observed due to the routing, pad, and bonding wire resistances so the results are not used here. Compared to reducing each unit resistor's variation with division and anti-parallel connection, the proposed averaging and interpolation method at circuit level has a much simpler layout but still can suppress the gradient errors' effect on nonlinearity. Four chips are measured in total and the INL results of the 8-bit output are summarized in Table 3.1. The INL errors range from 21 to 26 ppm. It's difficult to compare DAC designs as the specifications vary a lot. A performance

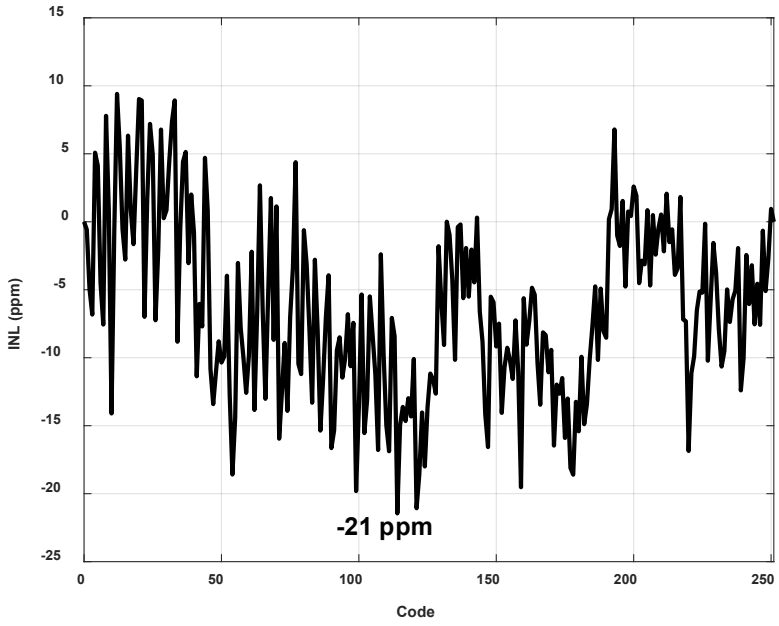

Figure 3.22 INL at the 8-bit interpolated output of four DACs.

Table 3.1 Measured INL results of four chips

Chip num	1	2	3	4
max(|INL_k|) (ppm)	21	25	25	26
max(INL_k)-min(INL_k) (ppm)	30	29	46	33

comparison table is provided in Table 3.2.The linearity is characterized in unit of ppm so that we can compare DACs of different resolutions. With the proposed gradient suppression structure, the achieved INL is six times smaller than [15]. Although TaN thin film resistors do have better matching and lower temperature coefficient than poly resistors, the difference is qualitatively in the same order of magnitude. High performance Si-Cr thin film resistor combined with dedicated calibration circuits is used in [7] to achieve the industry-leading sub-ppm INL. The proposed structure method apparently can be applied to any types of resistors and it doesn't require high performance analog processes or modern digital processes with dense memory. It can be combined with existing solutions to further improve linearity or reduce calibration efforts.

Table 3.2 Comparison of state-of-art

	ITC'05 [15]	JSSC'13 [7]	TCAS-I'15 [3]	This work
Technology	0.5µm	0.6µm	0.35µm	0.13µm
Resistor material	Poly	Si-Cr	Poly	TaN
Resolution	16-b	20-b	14-b	8-b
Output range (V)	0 to 5	-10 to 10	0 to 2.5*	0 to 3.3
INL (ppm)	122	0.33	244	21
DNL (ppm)	6.1	0.29	61	18
Die area (mm^2)	4	9.67	0.40	2.6
Current (mA)	0.75	4.2	0.37	0.19

*DAC's output before HV output driver.

3.7 Conclusion

This work presents a low-cost high-linearity string DAC structure with up to 2^{nd} order gradient errors suppression. A shift INL technique is introduced to suppress x^2 gradient error. The proposed averaging and interpolation methods eliminate the complex routing problem in resistor strings layout. Measurement results show significant reductions in INL at the averaged output and the 8-bit DAC prototype achieves 16-bit linearity performance without trimming and calibration. The residue INL error results from random mismatch which can be further reduced by increasing the unit resistor area.

The presented 8-bit string DAC can be utilized as the MSB stage and be combined with a fine ladder or an R-2R architecture to build a low-cost high-resolution RDAC with good linearity performance.

3.8 References

[1] C. C. McAndrew, "Layout Symmetries: Quantification and Application to Cancel Nonlinear Process Gradients," *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, vol. 36, no. 1, pp. 1–14, Jan. 2017, doi: 10.1109/TCAD.2016.2561970.

[2] *The 20-Bit DAC Is the Easiest Part of a 1-ppm-Accurate Precision Voltage Source | Analog Devices.* Accessed: Oct. 7, 2019. [Online]. Available: https://www.analog.com/en/analog-dialogue/articles/20-bit-dac-and-accurate-precision-voltage-source.html.

[3] R. Pierco, G. Torfs, J. Verbrugghe, B. Bakeroot, and J. Bauwelinck, "A 16 Channel High-Voltage Driver with 14 Bit Resolution for Driving Piezoelectric Actuators," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 62, no. 7, pp. 1726–1736, Jul. 2015, doi: 10.1109/TCSI.2015.2441963.

[4] C. Lu, P. Yin, and M. Lin, "A 10-bit Two-Stage R-DAC With Isolating Source Followers for TFT-LCD and AMOLED Column-Driver ICs," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 27, no. 2, pp. 326–336, Feb. 2019, doi: 10.1109/TVLSI.2018.2876638.

[5] M. J. M. Pelgrom, "A 10-b 50-MHz CMOS D/A converter with 75- Omega buffer," *IEEE J. Solid-State Circuits*, vol. 25, no. 6, pp. 1347–1352, Dec. 1990, doi: 10.1109/4.62178.

[6] A.- Annema, B. Nauta, R. van Langevelde, and H. Tuinhout, "Analog circuits in ultra-deep-submicron CMOS," *IEEE J. Solid-State Circuits*, vol. 40, no. 1, pp. 132–143, Jan. 2005, doi: 10.1109/JSSC.2004.837247.

[7] R. C. McLachlan, A. Gillespie, M. C. W. Coln, D. Chisholm, and D. T. Lee, "A 20b Clockless DAC With Sub-ppm INL, 7.5 nV/√Hz Noise and 0.05 ppm/°C Stability," *IEEE J. Solid-State Circuits*, vol. 48, no. 12, pp. 3028‑3037, Dec. 2013, doi: 10.1109/JSSC.2013.2278449.

[8] J. L. Brubaker, "Gradient insensitive split-core digital to analog converter," US7414561B1, 19-Aug-2008.

[9] J. J. Ashe, "Digital-to-analog converter with segmented resistor string," US5495245A, 27-Feb-1996.

[10] D. Dempsey and C. Gorman, "Digital to analog converter," US5969657A, 19-Oct-1999.

[11] W. C. Rempfer, "Digital to analog converter," US5396245A, 07-Mar-1995.

[12] A. Yilmaz, "LSB interpolation circuit and method for segmented digital-to-analog converter," US6246351B1, 12-Jun-2001.

[13] K. C. Dyer, J. P. Keane, and S. H. Lewis, "Calibration and Dynamic Matching in Data Converters: Part 1: Linearity Calibration and Dynamic-Matching Techniques," *IEEE Solid-State Circuits Mag.*, vol. 10, no. 2, pp. 46–55, Spring 2018, doi:10.1109/MSSC.2017.2771106.

[14] P. C. Kirby, C. G. Lyden, and T. M. Vinereanu, "Digital-to-analog converter structures," US7095351B2, 22-Aug-2006.

[15] K. Parthasarathy, T. Kuyel, Zhongjun Yu, Degang Chen, and R. Geiger, "A 16-bit resistor string DAC with full-calibration at final test," in *IEEE International Conference on Test*, 2005., pp. 10 pp. – 75, doi: 10.1109/TEST.2005.1583962.

[16] S. Su, T.-I. Tsai, P. K. Sharma, and M. S.-W. Chen, "A 12 bit 1 GS/s Dual-Rate Hybrid DAC With an 8 GS/s Unrolled Pipeline Delta-Sigma Modulator Achieving > 75 dB SFDR Over the Nyquist Band," *IEEE J. Solid-State Circuits*, vol. 50, no. 4, pp. 896–907, Apr. 2015, doi: 10.1109/JSSC.2014.2385752.

[17] G. A. M. V. D. Plas, J. Vandenbussche, W. Sansen, M. S. J. Steyaert, and G. G. E. Gielen, "A 14-bit intrinsic accuracy Q^2 random walk CMOS DAC," *IEEE J. Solid-State Circuits*, vol. 34, no. 12, pp. 1708–1718, Dec. 1999, doi: 10.1109/4.808896.

[18] Yonghua Cong and R. L. Geiger, "Switching sequence optimization for gradient error compensation in thermometer-decoded DAC arrays," *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 47, no. 7, pp. 585–595, Jul. 2000, doi: 10.1109/82.850417.

[19] M. S. Yenuchenko, A. S. Korotkov, D. V. Morozov, and M. M. Pilipko, "A Switching Sequence for Unary Digital-to-Analog Converters Based on a Knight's Tour," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 66, no. 6, pp. 2230–2239, Jun. 2019, doi: 10.1109/TCSI.2018.2890412.

[20] A. Yilmaz, "Resistor string integrated circuit and method for reduced linearity error," US6496133B1, 17-Dec-2002.

[21] N. Liu, C. Lash, J. Todsen, and D. Chen, "Practical linear and quadratic gradient errors suppression techniques in string DACs," in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems*, 2017, pp. 1176–1179, doi: 10.1109/MWSCAS.2017.8053138.

[22] A. Asenov, S. Kaya, and A. R. Brown, "Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness," *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May 2003, doi: 10.1109/TED.2003.813457.

[23] G. Gildenblat, Ed., *Compact Modeling: Principles, Techniques and Applications*. Amsterdam, The Netherlands: Springer, 2010.

[24] Y. Chen, R. C. Jaeger, and J. C. Suhling, "CMOS Sensor Arrays for High Resolution Die Stress Mapping in Packaged Integrated Circuits," *IEEE Sens. J.*, vol. 13, no. 6, pp. 2066–2076, Jun. 2013, doi: 10.1109/JSEN.2013.2247590.

[25] Chi-Hung Lin and K. Bult, "A 10-b, 500-MSample/s CMOS DAC in 0.6 mm^2," *IEEE J. Solid-State Circuits*, vol. 33, no. 12, pp. 1948–1958, Dec. 1998, doi: 10.1109/4.735535.

[26] M. J. M. Pelgrom, *Analog-to-Digital Conversion*. Amsterdam, The Netherlands: Springer, 2010.

[27] F. Khateb and T. Kulej, "Design and Implementation of a 0.3-V Differential Difference Amplifier," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 66, no. 2, pp. 513–523, Feb. 2019, doi: 10.1109/TCSI.2018.2866179.

[28] H. Alzaher and M. Ismail, "A CMOS fully balanced differential difference amplifier and its applications," *IEEE Trans. Circuits Syst. II Analog Digit. Signal Process.*, vol. 48, no. 6, pp. 614–620, Jun. 2001, doi: 10.1109/82.943332.

CHAPTER 4. A LOW-POWER AND AREA-EFFICIENT ANALOG DUTY CYCLE CORRECTOR FOR ADC'S EXTERNAL CLOCKS

Nanqi Liu*, Jim Todsen and Degang Chen*

*Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010

Texas Instruments, Tucson, AZ 85711

Modified from a manuscript to be submitted to IEEE Transactions on Circuits and Systems II: Express Briefs

4.1 Abstract

Using a clock without a large duty cycle error is important for some high performance mixed-signal circuits, and this chapter presents a low-power and area-efficient analog duty cycle corrector with feedback. A differential charge pump with fast startup is used to detect the duty cycle error and output a control voltage for current-starved inverters to adjust pulse width. The pump current is designed to be adaptive to input frequency, so a wide frequency range can be achieved without requiring a large load capacitor. The circuit is designed in a 130nm CMOS process and can correct 4-100 MHz clocks with a 25-75% duty cycle range. The simulation results show that output error is limited to 1.5% and the power consumption is 50 μW at 100 MHz.

4.2 Introduction

While digital clock speed has been increasing along with advances in scaling of CMOS processes, with more process variations, a clock generated from an oscillator can suffer from unpredictable duty cycle errors. Buffers and drivers are used to distribute a clock signal over an entire chip or to drive a heavy load, and mismatches between PMOS and NMOS transistors can result in unbalanced rise/fall times and change pulse widths. Asymmetric bias temperature instability (BTI) aging can also result in duty cycle error [1], causing a clock's duty cycle to significantly wander away from 50%.

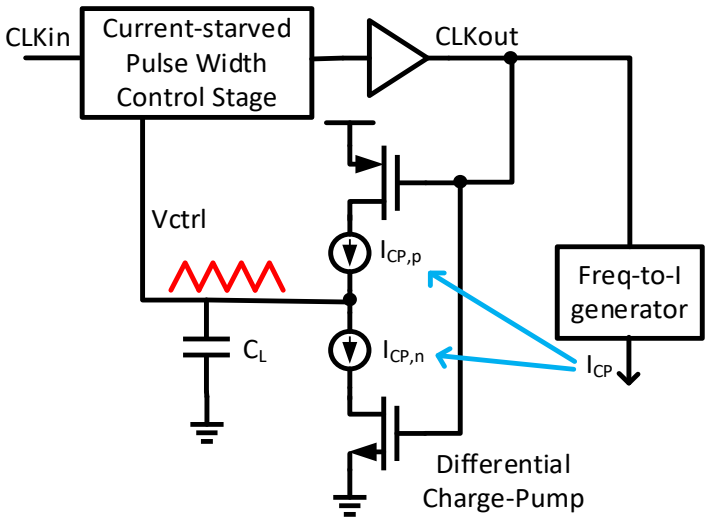

Figure 4.1 Proposed duty cycle corrector topology

Having a 50% duty cycle clock can be very important in many applications because events can occur on both edges. Double data rate-SDRAMs, for example, use both rising and falling edges and therefore require the duty cycle to be accurately maintained at 50%. In radio frequency (RF) communication systems, even order harmonics can arise from non-50% clocks and directly cause performance degradation [2]. These are high speed applications where clocks are at the GHz level and duty cycle correctors (DCCs) have been widely employed for on-chip calibration. Moreover, DCCs also have their uses in MHz applications. Double-sampling architectures have been widely implemented in analog-to-digital converters (ADCs) to reduce power consumption, and a pipelined ADC can save power and area by sharing an op-amp between two channels [3]. In a delta-sigma ADC [4], a double-sampling switched-operational-amplifier (SOP)-based integrator is adopted to reduce the clock frequency by half. When an external clock is sent to the ADC, a DCC will be useful to correct the duty cycle and guarantee sufficient and balanced settling time.

Many duty cycle correction circuits have been developed, and they can be broadly divided into two categories: analog methods with feedback and digital methods with or without feedback. In [5], [6], and [7], analog pulse width control loops (PWCLs) have been proposed in which a

single or differential charge pump is used as the duty cycle detector (DCD) that generates a feedback control voltage to stretch or shrink pulses. A differential amplifier and low pass filters have been used in [2], [8], and [9] to extract duty cycle information with reduced power consumption. In [10], a current-starving technique is presented for simultaneously adjusting rising and falling, which doubles the gain and accelerates the correction. Most digital methods without feedback use a half cycle delay line and an SR latch to regenerate a corrected clock [11]. Digital feedback methods have used time-to-digital converters and SAR controllers [12] to measure and adjust pulse width. In general, while digital techniques exhibit faster settling time and have no stability problems, they have a very limited frequency range and are more complicated than analog counterparts.

Precision ADCs often have external clocks with frequencies ranging from several MHz to hundreds of MHz. Digital DCCs are not attractive in this application because they require long delay chains to process the clock information, and given the slower clock, their settling time is no longer shorter than that of analog counterparts. In a charge pump or an integrator, the ripple voltage is inversely proportional to the switching frequency. For analog DCCs, either a small pump current or a large capacitor is required to keep the ripple small and maintain circuit functionality at low frequency. Since such a large capacitor consumes area, this makes the DCC less attractive. Smaller pump current can be more difficult to be matched if charge injection and clock feedthrough effects are considered, especially when the clock is fast. The DCC also requires a large settling time for high frequency clocks. To solve such problems, we propose an adaptive pump current generator that outputs a small current at low frequency and a larger current at high frequency, and combined with a fast startup circuit and a current-starved pulse width control stage, a low power and area-efficient duty cycle corrector will be introduced here.

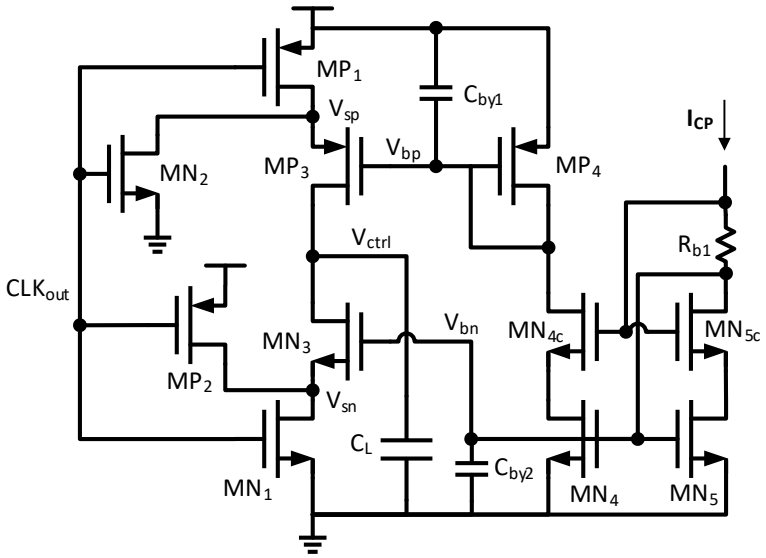

Figure 4.2 Differential charge pump.

The remainder of the chapter is organized as follows. Section 4.3 presents the structure of the proposed duty cycle corrector and explains the working mechanism of each block. Section 4.4 shows the simulation results of this design in a 130nm CMOS process. Finally, conclusions are drawn in Section 4.5.

4.3 Proposed Duty Cycle Corrector

Most analog DCCs use a charge pump to detect duty cycle error and generate a signal for the following control stage, as shown in Figure 4.1. With a differential charge pump [10], [13], an op-amp is no longer needed, so just one load capacitor C_L is required. It forms the dominant pole in the loop and determines the DCC's settling time and the V_{ctrl} ripple magnitude.

4.3.1 Differential Charge Pump with Fast Startup

Figure 4.2 shows the charge pump with power switching used here. CLK_{out} switching now will not couple to V_{ctrl} directly and this leads to smaller V_{ctrl} ripples [10]. A pump current I_{CP} is mirrored to both PMOS and NMOS sides, and the current matching performance directly impacts the duty cycle accuracy. In the locked state, the V_{ctrl} averaged value becomes stable and we have

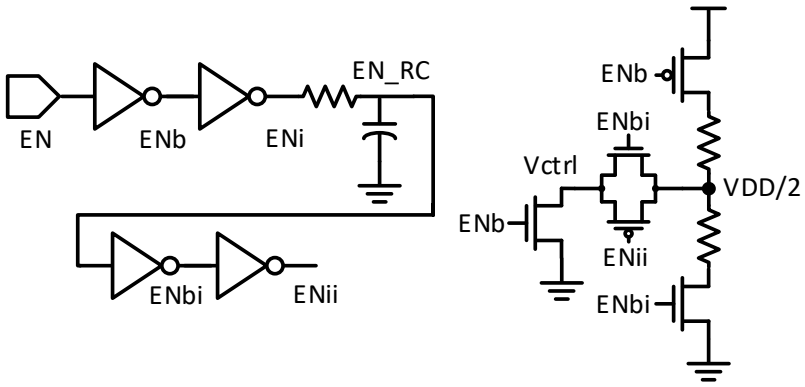

Figure 4.3 Fast startup circuit.

$$D * I_{CP,n} = (1-D) * I_{CP,p}, \tag{4.1}$$

where D is the duty cycle and $I_{CP,p}$ and $I_{CP,n}$ are top and bottom pump currents, respectively. From equation (4.1), the duty cycle is

$$D = \frac{I_{CP,p}}{I_{CP,p} + I_{CP,n}}. \tag{4.2}$$

To achieve a 50% duty cycle, current mirrors must have large sizes to achieve the good matching. Due to large parasitic capacitances C_{gd} and C_{gs}, biasing nodes V_{bp} and V_{bn} can now suffer from clock switching coupling, and when the input clock is fast, V_{bp} and V_{bn} cannot settle quickly. This will add more difference in averaged $I_{CP,p}$ and $I_{CP,n}$ values over time and result in a systematic duty cycle error. C_{by1} and C_{by2} are bypass capacitors used to mitigate coupling effects and stabilize the biasing voltages.

A minimum size transistor MN_2 is used to pull V_{sp} to ground during the discharging phase. Without MN_2, it takes time for V_{sp}, where large parasitic capacitance exists, to be discharged until MP_3 is fully turned off. This is the same reason for adding MP_2 to pull V_{sn} to the supply voltage during the charging phase. To reduce lock-in time, during startup V_{ctrl} is first pulled to $V_{DD}/2$ with the circuit shown in Figure 4.3. The enable signal (EN) of the DCC is delayed by 0.25 μs during

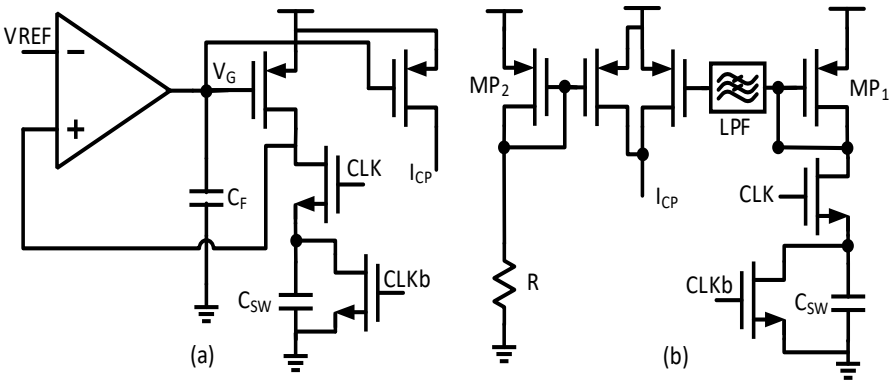

Figure 4.4 (a) Regulated bias, (b) proposed simple adaptive bias.

which the resistor-based divider is turned on to pull V_{ctrl} to $V_{DD}/2$. Small resistors are used here to produce fast settling. However, no static current will be consumed after startup because the divider only conducts during the delay between ENb and ENbi.

The V_{ctrl} ripple voltage is inversely proportional to the switching frequency f and C_L:

$$\Delta Vctrl = \frac{I_{CP}}{C_L * 2f}. \tag{4.3}$$

A large ripple can push MP_3 and MN_3 into their triode regions and result in large duty cycle error or even circuit failure. According to Equation (4.3), for a 4 MHz clock, a combination of 2 μA and 25 pF produces a 10 mV ripple. A large C_L not only consumes large area but also significantly increases the settling time. To solve this problem, the pump current is designed to be adaptive to the input clock frequency, i.e., is a smaller current for slower clocks.

4.3.2 Frequency to Current Converter and Adaptive Biasing

Switched capacitor circuits have been widely implemented in amplifiers, filters, and ADCs. A switched capacitor behaves like a frequency dependent resistor and can be used to generate a current inversely proportional to frequency, as given in Figure 4.4(a). The generated pump current is

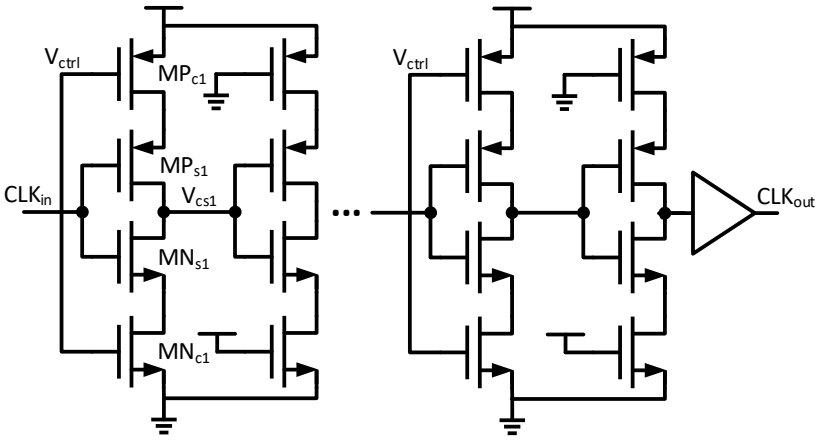

Figure 4.5 Current-starved pulse width control stage.

$$I_{CP} = V_{REF} * f * C_{SW}. \tag{4.4}$$

However, it is not realistic to have a pump current directly proportional to frequency because it may become too large in high frequency cases. Moreover, although the regulation loop can produce an accurate current, a large capacitor is required at the op-amp output to establish low bandwidth for ripple reduction. In the differential charge pump, an accurate I_{CP} is unnecessary because it doesn't affect the most important spec, i.e., the duty cycle. The new adaptive biasing circuit proposed in Figure 4.4(b) uses an square root relationship between current and frequency at min and max frequencies. The current can be calculated as:

$$I_{CP} = \frac{V_{DD} - V_{SG1}}{1/(f \cdot C_{SW})} + \frac{V_{DD} - V_{SG2}}{R} = (V_{DD} - \sqrt{\frac{2I}{K_1}} - V_{TH1}) \cdot fC_{SW} + I_{DC}. \tag{4.5}$$

where $K_1=\mu Cox(W/L)_1$. I_{DC} is a small fixed current chosen as 1 µA in this design. The first term is a frequency dependent term, and two design variables W/L and C_{SW} can be selected to achieve a square root current versus frequency relationship. In this design, C_{SW} has been chosen as 200 fF to make I_{CP} equal to 1.1 µA at 4 MHz and 5.5 µA at 100 MHz. C_L is set to be 7.5 pF, consuming

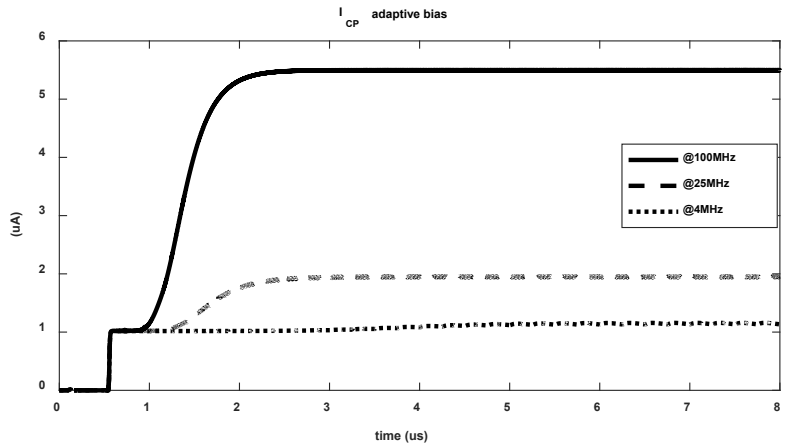

Figure 4.6 I_{CP} adaptive biasing at 4MHz, 25MHz and 100MHz.

a much smaller area than [10]. Since the proposed current generator suffers from process, supply and temperature variations, trimming bits can be added if accurate power consumption is required.

4.3.3 Current-starved Pulse Width Control Stage

Current-starved inverters have been widely used in voltage-controlled oscillators (VCOs), and with a simple modification, using V_{ctrl} as the gate bias of both PMOS and NMOS, such an inverter can adjust the duty cycle. As shown in Figure 4.5, the charge pump output V_{ctrl} determines the current of MP_{c1} and MN_{c1} and controls slew rate at the inverter output V_{cs1}. The adjusted pulse width in each current-starved inverter stage is

$$\Delta PW = t_{rise} - t_{fall} = \frac{C_p \cdot V_{DD}}{I_P} - \frac{C_p \cdot V_{DD}}{I_N}, \tag{4.6}$$

where I_P and I_N are saturation currents of MP_{c1} and MN_{c1} and C_p is the parasitic capacitance at switching node V_{cs1}. Assuming V_{ctrl} equal to V_{c0}, we have $I_P = I_N$ and the clock pulse width will not be adjusted. Applying small signal model, V_{ctrl} can be rewritten as $V_{c0}+\Delta V_c$ and the current difference can be calculated as:

$$I_N - I_P = gm_N \cdot \Delta V_c - gm_P \cdot (-\Delta V_c) = (gm_N + gm_P) \cdot \Delta V_c . \tag{4.7}$$

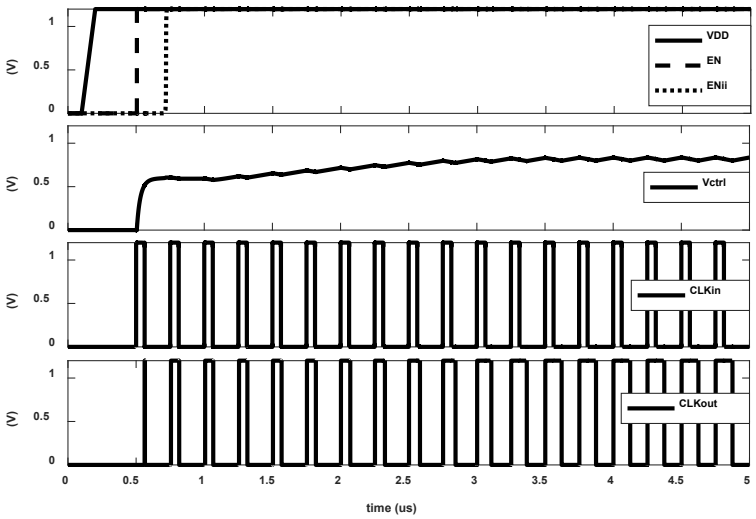

Figure 4.7 Duty cycle corrector timing with 4 MHz input.

The gain of a current-starved inverter is defined as:

$$K_{CS} = \frac{\Delta PW}{\Delta V_c} = \frac{(gm_N + gm_P) \cdot C_p \cdot V_{DD}}{I_P \cdot I_N} \tag{4.8}$$

The larger C_p contributes to higher gain but consumes more switching power. Using smaller W/L current device ratios can also help with the gain but may limit high frequency clock slewing. Moreover, the control stage is quite nonlinear and has much larger gain when V_{ctrl} is at two edges, i.e., supply and ground. The input duty cycle error range is dependent on the number of control stages and the size of current devices. Here three stages have been cascaded to achieve a 4-100 MHz frequency range. The overall system is a single pole system and the dominant pole is less than 1/10 of the input frequency, guaranteeing stable operation.

4.4 Simulation Results

The proposed duty cycle corrector was verified using a 130nm CMOS process. The adaptive pump current at different frequencies is given in Figure 4.6. As frequency is varied from 4 to 100 MHz, the pump current increases from 1.1 µA to 5.5 uA. The current ripple at 100 MHz is less than 1% after incorporation of a 2nd order low pass filter.

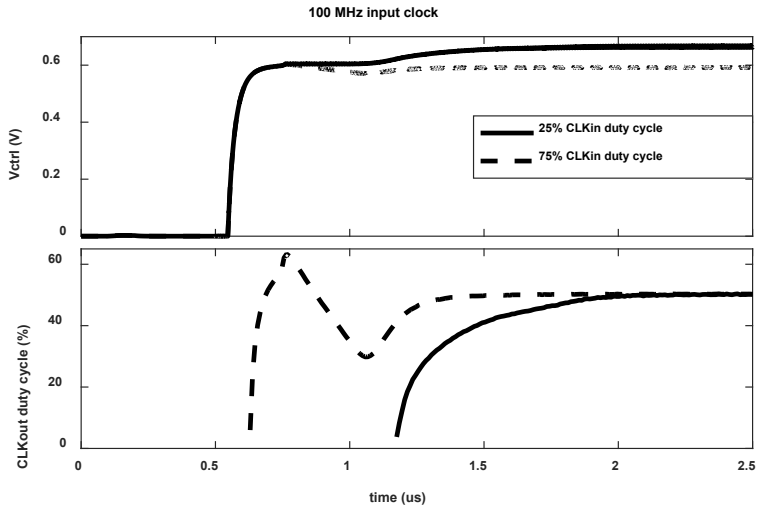

Figure 4.8 Duty cycle corrector timing with 100 MHz input.

The circuit's working mechanism is explained in Figure 4.7. During the time interval between EN and ENii, V_{ctrl} is first charged to $V_{DD}/2$ for quick startup, after which the unbalanced pulse width charges up V_{ctrl}, in turn increasing the CLK_{out} pulse width to 50 %. The lock-in time for a 4 MHz clock is ~ 3 µs, as shown in Figure 4.7. V_{ctrl} and the output duty cycle settling for the 100 MHz case are given in Figure 4.8. Because of the increased pump current, the lock-in time is reduced to 1.4 µs (EN starts at 0.5 µs).

The increased pump current at high frequency mitigates clock switching coupling effects and helps with the matching of $I_{CP,n}$ and $I_{CP,p}$. The 100 MHz input clocks with 25% and 75% duty cycles are sent into the duty cycle corrector. As given in Figure 4.9, Monte Carlo simulations reflect an output duty cycle error less than 1.5% in both cases.

4.5 Conclusion

This chapter presents a low power and area-efficient solution for correcting duty cycle errors in ADCs' external clocks. A differential charge pump with fast startup is used to detect the duty cycle error, and current-starved inverters are implemented to adjust the pulse width. The pump current is designed to be adaptive to the input frequency, so a wide frequency range can be

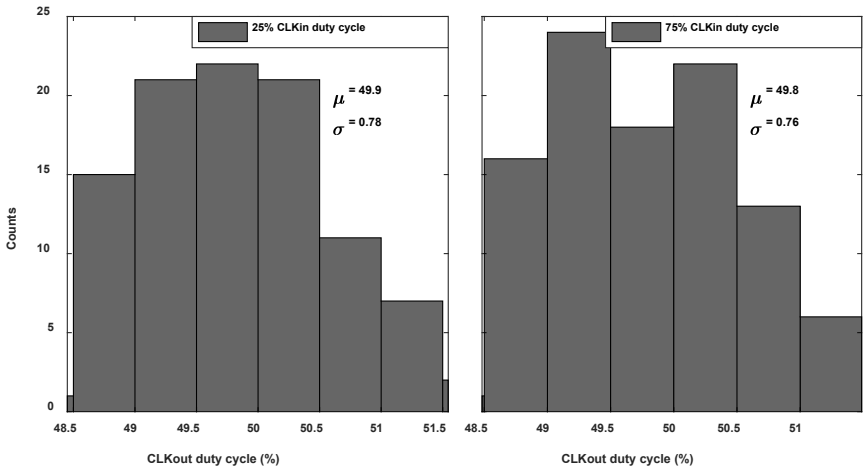

Figure 4.9 100 Monte Carlo runs with 100 MHz input clock.

achieved with minimal added hardware. The circuit, verified using a 130nm CMOS process, achieves comparable results with less power and area consumption.

4.6 References

[1] X. Wang, J. Keane, P. Jain, V. Reddy, and C. H. Kim, "Duty-cycle shift under asymmetric BTI aging: A simple characterization method and its application to SRAM timing," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, 2013, pp. 4A.5.1-4A.5.5.

[2] I. Raja, G. Banerjee, M. A. Zeidan, and J. A. Abraham, "A 0.1–3.5-GHz Duty-Cycle Measurement and Correction Technique in 130-nm CMOS," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 24, no. 5, pp. 1975–1983, May 2016.

[3] T. Miki, T. Morie, T. Ozeki, and S. Dosho, "An 11-b 300-MS/s Double-Sampling Pipelined ADC With On-Chip Digital Calibration for Memory Effects," *IEEE J. Solid-State Circuits, vol. 47*, no. 11, pp. 2773–2782, Nov. 2012.

[4] C.-H. Kuo, D.-Y. Shi, and K.-S. Chang, "A Low-Voltage Fourth-Order Cascade Delta–Sigma Modulator in 0.18-µm CMOS," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 57, no. 9, pp. 2450–2461, Sep. 2010.

[5] Fenghao Mu and C. Svensson, "Pulsewidth control loop in high-speed CMOS clock buffers," *IEEE J. Solid-State Circuits*, vol. 35, no. 2, pp. 134–141, Feb. 2000.

[6] K. Cheng, C. Su, and K. Chang, "A High Linearity, Fast-Locking Pulsewidth Control Loop with Digitally Programmable Duty Cycle Correction for Wide Range Operation," *IEEE J. Solid-State Circuits*, vol. 43, no. 2, pp. 399–413, Feb. 2008.

[7] Z. Zhu, M. Liu, J. Wang, and Y. Yang, "A Fast-Locking, Low-Jitter Pulsewidth Control Loop for High-Speed ADC," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, pp. 1–1, 2018.

[8] Poki Chen, Shi-Wei Chen, and Juan-Shan Lai, "A low power wide range duty cycle corrector based on pulse shrinking/stretching mechanism," in *2007 IEEE Asian Solid-State Circuits Conference*, 2007, pp. 460–463.

[9] Y. Qiu, Y. Zeng, and F. Zhang, "1-5 GHz duty-cycle corrector circuit with wide correction range and high precision," *Electron. Lett.*, vol. 50, no. 11, pp. 792–794, May 2014.

[10] Hong-Yi Huang, Chia-Ming Liang, and Shi-Jia Sun, "Low-power 50% duty cycle corrector," in *2008 IEEE International Symposium on Circuits and Systems*, 2008, pp. 2362–2365.

[11] J. Gu, J. Wu, D. Gu, M. Zhang, and L. Shi, "All-Digital Wide Range Precharge Logic 50% Duty Cycle Corrector," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 20, no. 4, pp. 760–764, Apr. 2012.

[12] C. Jeong, A. Abdullah, Y. Min, I. Hwang, and S. Kim, "All-Digital Duty-Cycle Corrector with a Wide Duty Correction Range for DRAM Applications," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 24, no. 1, pp. 363–367, Jan. 2016.

[13] S. Sofer, V. Neiman, and E. Melamed-Cohen, "Synchronous duty cycle correction circuit," in *2010 18th IEEE/IFIP International Conference on VLSI and System-on-Chip*, 2010, pp. 96–100.

CHAPTER 5. SUB-PPM/°C VOLTAGE REFERENCE WITH NATURAL BASE EXPANSION FOR CURVATURE CANCELLATION

Nanqi Liu and Degang Chen

Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010

Modified from a manuscript to be submitted to IEEE Transactions on Circuits and Systems I: Regular Papers

5.1 Abstract

This chapter presents a general approach based on natural base expansion to cancel the curvature in the base emitter voltage of a bipolar transistor. The proposed method can be implemented on any of the widely used Banba, Kuijk, or Brokaw structures to build references with sub-ppm/°C temperature coefficient. One embodiment based on Kuijk structure is used to illustrate how the cancellation can be realized. Error sources that can potentially affect the reference's performance are analyzed, and their effects on the temperature coefficient are minimized after trimming. Transistor level simulations show that the design can achieve a 0.25 ppm/°C temperature coefficient over a temperature range of -40 °C to 125 °C. A prototype realized both with on-chip devices in UMC 65nm process and off-chip components has been designed and fabricated. Measurement results reflect a 1.5 ppm/°C temperature coefficient using the proposed two temperature trimming method.

5.2 Introduction

Voltage references are required in most analog and mixed-signal circuits either as parts of internal biases or as targets of operations. Since electrical systems can be expected to operate under a wide variety of physical conditions, constancy of their references is of critical importance because it limits the maximum achievable absolute accuracy. The increasing demand for modern

high-performance circuits such as voltage regulators and high-resolution data converters poses a need for high precision, low temperature coefficient (TC) voltage references.

The primary method for generating a reference in integrated circuit applications is to use diodes. Over a 10-year interval (1964-1974) Hilbiber, Widlar, Kuijk, and Brokaw developed voltage reference structures that generate a voltage signal at the circuit output that is very close to the silicon bandgap voltage and has small temperature coefficients. This class of voltage references or others with similar properties are called bandgap references (BGRs). Over the four decades after that work, a standard solution has been to design circuits that can perform a linear combination of the base-emitter voltage V_{BE} that has a negative temperature coefficient and a proportional-to-absolute-temperature (PTAT) voltage. Unfortunately, V_{BE} has a well-known nonlinear term TlnT that results in an approximate 10 ppm/°C temperature drift at the output voltage over a 150 °C range [1], and for high precision BGRs, this nonlinear curvature must to be canceled which has been challenging to achieve.

Some curvature correction approaches [2] – [9] have been proposed to reduce or cancel this nonlinear temperature drift. A second-order Taylor approximation was implemented in [2] to compensate for V_{BE} curvature, and it achieved a typical TC of 1 ppm/°C. Piecewise-linear curvature compensation was investigated in [3], [4], and [5]. In [4], an extra logarithmic curvature term was added to the conventional exponential curvature to further reduce TC at higher temperatures. In [5], A five-piece linear compensation was applied, yielding measured results with an average TC of 3 ppm/°C. Some researchers cancel TlnT nonlinearity by generating a curve with temperature drift characteristics opposite to that of bipolar transistors. For example, [6] and [7] utilized MOSFETs operating in weak inversion and [8] utilized a current mirror, to create a curvature-up bandgap reference, then added it to a conventional curvature-down reference.

Because the TlnT term is not completely canceled in these works, there remain some levels of temperature dependency, and the achieved TC is so far limited to single-digit ppm/°C.

To cancel nonlinear curvature, [9] proposed a method for generating a current with TlnT dependence and used it to compensate for high order residues. However, the analysis assumed existence of a temperature independent current not fully achieved in their proposed structure.

In this chapter, we present a method for cancelling TlnT curvature based on natural base expansion, and current mode and voltage mode embodiments using three diodes and a self-bootstrapping concept will be provided. Error sources in the proposed structure are analyzed and a trimming approach for minimizing their effects on the temperature drift is introduced. Since op-amp offset and noise are critical to the reference's accuracy [1], chopping with notch filtering is typically used to reduce the offset, 1/f noise, and associated switching noise. We decided to liberate ourselves from the challenging job of designing precision op-amps, and a prototype realized with both on-chip devices and off-chip op-amps was designed, fabricated, and measured. The purpose here is to demonstrate the concept of TlnT curvature cancellation and V_{GOr} extraction for building sub-ppm/°C bandgap references.

The remainder of the section is organized as follows. Section 5.3 introduces the theory of natural base expansion and presents the proposed V_{GOr} extraction method and its analytical constraints. Section 5.4 describes a constant current reference based on Banba structure and a constant voltage reference based on Kuijk structure. Error sources are analyzed, and a novel two-temperature trimming approach is introduced to deal with them. Section 5.5 describes implementations, test setup, and measurement results. Finally, conclusions are drawn in Section 5.6.

5.3 Natural Base Expansion and V_{GOr} Extraction

Most precision references utilize diode-connected bipolar transistors (BJTs) rather than MOSFETs working in weak inversion to extract the silicon bandgap voltage [10] because the base-emitter voltage V_{BE} of a BJT is better characterized over temperature and varies less than +-2% over the process [11]. Part of the theory of the proposed method was first introduced in [12]. This section will first briefly review the V_{BE} temperature characteristic and explain natural base expansion, after which a general method using three diodes to cancel the curvature will be proposed.

5.3.1 Temperature Characteristic of V_{BE}

The I-V relationship for a diode-connected BJT can be expressed as

$$I_C(T) = I_S(T)e^{\frac{qV_{BE}}{kT}} = J_{SX}AT^{\eta}e^{-\frac{qV_G(T)}{kT}}e^{\frac{qV_{BE}}{kT}}, \tag{5.1}$$

where J_{SX} and η are process-dependent model parameters, A is the junction area, T is the absolute temperature in Kelvin, k and q are physical constants, and $V_G(T)$ is the bandgap voltage at temperature T. $V_G(T)$ does have small temperature dependence typically modeled as

$$V_G(T) = V_{GO} - \frac{\alpha T^2}{T+\beta} \tag{5.2}$$

in Spice models. Two sets of values have been found for V_{GO}, α and β. In the process used here, we have α = 7.02e-4 V/K, β = 1108 K, and V_{GO} = 1.16 V.

Based on Equation (5.1), Tsividis in [11] proposed an analytical expression for predicting V_{BE} temperature characteristics:

$$V_{BE}(T) = V_G(T) - \frac{T}{T_r}V_G(T_r) + \frac{T}{T_r}V_{BE}(T_r) - \eta\frac{kT}{q}\ln\left(\frac{T}{T_r}\right) + \frac{kT}{q}\ln\left[\frac{I_C(T)}{I_C(T_r)}\right], \tag{5.3}$$

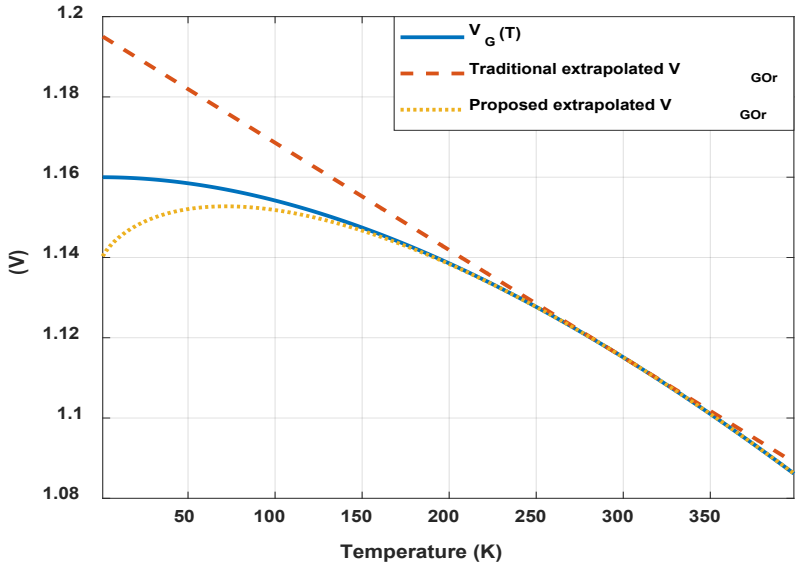

Figure 5.1 V_{GOr} extrapolation.

where T_r is the reference temperature. In [11], $V_G(T)$ is considered to be linear with temperature in the range interested:

$$V_G(T) = V_{GOr1} + a_1 T\,, \tag{5.4}$$

where V_{GOr1} is the extrapolated bandgap voltage at 0 K with first order fitting at T_r [13]. Strictly speaking, the value of V_{GOr1} is dependent on the reference temperature T_r. If the collector current is proportional to T^{δ}, equation (5.3) can be simplified to

$$V_{BE}(T) = V_{GOr1} - [V_{GOr1} - V_{BE}(T_r)]\frac{T}{T_r} - (\eta - \delta)\frac{kT}{q}\ln\left(\frac{T}{T_r}\right). \tag{5.5}$$

Equation (5.5) indicates that V_{BE} consists of a constant term, a CTAT term and a TlnT term and first-order bandgap references only compensate the CTAT part. The assumption of linear dependence for $V_G(T)$ is not accurate enough for sub-ppm/°C references. Since TlnT is the natural nonlinear term in the diode, it can be used as one base equation when modeling the $V_G(T)$ over the industry's standard temperature range of -40 °C to 125 °C. Using natural base expansion, $V_G(T)$ can be fitted as

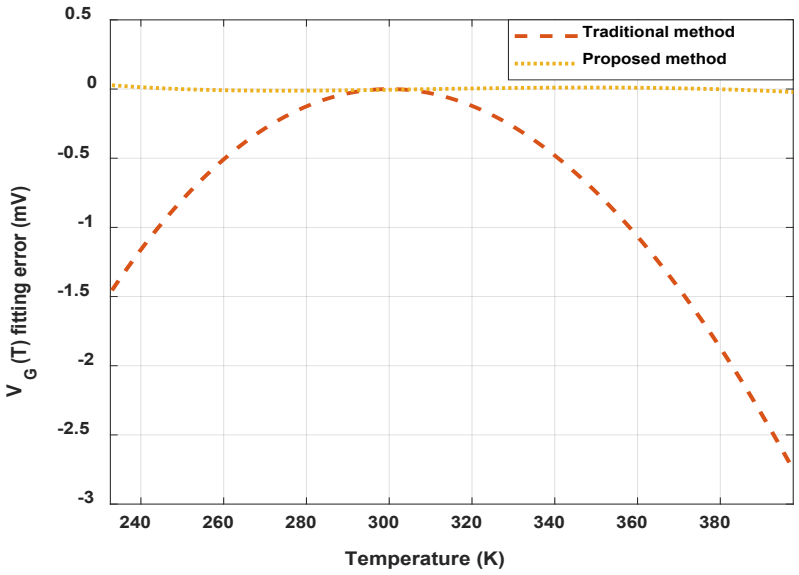

Figure 5.2 $V_G(T)$ fitting error in the interested temperature range.

$$V_G(T) = V_{GOr} + a_1 \frac{T}{T_r} + b_1 T \ln(\frac{T}{T_r}) \tag{5.6}$$

The $V_G(T)$ and two fitted curves, one using the tangent line at 300 K and one using our proposed natural base expansion, are shown in Figure 5.1, and residue fitting errors of the two methods are shown in Figure 5.2. With natural base expansion, the residue error is only 60 μV, corresponding to 0.3 ppm/°C over the 165°C temperature range. Substituting equation (5.6) into (5.3) while assuming the collector current proportional to T^{δ}, V_{BE} is expanded as

$$V_{BE}(T) = V_{GOr} - [V_{GOr} - V_{BE}(T_r)]\frac{T}{T_r} - [\eta - \delta - b_1 \frac{q}{k}]\frac{kT}{q} \ln\left(\frac{T}{T_r}\right). \tag{5.7}$$

Compared to equation (5.5), the nonlinearity in $V_G(T)$ results in a different gain coefficient for TlnT term and the constant voltage needs to extracted becomes V_{GOr}. The following section will show how to extract the new V_{GOr} from V_{BE} using three diodes.

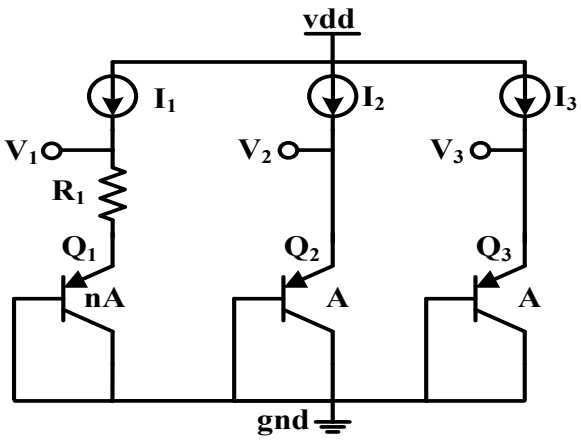

Figure 5.3 Illustration of using three BJT connected as diodes to extract the V_{GOr}.

5.3.2 V_{GOr} Extraction

PNP substrate BJT is available in most modern CMOS processes and will be used in the following analysis with the assumption that collector current is equal to emitter current. This assumption won't affect the correctness of the following derivations as long as current gain β is constant versus temperature. Errors resulted from the assumption of β and r_b will be addressed with the trimming approach described later. The V_{GOr} extraction method requires three diode-connected BJTs Q_1, Q_2, and Q_3 with an area ratio of n:1:1, as shown in Figure 5.3. Q_1 and Q_2 should have either the same or a ratioed current. Assuming that I_1=I_2 and an operational amplifier (op amp) keeps the voltages V_1 and V_2 equal, the V_{BE} difference between Q_1 and Q_2 can be calculated based on (5.3), which is a PTAT voltage designated as V_{PTAT}:

$$V_{PTAT} = \Delta V_{BE2,1} = V_T \ln(n) = \frac{kT}{q}\ln(n), \tag{5.8}$$

where V_T is the thermal voltage. The high accuracy of this PTAT voltage [14] allows us to obtain a good PTAT current if R_1 has a low temperature coefficient like SiCr thin film resistors, for example, that have been reported to have sub-ppm level TC [15]. V_{BE} of Q_2 is then:

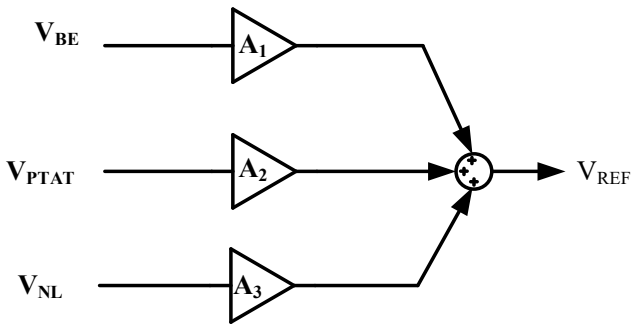

Figure 5.4 Weighted summation to extract the V_{GOr}.

$$V_{BE2}(T) = V_{GOr} - [V_{GOr} - V_{BE2}(T_r)]\frac{T}{T_r} - [\eta - 1 - b_1\frac{q}{k}]\frac{kT}{q}\ln\left(\frac{T}{T_r}\right). \tag{5.9}$$

The PTAT term in V_{BE2} can be cancelled utilizing the PTAT current across another resistor with low TC. To compensate the TlnT term, Q_3's current is forced to have the following property:

$$I_2 / I_3 \sim T^{\alpha}, \tag{5.10}$$

where $\alpha \neq 0$. I_3 needs to have a different temperature dependence from I_2, and its generation will be discussed later in the chapter. The V_{BE} difference between Q_2 and Q_3 is a superposition of a PTAT term and a TlnT term and it is designated as V_{NL}:

$$V_{NL} = \Delta V_{BE2,3} = \left(V_{BE2}(T_r) - V_{BE3}(T_r)\right)\frac{T}{T_r} + \alpha\frac{kT}{q}\ln(\frac{T}{T_r}). \tag{5.11}$$

Note here that $V_{BE2}(T_r)$ - $V_{BE3}(T_r)$ would not equal to zero if process variations and mismatches are considered, and this should not be ignored as [9] when deriving equations for V_{GOr} extraction. V_{PTAT} and V_{NL} can be combined linearly with V_{BE2} in either the voltage domain or the current domain to extract the V_{GOr}, as shown in Figure 5.4. To extract V_{GOr}, we would like

$$V_{REF} = A_1 V_{BE2} + A_2 V_{PTAT} + A_3 V_{NL} = V_{GOr}. \tag{5.12}$$

Solving (5.12), the weighted gains can be calculated:

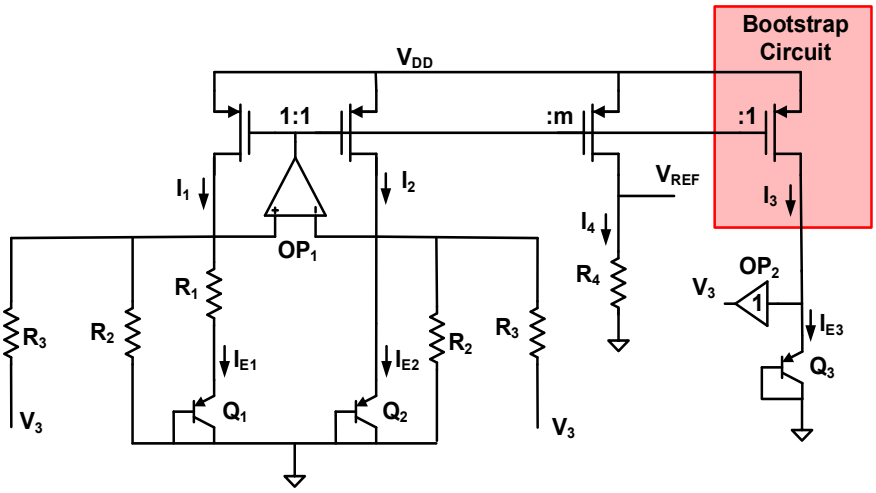

Figure 5.5 Constant current reference based on Banba structure.

$$A_1 = 1$$

$$A_2 = \frac{q}{kT_r \ln(n)}\left(-\frac{\left(\eta - 1 - b_1 \frac{q}{k}\right)}{\alpha} * [V_{BE2}(T_r) - V_{BE3}(T_r)] + [V_{GOr} - V_{BE2}(T_r)] \right). \qquad (5.13)$$

$$A_3 = \frac{\left(\eta - 1 - b_1 \frac{q}{k}\right)}{\alpha}$$

In this approach, the accuracy with which I_3 follows (5.10) will determine the voltage reference's performance. Some architectures [16], [17], utilized a first-order temperature independent current (V_{BG}/R) or a complementary-to-absolute-temperature (CTAT) current (V_{BE}/R) as I_3, but this is not sufficiently accurate to achieve sub-ppm/°C TC. The easiest way to fulfill (5.10) is to have $\alpha = 1$ which would require I_3 to be temperature independent. The self-bootstrapping concept introduced here leads to the generation of such a current and will be illustrated in the following embodiments.

5.4 Bandgap References Design

This section will describe two bandgap references using three diodes, natural base expansion, and a self-bootstrapping structure to extract V_{GOr}.

5.4.1 Constant Current Reference

A current model implementation of the proposed method based on Banba structure is given in Figure 5.5. While the basic structure of this implementation looks similar to that of [16], more constraints are necessary for the purpose of extracting V_{GOr}. In Figure 5.5, currents I_1 - I_3 are the same and all equal to

$$I_{1-3} = V_{BE2}\frac{1}{R_2} + \Delta V_{BE2,1}\frac{1}{R_1} + \Delta V_{BE2,3}\frac{1}{R_3} = \frac{1}{R_2}\left(V_{BE2} + \frac{R_2}{R_1}V_{PTAT} + \frac{R_2}{R_3}V_{NL}\right). \qquad (5.14)$$

The buffer at the emitter of Q_3 ensures that $I_{E3} = I_3$. When the resistor ratios are properly adjusted in accordance with Equation (5.13), I_3 will have little temperature dependence and equals to

$$I_{1-3} = \frac{1}{R_2}V_{GOr}, \qquad (5.15)$$

fulfilling the requirement of Equation (5.10). The buffer OP_2 must have small offset and low drift because it can directly add error to V_{NL}, the voltage used to cancel the curvature. I_4 is mirrored from I_{1-3} with a ratio of m and the voltage output, typically easier to measure during the trimming process, is

$$V_{REF} = R_4 I_4 = \frac{mR_4}{R_2}V_{GOr}. \qquad (5.16)$$

The actual TC of the circuit can be affected by non-ideal factors such as resistor, current mirror, and BJT mismatches, all of which must be trimmed out. Voltage measurements are generally easier to achieve than current measurements. In this current mode reference, the ideal trim target value of V_{GOr} is multiplied by the unknown mR_4/R_2 after fabrication when it appears at the V_{REF}. To trim gain coefficients of V_{PTAT} and V_{NL} for minimum temperature drift, three temperature measurements and a two-dimensional sweep of any two resistors in R_1, R_2, and R_3 are required. While such long trimming time can increase the overall cost of the circuit, in the voltage mode

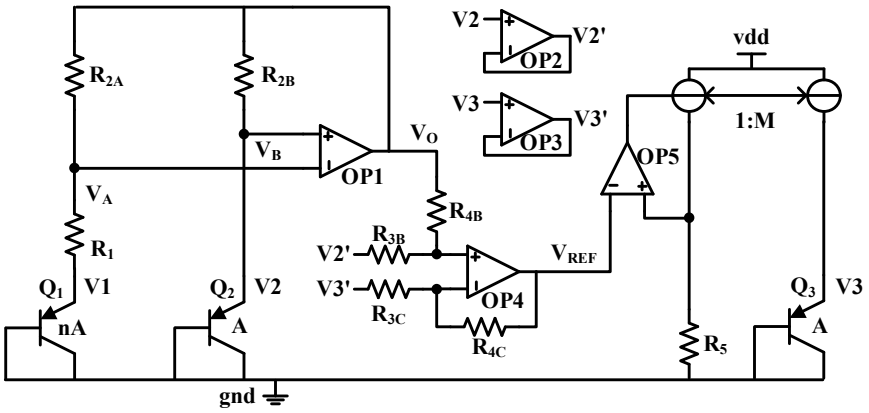

Figure 5.6 Constant voltage reference based on Kuijk structure.

bandgap reference implementation described in the next section, a two temperature trimming is sufficient.

5.4.2 Constant Voltage Reference

A voltage mode implementation of the proposed method based on Kuijk structure [18] is shown in Figure 5.6. Kuijk's output V_O already contains the first two terms of Equation (5.12). What we need here is to generate and add the third term to V_O.

A differential op-amp OP_4 is used to subtract V_{BE2} from V_{BE3} and add this V_{NL} to V_O to generate the reference voltage V_{REF}. V_{REF} is then regulated over a resistor R_5 to generate a current with different temperature dependence from Kuijk's core current. It is then mirrored to Q_3 to satisfy Equation (5.10). Here the resistors' temperature coefficient is canceled out as

$$\frac{I_2}{I_3} = \frac{V_{PTAT}}{R_1} / \frac{V_{REF}}{R_5} = \frac{V_{PTAT}}{V_{REF}} \frac{R_5}{R_1}. \tag{5.17}$$

In the proposed method, the resistor's TC appears only in the V_{BE} term in Equation (5.12), as will be discussed in the next subsection. Buffers are inserted between V_2' and V_2, V_3' and V_3 to guarantee that Q_3's current I_3 is equal to $M*V_{REF}/R_5$, a necessary condition to make (5.11) valid. In this structure, the new reference output V_{REF} is used to generate the current I_3 and produce V_{BE3}

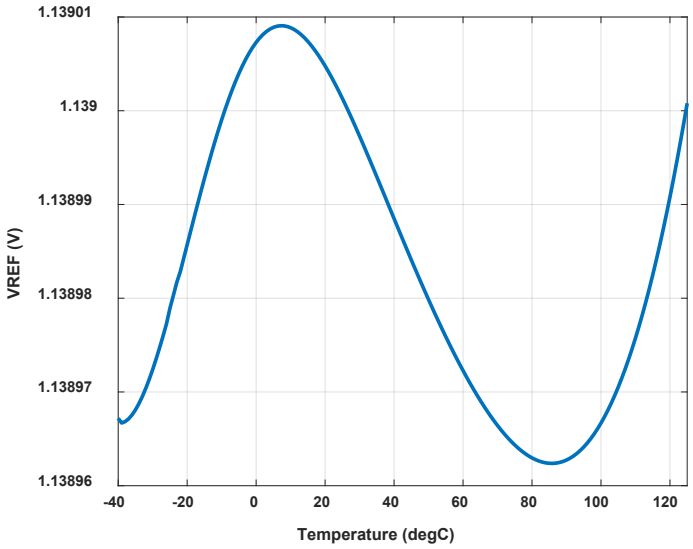

Figure 5.7 V_{REF} drift at typical corner.

and V_{NL} which is further feedback to V_{REF}. The self-bootstrapping concept here makes this V_{GOr} extraction method qualitatively superior to existing voltage reference generators.

Careful sizing and layout techniques, i.e., common centroid or interdigitation, should be used for creating matching resistors, i.e., R_{iA} and R_{iB}. The voltage reference's output is

$$V_{REF} = V_{BE2} + \frac{R_2}{R_1}(V_{BE2} - V_{BE1}) + \frac{R_4}{R_3}(V_{BE2} - V_{BE3}). \tag{5.18}$$

R_5 can be selected so that $V_{BE3}(T_r) = V_{BE2}(T_r)$, simplifying the equation and making the following trimming solution more straightforward. Compared to (5.13) with $\alpha = 1$, the resistor ratios can be calculated to make $V_{REF}=V_{GOr}$:

$$\begin{aligned} \frac{R_2}{R_1} &= -\frac{q}{kT_r \ln(n)}[V_{BE2}(T_r) - V_{GOr}] \\ \frac{R_4}{R_3} &= \eta - 1 - b_1 \frac{q}{k} \end{aligned}. \tag{5.19}$$

The temperature drift of the bandgap reference at the normal corner given in Figure 5.7 no longer has the typical quadratic shape of first-order bandgap references. The drift here comes from the

$V_G(T)$'s residue drift after excluding T and TlnT terms, as shown in Figure 5.2, and corresponds to 0.25 ppm/°C TC. The actual performance of the circuit after fabrication can be affected by many non-ideal factors discussed in the following section.

5.4.3 Error Sources and Trimming Approach

Possible error sources that could degrade the performance of voltage references mainly include process variations and mismatch of devices, temperature coefficient of resistors (TCR), base resistance spread, temperature dependence of limited current gain β, and op-amp offsets. V_{BE} spread or mismatch between devices result mainly in PTAT errors, and a single room temperature trim of R_2/R_1 would be able to compensate for such errors [1]. As proven earlier, TCR only appears in the V_{BE} term and not in V_{PTAT} and V_{NL}. The $I_{C2}(T)$ is no longer simply proportional to T in Kelvin:

$$\frac{I_{C2}(T)}{I_{C2}(T_r)} = \frac{T}{1+(T-T_r)*TCR}. \tag{5.20}$$

Substituting Equation (5.20) into (5.3) while applying Equation (5.6), V_{BE2} becomes

$$V_{BE2}(T) = V_{GOr} + \left[V_{BE2}(T_r) - V_{GOr}\right]\frac{T}{T_r} - [\eta - 1 - b_1\frac{q}{k}]\frac{kT}{q}\ln\left(\frac{T}{T_r}\right) - \frac{kT}{q}\ln[1+(T-T_r)*TCR]. \tag{5.21}$$

Applying the approximation $\ln(1+x) \approx x$ when $|x| \ll 1$ in equation (5.21), the TCR induced error can be expressed as

$$e_{TCR} = -\frac{kT}{q}\ln[1+(T-T_r)*TCR] \approx -\frac{kT}{q}(T-T_r)*TCR, \tag{5.22}$$

a 2nd order error. Other error sources, i.e., temperature dependence of base resistance rb, β, and offsets, are also not just PTAT errors. Existing trimming and chopping techniques have been used to reduce op-amp offset to the μV level with even less drift [1], so offset is not the focus here. Substrate PNP has very limited β (<10) compared to NPN and its base resistance can be large if

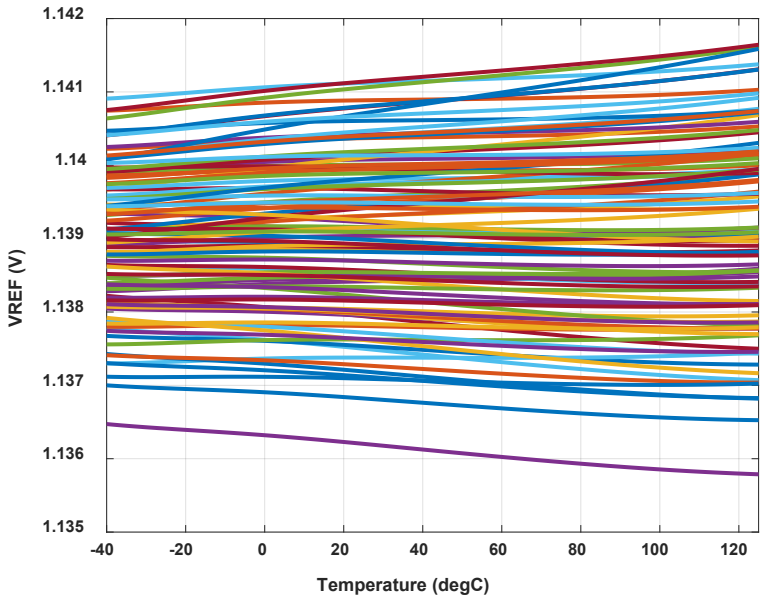

Figure 5.8 100 Monte Carlo runs of V_{REF}.

the layout is not optimized. Considering these effects, (5.8) becomes

$$\Delta V_{BE2,1} = V_T \ln(n) + \frac{I_1 r_{b1}}{\beta_1} - \frac{I_2 r_{b2}}{\beta_2}, \tag{5.23}$$

no longer a pure PTAT voltage. Moreover, the assumption of $V_G(T)$ changing linearly with T is not true. All these small error sources will propagate to the voltage reference's output following a transfer function. By again using the new natural base expansion coming from the natural equation of V_{BE}, each small error's effect on V_{REF} can be decomposed into constant, T, and TlnT bases:

$$e(T) = e(T_r) + aT + bT\ln(T/T_r) + \varepsilon(T), \tag{5.24}$$

where ε is the residue. Considering all error sources' effect, there exist R_2/R_1 and R_4/R_3 ratios resulting in lumped T and TlnT error terms at V_{REF} being canceled so that only high-order residue errors remain. R_2, R_4 and R_5 are digitally trimmable with 7-bit resolution over a +-2% range around their nominal values.

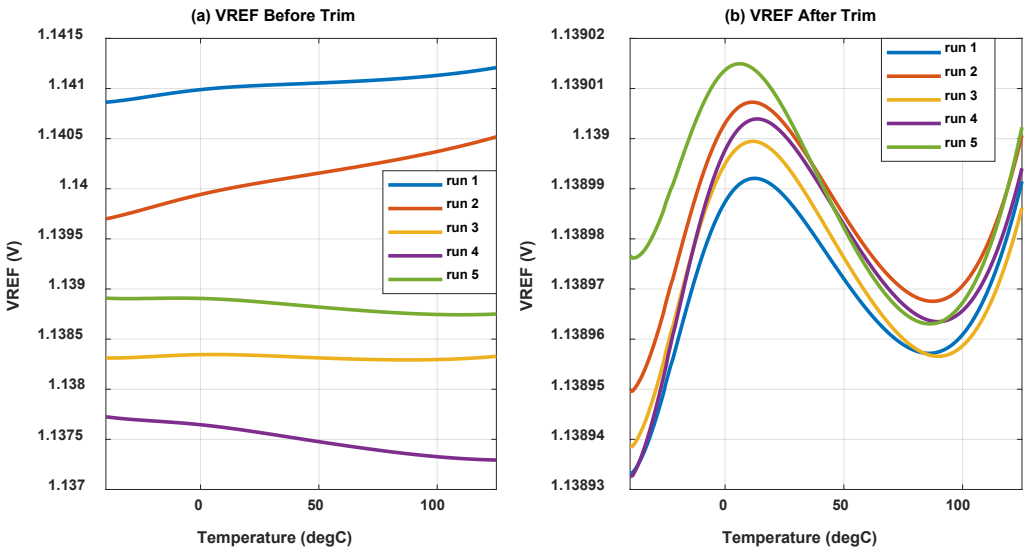

Figure 5.9 Randomly selected 5 runs: (a) V_{REF} before trimming, (b) V_{REF} after trimming.

A two-temperature trimming method based on the new base expansion approach is implemented here. First, at room temperature (Tr), R_5 is tuned to make $V_{BE3}(Tr) = V_{BE2}(Tr)$ so that $V_{NL} = -kT/q*\ln(T/T_r)$ and is 0 V at Tr. An analog test bus or an on-chip comparator may be used to measure or compare V_{BE3} and V_{BE2}. While still at room temperature, R_{2B}, R_{2C} are swept across their shared 7-bit trimming codes until $V_{REF} = V_{GOr} = 1.139$ V with R_{4B} and R_{4C} set to their normal values. While I_1 and I_2 will be changed slightly here, we assume that $V_{BE3}(Tr) = V_{BE2}(Tr)$ still holds because the I-V characteristic of a diode is exponential. After R_{2B} and R_{2C} are trimmed, the thermal environment will be changed to a hot temperature, e.g., 125 °C, and R_{4B} and R_{4C} are swept and, similarly, the right trimming code is chosen when V_{REF} is equal to V_{GOr}. V_{REF} at Tr is not changed here because V_{NL} is trimmed to be 0 V at T_r. In this trimming method, only two temperatures, room temperature and a chosen hot temperature, are needed which saves the cost of maintaining thermally regulated testing environments compared to methods requiring three temperatures. Since only a one-dimensional sweep of a single resistor is needed, sweeping time

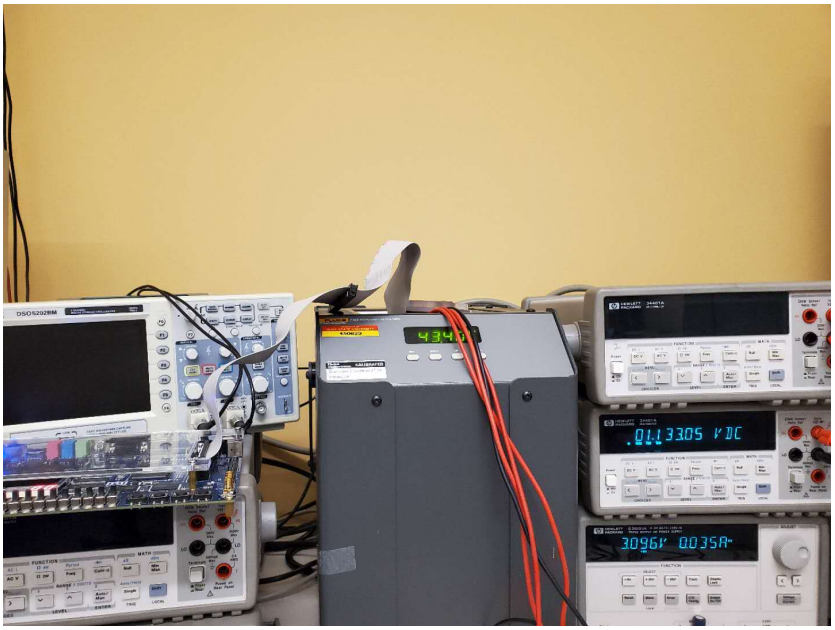

Figure 5.10 Lab setup.

can also be saved. Total measurement and trimming time may still be long, but that's the price for pursuing sub-ppm performance.

5.4.4 Simulation Results

The proposed voltage mode V_{GOr} extraction with two-temperature trimming was verified using a UMC 65nm process. To save the effort of implementing chopping in op-amps and filtering of switching noise, spice models of ADA4528 were used for simulation.

100 Monte Carlo runs are performed, and the temperature drift results are given in Figure 5.8. Before trimming, this reference circuit can achieve 10 ppm/°C typical temperature drift under process variations. Five runs are selected randomly to verify the trimming method proposed in Section 4.3.3. As shown in Figure 5.9, temperature coefficients of 5 runs are reduced from 8 ppm/°C or more to less than 0.4 ppm/°C. The post-trimmed curves all have similar shapes and the differences result from limited trimming resolution and $V_{BE2}(T_r)$ change after trimming of R_{2B} and R_{2C}. Iterations can be performed to make $V_{BE3}(T_r) = V_{BE2}(T_r)$ more accurate. Although the offset

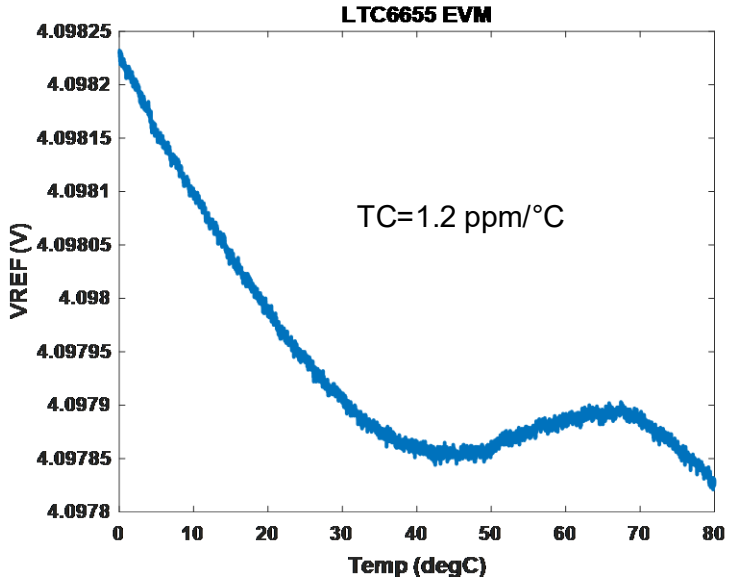

Figure 5.11 TC characterization data of LTC6655 EVM with the lab setup.

drift of op-amps is not included in the simulation, the proposed structure has been proven to be able to achieve sub-ppm/°C TC with the proposed two-temperature trimming solution.

5.5 Implementation and Measurements

This section will first describe the test setup used to characterize the references' temperature drift, followed by explanation of the implementations and measurement results for the two references.

5.5.1 Test Setup

Reference output measurements are made by immersing the test circuit in a Fluke Microbath Thermometer Calibrator to accurately control and monitor the temperature. The temperature characteristics can be measured in one of two ways. The first method is based on a ramp and soak approach whereby measurements are made at a fixed set of temperatures after the dielectric liquid had reached equilibrium (i.e., in the soak mode) at a target test temperature. After taking a series of measurements, the temperature is then ramped up to the next target test temperature and the next set of measurements are taken near the end of the soak mode. This ramp

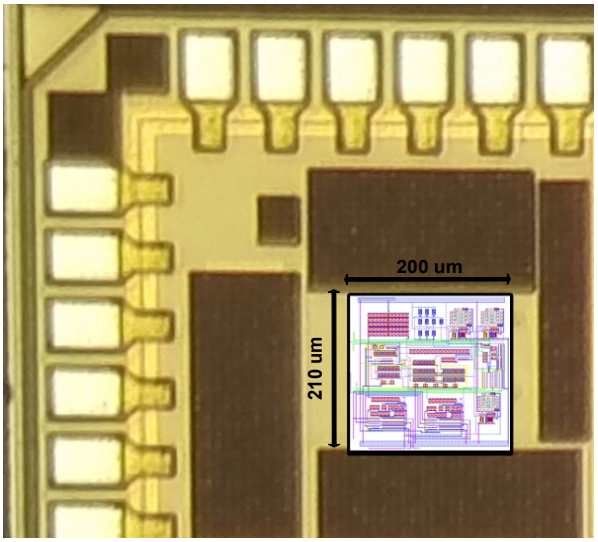

Figure 5.12 Die micrograph overlaid by the layout of the current mode BGR.

and soak approach is repeated until measurements at all predetermined temperatures have been obtained. The total time required to take a series of measurements using this approach can become quite long.

The second method is based on establishing a slow continuous temperature ramp with output voltage measurements taken throughout the ramp duration. With this method, measurements at a large number of temperatures can be obtained and the total test time is considerably shorter, i.e., between 30 and 40 minutes for a ramp from 0°C to 80°C. The temperature at the time the measurements are taken can be reasonably accurately obtained by monitoring the oil bath temperature. In this study, the 2nd method has been adopted for temperature drift characterization. In actual use, the customer sees exactly what we record here, real-time temperature drift and noise at the reference output.

NI Signal Express is used to control a 34401A DMM while sampling the reference output at 2 Hz, and a DE-115 FPGA board is used to generate control codes for resistor trimming. To

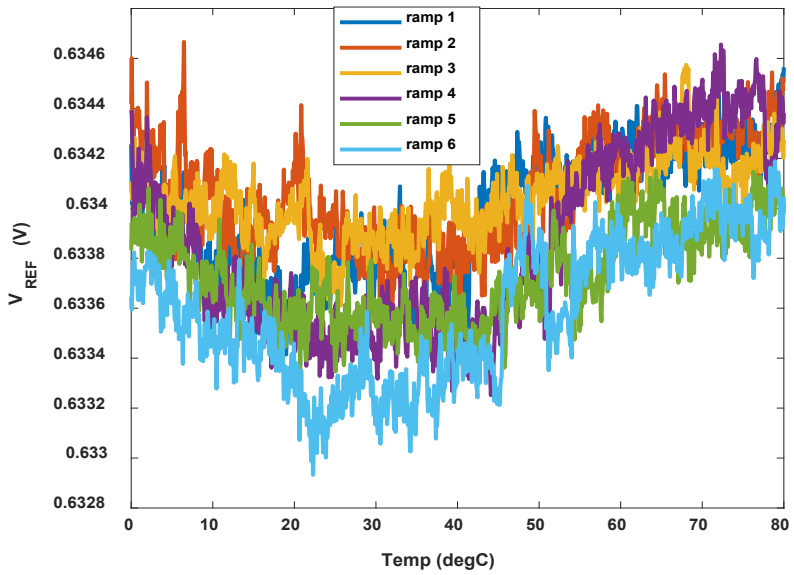

Figure 5.13 Measured drift curves in multiple fast temperature ramps.

verify effectiveness of the measurement instrumentations, a LTC6655 EVM is characterized with the test setup and it reflects a 1.2 ppm/°C TC, matching its datasheet value (1 ppm/°C typ).

5.5.2 Integrated Constant Current Reference

A complete current mode BGR whose schematic is shown in Figure 5.5 was designed and fabricated in GF130nm process. All blocks, including BJTs, two op-amps, and digital controlled trimmable resistors, were on-chip, and the die photo overlaid by the layout is shown in Figure 5.11. A two-parameter trim was required to cancel both the linear T term and the TlnT term. Resistors R_1 and R_3 were made digitally trimmable by placing switchable shunt resistors in series with small resistor segments. Conceptually, R_1 is used to trim the linear T part of the temperature dependence and R_3 is used to trim the TlnT part, even though these trims are interdependent. These resistors each had a 24-level trim, corresponding to about 5 bits. To reduce the temperature coefficient of the resistors, they were designed so that the positive coefficient associated with the resistor bodies were nominally cancelled by the negative temperature coefficient of the resistor tails. Since the emphasis was placed on performance and not on the trimming algorithm, trim codes were obtained

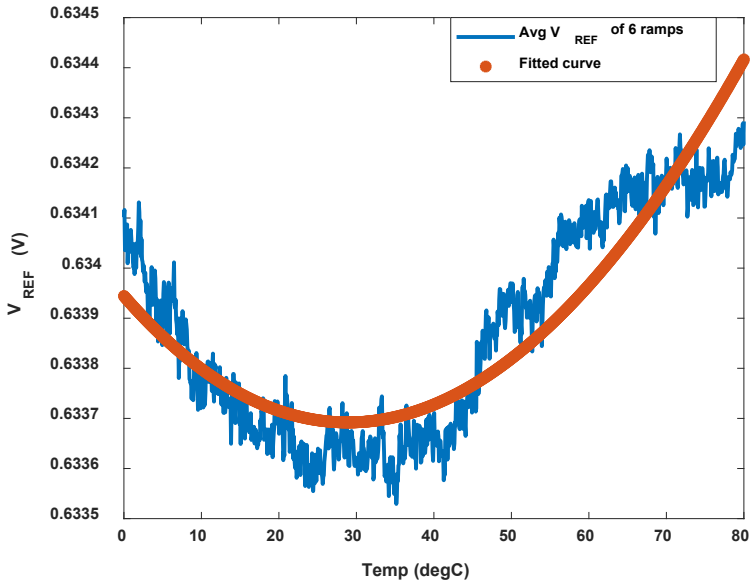

Figure 5.14 Averaged V_{REF} of 6 ramps and its fitted curve.

by scanning through all possible test code combinations at three temperatures (0 °C, 27 °C, and 80 °C), then selecting the trim code that experimentally minimized the TC of the reference output.

Since pin access to the two inputs and the output of the core op-amp OP_1 was available, an external commercial operational amplifier was placed in parallel with the core op-amp to reduce the offset voltage, although low-frequency flicker noise was still observed at the reference output. Six sets of measured results for one test circuit that used an external op-amp in parallel with the integrated op-amp are shown in Figure 5.13. These results are typical of what was seen for other test circuits. We believe the differences between ramps resulted from flicker noise in the BGR. The corresponding fit of the function to the average of these six sets of measured results is shown in Figure 5.14. This corresponds to a TC of about 14 ppm/°C. While using the external op amp reduced the effects of the offset voltage in the core op amp, the offset voltages in the buffer amplifier OP_2, an insufficient trim range in trimming resistors R_1 and R_3, and limited trim resolution made it difficult to accurately assess the performance potential of the V_{GOr} extraction capabilities of this circuit. At this point, three observations can be made. First, more attention needs

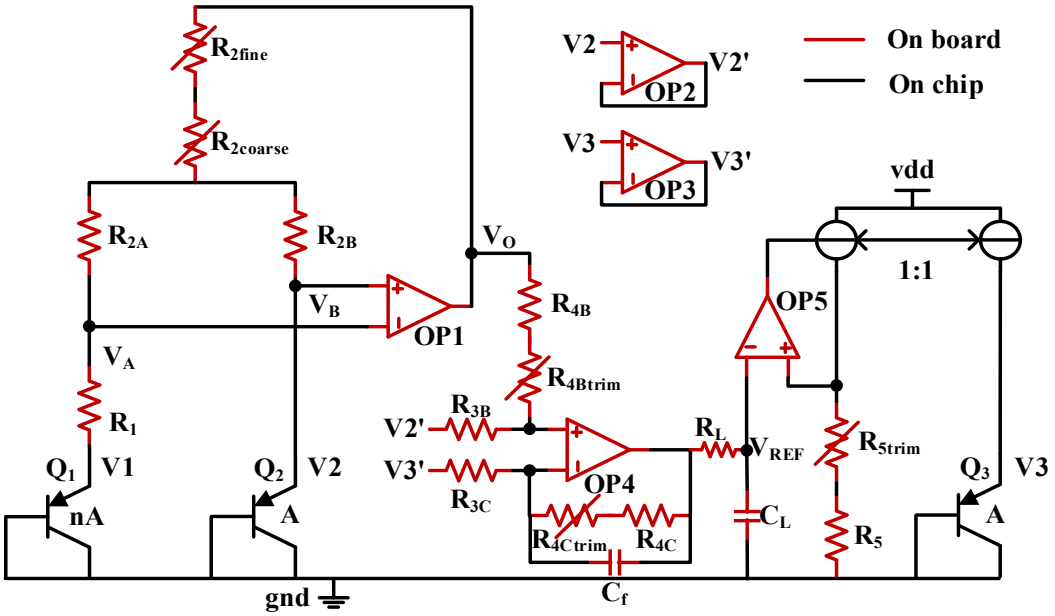

Figure 5.15 Schematic of the voltage mode BGR prototype.

to be placed on establishing adequate adjustment and trim ranges for the trimming resistors. Second, more emphasis should be placed on reducing the offset voltage of the operational amplifiers. Finally, more emphasis should be placed on reducing the 1/f noise in the operational amplifiers.

5.5.3 Partial Chip and Partial Board Constant Voltage Reference

The voltage noise and temperature drift of a voltage reference usually determine the measurement limits of a data acquisition system. Voltage reference datasheets generally provide low-frequency and wideband noise separately because the latter can be greatly reduced in a system application using lowpass filtering. Filtering of low-frequency noise, however, is cumbersome and not practical, as large capacitors are needed to filter out lower frequencies.

Low-frequency noise and temperature drift are lumped together during measurement, posing difficulty in characterizing the actual temperature coefficient. The op-amp in the Kuijk core, the dominant source of such problems, requires chopping to minimize flicker noise. To free

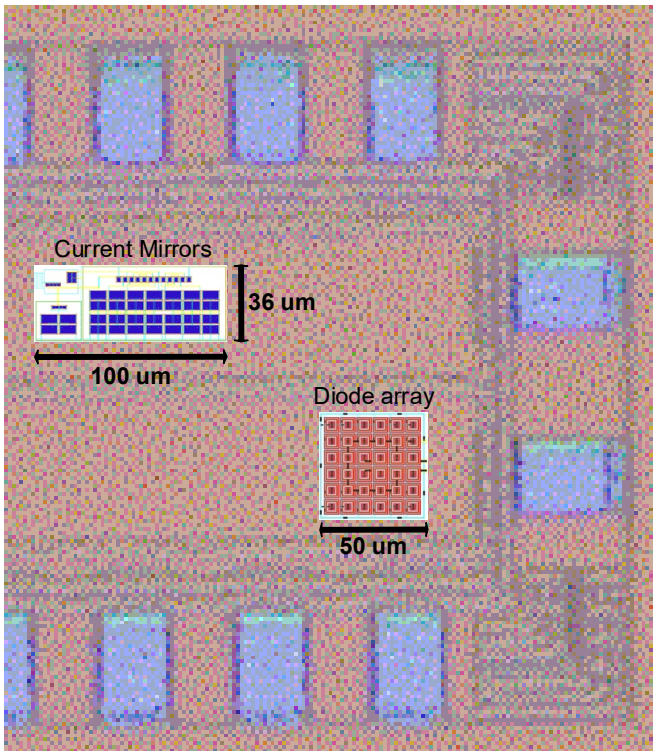

Figure 5.16 Die micrograph overlaid by the layout.

ourselves from the challenging job of designing a precision op-amp, a board-level voltage reference prototype was implemented without affecting the V_{GO} extraction concept using on-chip diodes, current mirrors, and off-chip commercial op-amps (OPA388). The schematic of the BGR prototype is shown in Figure 5.15. C_f is a feedback capacitor to avoid gain peaking at high frequency in the self-bootstrapping loop and C_L = 10 µF performs lowpass filtering at the reference output. Because the fabrication process only has resistors with hundreds of ppm/°C TC, resistors were also implemented with on-board components. R_1, R_{2A}, R_{2B}, R_{3B}, R_{3C}, R_{4B}, $R_{4C,}$ and R5 are fixed resistors, and trimmable parts $R_{2coarse}$, R_{4Btrim}, R_{4Ctrim}, and R_{5trim} were realized with digital potentiometer AD5144, with 10 kΩ range and 8-bit resolution. R_{2fine} built with digital potentiometer AD8400, with 1 kΩ range and 8-bit resolution, was provided for fine tuning of the V_{PTAT} gain.

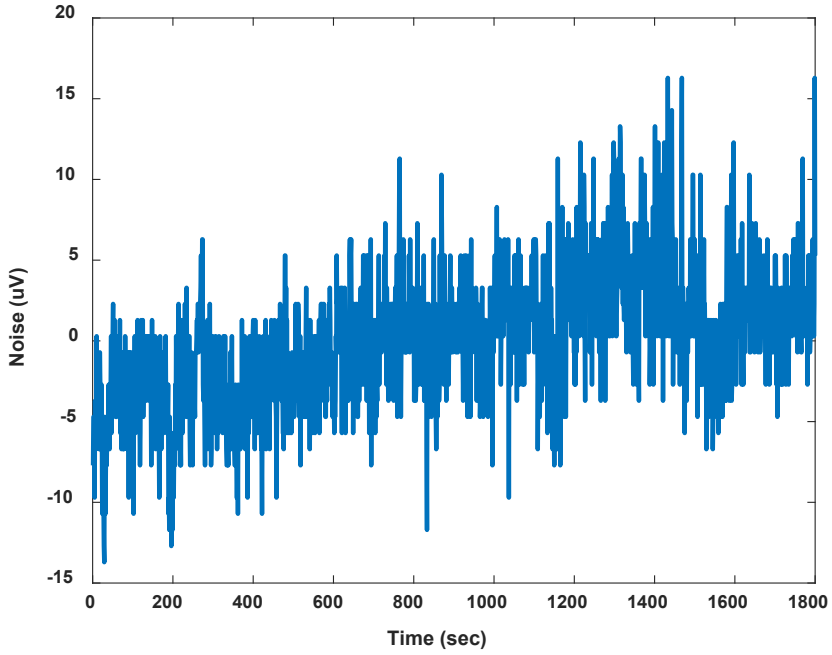

Figure 5.17 Noise (V_{REF}-mean(V_{REF})) at T_r in 30 minutes.

The die photo and layout of the on-chip diodes and current mirrors are given in Figure 5.16. Q_1, Q_2, and Q_3 have an emitter area ratio of 12:2:2, and Q_2 and Q_3 were placed in the middle of the 4*4 diode array and surrounded by dummy devices. Wide metals were used to connect base and connector terminals of the BJTs together to form a solid ground. Multiple wide metal wires were used in parallel to route the nodes to the corresponding pads to minimize parasitic resistances.

Figure 5.17 shows the noise at the reference output when the oil bath has settled at T_r = 27 °C. The white noise is below 5 µV and the low frequency 1/f noise is at the 10 µV level. Reference drift data of one chip trimmed after the proposed two-temperature trimming method is provided in Figure 5.18, and the curve suggests a 1.6 ppm/°C temperature coefficient. There still existed a PTAT shape, and additional trimming was carried out to find the optimal TC. The oil bath temperature was ramped from 20 °C to 60 °C while the reference drift direction was continuously monitored. Every 5 °C, if V_{REF} decreased, R_{2fine} was increased by 1 LSB, while if V_{REF} increased, R_{2fine} was reduced by 1 LSB. Drift curves during both up or down temperature ramping are

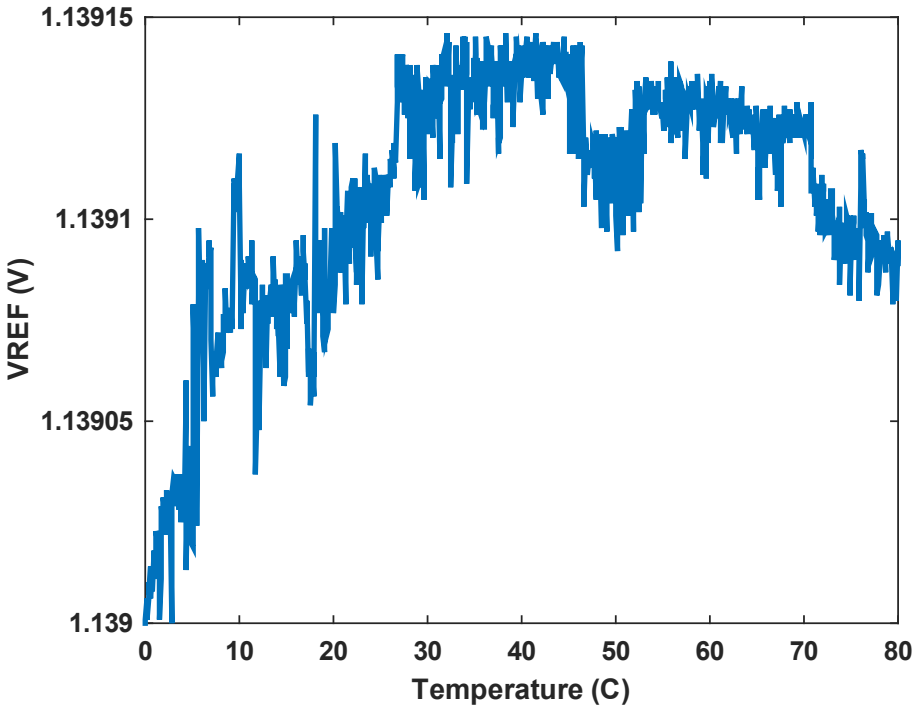

Figure 5.18 V_{REF} drift after the proposed two-temperature trimming.

provided in Figure 5.19. The additionally trimmed V_{REF} exhibited about 70 µV drift over a temperature range of 0 °C to 80 °C, corresponding to only an 0.8 ppm/°C temperature coefficient. There are several possible reasons that could explain why the proposed two-temperature trimming method didn't generate optimal results. The temperature range measured was narrow due to instrumentation limitations, which can result in a different drift shape compared to simulations. Moreover, the actual silicon bandgap voltage can vary from that of the spice model and a batch trimming may be required to find the optimal trimming target. Further investigation needs to be carried out to find a suitable trimming methodology for the proposed bandgap reference during high-volume production.

5.6 Conclusion

This chapter presented a general method for extracting V_{GOr} from the temperature characteristic of V_{BE}. Small error sources that could possibly affect the voltage reference's

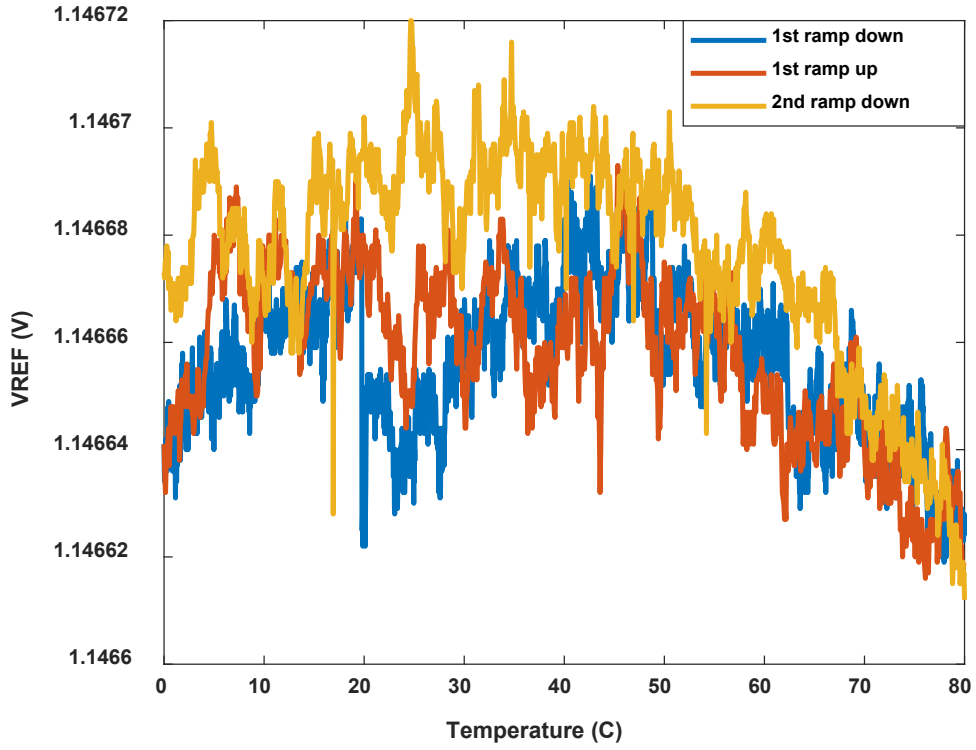

Figure 5.19 V_{REF} drift during temperature ramps after optimal trimming.

performance were analyzed using a new base expansion, and their effect on the temperature coefficient were minimized after trimming of the gains of V_{PTAT} and V_{NL}. The embodiment created in UMC 65nm process showed that the design achieved a sub-ppm/°C level temperature coefficient over the temperature range of 0 °C to 80 °C. The proposed method can be implemented on different BGR structures to achieve a sub-ppm/ºC temperature coefficient.

5.7 References

[1] G. Ge, C. Zhang, G. Hoogzaad, and K. A. A. Makinwa, "A Single-Trim CMOS Bandgap Reference With a 3σ Inaccuracy of ±0.15% From -40°C to 125°C," *IEEE J. Solid-State Circuits*, vol. 46, no. 11, pp. 2693–2701, Nov. 2011, doi: 10.1109/JSSC.2011.2165235.

[2] R. Oberhuber, R. Prakash and V. Ivanov, "A 1 ppm/°C bandgap voltage reference with new second-order Taylor curvature compensation, " *23rd IEEE International SOC Conference*, Las Vegas, NV, 2010, pp. 71-76, doi: 10.1109/SOCC.2010.5784635.

[3] J.-H. Li, X. Zhang, and M. Yu, "A 1.2-V Piecewise Curvature-Corrected Bandgap Reference in 0.5 μm CMOS Process," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 19, no. 6, pp. 1118–1122, Jun. 2011, doi: 10.1109/TVLSI.2010.2045519.

[4] Z.-K. Zhou et al., "A 1.6-V 25-μA 5-ppm/°C Curvature-Compensated Bandgap Reference," IEEE Trans. Circuits Syst. Regul. Pap., vol. 59, no. 4, pp. 677–684, Apr. 2012, doi: 10.1109/TCSI.2011.2169732.

[5] L. Liu, X. Liao, and J. Mu, "A 3.6 μV_{rms} Noise, 3 ppm/°C TC Bandgap Reference With Offset/Noise Suppression and Five-Piece Linear Compensation," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 66, no. 10, pp. 3786–3796, Oct. 2019, doi: 10.1109/TCSI.2019.2922652.

[6] B. Ma and F. Yu, "A Novel 1.2-V 4.5-ppm/°C Curvature-Compensated CMOS Bandgap Reference," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 61, no. 4, pp. 1026–1035, Apr. 2014, doi: 10.1109/TCSI.2013.2286032.

[7] Y. Huang, L. Zhu, F. Kong, C. Cheung, and L. Najafizadeh, "BiCMOS-Based Compensation: Toward Fully Curvature-Corrected Bandgap Reference Circuits," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 65, no. 4, pp. 1210–1223, Apr. 2018, doi: 10.1109/TCSI.2017.2736062.

[8] Q. Duan and J. Roh, "A 1.2-V 4.2- ppm/°C High-Order Curvature-Compensated CMOS Bandgap Reference," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 62, no. 3, pp. 662–670, Mar. 2015, doi: 10.1109/TCSI.2014.2374832.

[9] C. M. Andreou, S. Koudounas, and J. Georgiou, "A Novel Wide-Temperature-Range, 3.9 ppm/°C CMOS Bandgap Reference Circuit," *IEEE J. Solid-State Circuits*, vol. 47, no. 2, pp. 574–581, Feb. 2012, doi: 10.1109/JSSC.2011.2173267.

[10] K. Ueno, T. Hirose, T. Asai and Y. Amemiya, "A 300 nW, 15 ppm /°C, 20 ppm/V CMOS Voltage Reference Circuit Consisting of Subthreshold MOSFETs." in *IEEE Journal of Solid-State Circuits*, vol. 44, no. 7, pp. 2047-2054, July 2009, doi: 10.1109/JSSC.2009.2021922.

[11] Y. P. Tsividis, "Accurate analysis of temperature effects in I/SUB c/V/SUB BE/ characteristics with application to bandgap reference sources," *IEEE J. Solid-State Circuits*, vol. 15, no. 6, pp. 1076–1084, Dec. 1980, doi: 10.1109/JSSC.1980.1051519.

[12] N. Liu, R. Geiger, and D. Chen, "Bandgap Voltage V_{GO} Extraction with Two-Temperature Trimming for Designing Sub-ppm/°C Voltage References," in *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2019, pp. 1–4, doi: 10.1109/ISCAS.2019.8702697.

[13] W. Bludau, A. Onton, and W. Heinke, "Temperature dependence of the band gap of silicon," *J. Appl. Phys.*, vol. 45, no. 4, pp. 1846–1848, Apr. 1974, doi: 10.1063/1.1663501.

[14] M. A. P. Pertijs, G. C. M. Meijer, and J. H. Huijsing, "Precision temperature measurement using CMOS substrate pnp transistors," *IEEE Sens. J.*, vol. 4, no. 3, pp. 294–300, Jun. 2004, doi: 10.1109/JSEN.2004.826742.

[15] E. W. Beach, V. F. Drobny, and D. W. Robinson, “Thin film resistors integrated at two different metal interconnect levels of single die,” US20080272460A1, Nov. 06, 2008.

[16] P. Malcovati, F. Maloberti, C. Fiocchi and M. Pruzzi, “Curvature-compensated BiCMOS bandgap with 1-V supply voltage,” in *IEEE Journal of Solid-State Circuits*, vol. 36, no. 7, pp. 1076-1081, July 2001, doi: 10.1109/4.933463.

[17] J. L. Doorenbos, “Reference Voltage Generator System,” US20180059703A1, Mar. 01, 2018.

[18] K. E. Kuijk, “A precision reference voltage source,” *IEEE J. Solid-State Circuits*, vol. 8, no. 3, pp. 222–226, Jun. 1973, doi: 10.1109/JSSC.1973.1050378.

CHAPTER 6. A TRANSIENT-ENHANCED OUTPUT-CAPACITORLESS LDO WITH LOCAL FAST LOOP AND OVERSHOOT DETECTION

Nanqi Liu and Degang Chen

Dept. of Electrical and Computer Engineering, Iowa State University, Ames, IA, 50010

Modified from a manuscript published in IEEE Transactions on Circuits and Systems I: Regular Papers

6.1 Abstract

An output-capacitorless low-dropout regulator (OCL-LDO) with simple structure and fast transient response is proposed for system-on-chip (SoC) applications. A super source follower is inserted into a cascoded flipped voltage follower to drive the power transistor, which forms a fast-local loop for quick turn-on. A robust overshoot detection circuit consuming only leakage current is proposed for fast turn-off. The combination of these two techniques significantly reduce the overshoot/undershoot voltages and achieve fast settling time during load steps. A simple yet effective additional turn-around stage is added in the error amplifier to improve the positive phase slew rate for potential dynamic voltage scaling (DVS) function in battery-operated systems. The LDO is implemented in a 65 nm CMOS process and it can deliver a 20 mA load current with 0.9 V regulated output and 150 mV dropout voltage. It occupies an active area of 0.01 mm^2 and can work stably in a load range from 0 to 20 mA with 65 μA quiescent current. The measured results show a settling time about 100 ns for load steps from 100 μA to 20 mA as well as V_{REF} and V_{IN} steps.

6.2 Introduction

The power management ICs (PMICs) are critical modules in today's system-on-chip (SoC) applications. Various types of power management units are required for widely distributed voltage

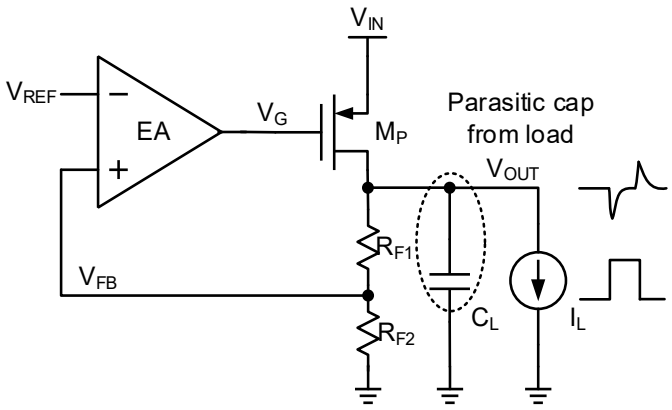

Figure 6.1 Output-capacitorless low-dropout regulator.

domains [1]. Low-dropout regulators (LDO) are usually cascaded after power-efficient switching regulators to filter the switching noise and provide a clean supply [2].

Conventional LDOs have an off-chip capacitor in the order of several micro-farads at the output to stabilize the system [3]. The removal of this bulky capacitor in output-capacitorless LDOs is beneficial in terms of integration. However, the power-supply rejection (PSR) and load transient response degrade [4]. The power line parasitic capacitor is several orders of magnitude smaller and cannot provide or absorb the instantaneous current when a sudden load step happens. In nano-scale technology, the current consumption of digital cells can change between near zero and maximum in a very short time [5]. The resulted supply voltage spikes as shown in Figure 6.1 can degrade the performance or even hurt the operation of the load circuits. Therefore, an LDO with ultra-fast response to load steps, especially low-to-high transition, is in demand. Low standby current is also desired for LDOs running on battery power. This is especially true in mobile medical applications that have extended use between charging cycles. Moreover, for blocks with different operating modes, LDOs can be designed to support dynamic voltage scaling function and then fast tracking of VREF is appreciated. In summary, for OCL-LDOs powering digital circuits in SoCs, fast transient responses are essential and low standby current is desired.

Stability is a basic requirement for an LDO. After getting rid of the large off-chip CL, the pole at the LDO output is no longer dominant and the stability control becomes more challenging. Different compensation techniques using small capacitors (~pF) have been proposed for OCL-LDOs [6], [7], [8] and result in increased circuit complexity. Performance in terms of accuracy, transient response and PSR is usually a more difficult task. This work is mainly focused on the improvement of transient responses of OCL-LDOs with a reasonable quiescent current budget. During large load steps, before the LDO loop responds, the spike at the output ΔVOUT can be estimated by:

$$\Delta V_{OUT} = \frac{1}{C_L}\left(\int_{t_0}^{t_0+T_{edge}} \Delta I_L(t)dt + \int_{t_0+T_{edge}}^{t_0+T_R} \Delta I_L dt\right) \approx \frac{\Delta I_L T_R}{C_L} \tag{6.1}$$

where Tedge is the load step's edge time and is assumed to be much smaller than the LDO's response time TR here [4]. Several design parameters determine TR. Besides the loop bandwidth (BW), the slew rate at the gate of MP also limits the response speed here. The finite response time can be approximated as:

$$T_R \approx \frac{1}{BW} + t_{SR} = \frac{1}{BW} + C_{par}\frac{\Delta V_G}{I_G} \tag{6.2}$$

where Cpar, ΔVG, and IG are the parasitic cap, required voltage change and slewing current at MP's gate, respectively. According to equation (6.1) and (6.2), a large bandwidth and high slew rate at the pass device's gate is necessary for an LDO to provide adequate power instantaneously.

In this chapter, we utilize a buffer with ultra-low output impedance and premium sink capability to drive the power transistor. The flipped voltage follower (FVF) topology with this buffer forms a local loop with large bandwidth which can quickly respond to the light-to-heavy load step. A simple and robust sensing circuit based on capacitor coupling is proposed to detect the overshoot and generate a dynamic current to turn off the power device. A folded-cascode error

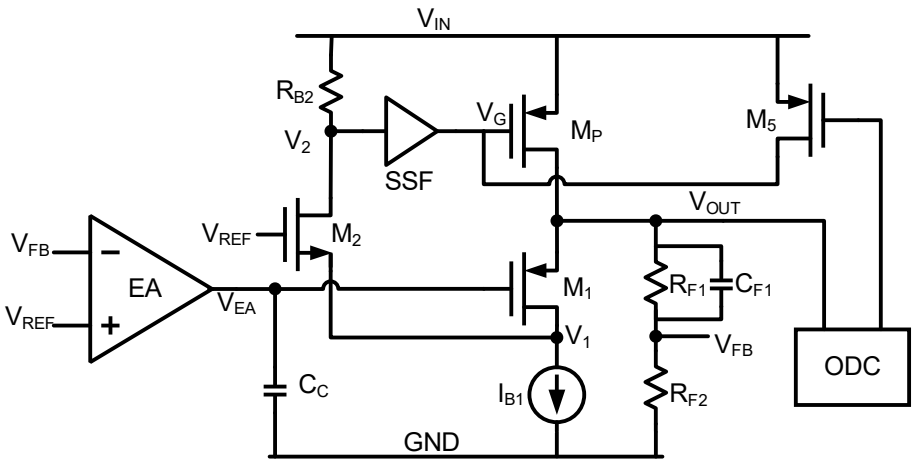

Figure 6.2 Proposed LDO architecture.

amplifier (EA) with improved slew rate is implemented to achieve good regulations and fast tracking of VREF at the same time. The local loop designed for large bandwidth and low gain is compensated by pushing parasitic poles to even higher frequency. The whole system is compensated by placing a capacitor at the EA's output to form the dominant pole. The overall structure of the OCL-LDO is simple and straightforward which implies robustness and reliability for analog circuits.

The rest of the chapter is organized as follows. In Section 6.3, the proposed OCL-LDO architecture is explained and compared to recently published transient enhanced structures. Both loops' frequency responses are then analyzed. Simulated load transient responses are given in Section 6.4 and measurement results are provided in Section 6.5. Finally, the conclusion is drawn in Section 6.6.

6.3 Proposed LDO Architecture

The proposed LDO architecture is given in Figure 6.2. The fast local loop dealing with load transients is analyzed first and a robust overshoot detection circuit is introduced. Finally, the global loop including feedback and the error amplifier which handles regulation is presented.

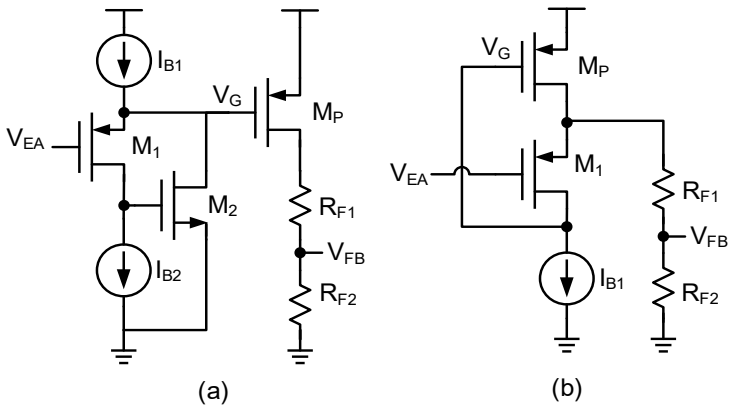

Figure 6.3 Buffer structures: (a) super source follower (SSF), (b) flipped voltage follower (FVF).

6.3.1 Fast Local Loop

To achieve good transient performance, a large unity gain frequency (UGF) is necessary to control the power device immediately. Multi-stage architecture [6], [9], [10], [11] was proposed in early 2000s to extend the bandwidth by inserting a buffer to decouple the EA's high output impedance and the pass transistor MP's large parasitic capacitance. A source follower was firstly used [12] and then the super source follower (SSF) as shown in Figure 6.3(a) was proposed [3]. It achieves a very low output impedance through shunt feedback:

$$r_o = \frac{1}{g_{m,M_1} * (g_{m,M_2} r_{ds,M_1})}, \tag{6.3}$$

and has a good sinking capability which can quickly turn on the MP. However, as a single loop system, the slewing speed at V_{EA} directly limits the transient response at V_G [13]. A large quiescent current is required to achieve fast responses and good stability at the same time.

LDOs based on flipped voltage follower (FVF) [14] provided stable voltage regulation with simple structures. The FVF based power stage as shown in Figure 6.3(b) achieves low output impedance and forms a local loop which can quickly respond to load changes [15]. An FVF with folded-cascode stage was further used in [7] to improve the loop gain. In both structures, although

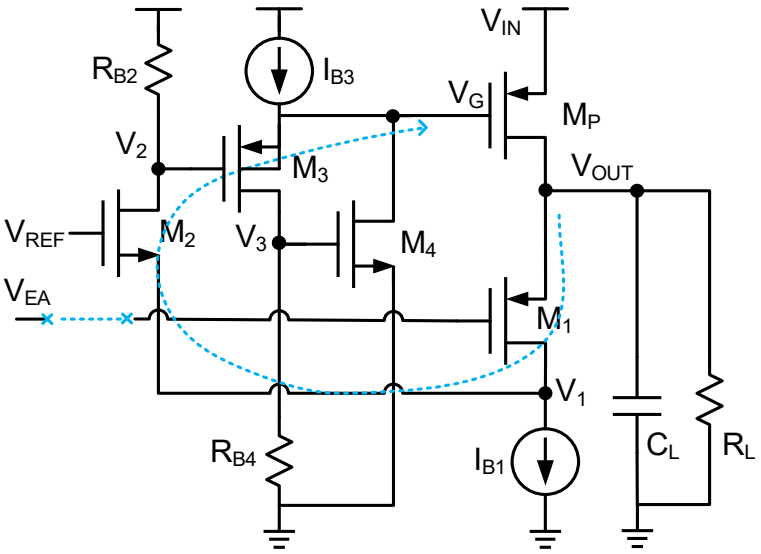

Figure 6.4 Fast local loop including load.

the loop bandwidth is large, the slew rate at V_G is limited by I_{B1} and with 8 uA quiescent current, the reported recovery time is in µs level [7]. A hybrid bias current generator was proposed in [16] to boost the buffer's bias current dynamically and adaptively. However, the response to low-to-high load transition is still limited by the speed of the detection and dynamic current scaling loop. A tri-loop LDO [2] was reported to be able to make one of its three loops obtain over 100 MHz UGF by setting the output pole as the dominant pole and pushing the internal poles to higher frequencies. Realized in a 65-nm process, its maximum output current is restricted to 10 mA and the PSR at low frequencies is only -20 dB.

To achieve quick transient responses without requiring large standby current, a fast local loop based on the cascoded FVF topology and the SSF buffer is implemented here. As shown in Figure 6.2, the EA's output is driving M_1 instead of M_P and a compensation capacitor C_C forms the dominant pole for the whole system which will be discussed later. During fast transients, M_1 and M_2 behave as two common gate transistors and can sense the undershoot at V_{OUT} and amplify it to the node V_2 with little delay. Transistor M_2 is biased with V_{REF} so that M_1 works in saturation

under different V_{REF} and V_{OUT} combinations. To get rid of the limitation of discharging rate at V_G due to I_{B1}, a super source follower with premium sink capability is added. The amplified voltage drop at V_2 can introduce a large transient current to discharge V_G and turn off the pass device. One drawback here is the limited supply input range. In order to keep transistors in the saturation region, the LDO's supply V_{IN} needs to satisfy:

$$V_{IN} > V_{SG,MP} + V_{SG,M3} + 2V_{DS,sat}, \tag{6.4}$$

which makes the proposed structure not suitable in super low supply applications.

A feedback loop with only one pole in its open-loop transfer function is unconditionally stable. Nevertheless, a two-stage system can be stable with any capacitive load by limiting the voltage gain of the first stage [17]. The direct dependence between the first stage gain and worst-case phase margin has been provided in [18] and the results are adopted here to control the stability when boosting the UGF of the local loop.

During a low-to-high load transition, the undershoot at V_{OUT} propagates as shown in Figure 6.4. To analyze the local loop's frequency response, the node V_{EA} is broken and considered as a DC voltage to M_1. The poles and zeros in this fast loop are derived blew to gain insight on how to boost the bandwidth and control stability. The pole at the LDO's output is dependent on both I_L and C_L and it equals to

$$P_{OUT} = \frac{1}{C_L R_{OUT}} = \frac{1}{C_L * (r_{ds,MP} \,||\, 1/g_{m,M_1} \,||\, R_L)} \propto \frac{I_L}{C_L}, \tag{6.5}$$

where R_{F1} and R_{F2} are ignored in the calculation of R_{OUT}. For large C_L and small I_L, P_{OUT} can locate at low frequency and become the dominant pole of the local loop. The miller capacitor at the pass transistor's gate equals to

$$C_G = C_{gs,M_p} + (1 + g_{m,M_p} R_{OUT}) C_{gd,M_p}. \tag{6.6}$$

The pole at M_P's gate is much less dependent on different I_L and C_L combinations compared to P_{OUT} and is calculated as

$$P_{VG} = \frac{1}{C_G R_{o,SSF}} \tag{6.7}$$

where $R_{o,SSF}$ is the output impedance of the SSF. For small C_L and large I_L, P_{OUT} is at high frequency and P_{VG} becomes the dominant pole and thus determines the UGF of the loop. Parasitic capacitance at V_G depends on the size of M_P and the technology node used. For this design in 65 nm process, the 20 mA load current doesn't require a super large M_P and the resulted C_G is not huge across the load range. In this design, M_4 is biased at near-threshold region by R_{B4} and can sink a large dynamic current when there is a rise at its gate. It consumes little quiescent current so that it doesn't further push the pole P_{VG} to too high frequency. Otherwise, the stability becomes difficult to control considering other parasitic poles in the loop.

The other parasitic poles are pushed far beyond the UGF for stability concern. The pole at the internal node V_1 is at high frequency since common gate transistor M_2 provides a low impedance:

$$P_{V1} = \frac{1}{(C_{gs,M_2} + C_{gd,M_1}) * (1/g_{m,M_2} \,||\, r_{ds,M_2})} \approx \frac{g_{m,M_2}}{C_{gs,M_2}}. \tag{6.8}$$

A resistor instead of a current mirror is used here to generate the biasing current and reduce the impedance at the node V_2. The pole P_{V2} is given by:

$$P_{V2} = \frac{1}{(C_{gs,M_3} + C_{gd,M_2}) * (R_{B2} \,||\, r_{ds,M_2})} \approx \frac{1}{C_{gs,M_3} R_{B2}}. \tag{6.9}$$

R_{B4} is used to bias M_4 at near-threshold region and helps push the pole PV3 beyond the UGF:

$$P_{V3} = \frac{1}{(C_{gd,M_3} + C_{gg,M_4}) * (R_{B4} \,||\, r_{ds,M_3})} \approx \frac{1}{C_{gg,M_4} R_{B4}}. \tag{6.10}$$

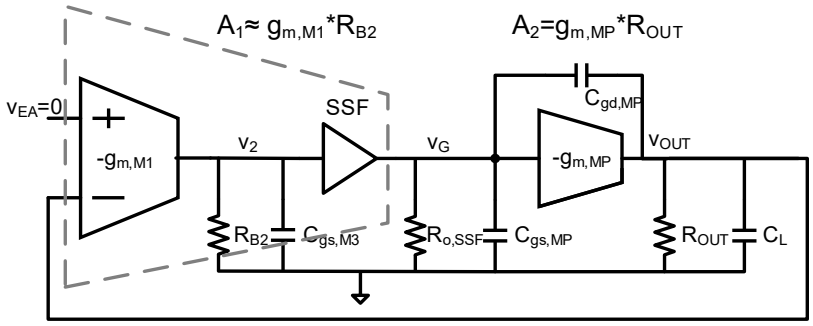

Figure 6.5 Small signal model of the local loop approximated as a two-stage amplifier.

With a large W/L ratio, M_4 can generate a large sourcing current to discharge VG when the V_{OUT} spike is amplified to V_3 and turn M_4 into saturation. However, the parasitic capacitor at the gate of M_4 directly affects the P_{V3} location and trade-off needs to be made between speed and stability.

As mentioned earlier, transistors with small channel length and resistor-based biasing with sufficient current are used to push P_{V1}, P_{V2}, and P_{V3} near GHz level. Then the local loop can be considered as a two-pole system as shown in Figure 6.5 with the node V_{EA} considered as ac ground. Its loop gain is

$$A_{ol,local} \approx g_{m,M_1} R_{B2} \cdot g_{m,M_P} R_{OUT} . \tag{6.11}$$

And the open loop transfer function is calculated as

$$TF_{ol,local} = \frac{A_{ol,local}}{(1+\frac{s}{P_{V_G}}) * (1+\frac{s}{P_{OUT}})} . \tag{6.12}$$

The first stage gain $A_1 = g_{m,M1}*R_{B2}$ is much less dependent on I_L and is designed to be 15 dB according to [18] to achieve good phase margin for the local loop across wide IL and CL range. This is not necessary for the stability of the whole system but is a good practice to make the internal closed loop transfer function to have stable poles and a low damping factor so that bode plot criteria is valid for the whole system.

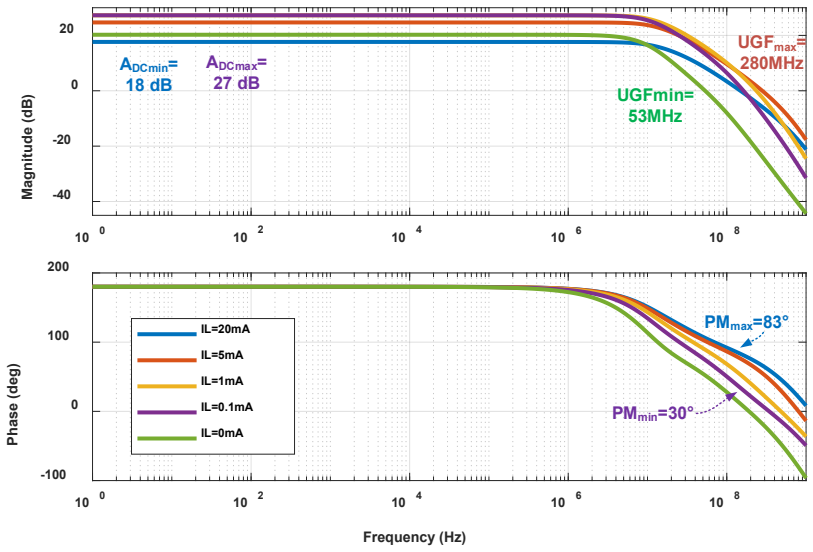

Figure 6.6 Simulated Bode plots of the local loop with C_L = 10 pF, V_{IN} = 1.1 V and V_{OUT} = 0.9 V under different I_L.

In digital circuits, the switching gates provide a low-resistance path from gates to one of the supply rails whereas the others' gate capacitances are connected to supply rails as bypass. With 20,000 gates/mm^2 gate density and 10 fF typical gate capacitance, the local bypass will be at 100 pF/mm^2 level assuming 50% activity. For OCL-LDOs powering digital loads, it's fair to assume the minimum load capacitor C_L is 10 pF. The stability of the local loop is more serious with small C_L due to the higher UGF.

To simulate the local loop's response, a DC voltage source representing V_{EA} is used to bias M_1 and the local loop is broken between V_{OUT} and M_1's source. With C_L = 10 pF, the simulated loop gains under different I_L conditions are provided in Figure 6.6. As we can see the loop dc gain is kept low and the UGF is 53 MHz at no load condition and higher than 100 MHz for I_L > 0.1 mA. The pole P_{OUT} is dependent on I_L and C_L and can vary a lot over the entire load current range. R_{OUT} can be approximately as R_L for large I_L. As I_L drops from 20 mA to 0.1 mA, DC gain keeps increasing according to (6.11), and P_{OUT} moves to lower frequency and hurts stability as shown in Fig. 6. When IL further decreases, RL becomes comparable to $r_{ds,MP}||1/g_{m,M1}$ and the loop gain

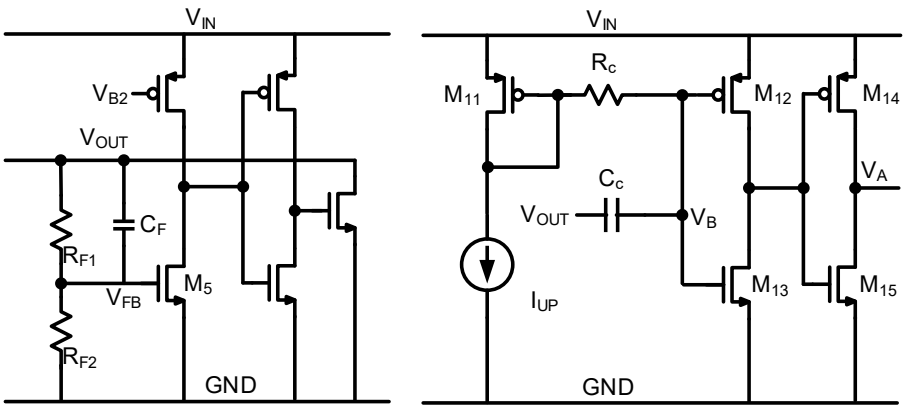

Figure 6.7 Inverter-based overshoot detection circuits.

drops as M_1 starts to go out of safe saturation region. The minimum phase margin is 30° when I_L = 0.1 mA. Due to the parasitic poles' effect, it is less than the simulated worst-case, 48°, reported in [18]. Larger C_L pushes the P_{OUT} to lower frequency and the UGF of the loop is always reduced no matter P_{OUT} is the dominant pole or not. The phase margin improves as the three parasitic poles P_{V1}, P_{V2}, and P_{V3} are farther away from the UGF.

6.3.2 Overshoot Detection Circuit

The large bandwidth of the local loop and good sinking capability of the SSF enables instantaneous turn-on of the pass device during low-to-high load transition. However, the charging up rate at the M_P's gate is limited by I_{B3}.

To deal with the other kind of load change, inverter-based overshoot detection circuit as shown in Figure 6.7 has been implemented by us in [19] and at the same time in [20]. The voltage overshoot is coupled to the inverter's input and then transferred into a dynamic current with little loop latency. However, the trip point here is sensitive to process variations and temperature changes. Therefore, a calibration scheme is needed which increases the design complexity. Moreover, we would like the transient detection consumes little current at steady state and still behaves fast when a transient occurs. A spike detection circuit based on quasi-floating gate

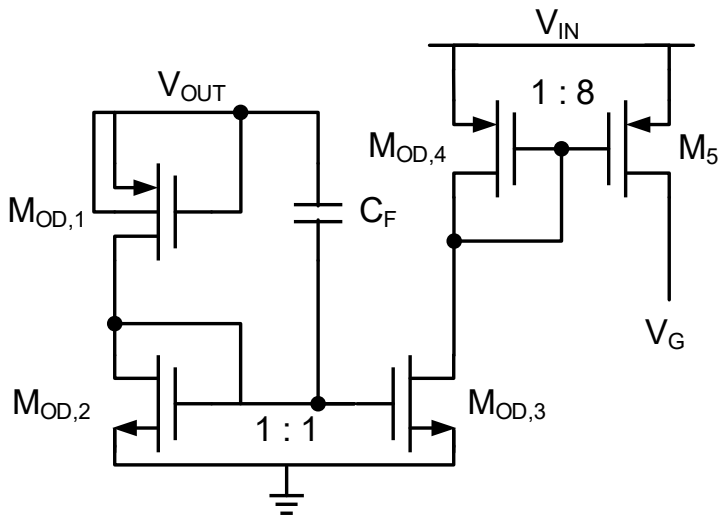

Figure 6.8 Robust overshoot detection circuit.

transistors was applied in [21]. It consumes no quiescent current, but the simulated output spikes were still large and took ~4.6 μs to settle.

A novel overshoot detection circuit consuming only leakage current is introduced in Figure 6.8. The structure is simple but it's robust under process and temperature variations. Transistor $M_{OD,1}$'s gate and source are connected so that it's always in off mode. Its leakage current through the diode connected transistor $M_{OD,2}$ sets the bias voltage of $M_{OD,3}$. A 200 fF coupling cap C_F is used to couple the V_{OUT} overshoot to the gate of $M_{OD,3}$. Since $M_{OD,3}$ is biased at near-threshold region, it will sink a large dynamic current when there is a rise at its gate. This current is then amplified using a current mirror ($M_{OD,4}$ and M_5) to charge up V_G and turn off the pass device. Although the leakage current varies a lot over corner and temperature, the resulted change on the bias voltage is significantly smaller than the dynamic overshoot in hundreds of mV level. The transition point of this circuit is based on the V_{TH} matching between $M_{OD,2}$ and $M_{OD,3}$. This makes it more robust against process variations than the ones in [19]-[20]. Moreover, when the LDO's load current is small, VG is high and makes the standby current (mostly M_5's leakage) even smaller.

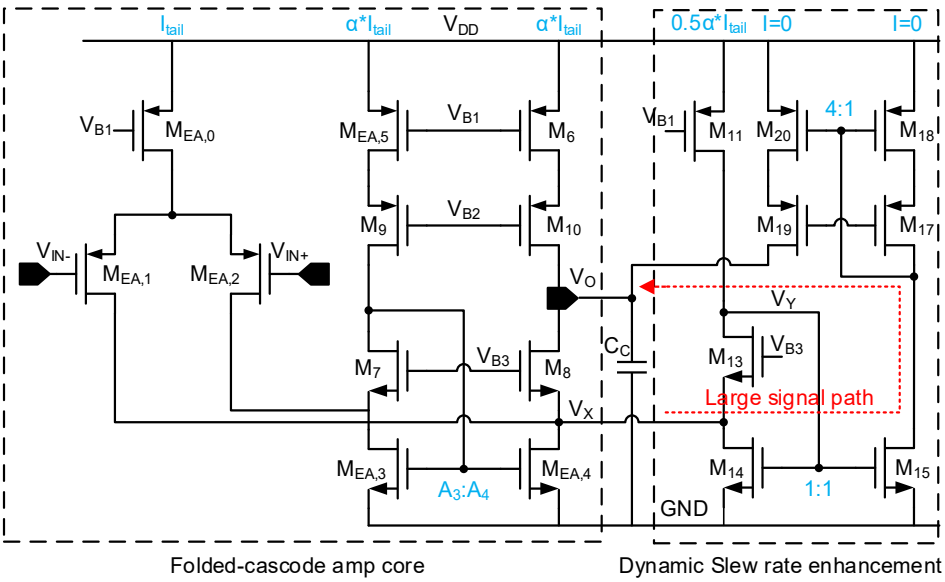

Figure 6.9 Error amplifier with improved positive slewing.

6.3.3 Global Loop for Regulation

The fast local loop is designed for high bandwidth and low gain. Then the regulation of the LDO relies on the global loop including the error amplifier (EA). For op amps, one of the most widely used architectures, is the folded cascode amplifier (FCA) because of its high gain and wide input common mode range (ICMR). The PMOS input FCA with differential-to-single-ended conversion on NMOS side is used here for its higher non-dominant poles [22] and lower input range. For stability, an on-chip capacitor CC = 1 pF is added at EA's output to form the dominant pole in the whole system:

$$P_{EA} = \frac{1}{C_C R_{o,EA}}, \tag{6.13}$$

where $R_{o,EA}$ is the output impedance of the EA.

Dynamic voltage scaling (DVS) has been a popular method to save the average power for microprocessors with hundreds of mA current consumption powered by buck converters [23]. It also has applications in digital blocks driven by LDOs in SoC integrations [24], [25]. To implement

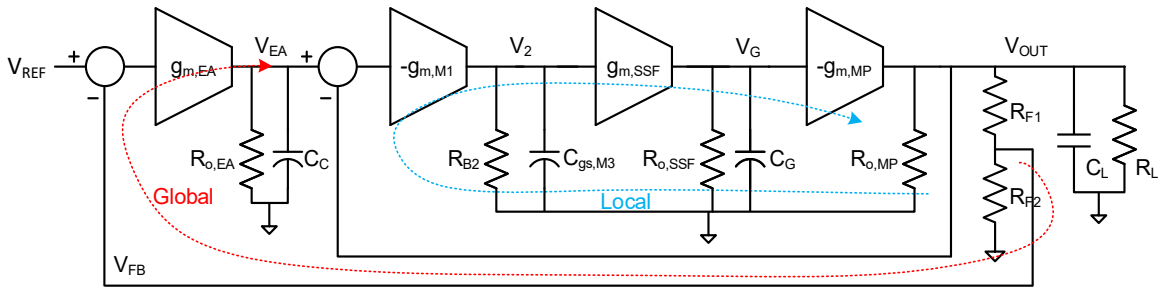

Figure 6.10 Small-signal model of the whole system.

the method, the LDO should have fast tracking of V_{REF}, especially the low-to-high transition when the load is waked up from idle mode [26]. A simple yet effective additional turn-around stage is added here to boost the large signal slewing. It consumes little current and has the potential for DVS implementation in low power applications like the Internet of Things (IoT) devices. As shown in Figure 6.9, the turn-around stage is normally off to preserve the amplifier's small signal performance but is activated in the positive slewing phase (i.e., V_{REF} low-to-high step).

Transistor M_{13} has the same aspect ratio and same V_{GS} as M_8 but with only half the bias current, therefore M_{13} works in the triode region in the dc operation, which leads to a low drain source voltage of M_{13} (VY close to VX). Upon application of a positive differential input signal, transistor M_{13} source voltage increases until it starts to work in the saturation region and transistor M_{14} turns on. M_{14}'s drain current is gained up by a factor of four via the current mirrors of M_{14}-M_{15} and M_{18}-M_{20}, and then the gained-up current is sent to the output to charge the load capacitor C_C. The turn-around stage largely enhances the amplifier's current conveyance capability in the positive slewing phase with minimal bias current consumption. As a result, the bias current of the FCA's cascode stage can be reduced to a current much smaller than Itail and a properly small α should be selected after trade-off with stability. In this design, α = 1/6 and corresponding A_3:A_4 is set to be 8:9 so that the input pair still has symmetric current.

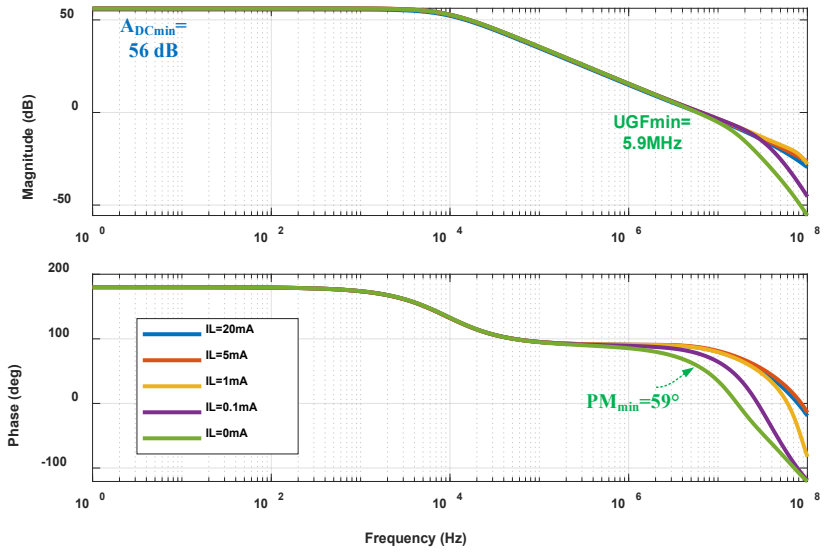

Figure 6.11 Simulated Bode plots of the global loop with C_L = 100 pF, V_{IN} = 1.1 V and V_{OUT} = 0.9 V under different I_L.

The overall block diagram is given in Figure 6.10. The open loop transfer function of the LDO is given below:

$$TF_{ol,global} = \frac{\beta g_{m,EA} r_{o,EA}}{1+\frac{s}{P_{EA}}} \cdot \frac{TF_{ol,local}}{1+TF_{ol,local}} = \frac{\beta A_{ol,global}}{(1+\frac{s}{P_{EA}}) * [A_{ol,local} + (1+\frac{s}{P_{V_G}}) * (1+\frac{s}{P_{OUT}})]}, \quad (6.14)$$

where $A_{ol,global} = g_{m,EA} r_{o,EA} \cdot A_{ol,local}$ and β = 2/3 is the feedback factor. The Routh-Hurwitz criterion is used to derive the stability conditions, which is widely used in control engineering for the stability analysis of closed loop systems. The stability condition of the global loop is fulfilled when

$$\beta \frac{g_{m,EA}}{C_C} < \frac{1}{C_G R_{o,SSF}} + \frac{1}{C_L R_{OUT}}. \quad (6.15)$$

This is used as the guideline to decide the compensation cap C_C. To simulate the loop gain, the global loop is broken at V_{FB}. The simulated bode plots with C_L = 100 pF under different IL conditions are provided in Figure 6.11. The dc loop gain is around 56 dB and the minimum phase margin is 59°. For an OCL-LDO, 60° phase margin is not necessary indeed [5] as this is not a

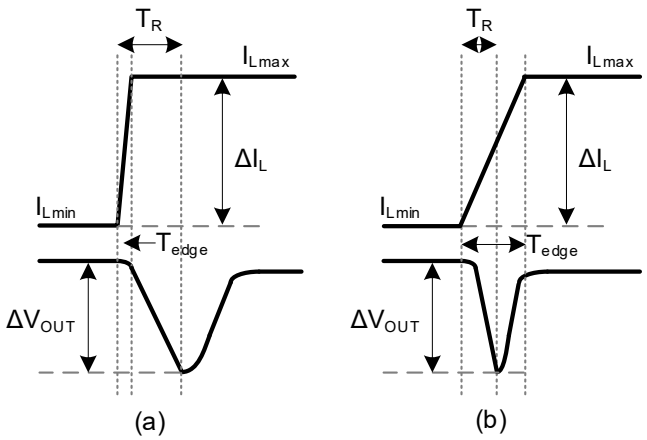

Figure 6.12 Transient responses (undershoot only) at the LDO's output with (a) $T_{edge} << T_R$ and (b) $T_{edge} > T_R$. (Redrawn from [4]).

signal amplifier and a ringing response is tolerated. Smaller C_L will increase the separation between the dominant pole at V_{EA} and P_{OUT} and improve phase margin. The 30° phase margin of the local loop makes sure there are no significant peaking in the global loop's frequency response and the bode plot criteria is straight-forward here. During large step load changes, the system gets nonlinear behavior and time domain simulations are needed to determine the stability [27].

6.4 Simulated Load Transient Responses

The proposed LDO is designed in a 65nm CMOS process. The input voltage range is from 1.05 to 1.2 V and the dropout voltage is 150 mV under 20 mA load current. The total quiescent current including biasing circuits is 65 µA. The FVF and SSF in the local loop consume 35 µA and 15 µA, respectively, to boost the bandwidth. The EA consumes 8.5 µA quiescent current and the overshoot detection circuit costs 1.5 µA. With 10-100 pF loading capacitance (C_L), the proposed OCL-LDO can work stably from zero to the maximum loading of 20 mA.

Response time is fixed for a specific LDO as a reflection of both small and large-signal responses and has been used to compare different LDOs. The approximation in (1) assumes the edge time Tedge of load step is much smaller than the response time TR of the LDO which

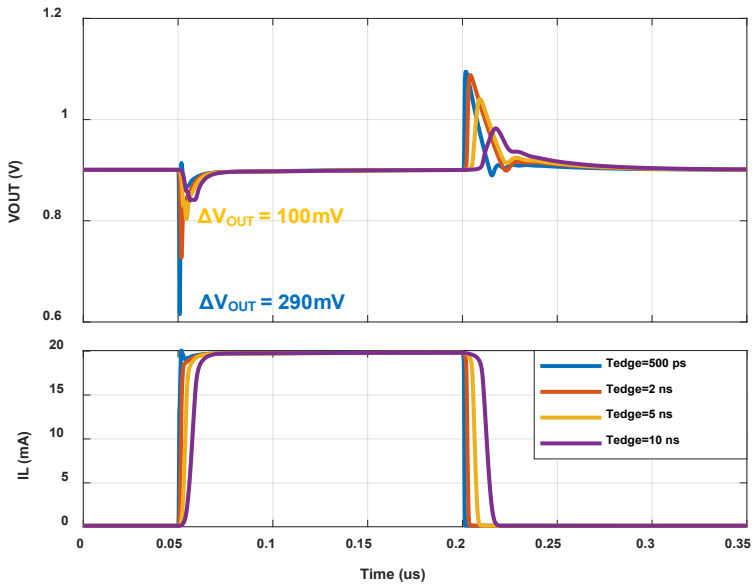

Figure 6.13 Simulated transient responses under load steps with different T_{edge} with C_L = 10 pF.

corresponds to the case described in Figure 6.12(a). However, this is not always true for ultrafast LDOs and Bu et al. proved that an inaccurate TR can be calculated using (3) when $T_{edge} > T_R$ [4]. For better estimation of T_R, they proposed a linear approximation model to calculate T_R using measured undershoot or overshoot results. A more explicit form was further developed in [8]:

$$T_R = \begin{cases} \sqrt{2C_L \Delta V_{OUT} T_{edge} / \Delta I_L}, T_R \le T_{edge} \\ C_L \Delta V_{OUT} / \Delta I_L + T_{edge} / 2, T_R > T_{edge} \end{cases}. \quad (6.16)$$

Equation (6.16) is used here to estimate T_R with measurement results to compare different LDOs' load transient responses.

As shown in Equation (6.1), C_L directly attenuates the voltage spikes, therefore the worst undershoot and overshoot during load steps happens when C_L = 10 pF. The simulated transient responses under load steps with different T_{edge} are provided in Figure 6.13. No bond wire or probe parasitic is included to make it correspond to the case when the LDO is used as an IP in a SoC. For large T_{edge}, the LDO starts to respond before the load step finishes and results in much smaller

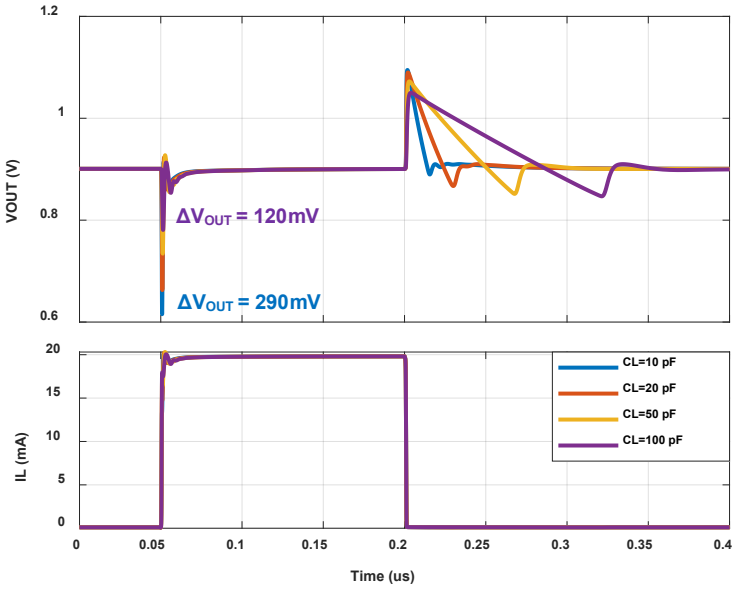

Figure 6.14 Simulated transient responses with different C_L under a load step with T_{edge} = 500 ps.

glitches at V_{OUT}. Although large C_L provides the instantaneous current for loads, it can also slow down the local loop and results in a larger response time. The simulated transient responses with different C_L under a load step with T_{edge} = 500 ps are provided in Figure 6.14. The voltage spikes at the LDO's output are not inversely proportional to C_L values and the overshoot settling time is much larger with C_L = 100 pF.

6.5 Measurement Results and Discussion

The proposed OCL-LDO has been implemented in a 65nm technology. The chip micrograph is shown in Figure 6.15. The total active area of the LDO itself is only 0.01 mm^2, making it suitable for SoC integrations. A R_{L1} = 45 Ω in series with a switch S_1 controlled by a buffered control signal V_{CLK1} to generate ultrafast load current steps is integrated on-chip as shown in Figure 6.16. The edge time T_{edge1} of the current transitions are about 500 ps in simulations. A C_L = 10 pF representing the bypass capacitance from load is added on-chip at the LDO's output for evaluation of worst-case V_{OUT} spikes during load transients. Two switches for V_{REF} switching

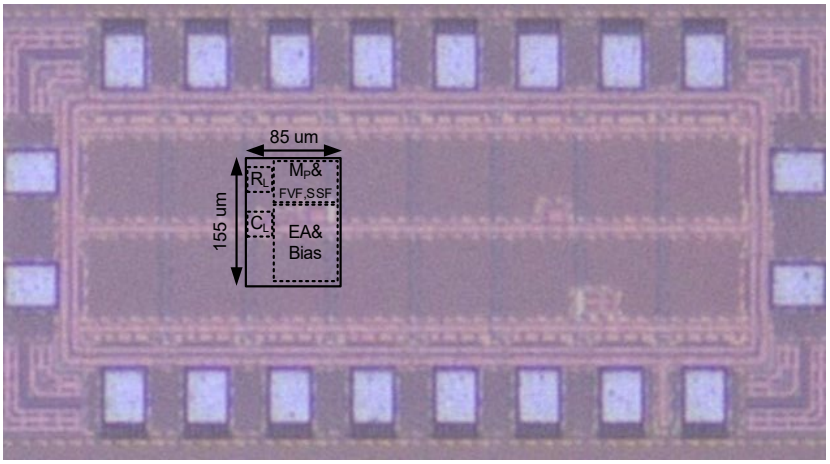

Figure 6.15 Chip micrograph.

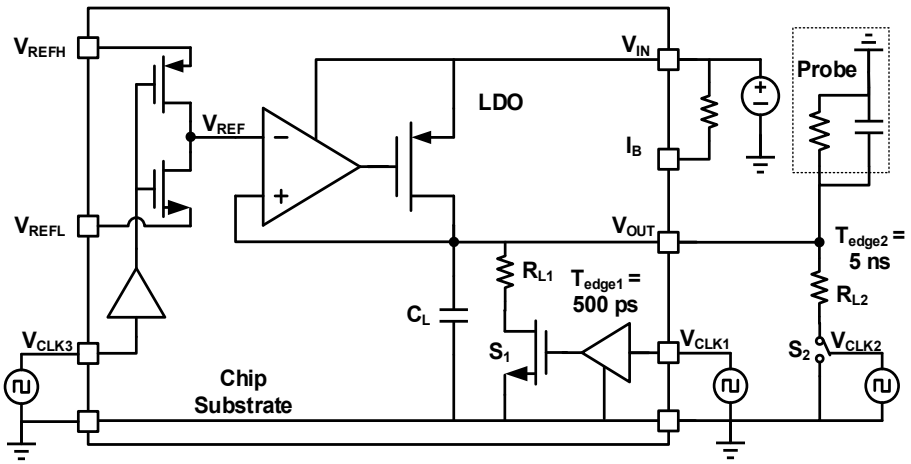

Figure 6.16 Testing setup for investigating the transient responses of the proposed OCL-LDO (architecture not redrawn here).

are also integrated and they are controlled by buffered V_{CLK3} to achieve quick V_{REF} steps. Note that the proposed LDO structure shown in Figure 6.2 is not drawn here for simplicity. An I_B = 500 nA is generated on the test board and provided to the LDO. There is another R_{L2} = 45 Ω in series with a switch S_2 controlled by V_{CLK2} to generate load steps with T_{edge2} = 5 ns. A 9 kΩ resistor is loaded on the board for light load condition (I_L = 0.1 mA).

A 4 GHz MSO9404 mixed-signal scope is utilized to capture the V_{OUT} transient waveforms. V_{CLK1} and V_{CLK2} are generated with a DE-115 FPGA board and has a rise/fall time

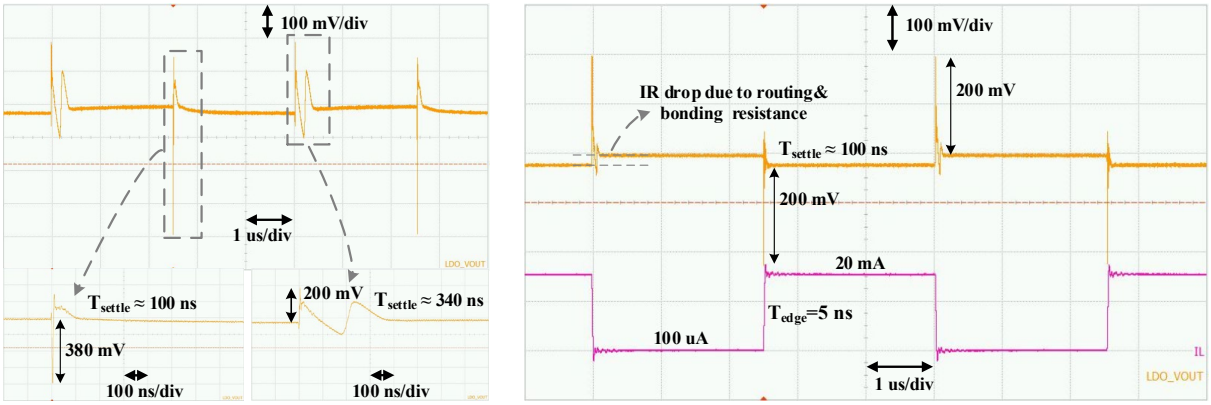

Figure 6.17 Measured load transient responses at V_{IN} = 1.1 V, V_{OUT} = 0.9 V with C_L = 10 pF under (a) $T_{edge1} \approx$ 500 ps and (b) T_{edge2} = 5 ns.

around 10 ns. An analog switch (TS3A4741) with 1 Ω on resistance and 5 ns turn-on/off time is used as S_2.

6.5.1 Load Transient Measurement

Figure 6.17 shows the load transient response with 500-ps and 5-ns edge time when C_L is 10 pF, whereas Figure 6.18 shows the results with C_L = 100 pF (an extra 90 pF capacitor is added on board). The load current step for 5-ns edge time is tracked with the method used in [28]. Due to the limited pad number during fabrication, we didn't add a separate sense pad for the LDO output. For on-board load steps shown in Figure 6.17(b) and Figure 6.18(b), settled V_{OUT} differences between 0.1 mA and 20 mA load are about 25 mV. This is mostly due to the pad resistance, the bond wire resistance from the LDO's output to the package and the trace resistance on the test board. For the on-chip load steps, the V_{OUT} difference is much smaller considering only the routing resistance of V_{OUT}. In real applications, however, bond wires do not exist in SoCs, and the load regulation will be therefore much better. In Figure 6.17(a), the overshoot response has double peaking and results in a longer settling time. The overshoot reduction in the proposed structure is a nonlinear activation and it can fight with the fast local loop for quick turn on once if

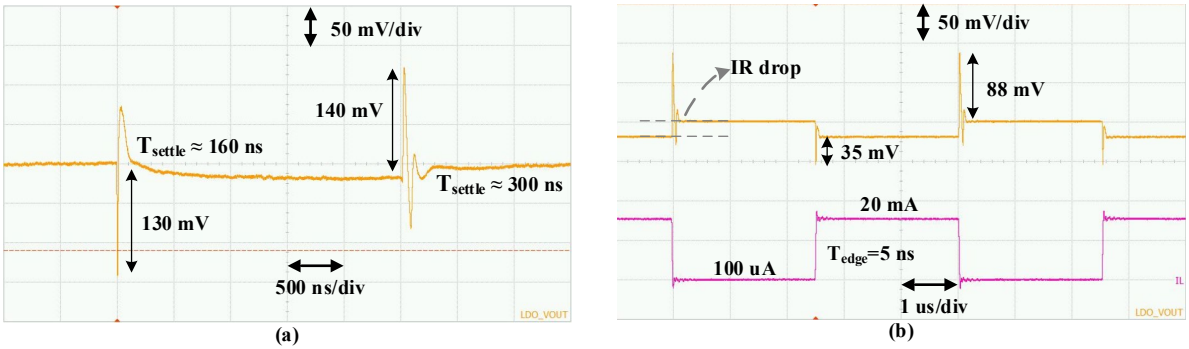

Figure 6.18 Measured load transient responses at V_{IN} = 1.1 V, V_{OUT} = 0.9 V with C_L = 100 pF under (a) $T_{edge1} \approx$ 500 ps and (b) T_{edge2} = 5 ns.

it overacts. However, its strength can be controlled by changing $M_{OD,1}$ leaking current or current mirror $M_{OD,4}/M_5$ ratio.

The undershoot with ~500-ps Tedge1 and 10 pF CL is 380 mV and reduced to 200 mV with 5-ns Tedge2. In both cases, the undershoot can settle within 100 ns. The 90mV increase compared to simulations can result from parasitic capacitor, gm and on-chip CL variations post fabrication. The bondwire inductance also makes the waveform worse than simulated, but this worsening is expected and was due to how we packaged it. Using (6.16) with pad and bond wire capacitance included in calculation, the corresponding T_R is 0.6 ns and 1.2 ns, respectively. The difference here results from the fact that the actual on-chip load step Tedge1 after fabrication can be larger than the simulated 500 ps. Since T_{edge2} is well defined in the analog switch TS3A4741's datasheet and also verified in measurement, the 1.2 ns response time is used in the comparison with state-of-the-art LDOs. The undershoot magnitudes are reduced with large C_L =100 pF as shown in Figure 6.18(b). The 0.1 mA to 20 mA load step with 5-ns edge time only causes 35 mV spike.

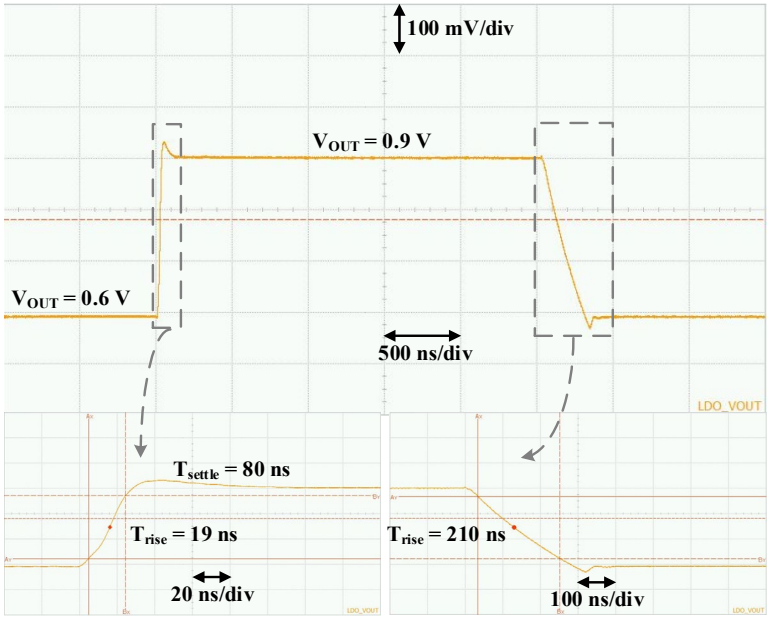

Figure 6.19 Measured V_{REF} transient responses at V_{IN} = 1.1 V, I_L = 0.1 mA with C_L = 10 pF.

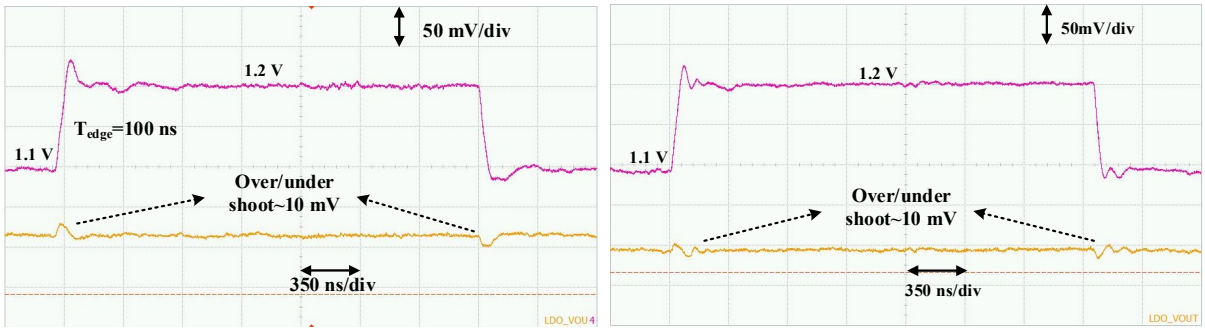

Figure 6.20 Measured line transient responses at V_{OUT} = 0.9 V, C_L = 10 pF with (a) I_L = 20 mA and (b) I_L = 0.1 mA.

6.5.2 V_{REF} and V_{IN} Transient Measurement

V_{REF} step responses are measured at light load condition (I_L = 0.1 mA) to evaluate the slew rate improvement of the error amplifier. As shown in Figure 6.19, the dynamic turn-around stage in the error amplifier improves the positive phase slew rate by 10 times compared to the negative phase where the fixed bias limits the slewing. The LDO's output settles within 80 ns when $V_{REF}$is

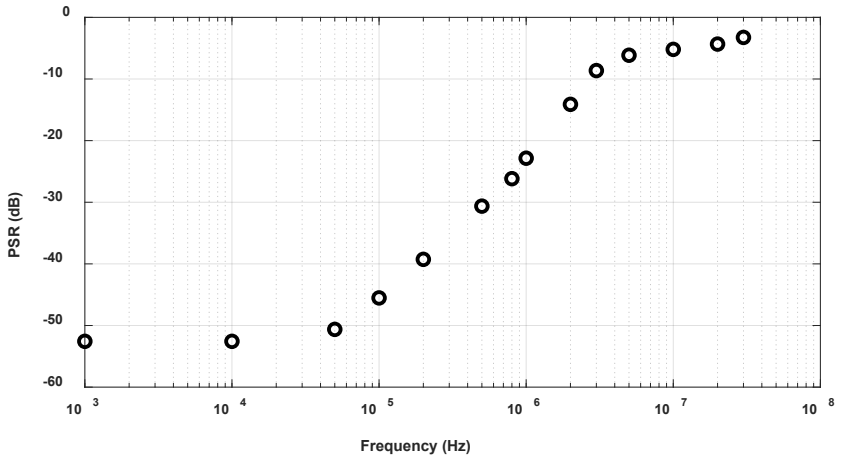

Figure 6.21 Measured PSR versus frequency at V_{OUT} = 0.9 V, I_L = 20 mA with C_L = 10 pF.

stepped from 400 mV to 600 mV with Tedge ≈ 500 ps. If needed, a flipped version can be added to boost the negative slewing and it consumes little current and area overhead.

The line transient responses at full and light loads are also measured as shown in Figure 6.20. A voltage waveform generator is used to generate the V_{IN} steps. The glitches at the LDO's output is small (~10 mV) and settles quickly for 100 mV V_{IN} steps with 100 ns edge time.

6.5.3 PSR Measurement

PSR is not the focus of this work as the proposed OCL-LDO hasn't specifically implemented any PSR improvement techniques. For completeness, the measurement results of the proposed LDO at 20 mA load is provided in Figure 6.21. The same voltage waveform generator is used to generate a 1.1 V voltage with 100 mV peak-to-peak sine wave ripples as the supply of the LDO. The PSR is calculated using the following equation:

$$PSR = 20\log\left(\frac{V_{P-P,V_{OUT}}}{V_{P-P,V_{IN}}}\right). \tag{6.17}$$

The low-frequency PSR is about -52 dB which is close to the simulated dc gain of the global loop. The used waveform generator has a limited bandwidth and the ripple magnitude starts dropping at

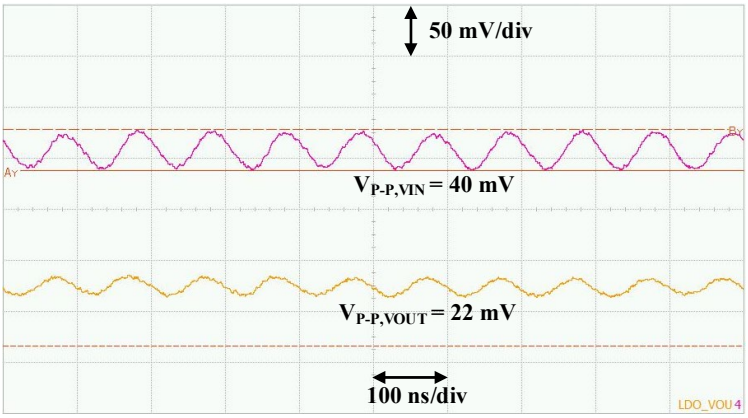

Figure 6.22 Measured transient waveforms at full load for PSR calculation at 10 MHz.

5 MHz, so the measurement is stopped at 20 MHz. The relevant transient waveforms at 10 MHz frequency are given in Figure 6.22 and the corresponding PSR is -5.2 dB.

6.5.4 Comparison with State-of-the-art

Table 6.1 has compared the performance of the proposed OCL-LDO with several state-of-the-art fast transient LDOs. The widely used figure-of-merit (FOM) proposed in [29]:

$$\text{FoM} = \frac{T_R I_Q}{I_{L(MAX)}}, \tag{6.18}$$

is adopted here to compare the transient performance. This FoM obviously favors low power subthreshold design which has natural disadvantages, e.g., sensitive to variations and mismatch. In the proposed design, the transient response is optimized with no signal carrying transistors working in subthreshold region. The proposed structure achieves 3.9 ps FoM with current efficiency η_c = 99.7 %, which is comparable to most state-of-the-art. The LDO in [4] has 4 times better FoM considering similar load current and PSR performance. Twice I_Q is burnt on boosting G_{m2} in the complex nested miller compensation to achieve unity-gain bandwidth > 100 MHz. In [8], a much smaller I_Q is used and the resulted response time and settling time are 10 times larger than the proposed structure.

Table 6.1 Performance Summary and Comparison with State-of-the-art OCL-LDOs

	JSSC'05 [29]	JSSC'14 [11]	TCAS-I'15 [2]	TPEL'18 [4]	TCAS-II'19 [13]	TPEL'20 [8]	This work
Technology	90nm	65nm	65nm	130nm	180nm	65nm	65nm
Dropout (mV)	300	200	150	200	200	150	150
V_{DD} (V)	1.2	0.75-1.2	1.2	1-1.4	1.2-1.8	0.95-1.2	1.05-1.2
V_{OUT} (V)	0.9	0.55	1	0.8	0.8-1.6	0.8	0.9
I_L(MAX) (mA)	100	50	10	25	100	100	20
I_Q (uA)	6000	15.9-487	50-90	120	10.2	14	65
Current efficiency	94.34%	99.04%	99.11%	99.55%	99.99%	99.99%	99.7%
C_L (pF)	600$^{\gamma}$	470-10000	130	0-25	100	0-100	0-100
On-chip cap* (pF)	0	4.1	10	0.73	0	6	1.4
UGB$^{\#}$ (MHz)	N/A	0.9-2.5$^{\eta}$	8-15$^{\eta}$	100.2	2.2-7.5	0.66-9.03	6/53$^{\zeta}$
PSR (dB@MHz)	N/A	-8@1	-18@10	-57@1	N/A	-33@0.01	-52@0.01
			-15.5@1000	-22@10		-6@1	-23@1
Area (mm^2)	0.098	0.013	0.023	0.008	0.031	0.01	0.01
T_R@T_{edge} (ns@ns)	0.54@0.1	14.58@100	1.15@0.2	0.2@0.3	61@100	10@220	1.2@5
Undershoot settling time (ns)	N/A	250	70$^{\eta}$	50	220	1000$^{\eta}$	100
FOM (ps)	32.4	4.67	5.74-10.35	0.9	6.1	1.4	3.9

$^{\gamma}C_L$ implemented on-chip
*On-chip capacitors for compensation, coupling, or bias, excluding C_L.
$^{\#}$Simulated results.
$^{\eta}$Estimated from reported figures.
$^{\zeta}$Global loop/local loop

The proposed OCL-LDO has good rejection to low frequency supply noise and still has -23 dB PSR at 1 MHz. Excluding the on-chip R_L and C_L integrated for measurement purpose, the stand-alone LDO occupies only 0.01 mm^2, benefited from the simplicity of the structure. With a maximum load current of 20 mA, the prototype is more area-efficient than the LDO in [2] which has narrower load range and requires large on-chip compensation capacitors.

6.6 Conclusion

This chapter introduces an output-capacitorless LDO with ultra-fast transient performance in a 65 nm CMOS process. The super source follower and flipped voltage follower based local loop achieves large bandwidth and high slew rate with reasonable current consumption. Combined

with an overshoot detection circuit, small voltage spikes and fast transient settling are achieved under load, reference, supply steps. The LDO requires only 0.01 mm^2 area with 1.4 pF total on-chip capacitance (excluding C_L) and consumes 65 µA quiescent current. The overall structure of the proposed LDO is simple and straightforward but achieves a FoM comparable to the state-of-the-art. The achieved specifications make the proposed LDO suitable for driving digital circuits in SoC applications.

6.7 References

[1] Q. H. Duong et al., "Multiple-Loop Design Technique for High-Performance Low-Dropout Regulator," *IEEE J. Solid-State Circuits*, vol. 52, no. 10, pp. 2533–2549, Oct. 2017, doi: 10.1109/JSSC.2017.2717922.

[2] Y. Lu, Y. Wang, Q. Pan, W. Ki, and C. P. Yue, "A Fully-Integrated Low-Dropout Regulator With Full-Spectrum Power Supply Rejection," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 62, no. 3, pp. 707–716, Mar. 2015, doi: 10.1109/TCSI.2014.2380644.

[3] M. Al-Shyoukh, H. Lee, and R. Perez, "A Transient-Enhanced Low-Quiescent Current Low-Dropout Regulator With Buffer Impedance Attenuation," *IEEE J. Solid-State Circuits*, vol. 42, no. 8, pp. 1732–1742, Aug. 2007, doi: 10.1109/JSSC.2007.900281.

[4] S. Bu, J. Guo, and K. N. Leung, "A 200-ps-Response-Time Output-Capacitorless Low-Dropout Regulator With Unity-Gain Bandwidth gt;100 MHz in 130-nm CMOS," *IEEE Trans. Power Electron.*, vol. 33, no. 4, pp. 3232–3246, Apr. 2018, doi: 10.1109/TPEL.2017.2711017.

[5] V. Ivanov, "Design Methodology and Circuit Techniques for Any-Load Stable LDOs with Instant Load Regulation and Low Noise," in Analog Circuit Design, M. Steyaert, A. H. M. van Roermund, and H. Casier, Eds. Dordrecht: Springer Netherlands, 2009, pp. 339–358.

[6] S. K. Lau, P. K. T. Mok, and K. N. Leung, "A Low-Dropout Regulator for SoC With Q-Reduction," *IEEE J. Solid-State Circuits*, vol. 42, no. 3, pp. 658–664, Mar. 2007, doi: 10.1109/JSSC.2006.891496.

[7] J. Guo and K. N. Leung, "A 6-µW Chip-Area-Efficient Output-Capacitorless LDO in 90-nm CMOS Technology," *IEEE J. Solid-State Circuits*, vol. 45, no. 9, pp. 1896–1905, Sep. 2010, doi: 10.1109/JSSC.2010.2053859.

[8] G. Li and Y. Lu, "Dual Active-Feedback Frequency Compensation for Output-Capacitorless LDO With Transient and Stability Enhancement in 65-nm CMOS," *IEEE Trans. POWER Electron.*, vol. 35, no. 1, p. 15, 2020.

[9] G. Rincon-Mora, Analog IC Design with Low-Dropout Regulators (LDOs), 1st ed. USA: McGraw-Hill, Inc., 2009.

[10] Ka Nang Leung and P. K. T. Mok, "A capacitor-free CMOS low-dropout regulator with damping-factor-control frequency compensation," *IEEE J. Solid-State Circuits*, vol. 38, no. 10, pp. 1691–1702, Oct. 2003, doi: 10.1109/JSSC.2003.817256.

[11] X. L. Tan, S. S. Chong, P. K. Chan, and U. Dasgupta, "A LDO Regulator With Weighted Current Feedback Technique for 0.47 nF–10 nF Capacitive Load," *IEEE J. Solid-State Circuits*, vol. 49, no. 11, pp. 2658–2672, Nov. 2014, doi: 10.1109/JSSC.2014.2346762.

[12] G. A. Rincon-Mora and P. E. Allen, "A low-voltage, low quiescent current, low drop-out regulator," *IEEE J. Solid-State Circuits*, vol. 33, no. 1, pp. 36–44, Jan. 1998, doi: 10.1109/4.654935.

[13] J. Tang, J. Lee, and J. Roh, "Low-Power Fast-Transient Capacitor-Less LDO Regulator With High Slew-Rate Class-AB Amplifier," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 66, no. 3, pp. 462–466, Mar. 2019, doi: 10.1109/TCSII.2018.2865254.

[14] T. Y. Man, K. N. Leung, C. Y. Leung, P. K. T. Mok, and M. Chan, "Development of Single-Transistor-Control LDO Based on Flipped Voltage Follower for SoC," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 55, no. 5, pp. 1392–1401, Jun. 2008, doi: 10.1109/TCSI.2008.916568.

[15] P. R. Surkanti, A. Garimella, M. Manda, and P. M. Furth, "On the analysis of low output impedance characteristic of flipped voltage follower (FVF) and FVF LDOs," in *2017 IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2017, pp. 17–20, doi: 10.1109/MWSCAS.2017.8052849.

[16] R. Magod, B. Bakkaloglu and S. Manandhar, "A 1.24 µA Quiescent Current NMOS Low Dropout Regulator With Integrated Low-Power Oscillator-Driven Charge-Pump and Switched-Capacitor Pole Tracking Compensation," in *IEEE Journal of Solid-State Circuits*, vol. 53, no. 8, pp. 2356-2367, Aug. 2018, doi: 10.1109/JSSC.2018.2820708..

[17] R. J. Reay and G. T. A. Kovacs, "An unconditionally stable two-stage CMOS amplifier," *IEEE J. Solid-State Circuits*, vol. 30, no. 5, pp. 591–594, May 1995, doi: 10.1109/4.384174.

[18] J. Hu, J. H. Huijsing, and K. A. A. Makinwa, "A Three-Stage Amplifier with Quenched Multipath Frequency Compensation for All Capacitive Loads," in *2007 IEEE International Symposium on Circuits and Systems*, 2007, pp. 225–228, doi: 10.1109/ISCAS.2007.378317.

[19] N. Liu, B. Johnson, V. Nadig, and D. Chen, "A Transient-Enhanced Fully-Integrated LDO Regulator for SoC Application," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1–5, doi: 10.1109/ISCAS.2018.8351236.

[20] Y. Huang, Y. Lu, F. Maloberti, and R. P. Martins, "Nano-Ampere Low-Dropout Regulator Designs for IoT Devices," *IEEE Trans. Circuits Syst. Regul. Pap.*, pp. 1–10, 2018, doi: 10.1109/TCSI.2018.2851226.

[21] J. Pérez-Bailón, A. Márquez, B. Calvo, and N. Medrano, "Transient-enhanced output-capacitorless CMOS LDO regulator for battery-operated systems," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2017, pp. 1–4, doi: 10.1109/ISCAS.2017.8050961.

[22] B. Huang, "Performance enhancement techniques for operational amplifiers," Grad. Theses Diss., Jan. 2017.

[23] J. M. Rabaey, A. P. Chandrakasan, and B. Nikolic´, *Digital Integrated Circuits*, 2nd ed. London, U.K.: Pearson, 2003.

[24] Y.-H. Lee et al., "A DVS Embedded Power Management for High Efficiency Integrated SoC in UWB System," *IEEE J. Solid-State Circuits*, vol. 45, no. 11, pp. 2227–2238, Nov. 2010, doi: 10.1109/JSSC.2010.2063610.

[25] Y.-H. Lee et al., "A Low Quiescent Current Asynchronous Digital-LDO With PLL-Modulated Fast-DVS Power Management in 40 nm SoC for MIPS Performance Improvement," *IEEE J. Solid-State Circuits*, vol. 48, no. 4, pp. 1018–1030, Apr. 2013, doi: 10.1109/JSSC.2013.2237991.

[26] M. Huang, Y. Lu, S.-W. Sin, U. Seng-Pan, and R. P. Martins, "A Fully Integrated Digital LDO With Coarse–Fine-Tuning and Burst-Mode Operation," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 63, no. 7, pp. 683–687, Jul. 2016, doi: 10.1109/TCSII.2016.2530094.

[27] H. K. Khalil, Nonlinear Systems, 3rd ed. London, U.K.: Pearson, 2002.

[28] P. Y. Or and K. N. Leung, "An Output-Capacitorless Low-Dropout Regulator With Direct Voltage-Spike Detection," *IEEE J. Solid-State Circuits*, vol. 45, no. 2, pp. 458–466, Feb. 2010, doi: 10.1109/JSSC.2009.2034805.

[29] P. Hazucha, T. Karnik, B. A. Bloechel, C. Parsons, D. Finan, and S. Borkar, "Area-efficient linear regulator with ultra-fast load regulation," *IEEE J. Solid-State Circuits*, vol. 40, no. 4, pp. 933–940, Apr. 2005, doi: 10.1109/JSSC.2004.842831.

CHAPTER 7. GENERAL CONCLUSIONS

Several techniques for designing and improving performance of common analog blocks in mixed-signal SoCs have been proposed, including string DAC, bandgap reference, and output-capacitorless low-dropout regulator. Both simulation and measurement results exhibited significant improvements with respect to linearity, temperature drift, and transient responses. The proposed string DAC with gradient errors suppression achieved 16-bit linearity performance which is the best linearity demonstrated in bulk CMOS technology. The developed BGR prototype realized with both on-chip devices and off-chip components reflected a sub-ppm/°C, or 0.8 ppm/°C to be exact, temperature coefficient over a temperature range of 0 °C to 80 °C after optimal trimming. The proposed dual loop OCL-LDO settled in 100 ns for load steps from 100 µA to 20 mA as well as V_{REF} and V_{IN} steps and achieved a FoM comparable to the state-of-the-art.

The proposed techniques are practical, straightforward, and independent of the technology node used. The presented string DAC can be utilized as an MSB stage and be combined with a fine ladder or an R-2R architecture to build a low-cost high-resolution RDAC with good linearity performance. The proposed curvature cancellation approach utilizing natural base expansion and self-bootstrapping can be implemented on any of the widely used Banba, Kuijk, or Brokaw structures to build sub-ppm/ºC references. The dual loop OCL-LDO structure introduced here achieves fast transient response and good regulation with reasonable current consumption, making it suitable for driving fast digital loads in SoCs.

In summary, the proposed techniques can be easily adopted by SoC or pure analog designers to help recover analog performance. Looking into the future, gradient insensitive patterns for binary weighted arrays are also worth investigations, such as capacitor arrays in successive-approximation-register (SAR) ADCs. With respect to the bandgap reference, more

dedicated filtering can be implemented to reduce the output noise and further improvements to reduce the number of op-amps in the structure to reduce cost seem possible. For the OCL-LDO, its PSR at high frequency can be improved so that it can also power noise sensitive precision blocks in SoCs.