Exploiting Co-Design & Computation for Energy-efficient Systems On Chip

Xun Sun

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Visvesh Sathe, Chair

Michael Taylor

Baosen Zhang

Program Authorized to Offer Degree:
Department of Electrical and Computer Engineering

University of Washington

Abstract

Exploiting Co-Design & Computation for
Energy-efficient Systems On Chip

Xun Sun

Chair of the Supervisory Committee:
Associate Professor Visvesh Sathe
Electrical and Computer Engineering

System-on-Chips (SoC) are the engines of modern computing. Unfortunately, with the plateauing of performance and energy efficiency benefits provided by each new process technology, power dissipation has become an overriding concern in computing. As an increasing proportion of worldwide energy resources continue to be devoted to high-performance computing (HPC), cloud computing, and pervasive mobile computing, energy efficient SoC design plays a key role thwarting an imminent resource crisis. In SoCs today, voltage margins due to supply noise or PVT variations are a dominant source of inefficiency. This dissertation describes research into two new approaches to address this challenge (1) clocking/voltage co-design and (2) exploiting advances in computing in service of more energy efficient SoC design. Both techniques directly address the problem of SoC voltage margins, the dominant source of inefficiency in CMOS-based computing. It is expected that continued advances in this field will enhance both the effectiveness and applicability of these design approaches.

The key idea behind clocking/voltage co-design is to unify traditionally independent clock and voltage regulation loops and thus minimize guardband requirements for supply noise and temperature variations. This thesis presents two versions of unified clock and power (*UniCaP*) architecture. For the first *UniCaP* work, the rationale for combining voltage and clock regulation loops into a single loop is discussed. This joint loop uses the

supply voltage of the domain (V_{dd}) as the control variable to adjust the operating clock frequency (f_{clk}) and thus lock system clock to *REFCLK* in order to drastically reduce voltage guardbands while providing performance guarantees. The *UniCaP* architecture is deployed on a 65nm buck-converter based test-chip, demonstrating temperature margin reductions of 40-55mV and 82% average V_{dd} guardband reduction across 0.6-1.0V without any performance loss from adaptive clocking.

While the first *UniCaP* prototype has demonstrated aggressive margin recovery while regulating average system performance, slow recovery of V_{dd} droop due to loop dynamics and control techniques in *UniCaP* has limited its applicability in certain applications requiring performance guarantees in sub-20 µs time frames. *UniCaP*-2, a dual-loop phase-locked adaptive clocking architecture was then proposed to reduce peak cycle loss ($\Delta\Phi_{max}$) and cycle-loss recovery time ($T_{recovery}$) while enabling higher V_{dd} margin reduction. Measurements on a 65nm test chip demonstrate 91-99% V_{dd} margin reduction and 38X $T_{recovery}$ improvement over first *UniCaP* prototype. The thesis also quantifies the impact of clock distribution delay (τ_{dist}) and V_{dd} sensitivity on V_{dd} margin reduction, an open and critical question surrounding adaptive clocking techniques.

Because supply voltage V_{dd} plays a dominant role in determining energy dissipation, V_{dd} scaling remains the salient approach to energy efficient CMOS computing. Integrated Voltage Regulation (IVR) – enabling components of an SoC to be partitioned into multiple V_{dd} domains, each regulated by an on-chip voltage regulator – dramatically enhances the effectiveness and applicability of voltage scaling. Fast and stable voltage regulator transient response helps minimize V_{dd} margins required to withstand load current transients that are common in digital systems. This dissertation discusses how computation can be exploited to achieve improved voltage regulation. Computational regulation is presented on different VR modalities to enhance transient performance. The concept is first demonstrated in a Digital LDO and relies on an accurate time-domain model, using low-latency computation to evaluate the state equations at run time for a rapid response. Measurement results from a 65nm CMOS test chip demonstrate a 2.9-cycle mean settling time for processor

load variations. Using this fundamental concept, a novel digital control architecture, low-complexity and low-latency Model Predictive Control (MPC), is deployed in integrated voltage regulators (IVRs). Control and datapath optimizations, combined with the advances in computational speed in modern CMOS technology nodes, are key enablers for realizing an IVR buck with optimal transient response. The proposed MPC-buck architecture was implemented in 65nm CMOS. The test-chip achieves a measured 2.49X settling-time (τ_{settle}) improvement over an optimally tuned Proportional Integral Differential (PID) controller.

Typical VR designs include large stability margins to ensure stable response even in worst case scenarios, resulting in overly conservative designs that degrade VR response time. Two methods are presented in this thesis to handle the worst-case stability margining issue. First, Autonomous Gain Tracking (AGT), a low-overhead, low-complexity technique for run-time loop gain adaptation is introduced to the computational LDO. The main idea is to track variations in loop gain by examining 1-bit statistics of the output voltage. Measurement results indicate that AGT successfully performs the broad gain adjustments required for consistently fast transient responses across varying V_{in}, V_{dd}, and temperature conditions. AGT is built upon the characteristics of computation LDO response so it cannot be directly applied to other controller designs. The operation of AGT will also be impacted by engineered current waveforms that break the assumption of random I_{load} noise. To address the limitations of AGT method, the thesis proposes to utilize a pseudo-random binary sequence (PRBS) based approach to excite the dynamics of a feedback loop and automatically adjust loop gain based on the cross-correlation between the injected PRBS noise signal and control output. Calculating the cross-correlation enables identification of the closed-loop impulse response and hence adjustment of control parameters accordingly.

TABLE OF CONTENTS

Page

List of Figures iii

Chapter 1: Introduction 1
1.1 Overview 1
1.2 Organization 4

Chapter 2: A Combined all-digital PLL-buck slack regulation system with autonomous CCM/DCM transition control 6
2.1 Introduction 6
2.2 Unified Clock and Power Architecture 8
2.3 *UniCaP* System Implementation 11
2.4 All-digital ZVS Operation and CCM/DCM Transition 17
2.5 Measurement Results 20
2.6 Conclusion 28

Chapter 3: Unicap-2: Phase-Locked Adaptive Clocking with Rapid Clock Cycle Recovery in 65nm CMOS 30
3.1 Introduction 30
3.2 Architecture 30
3.3 Silicon Measurements 33

Chapter 4: A computationally regulated digital LDO with 2.9-cycle mean settling time 37
4.1 Introduction 37
4.2 Challenges in Traditional DLDO Design 39
4.3 Computationally-Enhanced DLDO: Architecture 42
4.4 System Implementation 44
4.5 Measurement Results 50
4.6 Conclusion 56

Chapter 5: Model Predictive Control of Integrated Buck Converter for Digital SoC Domains 57
5.1 Introduction 57
5.2 MPC Buck Architecture 57
5.3 Test Chip and Measurement Results 61

Chapter 6: Auto-tuning of DLDO digital controllers 65
6.1 Introduction 65
6.2 AGT Architecture 65
6.3 AGT Implementation 68
6.4 AGT Measurement Results 71
6.5 Limitations of AGT Technique 74
6.6 PRBS-based Cross-correlation Method 75

Chapter 7: Conclusion and future work 79

Bibliography 82

LIST OF FIGURES

Figure Number — Page

1.1 Variations in Process, Temperature and Voltage require substantial performance/efficiency degrading frequency guardbands. 2

2.1 Conventional digital systems with (a) Independent voltage and clock regulation loops. (b) Dissipative voltage margins to maintain timing-slack under worst-case voltage droop. 7

2.2 (a) Proposed *UniCaP* architecture regulating operating frequency through V_{dd} control. (b) Elastic clock maintains correct function by compensating for supply noise induced critical path delay variation. 9

2.3 (a) Block diagram of regulating TRO frequency through a traditional PLL architecture. (b) Slow voltage droop causes timing failure. 10

2.4 Architecture overview of the buck-converter based *UniCaP* system. 11

2.5 Circuit diagram of TRO consisting of wire and logic dominated delay cells. . 12

2.6 (a) Block diagram of the frequency acquisition loop. (b) Root-locus plot demonstrates stable closed frequency-acquisition loop response. 14

2.7 Root-locus plot indicates an unstable phase loop for all K values with conventional PID control. 15

2.8 (a) Block diagram of composite control with frequency and phase path operating at different rates. (b) Equivalent phase loop model with model reduction technique applied. $G_{vd}(z)$ is the discrete-time representation of $G_{vd}(s)$ (c) Stable closed loop response for the reduced order system. 16

2.9 Circuit and timing diagram of the proposed DLL-enabled ZVS operation. . . 18

2.10 Circuit diagram of the V_X gradient based zero current detection method. . . 19

2.11 (a) CCM/DCM transition control finite state machine. (b) Timing diagram illustrates autonomous CCM/DCM transition through a combination of PWM and PSM. 19

2.12 Die micrograph. 21

2.13 (a) Measured locked ZVS-DLL code vs. load current. (DLL code proportional to ZVS dead time) (b) Measured CCM/DCM transition waveforms 22

2.14 Measured buck converter efficiency 23

2.15 Measured transient voltage response under load current step with phase and frequency lock configurations (a) step-up (b) step-down. 23

2.16 (a) Measured Cortex M0 f_{max} vs. V_{dd} illustrating 96% margin recovery through *UniCaP* at 320MHz. (b) Measured V_{dd} margin recovery loss vs. TRO calibration step size across 0.6-1.0V. 25
2.17 Measured V_{dd} vs. temp. with f_{targ} = 40MHz and 320MHz. 27
2.18 Measured oscilloscope capture of on-the-fly DVFS transitions by varying PLL divider ratio. 28
2.19 Measured total energy-per-cycle vs. operating frequency plot. 29
2.20 Normalized clock rate vs. droop event frequency for conventional, discrete clock stretching and *UniCaP* systems. 29

3.1 Construction and transient operation of (a) *UniCaP*, which is slow and leads to larger V_{dd} droop and (b) this work (*UniCaP*-2), which allows rapid V_{dd} recovery and phase lock. 31
3.2 *UniCaP*-2 test-chip architecture. A fast V_{dd} loop limits droop. A slower phase loop achieves phase lock. 32
3.3 Proposed TRO design achieves finer delay and V_{dd} sensitivity control from improved τ_{crit} and s_{crit} matching. 32
3.4 *UniCaP*-2 die photograph. 33
3.5 Measured V_{dd} margin reduction with *UniCaP*-2 under 100mV supply noise conditions @ V_{dd}=1V. 34
3.6 Measured V_{dd} response to I_{load} step. Runtime values of $\Delta\Phi$ and $V_{dd,\Phi}$ are captured by the BIST engine. *UniCaP*-2 asserts adjusted V_{ref} adjustments to target rapid phase-lock. 35
3.7 Empirical setup to evaluate the impact of τ_{dist} and s_{dist} on *UniCaP* V_{dd} margins. 35
3.8 Margin recovery vs. τ_{dist} across (a) s_{dist}-s_{crit} mismatch and (b) $f_{v_{DD}}$. 36
3.9 Comparison with related work. 36

4.1 Overview of the proposed computationally enabled LDO architecture (CLDO). 37
4.2 (a) Simplified synchronous DLDO schematic. (b) DLDO voltage response to an I_{load} step. Discrete sampling of the error voltage (V_{ref}-V_{dd}) leads to a delayed response, which also depends on relative timing between the I_{load} step and the error-voltage sampling edge. (c) Simulated peak droop (normalized) resulting from a unit step current at various times within a DLDO clock cycle. 40
4.3 Ideal CLDO response to an I_{load} step with single cycle recovery. 43
4.4 Block-level description of the DLDO implementation. 44
4.5 DLDO timing diagram. 45

4.6 Comparator bank architecture used to quantize V_{dd}. A matching resistive network connects V_{dd} inputs to the comparators to mitigate kick-back noise. Suitable tap-off point selection from a chain of identical unit resistors provides non-uniform reference levels. Non-uniform quantization presents a suitable trade-off between the voltage resolution requirements and voltage capture range. 46
4.7 Loop delay results in a delayed update of $k[n]$. The timing model samples V_{dd} at the rising clock edge, but adjusts the measurement to determine $\widetilde{v}_o[n]$, the sampled voltage at the onset of applying $k[n]$, chosen to set $\widetilde{v}_o[n+1] = 0$. 48
4.8 (a) Datapath for computing $k[n]$. Modules marked in red are required to model loop delay. (b) Optimized signal-flow graph of the datapath, reducing the loop-delay contribution to one multiply-accumulate operation. 50
4.9 Die micrograph of proposed DLDO test-chip. 51
4.10 Measured voltage response to a load-step (a) without loop-delay modeling and (b) with loop delay modeling. Settling times are measured from the clock edge when the droop is first detected till V_{dd} recovers to within 1% of V_{ref} 52
4.11 Measured voltage response to a load-step under low V_{in}. 53
4.12 Histogram of LDO recovery time. Measurements were obtained at V_{in}=1.1V, V_{dd}=1.0V. 53
4.13 Measured voltage regulator response under random load conditions at V_{in}=1.1V, V_{dd}=1.0V, f_s=100MHz. 54
4.14 Measured load regulation and current-efficiency curves for the CLDO. 55
4.15 Comparison with related works. 56

5.1 MPC Architecture and non-linea buck model used by Solver. 58
5.2 MPC Constraint Equations of $d^*[n]$ for 2-cycle V_{dd} restoration. 58
5.3 Stable MPC response under $d[n]$n saturation through Q_s tracking. 59
5.4 Dataflow evaluating stability criterion with saturated d_n. 60
5.5 Dataflow graph for $d^*[n]$ computation. 60
5.6 MPC Test chip architecture. Delayed clocks provide time for Comparator and Solver evaluation with glitch-free operation. 61
5.7 Die Photograph. 62
5.8 Measured V_{dd} response to a load-step for MPC and optimally tuned PID controllers. Both incur $d[n]$ saturation resulting in similar V_{droop}. 63
5.9 Measured MPC V_{dd} waveforms under random loading conditions. 63
5.10 Measured MPC V_{dd} responses to I_{load} pulse with step-down occurring during τ_{settle}. 64

5.11 Comparison with related work. 64

6.1 Motivation for using the sign of $R_{vo}[1]$ to guide gain-tuning. Varied polarities of $R_{vo}[1]$ are indicative of varied conditions of regulator transient response. . 66

6.2 MATLAB simulation waveforms of calculated $R_{vo}[1]$ under ideal conditions, and under over-damped and under-damped conditions. $R_{vo}[1]$ is observed to be close to 0 under ideal tuning conditions. 67

6.3 Matlab simulation waveforms corresponding to full-precision and 1-bit precision tracking of $R_{vo}[1]$. The offset term corresponding to Equation 6.3 is verified to be sufficiently small to justify 1-bit characterization of sign($R_{vo}[1]$). 69

6.4 Block diagram of the AGT implementation relying on the sign-bit of the quantized v_o signal to auto-tune $G_{S,model}$. 70

6.5 CLDO voltage response to I_{load} step (a) with AGT disabled and (b) with AGT enabled. 71

6.6 Recorded values of $G_{S,model}$ with AGT enabled across target operating range of V_{dd}, V_{in} and temperature. 72

6.7 Run-time evolution of $G_{S,model}$ by AGT starting from three distinct and incorrect initial values. $G_{S,model}$ is found to converge in all three cases. 73

6.8 Matlab simulation of engineered current waveform to break the AGT loop and track solver gain $G_{S,model}$ incorrectly. 74

6.9 Architecture overview of PRBS-based cross-correlation technique. 75

6.10 Block-level description of the cross-correlation based auto-tuning method. . . 78

6.11 Simulation result of cross-correlation based impulse response adaption. 78

ACKNOWLEDGMENTS

First of all, I would like to express my sincere thanks to my advisor, Prof. Visvesh S. Sathe. I would like to thank him for offering me the opportunity to join his group and pursue my Ph.D. degree at UW starting in 2016. I would not have been able to finish all of these projects within five years without his guidance and support. During my first year at UW, I was fresh to the area of DC-DC converter design, had limited tape-out experience and could not imagine being able to tape out my own chip at the end of my first year. Professor Sathe was closely involved in my first project and provided a lot of useful insights and even direct help with debugging my code when I encountered any problems. It was the success of my first chip that boosted my confidence in tackling challenges of other research projects. In addition to giving valuable advice for my research, Prof. Sathe also helped me develop essential communication, writing and presentation skills needed for a successful career after graduation.

I would also like to thank Prof. Zhang not only for serving as the committee for both my general exam and final exam, but also for his willingness to utilize his expertise in control theory and share with us his knowledge about Model-Predictive Control which helped speed up the writing process of the MPC-buck paper. My sincere thanks to Prof. Taylor and Prof. Fuller for being the defense committee and providing insightful feedback and comments.

My PhD projects would not have been possible without the support from the project funding sources. I would like to thank Semiconductor Research Corporation, Intel Corporation and Qualcomm for the financial grants. Also, thank you to all industry liaisons for providing valuable advice for our SRC projects. My sincere thanks to Nasser Kurd's group at Intel for giving a lot of useful insights and comments from the industry point of view during our regular project meetings. Appreciation is also due to Keith Bowman from Qualcomm. His expertise in clock-power management and commitment to excellence helped my

research in this area and its subsequent publications and conference presentations.

I find myself lucky to have the chance to work with the outstanding engineers in industry. I am very grateful to Prof. Sathe who introduced me to my internships at Circuit Research Lab of Intel. During my three internships at CRL, I was fortunate to have some amazing managers/mentors. Carlos Tokunaga was my mentor and manager for my two internships at Intel. He encouraged me to explore various directions for my internship project and provided support for me when I encountered difficulties in the area that I was not familiar with. Harish K Krishnamurthy was my mentor when I interned at the power delivery circuit and research team at CRL. Regular discussion with Harish helped accelerate my internship project and without his guidance and support, it would have not been possible to submit a patent application for my internship project. I would also like to thank my internship managers, Muhammad Khellah and Krishnan Ravichandran, and collegues, Suyoung Bang, Minki Cho and Saurabh Kumar for their support during my internship.

I would like to thank Dr. Rajesh Pamula who had worked as a postdoc in our group. He is a great person – very helpful, friendly and passionate about circuit design. When I have any tapeout process or circuit design related questions, he would be my go-to person for advice and he is always patient even when he is on a tight tapeout schedule and able to explain complicated things in an intuitive way. Thank you to my fellow labmates. Sung Kim, Naveen John, Xi Li and Fahim ur Rahman helped me a lot for my first *UniCaP* chip. Akshat Boora took the responsibility to design the ADC for two of my chip tapeouts. My two computational regulation chips would not have been successful without the properly designed ADC by Akshat. I would also like to thank Chi-hsiang Huang and Yidong Chen for their significant contributions to my *UniCaP*-2 project. Many thanks to Diego Pena, Arindam Mandal, Kevin Patino Sosa and Huwan Peng and Cindy Liu from Prof. Richard Shi's lab for all the fun we had in the past years. I wish the very best to all of them in the future.

Finally, I would like to thank my family: my parents, Yongbin Sun and Chunfang Gao, and my husband Jiabo Li for their understanding, unconditional love and support for me

to struggle through the long journey of graduate school.

DEDICATION

To my parents, Yongbin Sun and Chunfang Gao,

my husband, Jiabo Li

Chapter 1
INTRODUCTION

1.1 Overview

System-on-Chip (SoC) technology has seen rapid growth over the last few decades driven by Moore's law and market economics. In the desktop PC era, the primary goal was to increase processor speed, leading to high-performance SoC capable of integrating multi-core processors and multiple graphic processing units into a single system. Recently energy efficiency has become the key constraint in modern SoCs as a result of Moore's law slowing down and the power wall issue. Server-class chips are thermally limited and improved energy efficiency can significantly lower overall data center power consumption and thus lower the cost for cooling requirements [1]. For portable and wearable electronics, the longer battery life is one of the key elements to attract consumers. Optimizing for higher SoC system efficiency is effective for extending battery life given the constraint of fixed battery capacity.

In order to make SoC systems more efficient, various techniques have been proposed. Clock gating and power gating [2, 3] has become a standard feature in SoCs. Multi-voltage designs are common in digital SoC domains, where each block can operate at a different voltage according to its performance requirements. Power consumption can be significantly reduced by this method as dynamic power is proportional to supply voltage squared. Dynamic Voltage and Frequency Scaling (DVFS) is one of the most widely used multi-voltage design strategies where the voltage level for each block/each core is dynamically configured to follow varying workloads and performance constraints [4]. Near threshold voltage (NTV) operation, where supply voltage is lowered close to the transistor threshold, has demonstrated the potential to improve energy efficiency by an order of magnitude [5]. With the application of these energy-efficient design approaches, especially the ones scaling down V_{dd}, variations become a severe challenge for timing across different operating conditions. Consequently, timing margins are added to the operating clock frequency for phenomena

including clock jitter, on-chip temperature variations, V_{dd} noise, process variations and all other possible effects for safety (Figure 1.1). Worst-case timing margins lead to degraded system energy efficiency as significant power and performance will be wasted for normal operating conditions. These margins, especially voltage margin – the largest contributor to the overall margins, are becoming the dominant source of inefficiency in modern SoCs.

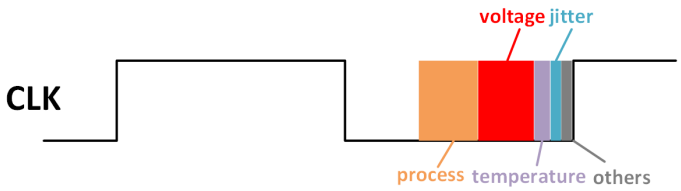


Figure 1.1: Variations in Process, Temperature and Voltage require substantial performance/efficiency degrading frequency guardbands.

Solutions to reducing SoC voltage margin can span a variety of areas and this thesis will focus on two categories of techniques to minimize voltage margin in SoCs. First is through clocking/voltage co-design where the core idea is to unify the design of traditionally independent clock and voltage regulation loops. A new unified clock and power architecture *UniCaP* is proposed and the idea is implemented on a buck-converter based prototype to address the issue of V_{dd} margins [6, 7] in IVR systems. This *UniCaP* work successfully demonstrated a stable adaptive clocking with a compensation architecture applicable to any VR modality. However, in this implementation, slow recovery of V_{dd} droop (and thus f_{clk}) increases peak cycle loss ($\Delta\Phi_{max}$), leading to increased pointer separation and even saturation in clock-domain crossing FIFOs. Furthermore, adaptive clocking poses another challenge to real-time systems, which is to regulate performance within application-dependent time frames. In some real-time applications, this window can be as narrow as 4-5 microseconds, requiring these adaptively clocked systems to not only phase lock, but also restore lock rapidly and recover any lost cycles. Existing adaptive clocking implementations either don't recover lost cycles, or when they do, they do so slowly, over much longer periods than 5µs, making the technique unsuitable for real-time applications. Phase-locked adaptive clocking with rapid cycle recovery – *UniCaP*-2 – is proposed to tackle these outstanding

adaptive clocking challenges [8].

The other category of solutions for voltage margin minimization focuses on improving voltage regulation (VR) performance through computation-enabled techniques. Enhanced VR transient performance can reduce V_{dd} noise in SoCs and thus reduce corresponding voltage margin. This thesis will demonstrate how computation can be used to improve VR performance and enable fast and stable transient responses. Conceptualization of utilizing computation in VR control is first applied to digital low dropout regulators (DLDO) [9], as there is significant interest in viable DLDO solutions due to their ease of portability across designs, low-voltage capability and scalability across technologies. We construct a time-domain regulator model, which provides enhanced accuracy over a linear discrete-time transfer function model, and evaluate it at run-time by a Solver to ensure rapid response. This idea is then extended for VR τ_{settle} minimization, which aims to minimize the amount of time it takes to restore V_{dd} to its reference value in adaptive clocking systems. In this dissertation, a low-complexity, low-latency model predictive control (MPC) architecture is applied to realize an IVR buck regulator with optimal transient response [10].

In conventional offline VR controller designs, the controller parameters are usually tuned based on estimated analytical models and assumptions about expected operating range and component value uncertainties. The result would be a stable, but overly conservative design, as significant margins are applied in the design in order to ensure stability under worst-case scenarios, which adversely affects VR response times and overall system efficiency. This thesis will discuss how computation can be utilized to enable auto-tuning of VR control parameters to address the worst-case stability margining issue. Automatic Gain Tracking (AGT) is employed in the computational controlled DLDO by examining 1-bit statistics of the output voltage to track variations in loop gain and adjust solver gain accordingly. As the AGT technique is solely designed for the computational LDO and is susceptible to specially designed load current patterns, a more generic, impulse response adaptation method, which calculates the cross-correlation between the control output and the injected noise signal, is proposed for auto-tuning the loop gain.

1.2 Organization

The rest of this Dissertation is organized as follows.

Chapter 2 presents a Unified Clock and Power (*UniCaP*) architecture to address excessive voltage guardband problems caused by supply noise or temperature variations in conventional systems. In this architecture, f_{clk} instantly adapts to V_{dd} and T variations, requiring minimal guardbands. With *UniCaP*, V_{dd} control is absorbed into the clock regulation loop, so f_{clk} can be adjusted and locked to *REFCLK* through V_{dd} change. *UniCaP* allows for all-digital construction, reduced V_{dd} guardbands, and on-the-fly dynamic voltage and frequency scaling (DVFS). We deployed the *UniCaP* architecture on a 65nm buck-converter test-chip powering a 0.6–1.0V Cortex-M0 microprocessor with autonomous transition between continuous and discontinuous conduction modes (CCM and DCM) in order to support a wide load current range. The *UniCaP* test-chip demonstrates 82% average V_{dd} guardband reduction, without any performance loss from adaptive clocking, and temperature margin reductions of 40-55mV.

Chapter 3 discusses the limitations of the *UniCaP* implementation in Chapter 2 and proposes a new, improved Unified Clock and Power architecture (*UniCaP*-2) is presented. A dual-path feedback is utilized in *UniCaP*-2 to further reduce supply-voltage (V_{dd}) margins in an ARM Cortex-M0 processor while minimizing both peak cycle loss ($\Delta\Phi_{max}$) and cycle-loss recovery time ($T_{recovery}$) associated with adaptive clocking. Measurements on a 65nm test chip demonstrate 91–99% V_{dd} margin reduction and 38X $T_{recovery}$ improvement over [6]. This chapter also includes measurement results to quantify the impact of clock distribution delay (τ_{dist}) and V_{dd} sensitivity on V_{dd} margin recovery.

Chapter 4 describes a Digital Low-Dropout (DLDO) regulator architecture designed for fast and stable transient response. The proposed technique relies on a time-domain model that accurately describes regulator behaviour, using low-latency computation to evaluate the equations that describe this model. These computations are performed every DLDO cycle by a solver that determines the optimal number of PMOS headers to be turned on in order to minimize regulator settling time. The proposed DLDO is implemented to regulate an ARM Cortex-M0 processor and an integrated linear-algebra co-processor in 65nm CMOS.

Transient response measurements arising from processor load variation demonstrate a 2.9-cycle mean settling time.

Chapter 5 describes a digital control architecture for integrated voltage regulators (IVRs) that achieves time-optimal transient supply-voltage (V_{dd}) response under random load-current (I_{load}) fluctuation. Implementing low-complexity low-latency Model Predictive Control (MPC) in a 65nm CMOS test chip is key to achieving a measured 2.49X settling-time (τ_{settle}) improvement over optimally tuned Proportional Integral Differential (PID) control.

In Chapter 6, two methods are presented to track loop gain variations at run-time and adjust control parameters accordingly for an optimal design across the system operating range. AGT, a technique that relies on computing the auto-correlation of the sampled V_{dd} with single-bit precision in order to classify regulator response and enable loop gain auto-tuning across $PV_{in}V_{dd}T$ variations, is implemented on the computation LDO test-chip [9]. Silicon measurements indicate that the AGT mechanism performs broad loop-gain adjustments to track a wide range of $PV_{in}V_{dd}T$ configurations: even individual variation in V_{in} (0.9V–1.1V), V_{dd} (0.8V–1.0V) and temperature (-15°C–105°C) requires loop-gain modulation over a range of approximately 2×. The more generic, impulse response adaptation method is proposed in this chapter to address the limitation of AGT technique. A PRBS signal, a digital approximation of white noise signal, is added as a small perturbation to the controller output. Cross-correlation between the injected noise and controller output identifies the closed-loop impulse response. Comparison results of the measured and expected impulse response values can then be used to adjust the control parameters.

Chapter 2

A COMBINED ALL-DIGITAL PLL-BUCK SLACK REGULATION SYSTEM WITH AUTONOMOUS CCM/DCM TRANSITION CONTROL

2.1 Introduction

Dynamic Voltage and Frequency Scaling (DVFS) remains a salient technique for energy-efficient digital design. Integrated voltage regulation (IVR) significantly enhances DVFS through finer spatio-temporal V_{dd} control. Buck converter based IVR designs in particular [11], present advantages over their switched-capacitor (SC) converter and linear regulator (LDO) counterparts in terms of V_{dd} resolution and regulator-efficiency respectively. However, several IVR-related challenges persist. Independently regulated fine-grained voltage domains fragment shared decoupling capacitance (decap) into multiple smaller domains. The resulting per-domain decap reduction worsens V_{dd} droop due to load-current (I_{load}) disturbances which are common in digital domains. This supply droop, along with variations in temperature (T), require the addition of wasteful V_{dd} guardbands to ensure that digital logic continues to meet timing under worst-case operating conditions.

Conventional digital systems, with independent V_{dd} and clock regulation loops (Figure 2.1a) are not well-suited to address mounting guardband requirements driven by increased V_{dd} droop and higher sensitivity to PVT variation. In such systems a Phase-locked Loop (PLL), powered by a clean dedicated supply voltage, delivers a clock that is phase-locked to *REFCLK* with frequency f_{clk}. A separate voltage-regulator loop is tasked with maintaining the target V_{dd}. A V_{dd} droop arising from a line-side disturbance at the input supply (V_{in}), or a load transient slows down the logic. If f_{clk} is fixed, this slowdown degrades timing slack (Figure 2.1b). Droop suppression through faster regulation alone is energy-inefficient, and a V_{dd} guardband is introduced to margin for the worst-case droop. Buck converters for SoCs must also maintain efficient operation over a large I_{load} range. Consequently, these converters must be capable of both, autonomous CCM/DCM transition, and enabling

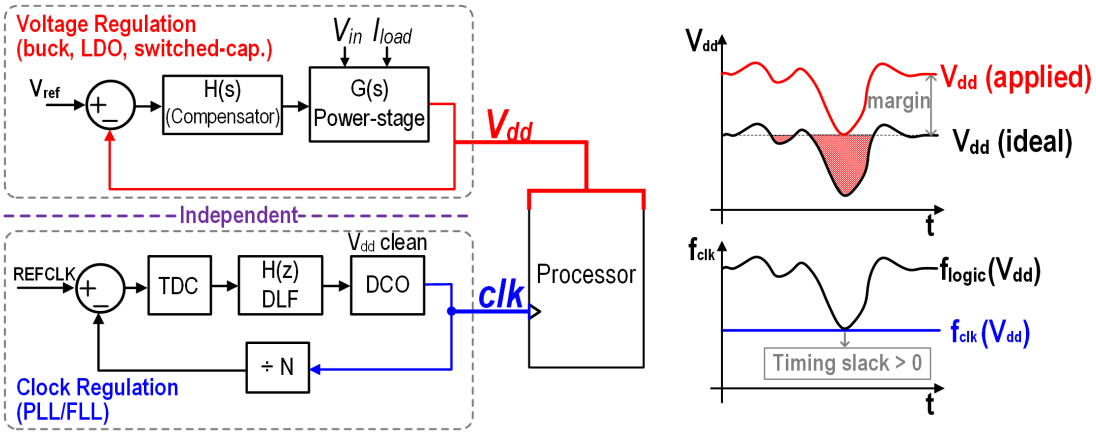


Figure 2.1: Conventional digital systems with (a) Independent voltage and clock regulation loops. (b) Dissipative voltage margins to maintain timing-slack under worst-case voltage droop.

efficient zero-current-switching in DCM, preferably by all-digital circuit techniques.

Prior work has addressed some challenges in fine-grained IVR design for digital SoCs. Transitions between CCM/DCM operation based on I_{load} sensing relying on an analog zero-crossing detector [12] have been presented. In addition to precluding all-digital design, the technique does not account for bridge driver delays for zero voltage switching (ZVS), degrading converter efficiency. Ring-oscillator based methods have also been proposed [13, 14] in place of ADCs to quantize V_{dd} error. However, these techniques do not provide adaptive clocking [15, 16, 17, 18, 19, 6, 20] and therefore require significant guardbands for V_{dd} droop. More recent works have combined clock and V_{dd} regulation into a single system [19, 20, 21]. However, none of the works address the problem of maintaining phase-lock, essential for cross-domain communication in real-world implementations. Combining clocking and V_{dd} regulation requires addressing important challenges pertaining to stability which are addressed by this dissertation. Also absent in prior work is a UniCaP loop controller design methodology that is broadly applicable to the general class of integrated voltage converters (buck, switched-capacitor and low-dropout regulators).

This chapter presents a *UniCaP* framework that is applicable across the general class of regulator types including LDOs, SC and buck converters [6]. It also addresses the critical

need for reliable phase and frequency lock to *REFCLK* while providing adaptive clocking under noisy V_{dd} conditions. This chapter also presents an all-digital delay-locked loop (DLL) based technique for autonomous CCM/DCM transition and Zero Voltage Switching (ZVS) under various I_{load} conditions. A buck converter-based *UniCaP* system capable of supporting a wide I_{load} range was implemented in a 65-nm CMOS test chip. An on-chip ARM Cortex-M0 processor simultaneously serves as a load and enables characterization of the V_{dd} margin reduction benefits of the *UniCaP* architecture. Measurements indicate an average V_{dd} margin recovery of 82% over a range of 0.6V-1V due to supply noise and 55mV temperature guardband recovery at 0.6V. The ability to recover any lost or gained cycles due to V_{dd} disturbances is also validated.

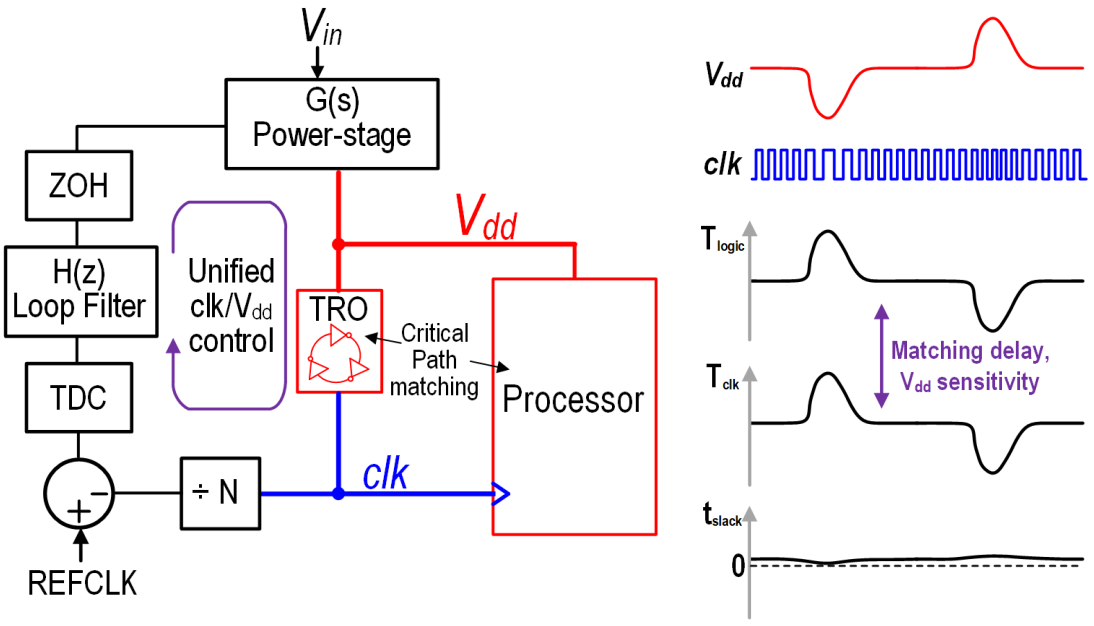
The remainder of this chapter is organized as follows. *UniCaP* architectural details are described in Section 2.2. Key *UniCaP* modules including the TRO, a wide-range coarse-grained cycle-counting time-to-digital converter (TDC) and compensator design are explained in Section 2.3. Section 2.4 provides circuit implementation details of the all-digital ZVS operation and automatic CCM/DCM transition. Section 2.5 shows the test-chip measurement results.

2.2 Unified Clock and Power Architecture

The primary goal of V_{dd} control in SoC domains is timing slack regulation across PVT conditions at a guaranteed performance target. In view of this observation, the *UniCaP* architecture (Figure 2.2a) implements adaptive clocking, using a V_{dd}-powered tunable-replica oscillator (TRO) to mimic both, the delay and the droop or temperature sensitivity of the critical path in the Cortex-M0 core. As a result, any slow-down (speed-up) in the critical path is accompanied by a clock stretching (compression) event that avoids any impact on timing slack. Thus, to the first order [1], use of a V_{dd}-powered TRO guarantees timing-slack regardless of supply droop events due to PVT variation and maintains correct function.

Adaptive clocking provides supply noise and temperature variation tolerance. However, real-world applications require performance guarantees, ensuring that any lost (or gained)

[1] Non-idealities such as within-die VT variation, insertion delay and imperfect TRO tracking of the critical path will require additional margin.



(a) (b)

Figure 2.2: (a) Proposed *UniCaP* architecture regulating operating frequency through V_{dd} control. (b) Elastic clock maintains correct function by compensating for supply noise induced critical path delay variation.

cycles are recovered. *UniCaP* achieves performance regulation by using a PLL that locks TRO clock to *REFCLK*, thus avoiding any gain or loss of cycles due to adaptive clocking.

Adaptive clocking architectures that inject V_{dd} noise into a traditional PLL architecture (block diagram illustration in Fig. 2.3a) have been previously implemented [16, 22, 23]. These techniques provide the desired f_{clk} adaptation in response to a rapid voltage droop but face two drawbacks. The first is the difficulty in matching V_{dd} sensitivity between the VCO or DCO, and the critical path. More critical, however, is that the PLL operates independently of V_{dd}: any control voltage or oscillator code changes made by the PLL to lock f_{clk} to REFCLK do not track the V_{dd}-governed critical path delay, leading to timing failure as shown in simulation waveforms in Fig. 2.3b. UniCaP avoids this limitation by providing a single point of f_{clk} control: the shared V_{dd}. Any change in V_{dd} that tunes f_{clk} to track *REFCLK* also affects the critical path identically, leaving timing-slack unaffected. In order to use V_{dd} control for phase-lock, the voltage converter stage (a buck converter in this work) is absorbed into the PLL loop, leading to a unified control loop. Such a

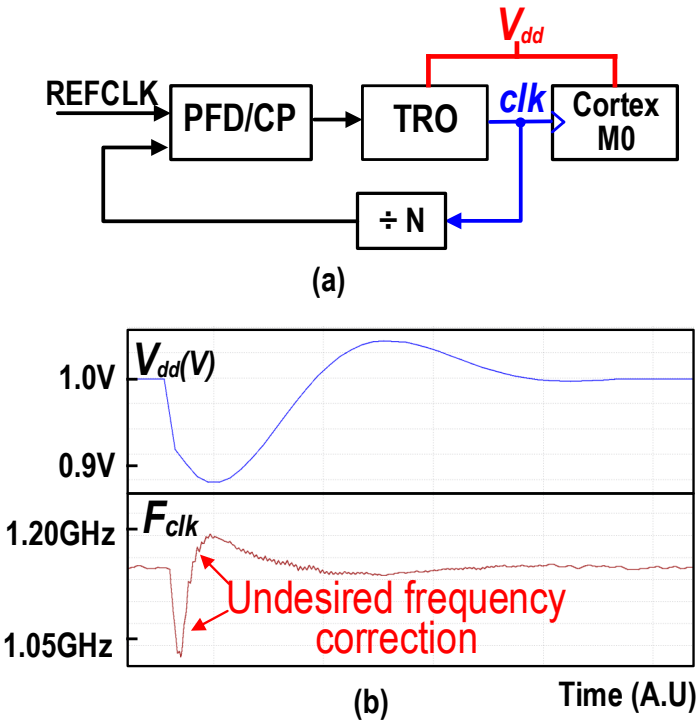


Figure 2.3: (a) Block diagram of regulating TRO frequency through a traditional PLL architecture. (b) Slow voltage droop causes timing failure.

unified approach enables a synergistic coupling between power and clock, allowing f_{clk} to vary during a V_{dd} drop to avoid timing failure, while at the same time controlling V_{dd} to enable timing-neutral REFCLK tracking.

Unifying clock and power control affords several SoC design benefits, including voltage-reference free all-digital construction and effective tracking of chip-mean temperature variation. In particular, *UniCaP* provides two critical benefits:(1) To the first order, *UniCaP* decouples regulator bandwidth requirements from V_{dd} droop margin, enabling a relaxation of bandwidth requirements, a simplification of voltage regulation, all while improving energy-efficiency. (2) Adaptive clocking employed to maintain timing-slack despite V_{dd} droop or temperature variation remains within PLL control, ensuring that any cycles lost (gained) during a V_{dd} droop (surge) event are fully recovered. The duration over which these cycles are recovered depends on the PLL-regulator bandwidth which must be set in accordance with system and application phase-tracking performance requirements.

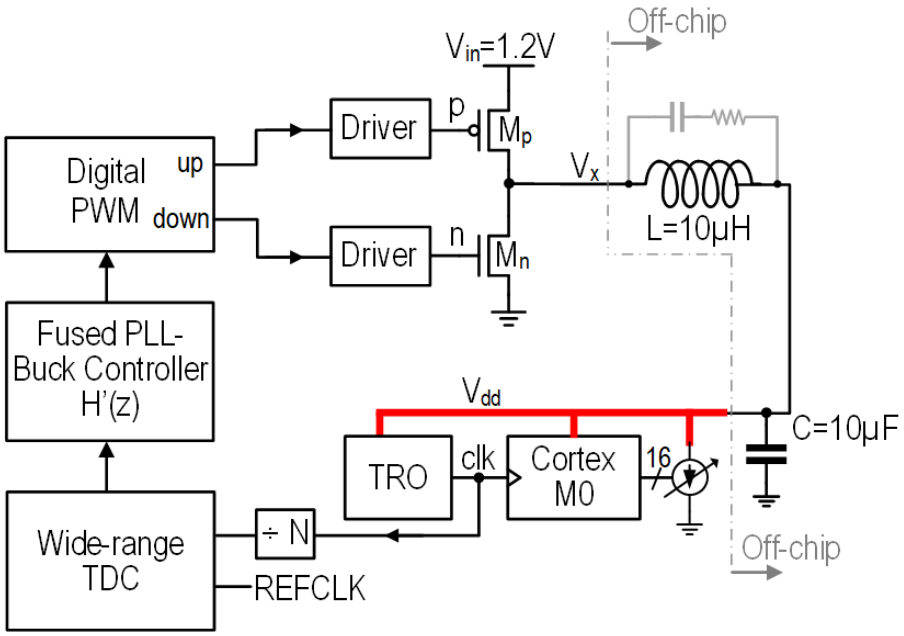


Figure 2.4: Architecture overview of the buck-converter based *UniCaP* system.

2.3 UniCaP *System Implementation*

Figure 2.4 shows a UniCaP-buck system. All modules, with the exception of the bridge, TRO, and ZVS comparators were designed using a standard synthesis, place-and-route flow. The buck converter output (V_{dd}) powers both, a Cortex-M0 processor and the TRO. The TRO clock is divided and sampled by a cycle-counting TDC that quantizes the phase-error for use by the compensator. The compensator provides the digital pulse-width modulator (DPWM) with the duty-cycle for signals M_p and M_n. Because UniCaP effectively de-couples required V_{dd} margins from regulator response time, the need for rapid V_{dd} transient response is diminished. Reduced bandwidth requirements allow for lower switching frequencies and improve efficiency through decreased switching loss. Although UniCaP is ideally suited to IVR technology, we selected off-chip inductors for lack of access to the necessary technology. The selected off-chip inductor and filter capacitor values were 10μH and 10μF. A switching frequency of 1-MHz was chosen for this design.

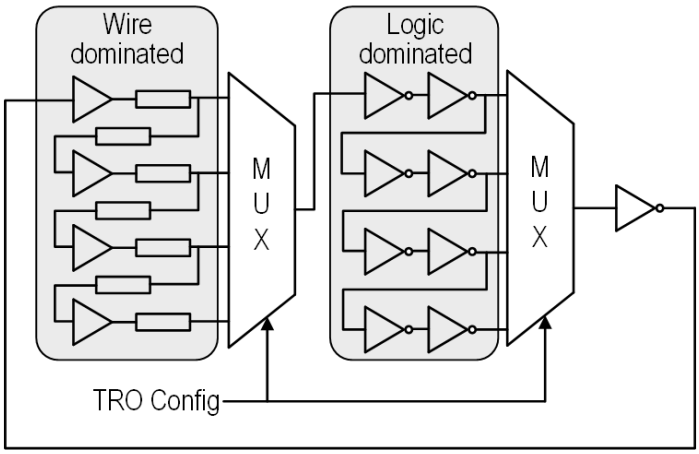


Figure 2.5: Circuit diagram of TRO consisting of wire and logic dominated delay cells.

2.3.1 Tunable Replica Oscillator(TRO)

The TRO is implemented as a ring oscillator with run-time programmable T_{clk} and V_{dd}-delay sensitivity to track worst-case critical path delay across the operating V_{dd} range. The approach resembles that used in [18], consisting of two building blocks that model wire (less V_{dd} sensitive) and logic (more V_{dd} sensitive) dominated delays (Figure 2.5). Judicious choice of the number and proportion of each type of element provides the desired critical path matching. Characterizing $f_{max}(V_{dd})$ on production silicon allows UniCaP to compensate for both chip-mean, and within-die process variation. However, cost and tester time considerations limit the number of characterization points, resulting in a modest V_{dd} guardband. Additional sources of required margin stem from the delay-sensitivity precision afforded by the TRO design, clock-insertion delay effects, and within die variation in V_{dd} and T.

2.3.2 Loop Compensation

Loop compensation is a critical consideration in enabling a UniCaP architecture applicable across different regulator types. Incorporating V_{dd} control into the PLL can introduce stability challenges that need to be addressed, as the dynamics of the PLL and the voltage regulator are actively combined into a single loop. This combination is especially challeng-

ing in the case of a buck-converter, whose higher-order nature makes compensation more complicated (Section 2.3.2). Other common challenges with using V_{dd} as the PLL oscillator supply are the risk of cycle slipping (requiring a wide-range TDC), and stability margining to account for the significant TRO gain variation resulting from V_{dd} scaling. Notably, TRO gain varies by 32% over a 0.6-1.0V range.

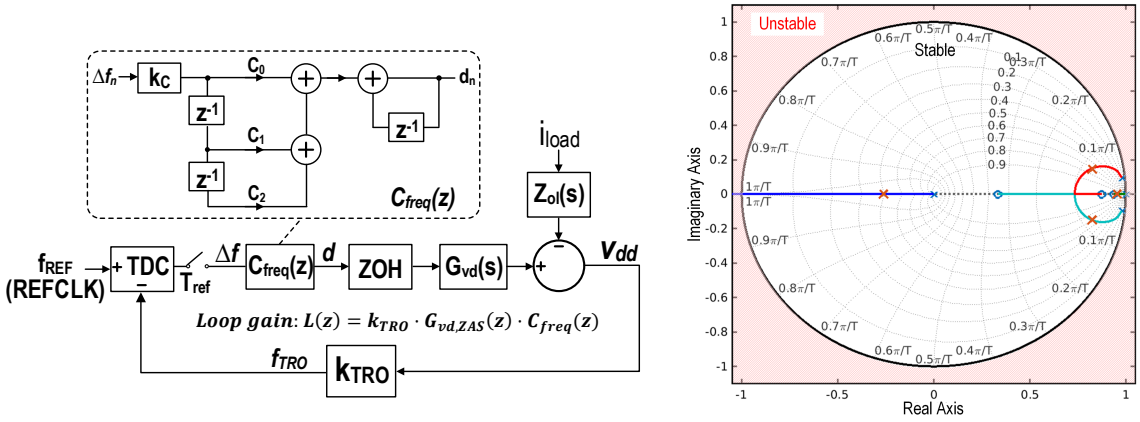
Time-to-digital Converter(TDC)

Use of a V_{dd}-powered TRO introduces significant (V_{dd} correlated) clock jitter, leading to phase-error potentially exceeding 2π detection range of traditional phase-detectors. Consequently, a coarse-grained cycle-counting TDC is employed to provide a wide capture range [21]. A 2-9 bit counter (2-bit coarse, 9-bit fine counter) counts TRO clocks within a *REFCLK* cycle. Phase-error in excess of 2π, corresponding to a divider-ratio count N no greater than 512, causes the fine-count ($F[n]$) to reset to 0 and the coarse counter ($C[n]$) to increment. A *REFCLK* rising edge (suitably re-timed with TRO *clk*) samples both counter values before they both resume from 0 in the next cycle. The difference between sampled counter values corresponds to PLL frequency error: $\Delta f = F[n] - F[n-1] + (C[n] - C[n-1] - 1)N$, which is subsequently accumulated to yield the phase error. The 2-9 bit counter provides the PLL with a detection range of at least 4 *REFCLK* cycles, effectively avoiding cycle-slipping. Increasing the bit-width of the coarse-counter can provide additional range, but was not necessary in this design.

Frequency Acquisition

Figure 2.6a shows the frequency-acquisition portion of the UniCaP system, which uses feedback to compensate the open-loop output impedance of the buck converter (Z_{ol}) . The goal of frequency-acquisition in *UniCaP* is similar to voltage-regulation: maintaining a target V_{dd}. The key difference is that *UniCaP* determines this V_{dd} target to ensure timing-slack across PVT variation. In the figure, k_{TRO} denotes the linearized TRO voltage-to-frequency gain around its steady state V_{dd} value. The cycle-counting TDC produces Δf which is provided to a PID compensator ($C_{freq}(z)$). The effect of the duty cycle (d)

produced by the compensator on V_{dd} is modeled as G_{vd}. The combined system incorporates the dynamics of both the buck-converter and the TRO. Achieving frequency-lock for the resulting system, however, requires the presence of only one open loop pole at z=1. This requirement greatly simplifies loop-compensation: A PID controller operating on Δf is sufficient to compensate the frequency acquisition loop. The controller gain k_C is chosen using a f_{clk}-indexed table, similar to the TRO sensitivity and delay configuration, to provide maximally rapid but stable response as V_{dd} varies from 0.6V to 1.0V.



(a) (b)

Figure 2.6: (a) Block diagram of the frequency acquisition loop. (b) Root-locus plot demonstrates stable closed frequency-acquisition loop response.

The PLL *REFCLK* also serves as the bridge clock. The applied d is determined a short delay (τ_d) after the rising edge of the buck converter bridge clock (*REFCLK*) due to TDC resolution-time and compensator latency. Waiting for τ_d for PMOS turn-on would limit maximum achievable d, and hinder the ability to address V_{dd} droop. Instead, the PMOS turns on at the start of every switching cycle. Once the duty cycle for the current cycle is determined, the PMOS is immediately disabled if $d \cdot T_{ref} < \tau_d$. This approach enables PID compensation without the need to model for loop-delay, but at the cost of limiting minimum achievable d. Therefore, the PID digital controller is modeled with no transport delay in Figure 2.6a. Simulations combining suitably chosen k_C with judiciously chosen controller coefficients $c_0 = 1, c_1 = -1.8125, c_2 = 0.8203$ demonstrate stable closed-loop

response (Figure 2.6b).

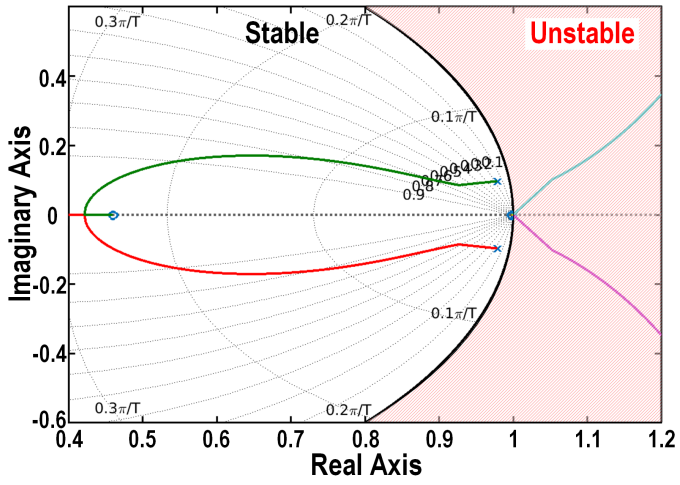
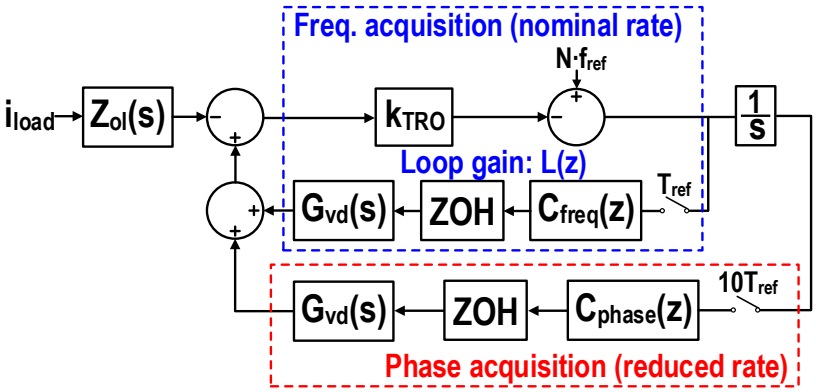


Figure 2.7: Root-locus plot indicates an unstable phase loop for all K values with conventional PID control.

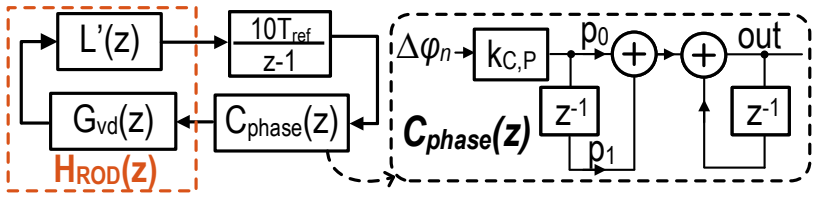
Phase Acquisition

Tracking *REFCLK* phase is considerably more involved compared to frequency-lock. The need for two discrete-time open-loop poles at $z = 1$, combined with the complex poles corresponding to the buck converter output impedance makes traditional PID compensation inviable: Figure 2.7 shows the root-locus plot resulting from PID compensation, indicating unstable operation.

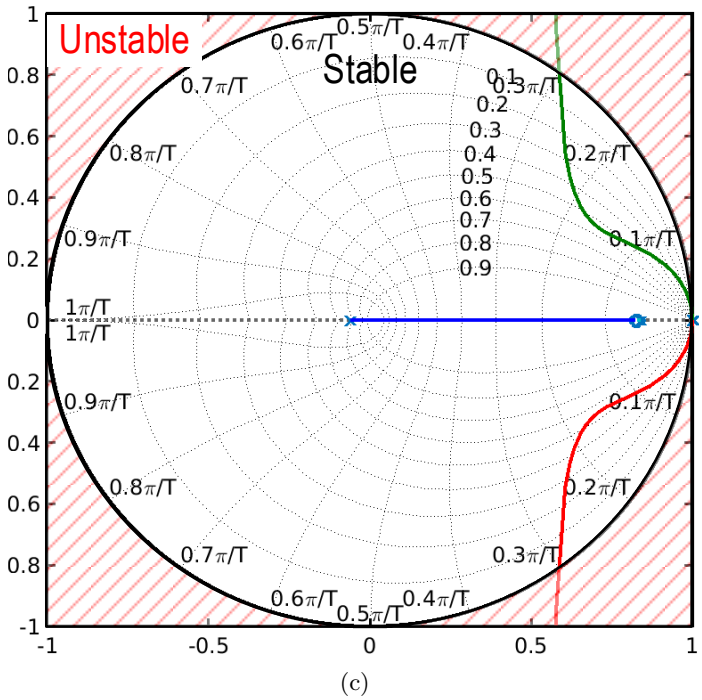
In this work, we exploit the idea of multi-rate control [24] to achieve stable loop compensation. In this approach, different sections of the feedback-loop operate at different rates to maintain an overall stable loop response. The frequency path works at the nominal rate (T_{REFCLK}) to track TRO frequency (and therefore V_{dd}). A phase acquisition loop operating at 10x lower rate ($10T_{REFCLK}$) to maintain phase-lock (Figure 2.8a). At this lower rate, the phase loop regulates a closed-loop frequency-compensated system, and is effectively tasked with removing residual phase-offset errors that the fast frequency loop does not eliminate: any build up of phase-error is precluded by the action of the phase-locked loop.



(a)



(b)



(c)

Figure 2.8: (a) Block diagram of composite control with frequency and phase path operating at different rates. (b) Equivalent phase loop model with model reduction technique applied. $G_{vd}(z)$ is the discrete-time representation of $G_{vd}(s)$ (c) Stable closed loop response for the reduced order system.

Figure 2.8b shows the resulting reduced-order frequency-compensated model which presents a much simpler compensation problem to the phase-acquisition loop. This approach provides a suitable strategy for stable phase-locked operation even with voltage converters posing more complicated transient responses. The phase compensator is designed accordingly, using the reduced order model $H_{ROD}(z)$, extracted from the balanced state-realization of the sub-sampled frequency loop (a fourth-order model). Significant poles that determine system dynamics were subsequently selected based on eigenvalue magnitudes in the state-matrix to arrive at a reduced second-order model that preserves most of the original system dynamics. Use of a PI controller with $p_0 = 1, p_1 = -0.78$ readily provides stable closed-loop compensation for phase-lock, as seen in Figure 2.8c.

2.4 All-digital ZVS Operation and CCM/DCM Transition

2.4.1 All-digital ZVS

Significant losses are incurred in switching the considerable output capacitance C_{bridge} (Figure 2.9). ZVS can reduce these losses and improve efficiency by recycling the charge stored in C_{bridge} through L [25]. Furthermore, DCM operation is often employed under light load conditions to help reduce conduction losses by turning off power transistors when the minimum inductor current $I_{L,min} < 0$ [26]. Autonomous CCM/DCM transition is particularly desirable for SoCs which typically exhibit wide data-dependent load fluctuation. This section elaborates upon the circuit implementation of ZVS and autonomous CCM/DCM transition.

Figure 2.9 shows the circuit schematic and timing diagram of the proposed ZVS operation. C_{bridge} charges to V_{in} during the charging phase. A deliberate dead-time is introduced between M_p turn-off and M_n turn-on, resulting in both FETs not conducting, allowing capacitor charge transfer to the load through the inductor. Ideally, M_n must be turned on approximately when V_x crosses 0 to maximize switching loss reduction by $1/2 C_{bridge} V_{in}^2$ per cycle. An all-digital DLL is used to track the $V_x = 0$ transition, and subsequently tune the delay between rising edges of p and n so that the turn-on of M_n aligns with the condition that drain voltage of M_n is zero. A p-type Strong-ARM latch is used as a phase-detector

to evaluate the polarity of V_x at the instant that n transitions to "1". The phase-detector output is used to control a thermometer-coded programmable telescopic delay chain [27] with a coarse delay of 80ps and a fine resolution of 20ps achieved by tunable MOS-based capacitor load at the output buffer to adjust t_{dead} to enable the alignment of $V_x = 0$ and the rising edge of n. The DLL tracking loop is independent from the *UniCaP* loop.

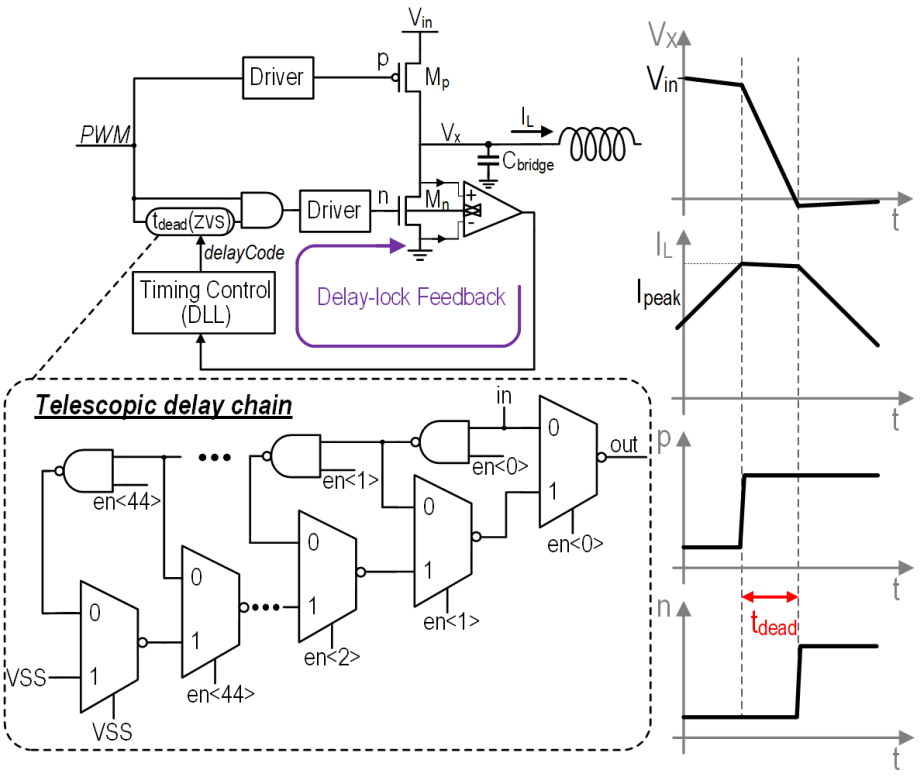


Figure 2.9: Circuit and timing diagram of the proposed DLL-enabled ZVS operation.

2.4.2 Autonomous CCM/DCM Transition

Figure 2.10 shows a block diagram of the proposed CCM/DCM controller. DCM operation requires Zero Current Switching (ZCS) for higher efficiency which ensures $I_L = 0$ when M_n turns off. Prior work has examined techniques to detect the $I_L = 0$ event [12, 28]. In [12], this detection is achieved by sensing the direction of current flow based on the polarity of the V_X node voltage. A low-offset analog comparator consuming significant area and power

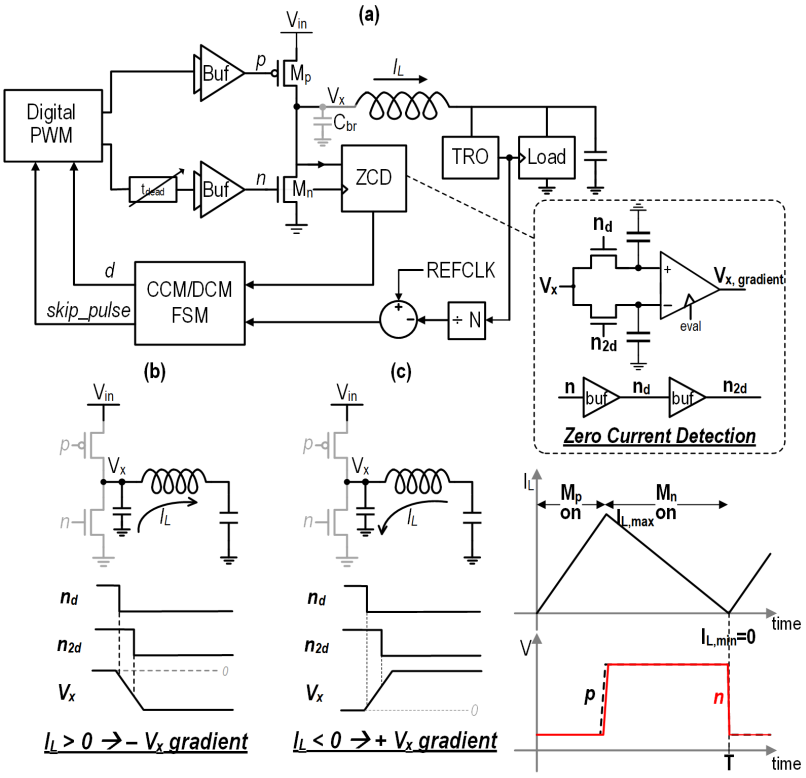
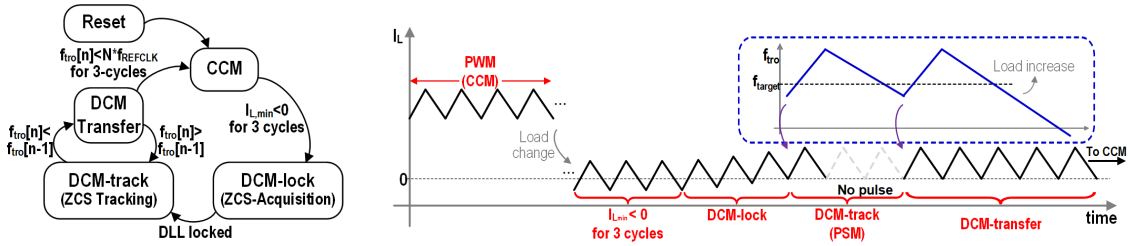


Figure 2.10: Circuit diagram of the V_X gradient based zero current detection method.



(a) (b)

Figure 2.11: (a) CCM/DCM transition control finite state machine. (b) Timing diagram illustrates autonomous CCM/DCM transition through a combination of PWM and PSM.

is required to determine the V_X polarity change. A digital zero current detection (ZCD) method is proposed in [28]. This approach connects V_X and V_{out} through a $2V_{in}$ gate bootstrapped switch, and relies on a comparator to determine the polarity of the switch. We propose a low-overhead, low-complexity approach for ZCD which indirectly measures I_L

direction by evaluating the polarity of the V_X gradient after M_n turns off to determine the direction of charge flow. The V_X node voltage is sampled successively on the falling edge of n_d and n_{2d}, delayed versions of n. Once sampled, a clocked comparator performs comparison on the two voltages. Double sampling of the V_X node to determine I_L direction is preferred to measuring the polarity of V_X to avoid having to detect a low voltage drop across the strong M_n switch. Figure 2.10 (b) and (c) illustrate inductor current direction evaluation based on the comparator output. Adjusting the pulse width, and locking to the transition edge of the comparator is indicative of achieving M_n turn-off at the $I_{L,min} = 0$ condition (Figure 2.10 (a)).

Figure 2.11 shows a timing diagram illustrating the operation of the CCM/DCM transition managed by a finite-state machine (FSM). The converter works in PWM mode during CCM operation when $I_{L,min}$ – as detected by the ZCD comparator – is positive under normal load conditions. If the comparator detects an $I_{L,min} < 0$ condition for three consecutive cycles during steady state, the FSM transitions into *DCM-lock* state, adjusting PWM duty cycle to lock to the $I_{L,min} = 0$ condition. Once the system enters the DCM tracking state, pulse skipping modulation (PSM) [29] is used to reduce the switching loss of the driver for M_p and M_n. The TRO clock frequency f_{tro} is compared with the target frequency f_{targ} at the onset of every switching cycle. A switching pulse is generated to charge the output capacitor only when f_{tro} is below f_{targ}. If the load current exceeds values achievable by DCM operation, V_{dd} and f_{tro} both decline, with f_{tro} falling below f_{targ}. After three consecutive cycles with $f_{tro} < f_{targ}$, indicative of a current load that exceeds the capabilities of DCM at the given V_{dd}, the FSM transitions back to CCM operation. The three-cycle delay introduces necessary hysteresis in the mode-transition to avoid dithering between the operating states. Phase lock was not implemented in the DCM configuration in this test-chip for logistical reasons. Per our analysis and simulations, DCM does not preclude achieving UniCaP phase-lock.

2.5 Measurement Results

A UniCaP-buck test-chip was implemented in 65 nm CMOS (Figure 4.9). The inductor and decap for this prototype were placed off-chip. All characterization tests, and reported

numbers are qualified by correct operation of the f_{max} and speed-indicative benchmarks of the Cortex-M0. Characterization of f_{max} along with corresponding TRO settings was performed in steps of 50mV in the range of 0.6 to 1V. At each calibration point, the TRO is configured to match both the delay and the V_{dd} sensitivity of the critical path.

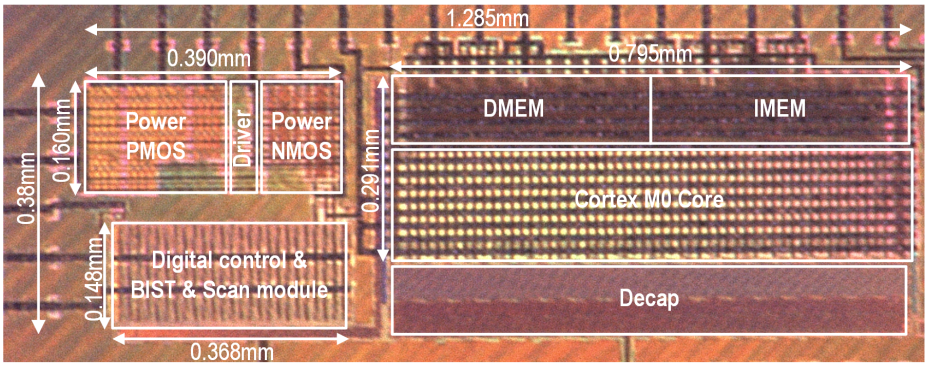


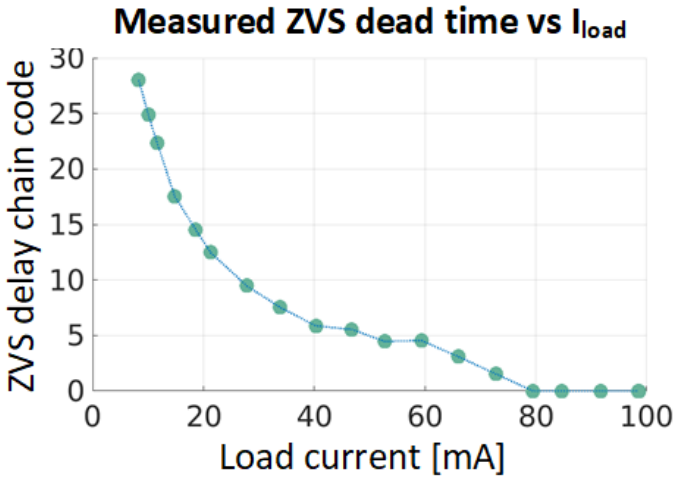
Figure 2.12: Die micrograph.

2.5.1 ZVS and CCM/DCM Transition

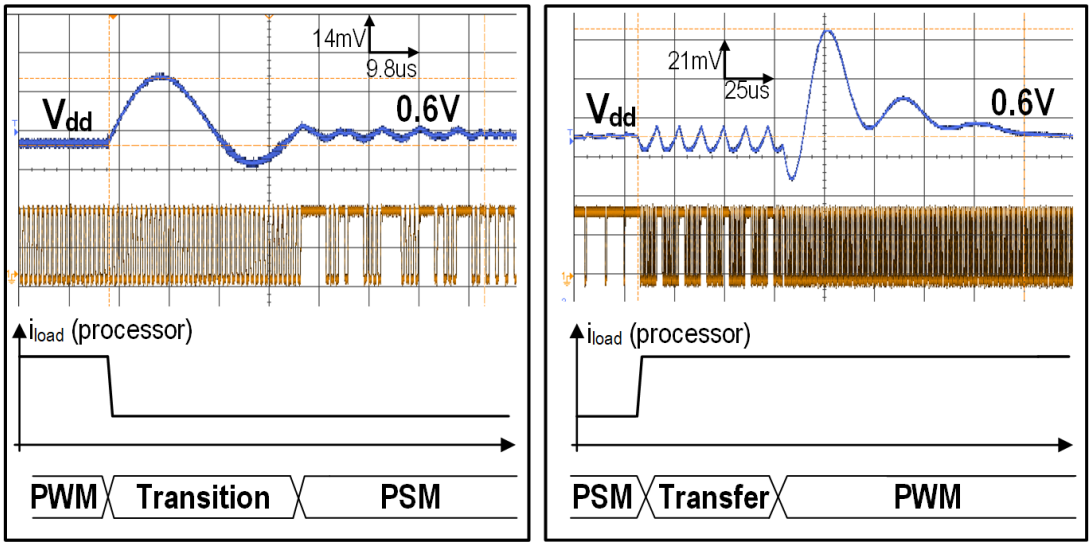
Figure 2.13a shows measured DLL delay-line codes under various load current conditions during steady-state ZVS operation. The modulation of the required delay line in response to I_L matches analytical expectations of dead-time delay for ZVS, and demonstrates the tracking operation of the DLL. Oscilloscope traces captured in Figure 2.13b illustrate CCM/DCM transition in response to load changes at $V_{dd} = 0.6$V. The system autonomously transitions to DCM as load current drops due to reduced processor activity, and changes back to CCM after I_{load} increases again. Measured buck converter efficiency at $V_{dd} = 1$V across 7-105mA is shown in Figure 2.14. Efficiency over 88% was observed across the range, with a peak efficiency of 96%.

2.5.2 Transient Response

In addition to the load provided by the Cortex-M0, the chip employs an integrated synthetic load that is memory mapped to the processor. Load current magnitude is configured by



(a)



(b)

Figure 2.13: (a) Measured locked ZVS-DLL code vs. load current. (DLL code proportional to ZVS dead time) (b) Measured CCM/DCM transition waveforms

performing store operations to the memory location mapped onto the synthetic load. This capability to provide a direct, program-controlled time-varying load can be used to generate fast, user-programmable I_{load} patterns. Figure 2.15 shows V_{dd} oscilloscope captures in response to I_{load} step-up and step-down (90mA/1ns). Responses for both, frequency-lock (*f-lock*) and phase-lock configurations, are included to highlight their differing behaviors.

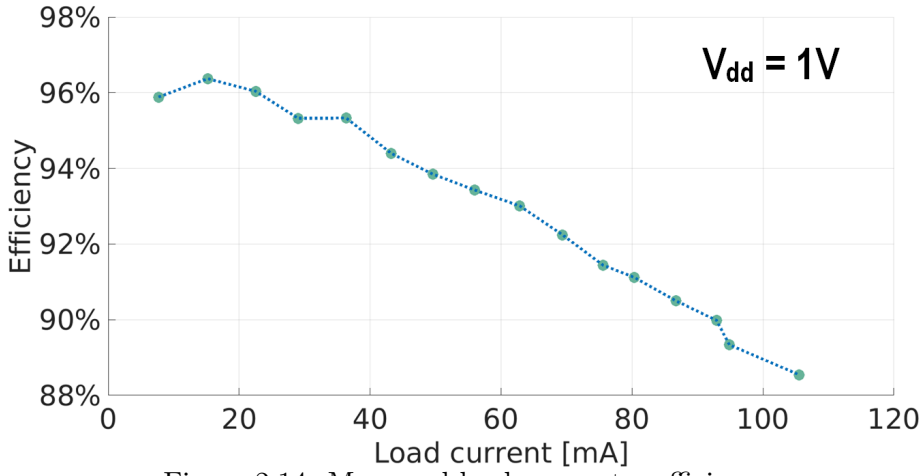
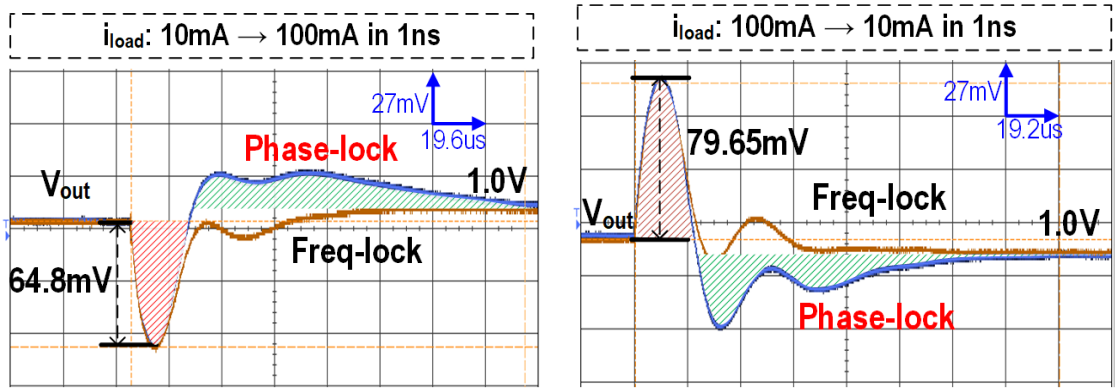


Figure 2.14: Measured buck converter efficiency

The *f-lock* response, measured by disabling the phase-control path, resembles a typical voltage regulator response since it involves V_{dd} recovery to restore f_{clk}. Phase-lock not only involves f_{clk} restoration but also recovers any cycles lost during the droop by temporarily driving V_{dd} above V_{target}. The corresponding behavior is observed for a current step-down. The increased (decreased) settling voltage resulting from a step up (down) is due to the IR drop between the inductor pin (our V_{dd} probe point) and the actual V_{dd} at the Cortex-M0, which is the voltage being regulated to enable phase lock.



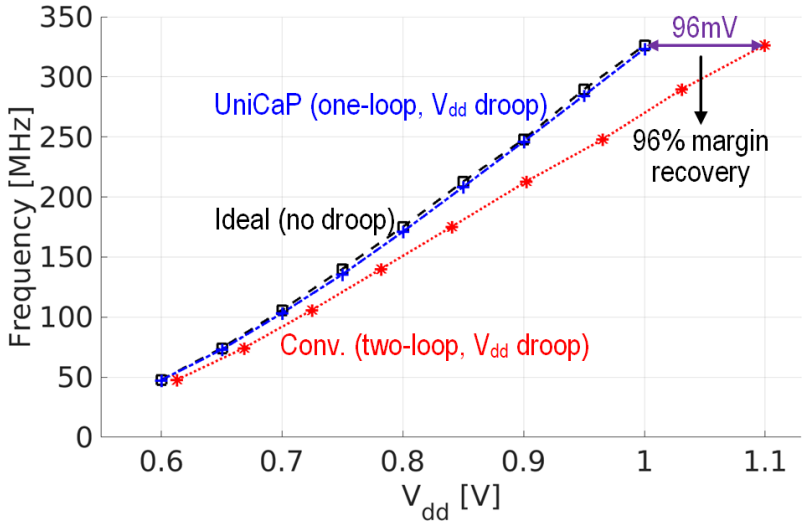
(a) (b)

Figure 2.15: Measured transient voltage response under load current step with phase and frequency lock configurations (a) step-up (b) step-down.

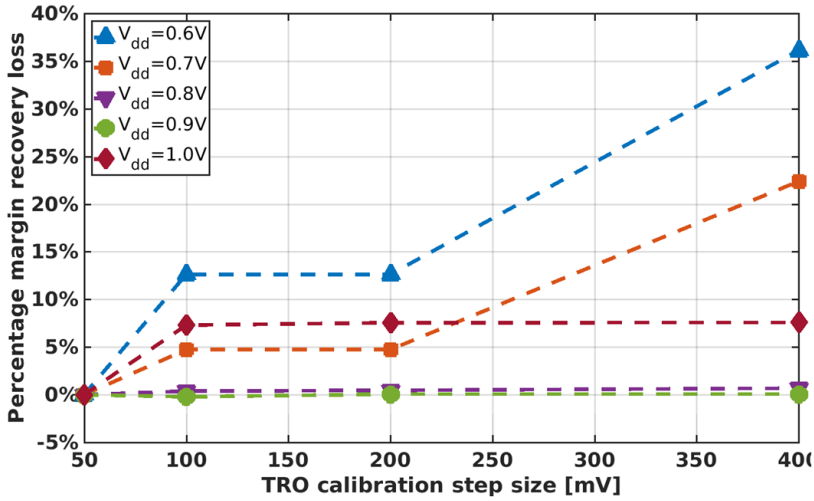
2.5.3 V_{dd}-margin Reduction

One of the key benefits afforded by UniCaP is its capability to track supply noise, temperature and inter-die process variations and aggressively reduce related V_{dd}-margins. We ran Cortex-M0 f_{max} benchmarks in the presence of supply droop to investigate UniCaP benefits on supply noise margin reduction. Experiments are performed under *ideal*, *conventional* (no elastic TRO) and *UniCaP* (elastic TRO) modes. The f_{max} curve corresponding to the *ideal* mode is obtained without any V_{dd} droop injection. Conventional mode measurements are made by configuring the system to operate conventionally, with independent V_{dd} and clock regulation loops. The injected V_{dd} droop results from a relatively conservative 100% change in I_{load}. At a target frequency of 320MHz, the I_{load} causes a 100mV supply droop during *conventional* operation. Under *UniCap*, elastic clocking allows the recovery of 96% of the droop margin, providing 96mV of V_{dd} reduction (Figure 2.16a) over the conventional approach. To the first order, the timing margin is independent of transient V_{dd} response because the TRO immediately stretches the clock in response to a droop. Thus, unlike conventional regulation, required V_{dd} margins are not determined by the transient V_{dd} response to load variation, but rather by I/O and buffering considerations. For this reason, the measured 96% margin recovery was achievable in spite of loop bandwidth well below 1MHz.

One of the several sources that contribute to reduced margin recovery in the *UniCaP* architecture (a summary of which is provided in the discussion section of [30]), is the mismatch between the TRO and the critical path. The number of V_{dd} points at which the TRO is calibrated to the critical path impacts the amount of mismatch margin that must be applied. Figure 2.16b shows the percentage loss in V_{dd} margin-recovery vs. the V_{dd} step-size used for TRO calibration. A 50mV step-size baseline is assumed. In the worst-case, a single TRO setting across the V_{dd} range would result in a 36% reduction in margin recovery at 0.6V. TRO calibration for production silicon is expected to be similar to that for any canary-based system [31] used in industry. The number of calibration points and TRO settings will be determined based on the trade-off between testing cost and the impact of calibration points on V_{dd} margin recovery.



(a)



(b)

Figure 2.16: (a) Measured Cortex M0 f_{max} vs. V_{dd} illustrating 96% margin recovery through *UniCaP* at 320MHz. (b) Measured V_{dd} margin recovery loss vs. TRO calibration step size across 0.6-1.0V.

The remaining margins associated with UniCaP are due to non-zero clock insertion delay, and within-die V_{dd},T variation not captured by calibration. A detailed analytical treatment of the impact of insertion delay is outside the scope of this work. To the first order, *UniCaP* transient response will be determined by how quickly frequency and cycle-count loss needs to be recovered rather than the worst-case V_{dd} droop magnitude.

The ability of UniCaP to track temperature variation through V_{dd} control in the range of -10-to-100°C also provides opportunistic margin reduction. Figure 2.17 shows measured V_{dd} value with target frequency of 40MHz and 320MHz across the temperature range. Conventional systems which require V_{dd} margining based on worst-case temperature conditions, consistently operate at the highest required V_{dd}. In contrast, UniCaP autonomously tunes V_{dd} to maintain phase-lock to *REFCLK* while meeting timing-slack, allowing the system to trade-off positive timing-slack from favorable temperature conditions for V_{dd} reduction. The temperature-induced V_{dd} margin reduction was measured to be 55mV at 40MHz and 40mV at 320MHz. The differing slope polarities of the two curves are the result of operation on either side of the temperature-inversion point in the 65nm process.

2.5.4 On-the-fly DVFS

The ability for on-the-fly DVFS, without interrupting processor operation is another significant advantage of *UniCaP*. Figure 2.18 shows *UniCaP* performing uninterrupted, on-the-fly DVFS by providing a sequence of new PLL divider ratios to the system to assert a new target frequency. In conventional systems, DVFS operation requires independent voltage and frequency codes (voltage-frequency look-up table) to be provided to VR and PLL while in *UniCaP*, a new divider ratio and TRO setting are asserted for a new transition. Importantly, the system automatically transitions V_{dd} to the voltage level required for correct operation at the new target frequency. As f_{clk} and V_{dd} are tuned simultaneously without timing slack degradation, uninterrupted processor operation is possible during DVFS events and has been verified in our experiments.

Mismatch in TRO sensitivity at the current vs. the target f_{clk} across a large f_{clk} range can be accounted for by a two-step TRO configuration process. At the onset of DVFS, the TRO is configured to the V_{dd} sensitivity based on the target f_{clk}, but with an additional time-period margin to guarantee timing-slack regardless of current operation away from its corresponding V_{dd}. This added TRO time-period margin can be removed once the PLL re-locks to the target f_{clk}. Use of an additional margin during DVFS transitions has the effect of adding a temporary V_{dd} margin to the system to account for V_{dd} sensitivity mismatch.

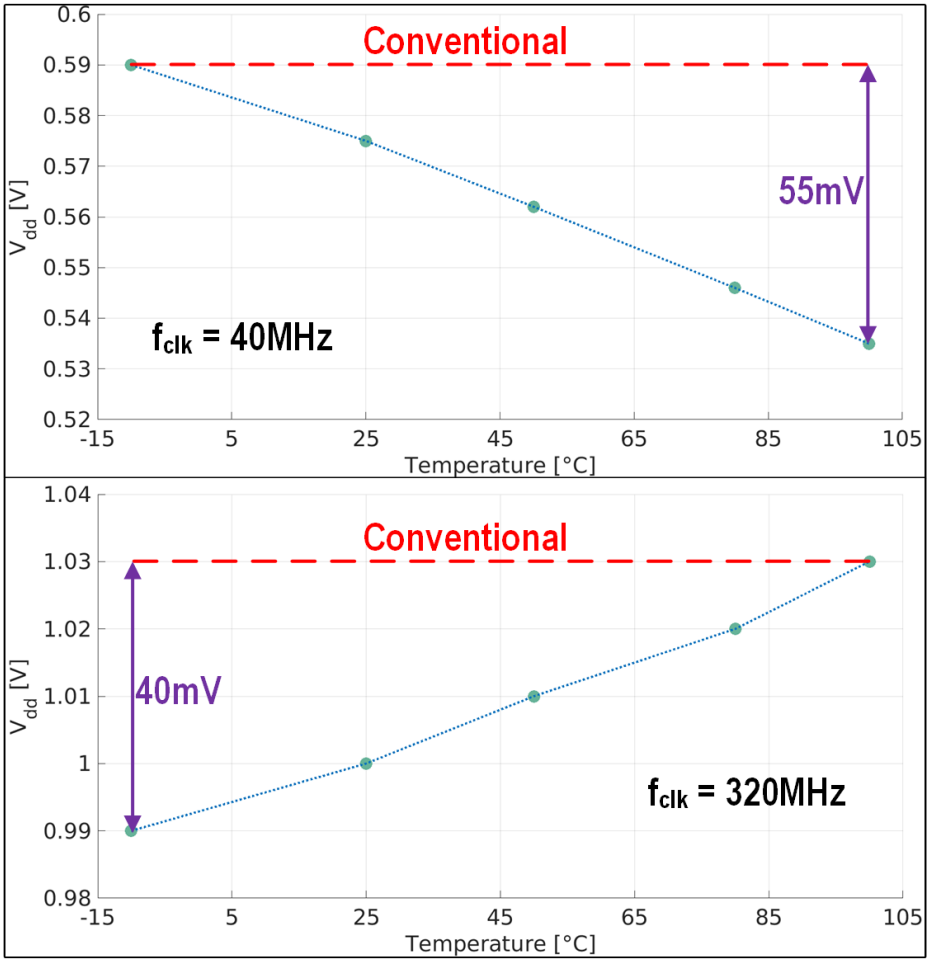


Figure 2.17: Measured V_{dd} vs. temp. with f_{targ} = 40MHz and 320MHz.

2.5.5 Total Energy-per-cycle (EPC)

The total energy-per-cycle (EPC) metric accounts for the total dissipation incurred in computation, inclusive of the clocking and power-delivery and regulation system. Total EPC therefore accounts for inefficiencies of supply noise and temperature-induced margins. At each frequency point, the conventional system requires regulating V_{dd} to a higher-than-ideal value, margining for worst-case supply droop and temperature variation in (-10°C, 100°C) range, resulting in increased energy dissipation. Figure 2.19 demonstrates the EPC benefit afforded by UniCaP, allowing a system energy consumption reduction of 35% at 1.0V, and 48% at 0.6V.

2.5.6 Performance Benefit

Maintaining phase-lock is one of the key advantages of *UniCaP*, allowing it to recover any cycles lost or gained during a V_{dd} transient. Figure 2.20 illustrates this capability, showing the impact of increasingly frequent fixed-amplitude voltage droop events on the effective system clock frequency indicated by the normalized clock rate, for conventional, discrete clock stretching [18, 17] and *UniCaP* systems. A fixed 15% frequency guardband is required by conventional systems, which remains steady regardless of droop frequency. Using discrete clock stretching approaches reduces the voltage margin required for low droop occurrence, but results in a normalized clock rate degradation due to more frequent clock stretching events. *UniCaP* not only requires a much lower guardband but also ensures constant clock rate throughout the droop event frequency sweep, regardless of the number of droop events since any cycle losses are compensated for through phase-lock.

2.6 Conclusion

This section describes a unified clock and power architecture to allow IVR enabled systems to aggressively reduce V_{dd} margins induced by supply noise and temperature variation. A buck converter based *UniCaP* system with all-digital ZVS and autonomous CCM/DCM

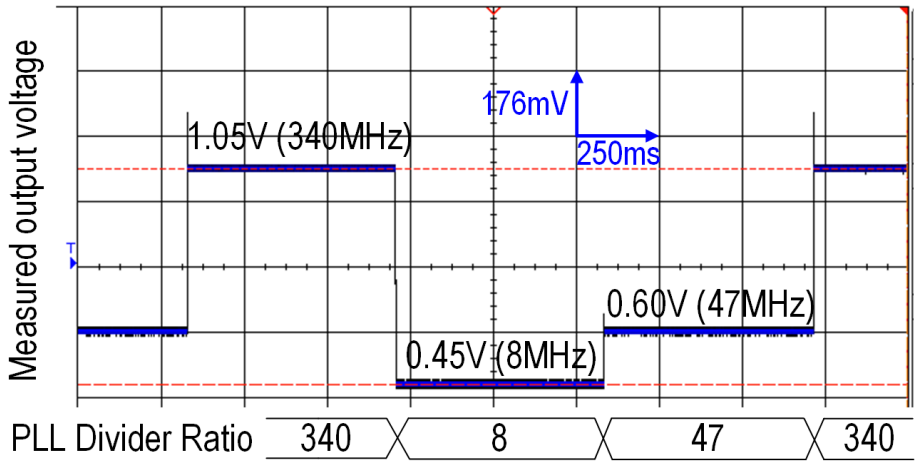


Figure 2.18: Measured oscilloscope capture of on-the-fly DVFS transitions by varying PLL divider ratio.

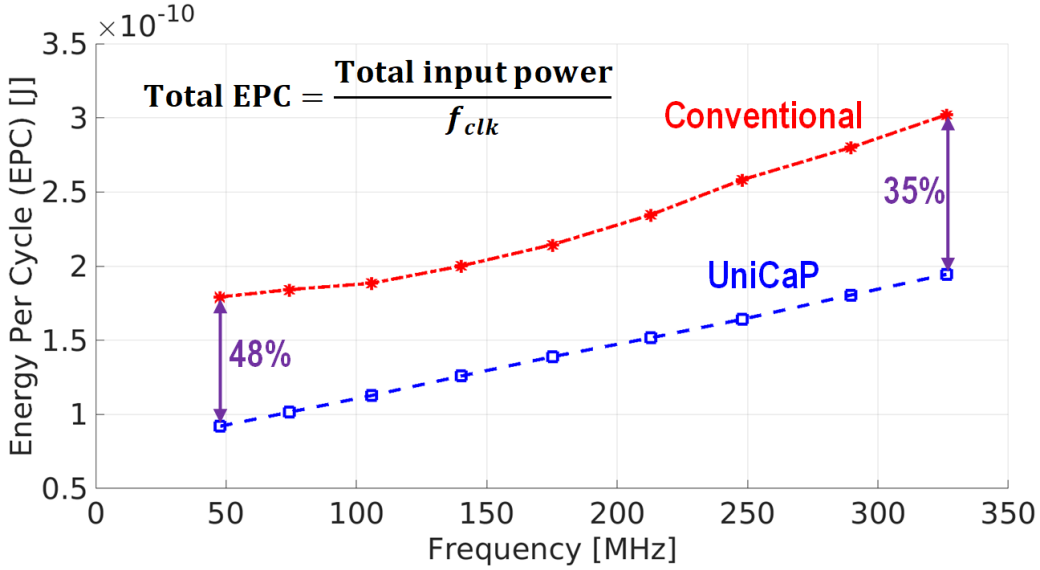


Figure 2.19: Measured total energy-per-cycle vs. operating frequency plot.

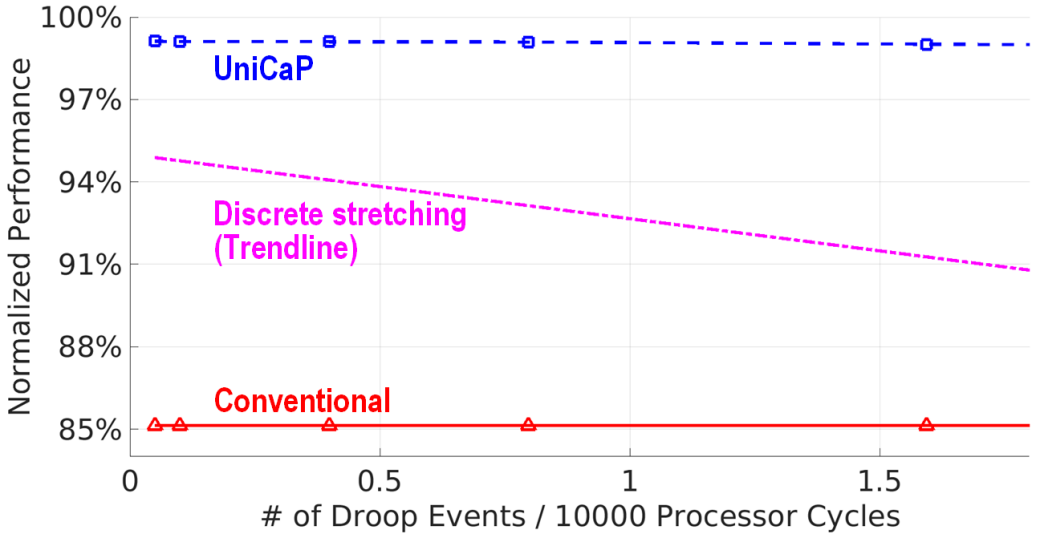


Figure 2.20: Normalized clock rate vs. droop event frequency for conventional, discrete clock stretching and *UniCaP* systems.

transition control was prototyped in 65nm CMOS. The test-chip achieves 10-13.5% V_{dd} reduction in 0.6-to-1.0V range by recovering an average of 82% V_{dd} margin due to supply droop and 40-55mV due to temperature variation. PLL phase-lock action ensures any cycles lost or gained resulting from supply noise are fully recovered.

Chapter 3

UNICAP-2: PHASE-LOCKED ADAPTIVE CLOCKING WITH RAPID CLOCK CYCLE RECOVERY IN 65NM CMOS

3.1 Introduction

UniCaP adaptive clocking architectures [7, 19, 20, 30], have recently demonstrated aggressive margin recovery while regulating average system performance. UniCaP employs a tunable V_{dd}-powered critical path replica oscillator (TRO) to produce a clock (*clk*) whose period (T_{clk}) tracks both the delay and V_{dd} sensitivity of the critical path across operating conditions (Figure 3.1(a)). Any V_{dd} change thus identically modulates T_{clk} and the critical path delay (τ_{crit}) to maintain timing slack. This property allows V_{dd} to be used as the sole control variable within the clock loop to phase-lock *clk* to a reference (*REFCLK*) without impacting timing. Unfortunately, using phase error ($\Delta\Phi$) – a time-averaged metric of V_{dd} error – inherently limits V_{dd} regulation bandwidth. Slow recovery of V_{dd} droop (and thus f_{clk}) increases peak cycle loss ($\Delta\Phi_{max}$), leading to increased pointer separation and even saturation in clock domain-crossing FIFOs. Furthermore, a combination of loop dynamics and existing control techniques limit phase-lock bandwidth to $f_{REFCLK}/100$ for stability [7], thereby degrading $T_{recovery}$ and rendering *UniCaP* unfeasible for real-time systems requiring performance guarantees in sub-20µs time frames. This paper introduces *UniCaP*-2 (Figure 3.1(b)), a dual-loop phase-locked adaptive clocking architecture that reduces $\Delta\Phi_{max}$ and $T_{recovery}$ while enabling higher V_{dd} margin reduction. We also leveraged the 65nm test chip to empirically assess the impact of clock distribution delay on attainable margin reduction.

3.2 Architecture

Figure 3.2 outlines the *UniCaP*-2 architecture as applied to a buck converter. V_{dd} remains the only control variable in the clock loop but is determined by two distinct feedback paths:

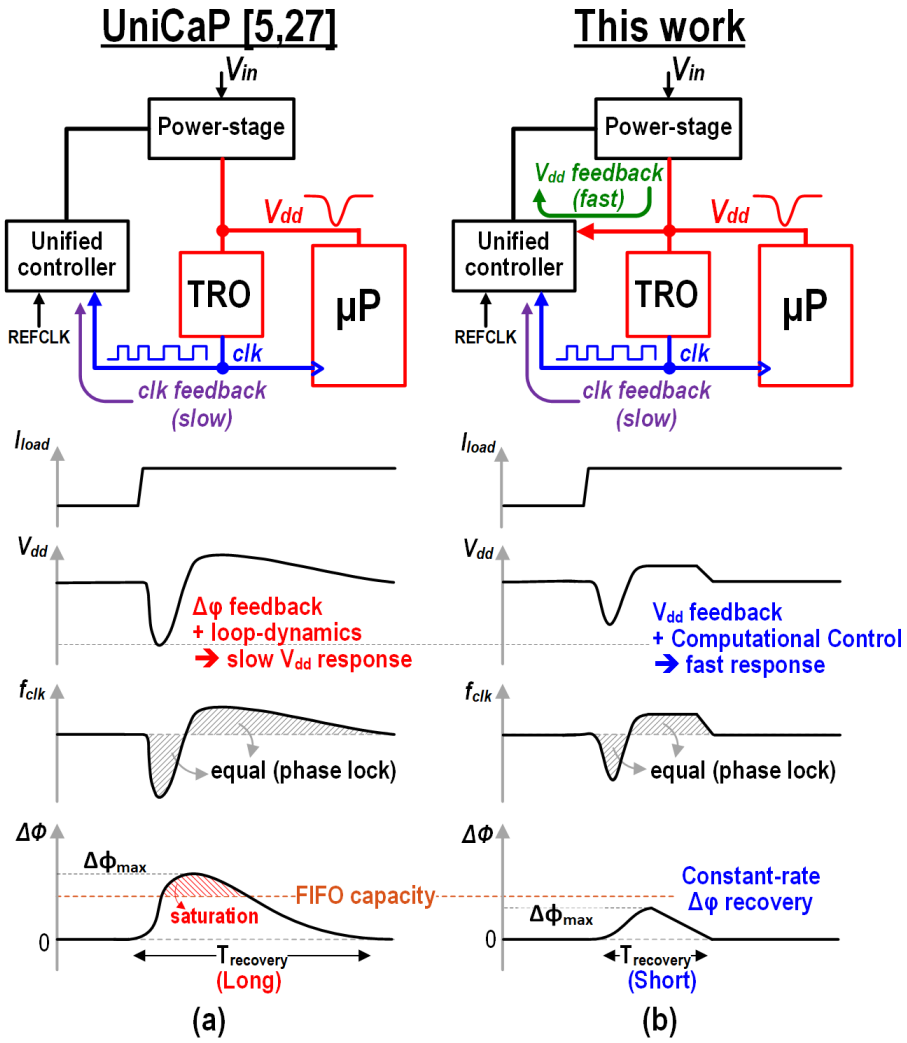


Figure 3.1: Construction and transient operation of (a) *UniCaP*, which is slow and leads to larger V_{dd} droop and (b) this work (*UniCaP*-2), which allows rapid V_{dd} recovery and phase lock.

(1) a faster V_{dd}-path that quickly senses and recovers from a V_{dd} droop (or surge) through Model Predictive Control (MPC) [10]; and (2) a slower phase-path that rapidly eliminates $\Delta\Phi$. The phase-path applies, in each *REFCLK* cycle, a calculated offset ($V_{dd,\Phi}$) to the reference voltage (V_{ref}). The corresponding f_{clk} adjustment after V_{dd} settles to $V_{ref}+V_{dd,\Phi}$, recovers any cycles gained (or lost) during a V_{dd} transient by the end of the *REFCLK* cycle. Restoring phase lock through a fixed V_{dd} (f_{clk}) offset achieves the fastest recovery under random I_{load} conditions and a given $\Delta\Phi_{max}$ constraint.

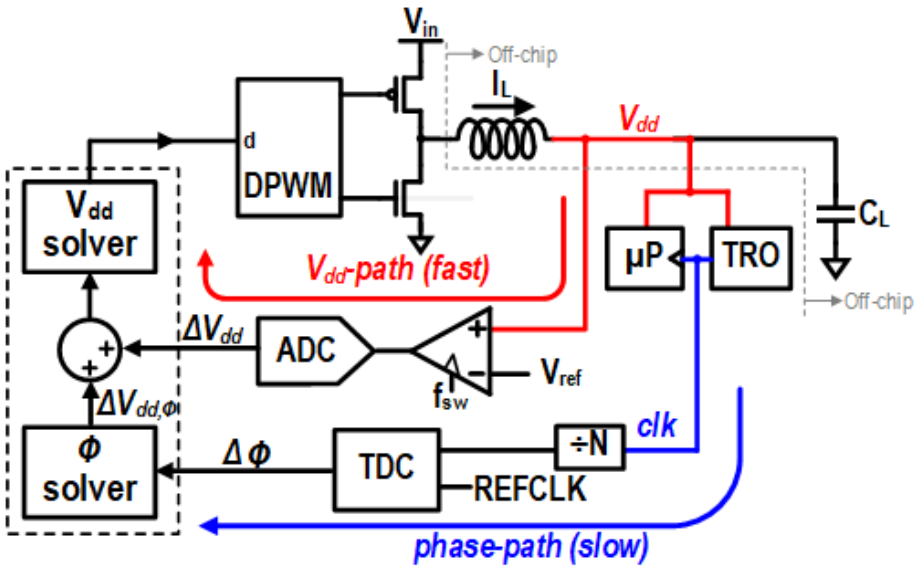


Figure 3.2: *UniCaP*-2 test-chip architecture. A fast V_{dd} loop limits droop. A slower phase loop achieves phase lock.

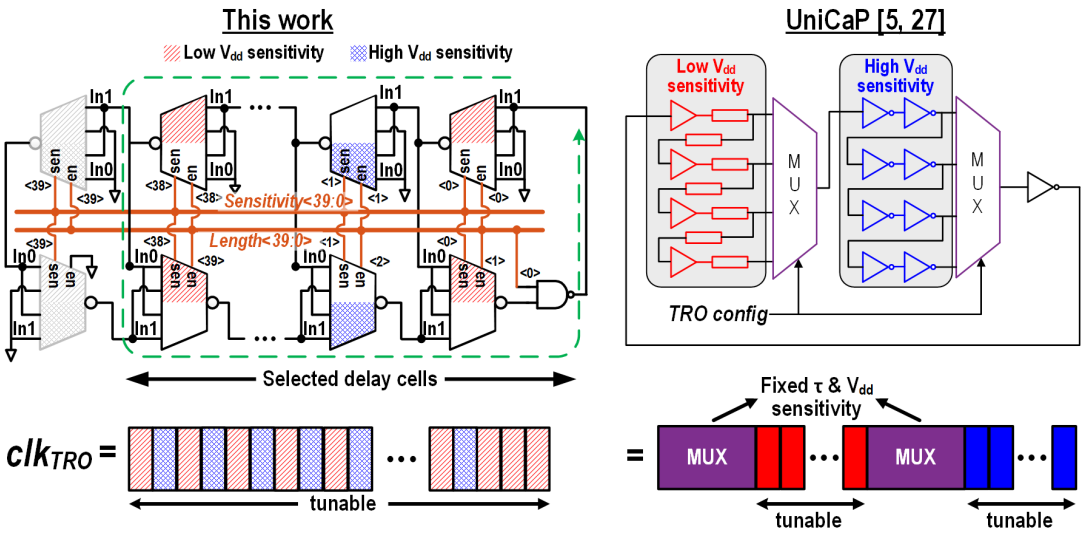


Figure 3.3: Proposed TRO design achieves finer delay and V_{dd} sensitivity control from improved τ_{crit} and s_{crit} matching.

Critical path replicas, designed to match both τ_{crit} and V_{dd} sensitivity ($s_{crit} = \partial\tau/\partial V_{dd}$), are essential to minimizing V_{dd} droop margins. Traditional TRO designs [7, 30, 18] rely on wide MUXes to configure the appropriate combination of high- and low-sensitivity delay cells, thus ensuring T_{clk} and its V_{dd} sensitivity (s_{TRO}) match τ_{crit} and s_{crit} respectively.

The proposed telescopic TRO (Figure 3.3) avoids the fixed-sensitivity delay overhead of these MUXes while offering superior per-cell V_{dd}-sensitivity control and precise τ_{crit} and s_{crit} matching.

3.3 Silicon Measurements

For V_{dd} margin characterization, the 65nm *UniCaP*-2 test chip comprises a Cortex-M0 processor and a configurable τ_{dist} of up to $8T_{clk}$ with programmable sensitivity. A programmable synthetic load (I_{load}) draws 350mA of peak current in 0.11ns at V_{dd} = 1V. These features allow the test chip to emulate realistic designs with larger τ_{dist} and power consumption. A die photograph of the test chip is shown in Figure 3.4.

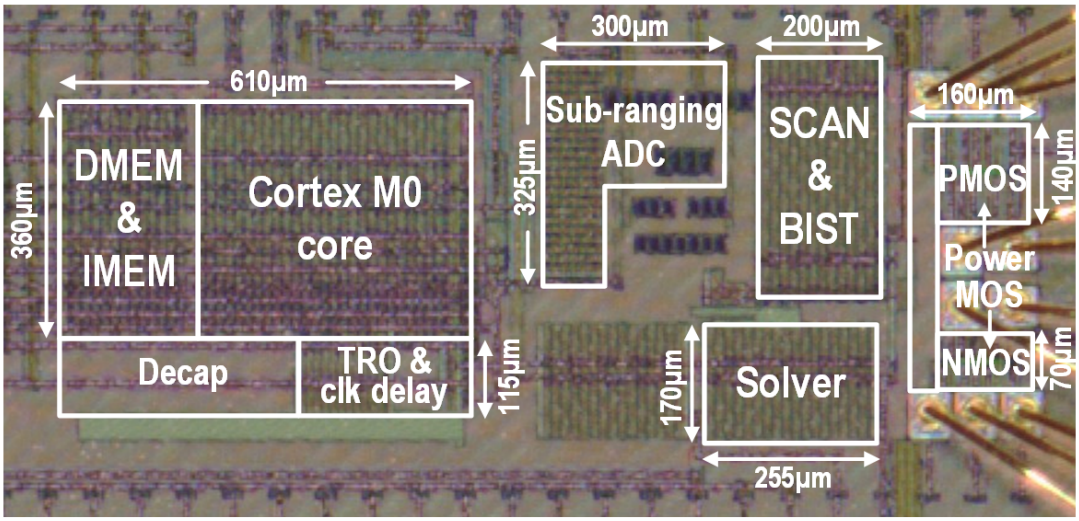
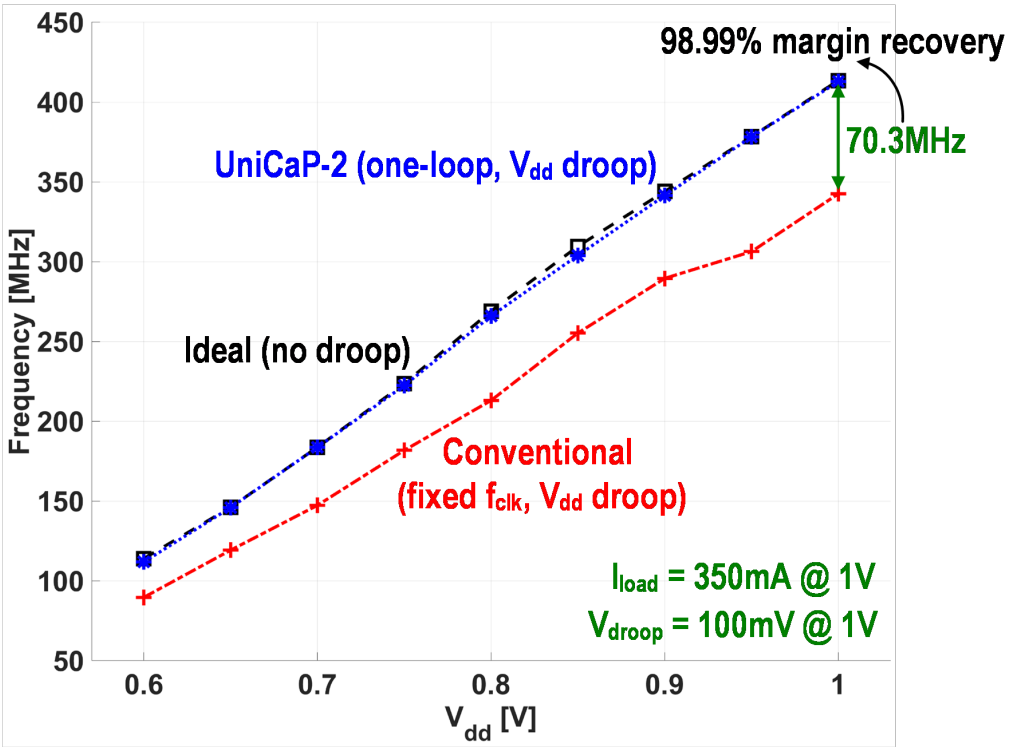


Figure 3.4: *UniCaP*-2 die photograph.

To measure the V_{dd} margin reduction achieved by *UniCaP*-2, maximum clock frequency (f_{max}) experiments were run on the Cortex-M0 core using standard benchmarks under various conditions (Figure 3.5). Baseline measurements were taken under nominal conditions without additional V_{dd} noise. The synthetic load was then used to draw a 350mA current step to produce 100mV droop at 1V, typical in a product scenario. The f_{max} measurements were made under fixed f_{clk} conditions (Conventional) and with *UniCaP*-2 enabled. *UniCaP*-2 achieves an average margin reduction of 96.2% under noisy V_{dd} conditions in the 0.6–1.0V range.



Note: Droop magnitude scales along with I_{load} at lower V_{dd}

V_{dd}	0.6	0.7	0.8	0.9	1.0
Margin recovery %	91.07%	99.81%	94.51%	95.32%	98.99%

Figure 3.5: Measured V_{dd} margin reduction with *UniCaP*-2 under 100mV supply noise conditions @ V_{dd}=1V.

Figure 3.6 shows measured V_{dd} step-up and -down responses (including snapshots of at every *REFCLK* cycle) in response to a 123mA/0.11ns I_{load} step. *UniCaP*-2 handles this transient by rapidly re-locking within a few *REFCLK* cycles. In the case of an I_{load} step-up, the Φ-solver in the phase-path uses the sensed $\Delta\Phi$ to make the required V_{ref} adjustment by $V_{dd,\Phi}$ to target phase lock by the end of the *REFCLK* cycle. MPC rapidly transitions V_{dd} to this adjusted voltage. Designs targeting a faster transient response through advanced (more integrated) package technologies can operate at higher buck switching frequencies (f_{sw}) to further reduce $T_{recovery}$.

The attainable margin reduction depends on matching between the generated T_{clk} and

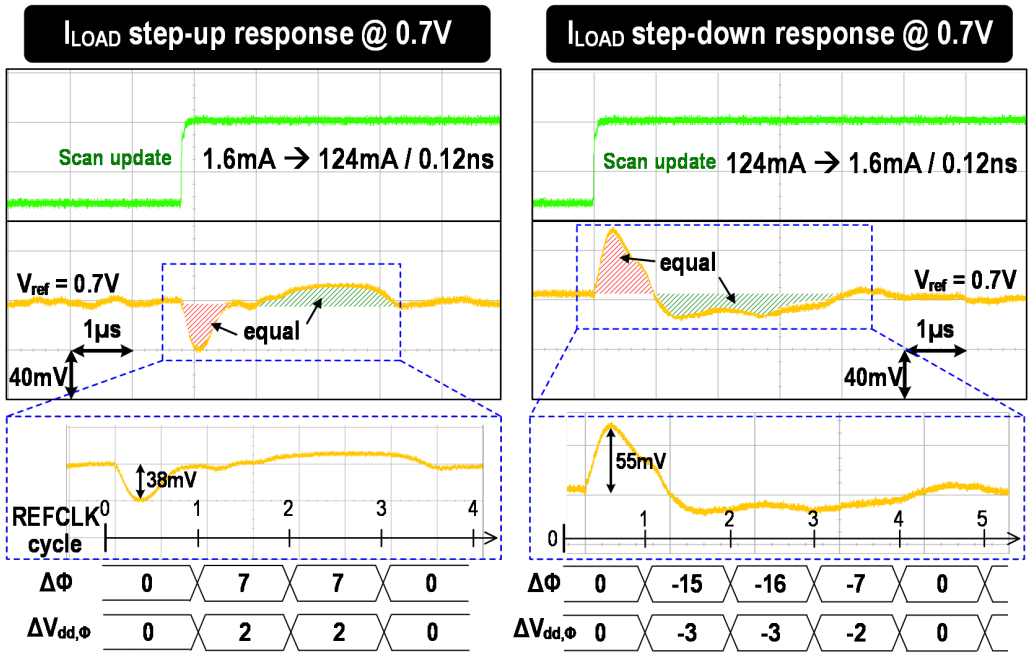


Figure 3.6: Measured V_{dd} response to I_{load} step. Runtime values of $\Delta\Phi$ and $V_{dd,\Phi}$ are captured by the BIST engine. *UniCaP*-2 asserts adjusted V_{ref} adjustments to target rapid phase-lock.

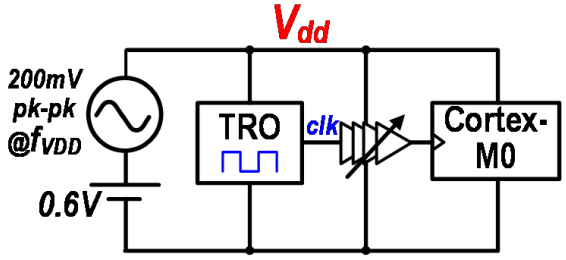


Figure 3.7: Empirical setup to evaluate the impact of τ_{dist} and s_{dist} on *UniCaP* V_{dd} margins.

τ_{crit} at the ends of the clock distribution. However, T_{clk} is generated by the TRO at the clock distribution source and is subsequently delayed, V_{dd}-modulated, and thus modified before reaching the sink nodes of this distribution. Ideal performance is thus achieved either when $\tau_{dist} \rightarrow 0$ or when the V_{dd} sensitivity of the distribution (s_{dist}) matches s_{crit}; any deviation from this ideal scenario degrades margin reduction. In this work, we empirically quantify the impact of dist and s_{dist}–s_{crit} mismatch. Margin reduction experiments at V_{dd}=0.6V were performed while varying τ_{dist} up to $8T_{clk}$ and across a variety of s_{dist}–s_{crit} mismatch scenarios (Figure 3.7). A programmable frequency ($f_{V_{DD}}$) sine-wave generator emulated

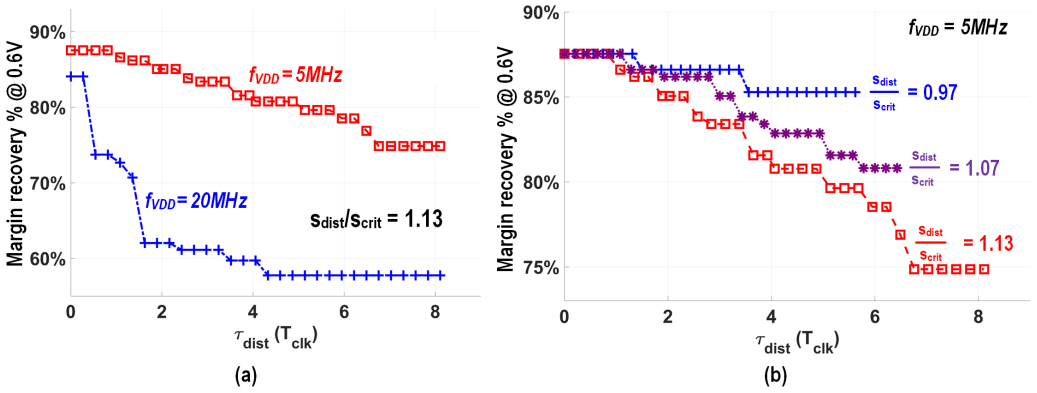


Figure 3.8: Margin recovery vs. τ_{dist} across (a) s_{dist}-s_{crit} mismatch and (b) $f_{v_{DD}}$.

the V_{dd} noise seen by the Cortex-M0. Measurements indicate a 2.5% (22%) reduction in recovered V_{dd} margin at $f_{V_{DD}}$ = 5MHz (20MHz) when $\tau_{dist} = 2T_{clk}$ (Figure 3.8 (a)). Measurements at higher $f_{V_{DD}}$ (Figure 3.8 (b)) confirm expectations of more rapidly decreasing margin reduction down to a plateau. Figure 3.9 compares the proposed architecture with related work.

Paper	[17]	[5]	[27]	[28]	**This work [6]**
Process	130nm	65nm	65nm	7nm FinFET	**65nm**
Topology	LDO	Buck	Sw.-cap	Buck	**Buck**
V_{in} [V]	0.6 – 1.0	1.2	1.2	1.1 – 1.3	**1.5 – 1.8**
V_{out} [V]	0.38 – 0.81	0.6 – 1.0	0.44 – 0.56	0.45 – 0.9	**0.6 – 1.0**
Frequency / Phase locked to REFCLK	Yes / No	Yes / Yes	Yes / No	Yes / No	**Yes / Yes**
Real-time compatibility (sub-5µs time-frame)	No	No	No	No	**Yes**
I_{LOAD} [mA]	0.1 – 6	1 – 100	N.A.	1 – 900	**1 – 400**
L [H] / C_L [F]	N.A.	10µ / 10µ	N.A.	1.6n / 0.47µ	**1µ / 1µ**
F_{sw} [MHz]	N.A.	1	N.A.	250[1] (F_{REF}/8)	**10**
Peak efficiency [%]	99.4 (current efficiency)	96.3	82.4	64	**91.9**
ΔV_{out} [mV] @ ΔI_{LOAD} / T_{EDGE}	160 @ 3mA/0.5ns	64.8 @ 90mA/1ns	50 @ 1mA/<1ns	55 @ 177mA/0.5ns	**38 @ 123mA/0.12ns**
$T_{recovery}$ (in REFCLK cycles)	N.A.	152	N.A.	N.A.	**4**
Overshoot [2]	28%[1]	0%[3]	0%	72.7%[1]	**6.7%[3]**

[1] Deduced from reported waveform measurement
[2] Overshoot percentage defined as peak overshoot/peak droop in response to a load step – stability indicator
[3] Based on measurement data with phase tracking path disabled

Figure 3.9: Comparison with related work.

Chapter 4

A COMPUTATIONALLY REGULATED DIGITAL LDO WITH 2.9-CYCLE MEAN SETTLING TIME

4.1 Introduction

Low dropout regulators (LDOs) play an important role in enabling fine-grain Dynamic Voltage and Frequency Scaling (DVFS) [32] for enhanced energy efficiency. Compared to their switched-capacitor and switched-inductor counterparts, the low area overhead and rapid transient response offered by LDOs make them particularly well-suited for use in multi-voltage domain SoC designs. Recently, DLDOs have attracted much interest due to their capability for use in low-voltage designs, ease of portability across designs, and scalability across process technology nodes.

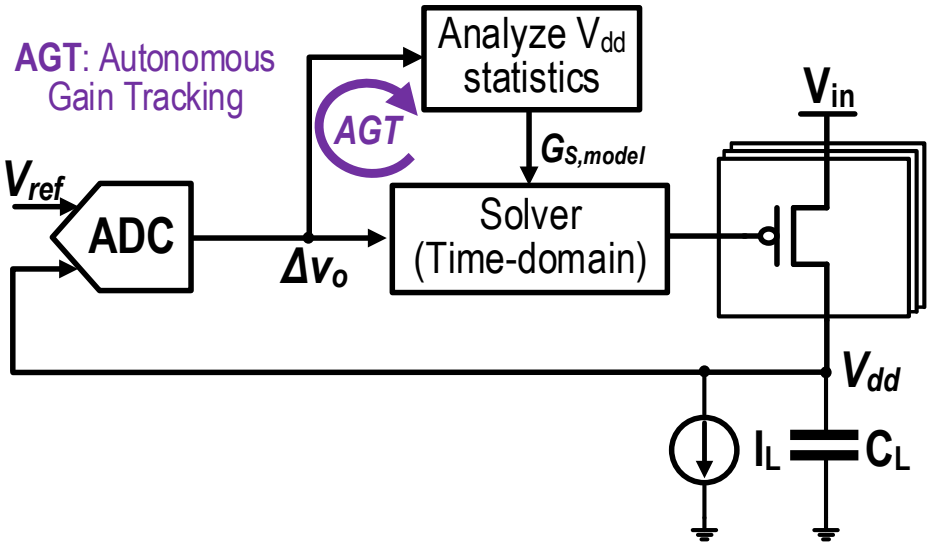


Figure 4.1: Overview of the proposed computationally enabled LDO architecture (CLDO).

Despite significant integration and scalability advantages, DLDOs face several challenges which currently hinder broad applicability across SoCs. These challenges, elaborated upon

in Section 4.2, include (1) delayed regulator response to load current (I_{load}) due to discrete-time V_{dd} sampling; (2) degraded response time and loop-stability due to the significant loop delay incurred between sensing V_{dd} deviation from a reference voltage (V_{ref}) and adjusting regulator output current; and (3) significantly higher loop gain sensitivity to process, regulator input voltage (V_{in}), V_{dd} and temperature ($PV_{in}V_{dd}T$) which forces stability margins and degrades response time.

Recent efforts have focused on improving DLDO transient response through non-linear control techniques and event-driven architectures. Event-driven DLDOs [33, 34, 35] overcome the limitations of discrete-time sampling by triggering the regulator in response to threshold crossings in V_{dd}. Digital loads, however, exhibit significant run-time I_{load} variation which can trigger frequent threshold crossings and lead to prohibitive switching losses. Introduction of hysteresis [35] overcomes this drawback, but at the cost of transient response degradation. A combination of linear and non-linear [36] techniques reduces DLDO settling time, but faces the challenge of stable and seamless transition between the two control modes under random I_{load}. Logarithmic search-based [37, 38] architectures similarly reduce DLDO settling time but can result in excessive droop in the presence of random I_{load} profiles.

In this chapter, we present a DLDO architecture that employs low-latency computation for rapid, stable response across $PV_{in}V_{dd}T$ (Figure 4.1) [9]. Regulator design involves deriving a time-domain model of the regulator that provides enhanced accuracy over a linear discrete-time transfer function model. Such a model is capable of capturing non-linear effects and loop-delay more effectively. A *Solver* is used for run-time evaluation of the resulting system of equations to achieve rapid response. Computation also enables the implementation of autonomous gain-tuning (AGT), a technique that relies on computing the auto-correlation of the sampled V_{dd} with single-bit precision in order to classify regulator response and enable loop gain auto-tuning across $PV_{in}V_{dd}T$ variations (Details will be discussed in Chapter 6).

A test-chip implementation of the proposed computationally regulated LDO (CLDO) powering a Cortex-M0 processor and a linear-algebra co-processor was fabricated in a low-power 65nm CMOS process. Silicon measurements indicate a 2.9-cycle mean settling time

under a random current loading configuration.

The remainder of this chapter is organized as follows: Section 4.2 outlines specific challenges and limitations of synchronous DLDO designs. Section 4.3 provides an architectural overview of the DLDO architecture, explaining how computation is used to achieve rapid response. System implementation details of the test-chip are presented in Section 4.4. Measurement results are presented in Section 4.5.

4.2 Challenges in Traditional DLDO Design

DLDOs face three significant challenges that adversely impact their transient response and stability. This section outlines these challenges briefly, particularly in contrast to analog LDOs.

Figure 4.2a shows an illustration of a simplified synchronous DLDO [39]. The error voltage (V_{ref}-V_{dd}) is sensed using an analog-to-digital converter (ADC) to sample and quantize V_{dd}. This quantized V_{dd} is processed digitally to control the number of PMOS headers that are turned on. Notably, header transistors operate with full gate overdrive ($V_{gs} = V_{in}$). ADC, control logic, and buffering required to drive the headers account for a loop delay (τ_d).

Sampling V_{dd} at discrete times inherently limits the transient response of synchronous DLDOs. As seen in Figure 4.2b, this discrete sampling leads to a delay between the I_{load} step and sampling V_{dd}. Without any response from the regulator, V_{dd} continues to drop during this time, worsening droop. Another consequence of the impact of the time delay between the I_{load} step and sampling V_{dd} is that V_{dd} droop magnitude and regulator settling time are become dependent on arrival time of the I_{load} step relative to the DLDO sampling clock edge (Figure 4.2c). For such designs, empirical measurement of the load-step response requires performing multiple iterations of the droop experiment and extracting the relevant statistics from the measurement of interest (e.g. voltage droop). While transient response degradation in synchronous DLDOs is unavoidable due to discrete sampling, it places a further emphasis on suppressing other contributors to droop and settling time.

Loop delay, modeled by τ_d in Figure 4.2 delays regulator response-time and reduces the phase-margin of the regulator. One source of delay is the latency incurred in the ADC, control logic and header buffers. A dominant delay contributor, however, is clocked timing-

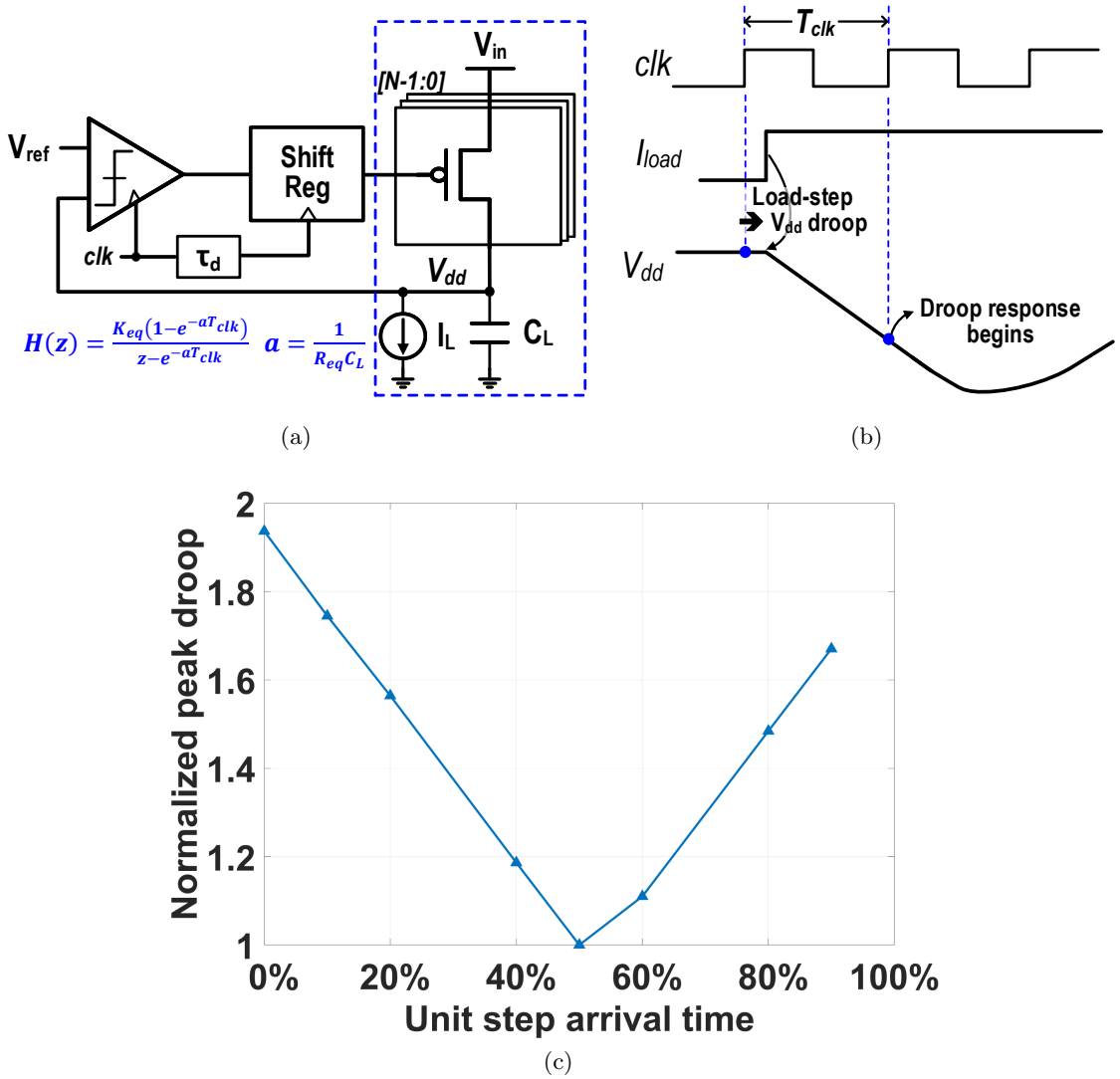


Figure 4.2: (a) Simplified synchronous DLDO schematic. (b) DLDO voltage response to an I_{load} step. Discrete sampling of the error voltage (V_{ref}-V_{dd}) leads to a delayed response, which also depends on relative timing between the I_{load} step and the error-voltage sampling edge. (c) Simulated peak droop (normalized) resulting from a unit step current at various times within a DLDO clock cycle.

elements that are inserted in the regulator control path to filter any glitches that arise either from the ADC or the logic. Preventing these glitches from arriving at the PMOS headers is critical to avoiding exacerbating both, switching loss and voltage droop. Traditional implementations employed flip-flops [39, 40], introducing a full-cycle of loop delay.

Designing a stable regulator with this significant additional loop delay requires conservative compensator design that results in a slower response. Negative edge triggered flip-flops have been employed more recently in an effort to reduce this delay [41]. Unfortunately, a half-cycle latency remains a significant factor in degrading the regulator response time-stability trade-off. Achieving rapid response in DLDOs requires addressing these contributors to τ_d.

An important aspect that distinguishes DLDOs from analog LDOs is that they operate by varying the number of conducting PMOS headers (or NMOS footers, in the case of NMOS based LDOs), each operating in the linear region with full gate overdrive. The expression for the current provided by a unit-header device can be written as [42]:

$$I_{lsb} = \frac{\mu C_{ox} W}{L}[(V_{in} - V_{th})V_{dropout} - \frac{1}{2}V_{dropout}^2], \tag{4.1}$$

where $V_{dropout} = V_{in} - V_{dd}$. The linear-mode operation of these header devices has major implications on system-linearity and $PV_{in}V_{dd}T$ sensitivity of the design that are in stark contrast with analog LDOs.

Varying the number of devices, each with a VI characteristic of Equation 4.1 yields a system whose circuit parameters are not time-invariant: DLDO circuit parameters vary from cycle to cycle. Thus, although all prior work in DLDO design that the authors are aware of has relied on transfer function-based analysis, this approach is at best an approximation: the resulting closed-loop DLDO system does not meet the Linear and Time-Invariant (LTI) requirement needed for frequency domain analysis. The non-linear behavior of DLDOs exposes a key limitation of traditional frequency-domain techniques (such as PID control) which can be avoided using time-domain approaches. Frequency domain approaches rely on a transfer-function formulation of the system, based on approximating the header resistance as a current source whose magnitude is varied at a resolution of I_{lsb}. Time-domain modeling – repeatedly evaluating (non-linear) RC expressions over one DLDO cycle – provides an accurate alternative.

Equation 4.1 also highlights the vulnerability of DLDOs to $PV_{in}V_{dd}T$ variation compared to analog LDOs for a given I_{load}. To elaborate, we first note that loop-gain contribution by the transconductance (g_m) of the header device of an analog LDO depends on the ratio of

I_{load} and gate-overdrive ($V_{gs} - V_{th}$). Because analog LDOs modulate header drive current through gate-to-source voltage (V_{gs}) control using feedback, scaling V_{in} at run-time causes the feedback loop to adjust the header gate voltage, thus maintaining V_{gs} and therefore header transconductance. Similarly, modulating the gate overdrive of a saturated device feedback allows the loop to compensate for threshold voltage (V_{th}) variation, and diminish the sensitivity of drain current to mobility variation in the header: both effects result from temperature variation.

In contrast with analog LDOs, the loop-gain of the DLDO header is determined by I_{lsb}, which varies strongly with V_{in} and $V_{dropout}$. Equation 4.1 also highlights DLDO sensitivity to temperature variation (through V_{th} and μ) on I_{lsb}, which directly impacts DLDO loop-gain. Ensuring stable response across wide variations in loop gain through traditional control mechanisms requires designers to robustly margin for the worst-case loop gain. Such margining degrades transient response and is particularly significant in the context of power-management features such as DVFS, where V_{dd} and V_{in} vary over a wide range.

4.3 Computationally-Enhanced DLDO: Architecture

This section describes the architecture for the proposed CLDO. The operation of the *Solver* in evaluating a time-domain model of the regulator is discussed in Section 4.3.1. The rationale and operation of the proposed adaptive gain tuning mechanism will be discussed in details in Chapter 6.

This work proposes a computational approach to DLDO control as an effective alternative to traditional PID based control. The approach relies on a time-domain model of the regulator (at the desired level of accuracy) to solve for, at the beginning of each cycle, the header width that will minimize settling time. The time-domain approach exploits continued advances in computing delay and energy-efficiency to perform more complex computations that allow for the time-varying circuit parameters (header resistance) and loop-delay effects in the DLDO to be modeled accurately, and thus enable rapid response. Furthermore, we also demonstrate the use of computation – low-precision analysis of quantized V_{dd} – to enable auto-tuning of the loop gain, avoiding worst-case margining of regulator loop gain that leads to transient response degradation.

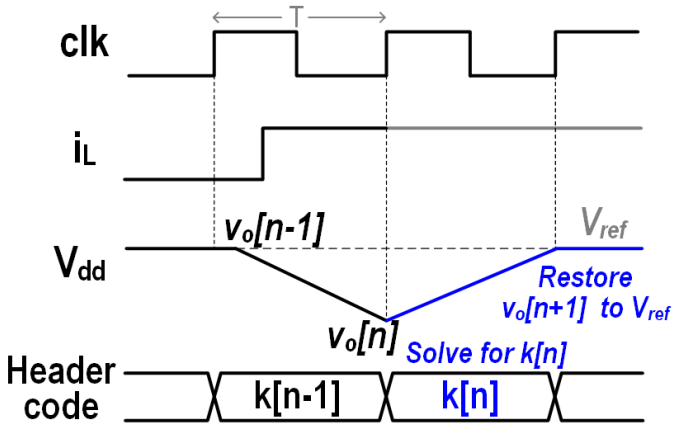


Figure 4.3: Ideal CLDO response to an I_{load} step with single cycle recovery.

4.3.1 Time-domain Model

Figure 4.3 illustrates the ideal response of the CLDO to an I_{load} step. The CLDO periodically quantizes V_{dd} in both voltage and time to obtain v_o. This ADC function is performed by a bank of comparators. Successive differences in v_o are indicative of a mismatch between the load current, I_{load} and the regulator current in the current cycle. Thus the average load current over cycle $n-1$ can be written as:

$$\overline{I_l}[n-1] = k[n-1] \cdot I_{lsb} + \frac{C_L}{T} \cdot (v_o[n-1] - v_o[n]), \tag{4.2}$$

where $k[n-1]$ represents the number of PMOS headers turned-on in the prior cycle. The number of headers to be turned on in cycle n is governed by the need for the CLDO to provide $\overline{I_l}$[n-1], and deliver the additional charge required to restore V_{dd}:

$$k[n] = \frac{\overline{I_l}[n-1]}{I_{lsb}} + \frac{C_L}{I_{lsb} \cdot T}(V_{ref} - v_o[n]). \tag{4.3}$$

Combining Equations 4.2 and 4.3 yields an expression for $k[n]$ as a function of v_o:

$$k[n] = k[n-1] + \frac{C_L}{I_{lsb} \cdot T}(V_{ref} + v_o[n-1] - 2v_o[n]). \tag{4.4}$$

Denoting $C_L/(I_{lsb}T)$ to be the solver gain (G_S), the expression in Equation 4.4 simplifies to:

$$k[n] = k[n-1] + G_S(V_{ref} + v_o[n-1] - 2v_o[n]). \tag{4.5}$$

Thus, the *Solver* effectively solves the model to determine $k[n]$ for a single-cycle response: ideally restoring V_{dd} after one CLDO cycle. This simplified model–which does not account for loop delay or the non-linear dependence of $V_{dropout}$ on I_{lsb}–can be shown to be identical to dead-beat control [43]. In dead-beat control, closed loop z-domain poles are positioned at the origin [43] to provide the fastest possible response. Loop-delay, however, is an important factor in determining stability, particularly in applications seeking rapid transient response. Although significantly suppressed in the CLDO architecture (discussed in Section 4.4), this delay must be accounted for in the regulator model.

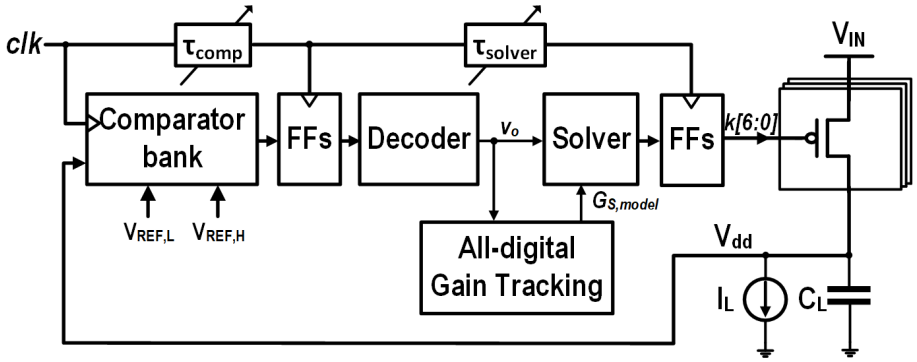


Figure 4.4: Block-level description of the DLDO implementation.

4.4 System Implementation

This section outlines the design and implementation of key aspects of the DLDO design (Figure 4.4) for the test chip, including circuit-architectures such as the voltage quantizer and datapath design for low-latency DLDO header update. We also discuss the implemented extension to the time-domain model that incorporates the effect of loop delay. Timing optimizations on the implementation of the resulting control law are discussed.

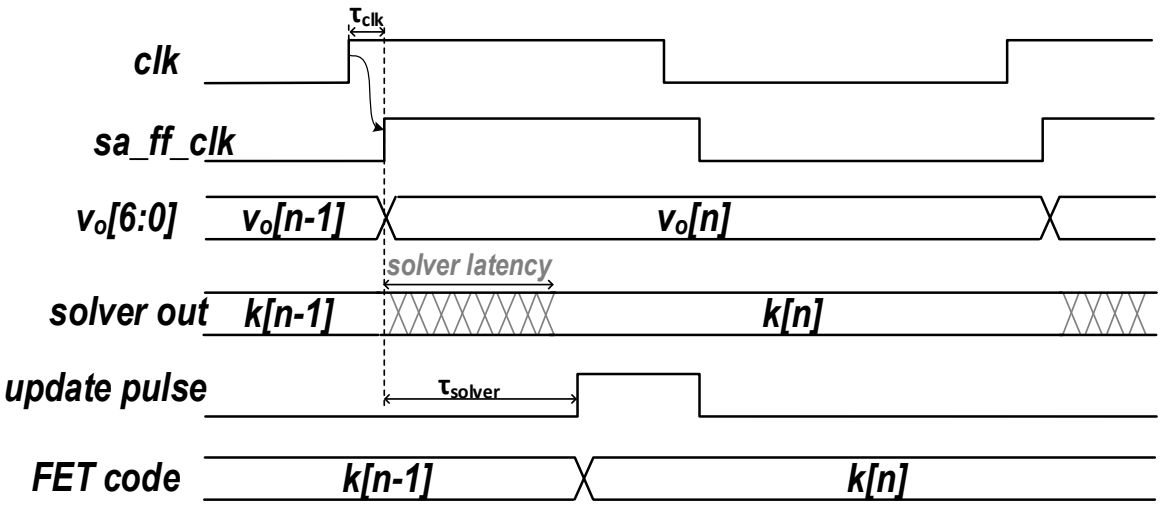


Figure 4.5: DLDO timing diagram.

4.4.1 Low-latency Header Update

The CLDO achieves low-latency, same-cycle header update while continuing to filter glitches that occur in the digital logic. Figure 4.4 illustrates the sequence of operations after the regulated voltage, V_{dd}, is first sampled and non-uniformly quantized by a bank of comparators. A decoder interprets the comparator-bank output to determine the error voltage (Δv_o), which is used by the *Solver* to evaluate a loop-delay aware time-domain model of the regulator. The *Solver* output is provided to a binary-weighted bank of PMOS header devices to control the output resistance of the DLDO.

Filtering glitches is necessary for DLDO implementations because both, the comparator-bank and the compensator (the *Solver* in the case of the CLDO) produce outputs that glitch. Process variation, and the dependence between resolution-time and error-voltage causes a spread in comparator evaluation time, resulting in glitches at the decoder input. Glitching at the *Solver* output results from the evaluation of static digital logic. State elements such as flip flops and latches are traditionally used to filter these glitches [42] but add excessive loop-delay.

The CLDO architecture fulfills the competing objectives of glitch filtering and reduced loop-delay by relying on flip-flops that are triggered by suitably skewed clocks. This is

achieved by inserting delays of τ_{comp} and additionally τ_{solver} to the flip-flops that sample the comparator and *Solver* outputs respectively. These delays represent a conservative bound on comparator and *Solver* evaluation times, avoiding any timing violations. Figure 4.5 illustrates the timing diagram of the resulting system, which allows $k[n]$ to be updated at a time αT $(0 < \alpha < 1)$ after the rising edge of the DLDO clock. Thus, the DLDO allows a header code update corresponding to the V_{dd} sampled in the same cycle: the loop-delay contribution is limited to the sum of the worst-case comparator resolution time and the *Solver* critical path delay.

4.4.2 Comparator Bank

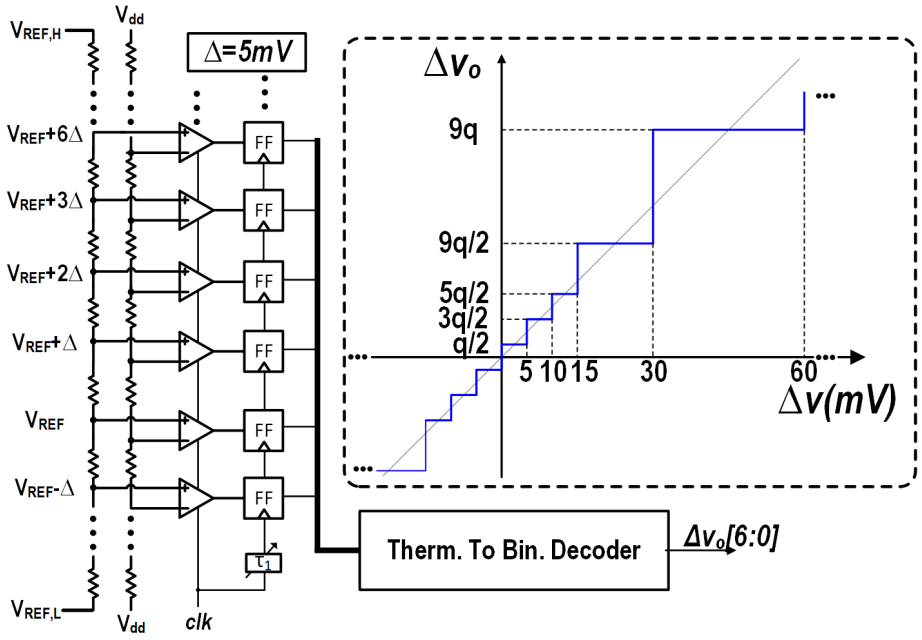


Figure 4.6: Comparator bank architecture used to quantize V_{dd}. A matching resistive network connects V_{dd} inputs to the comparators to mitigate kick-back noise. Suitable tap-off point selection from a chain of identical unit resistors provides non-uniform reference levels. Non-uniform quantization presents a suitable trade-off between the voltage resolution requirements and voltage capture range.

The CLDO relies on a bank of 13 comparators for direct flash conversion of V_{dd} every DLDO cycle. Covering a target range of 240mV with uniform quantization of 5mV would

require a 5.6-bit flash ADC. To relax this dynamic range requirement, we therefore implement a 13-level flash ADC with non-uniform quantization (Figure 4.6): Successive codes in each direction around the zero-code occur at Δ (5mV), 2Δ, 3Δ, 6Δ, 12Δ and 24Δ. Setting Δ to correspond to 5mV provides a converter range of 240mV, necessary to cover a worst-case voltage surge or droop. Such an approach avoids the complexity and power dissipation of a wide-dynamic range ADC while minimizing the impact of quantization error [44]. Non-uniform quantization provides wider conversion range but degrades settling time, a favorable trade-off that avoids the significant complexity and power dissipation of a uniformly quantized high-speed ADC.

A preampifier-free strong-ARM latch was used for the clocked comparator. Reliably achieving the required 5mV ADC resolution requires specific efforts to compensate for random variation between the differential MOS devices in the comparator. Suppressing this variability by increasing the size of the comparator is undesirable due to area and power overhead. The comparators employed in this ADC consist of an integrated bank of digitally tunable load capacitors that offers a offset cancellation resolution of 1mV (worst-case), and a range of 40mV (worst case). These offset cancellation specifications were found to be sufficient for the test chip. Intermediate reference voltages for the comparator were generated using a resistor chain made up of identical 30Ω salicided polysilicon resistors. The comparators allow for digitally-controlled offset cancellation at a resolution of approximately 1mV. Kick-back noise was mitigated by matching the impedance looking back from the comparator inputs. This matching was performed by using a resistor chain for the V_{dd} input of the comparator identical to that of the reference voltage as shown in Figure 4.6.

4.4.3 Modeling Loop Delay

Incorporating loop-delay into the simple time-domain model of the DLDO from Section 4.3 is an essential component of ensuring rapid regulator response that remains stable across $PV_{in}V_{dd}T$. Stability considerations constrain DLDO transient response time, motivating effective modeling of loop-delay, a key contributor to instability. The effect of loop delay in DLDOs is that the the number of headers enabled in cycle n is $k[n-1]$ for a duration

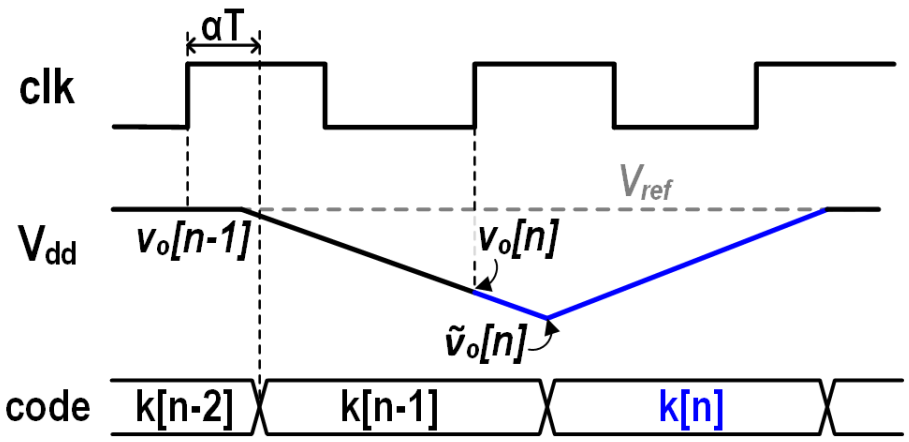


Figure 4.7: Loop delay results in a delayed update of $k[n]$. The timing model samples V_{dd} at the rising clock edge, but adjusts the measurement to determine $\widetilde{v}_o[n]$, the sampled voltage at the onset of applying $k[n]$, chosen to set $\widetilde{v}_o[n+1] = 0$.

of αT (Figure 4.5), and $k[n]$ for the remaining time $(1-\alpha)T$. Consequently, the effective header-drive depends on both $k[n-1]$ and $k[n]$. This two-cycle dependence must either be accounted for in the timing model, or must be compensated for by over-damping the regulator response. To maintain fast transient response, we incorporate the effect of this αT delay.

Figure 4.7 shows a simplified timing waveform that illustrates adjustments applied to the simple timing model discussed in Section 4.3.1. The adjustment adopted to incorporate loop-delay has the effect of effectively shifting the time-reference of the analysis to αT after the rising edge of clock. Once the error voltage, $\widetilde{v_o}[n]$ is determined in this shifted reference, the *Solver* solves for $k[n]$ using an approach identical to that discussed in Section 4.3.1.

At the start of cycle n, the *Solver* first samples v_o at the rising edge, and estimates the average load current over the previous cycle:

$$\begin{aligned}\overline{I_l}[n-1] =& \frac{C_L}{T}(v_o[n-1] - v_o[n]) \\ &+ (\alpha k[n-2] + (1-\alpha)k[n-1])\cdot I_{lsb}.\end{aligned} \tag{4.6}$$

This loop-delay aware load current evaluation allows for the evaluation of $\widetilde{v_o}[n]$, the projected

value of v_o at time αT after the rising edge of the clock when $k[n]$ is asserted at the headers:

$$\widetilde{v_o}[n] = v_o[n] + \frac{\alpha T}{C}(k[n-1] \cdot I_{lsb} - \overline{I_l}[n]). \tag{4.7}$$

With $\widetilde{v_o}[n]$ determined, $k[n]$ can be determined by evaluating the expression for restoring v_o at time αT into the next cycle:

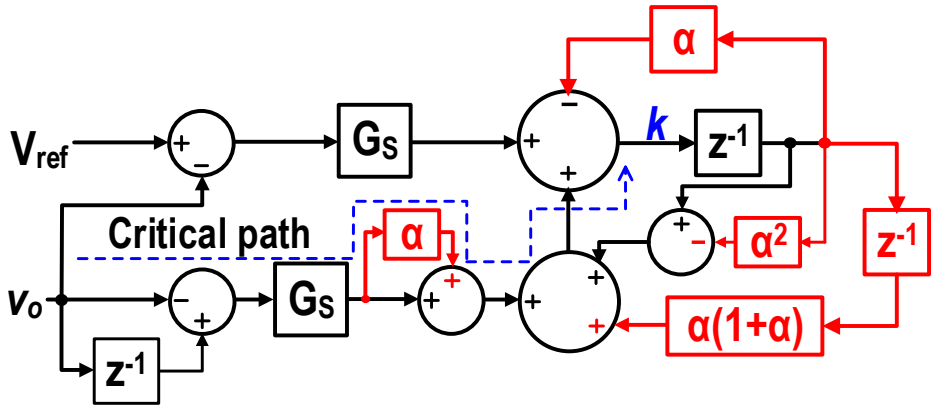
$$\frac{C}{T}(V_{ref} - \widetilde{v_o}[n]) = k[n] \cdot I_{lsb} - \overline{I_l}[n-1]. \tag{4.8}$$

Equations 4.6, 4.7 and 4.8 can be combined into a single expression that relies on v_o, which is evaluated by the *Solver* in each DLDO cycle to solve for $k[n]$.

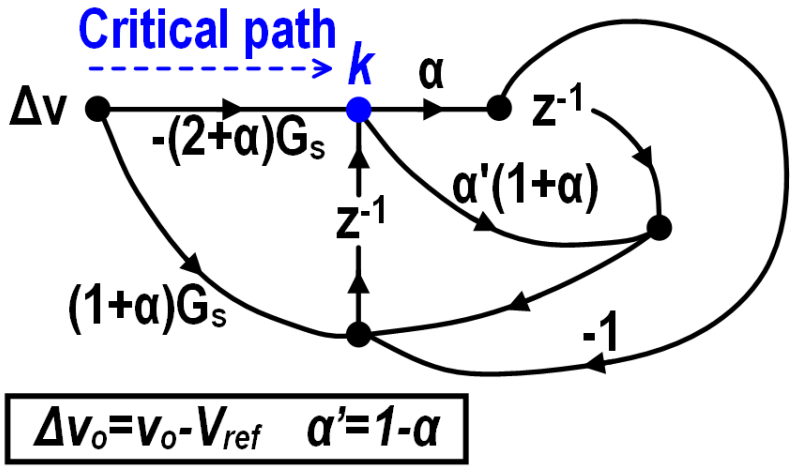
$$\begin{aligned} k[n] = G_S[(V_{ref} - v_o[n] + (1+\alpha)(v_o[n] - v_o[n-1])] \\ + (\alpha^2 - \alpha - 1)k[n-1] + (\alpha^2 + \alpha)k[n-2] \end{aligned} \tag{4.9}$$

Figure 4.8a illustrates the datapath for computing $k[n]$ according to Equation 4.9. Additional multiplications are associated with the loop-delay augmented model (in red) but these involve scaling by constant factors. Note that although the augmented model accounts for loop-delay, the uncertainty in estimating $\widetilde{v_o}[n]$ and need to minimize the response time in reacting to a droop motivates minimizing α. The critical path of the datapath is therefore determined by the combinational delay from v_o to $k[n]$. In order to reduce this delay, we optimized the signal-flow graph of the datapath, consolidating terms and scheduling computations corresponding to sequential paths (i.e. from or to flip-flops). The resulting design is illustrated in Figure 4.8b and achieves significant critical path delay reduction, reducing it to one multiply-accumulate operation.

The combined effect of using CLDO flip-flops with delayed data-capture (Figure 4.4) clocked at $f_s = 100MHz$, and an optimized loop-delay aware datapath delivers $\alpha = 0.35$, inclusive of buffer delay incurred in driving the header devices.



(a)



(b)

Figure 4.8: (a) Datapath for computing $k[n]$. Modules marked in red are required to model loop delay. (b) Optimized signal-flow graph of the datapath, reducing the loop-delay contribution to one multiply-accumulate operation.

4.5 Measurement Results

The proposed computationally-enabled DLDO was implemented in 65 nm CMOS (Figure 4.9). The design uses no off-chip decoupling capacitance (decap), and relies on 0.25nF of explicit decap. Loading for the DLDO was provided by a Cortex-M0 processor and a linear-algebra co-processor. We also implemented a synthetic arbitrary load-current generator that is memory-mapped to the processor in order to offer programmable support for software-

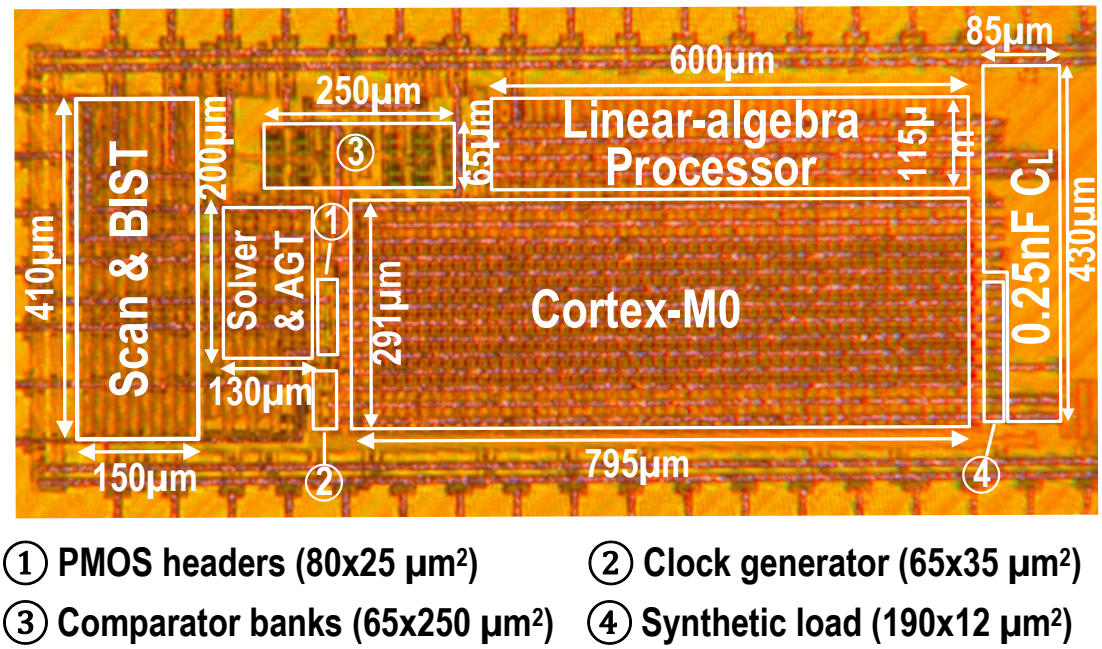
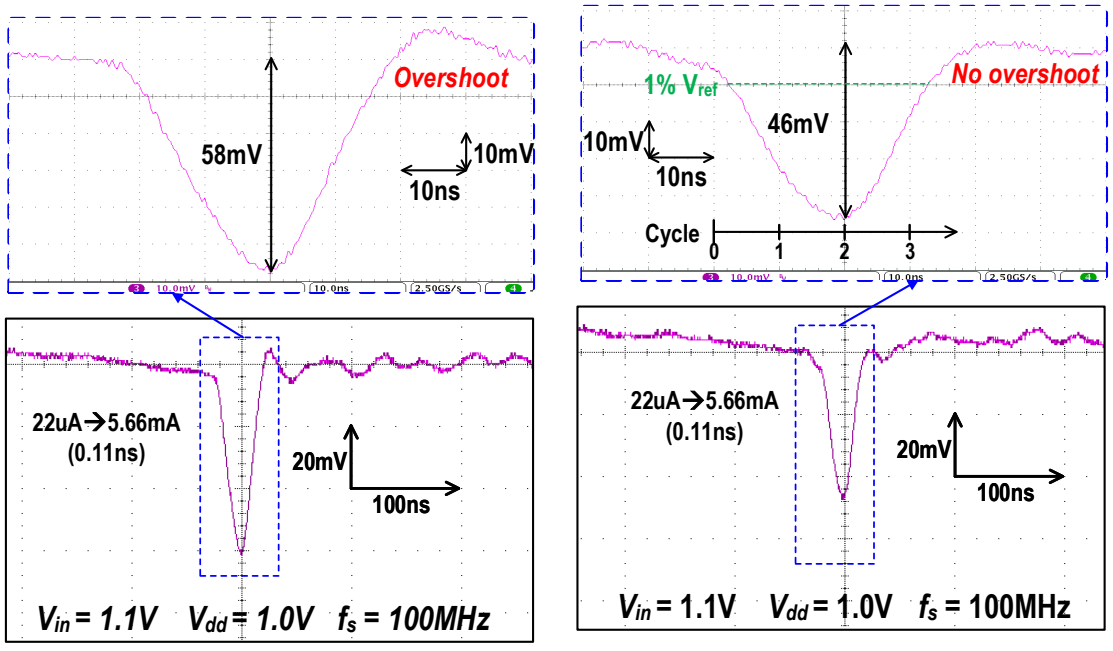


Figure 4.9: Die micrograph of proposed DLDO test-chip.

defined load-current waveform shape and magnitude.

Figure 4.10 illustrates the voltage response of the DLDO under a rapid load step from 22μA to 5.66mA with a 110ps rise-time. Oscilloscope waveforms are shown both, with and without loop-delay modeling enabled. Evaluating the loop-delay model affords the DLDO significantly rapid response, which provides an additional 20% droop improvement. This droop comparison also provides insight into the significance of accurately modeling α. Effectively, not modeling the loop delay is equivalent to setting $\alpha = 0$. Voltage droop progressively worsens as the error in modeling α continues to increase. Measured transient response under low V_{in} is shown in Figure 4.11.

The settling time of DLDOs is an inherently statistical parameter as explained in Section 4.2. Consequently, multiple iterations of load-step response experiments were conducted to obtain a better understanding of DLDO capabilities. The histogram produced from data obtained over 3000 iterations is presented in Figure 4.12. Measurements indicate that the mean (max) settling time of the DLDO for a load-step response is 2.9 and 5 respectively. While both mean and maximum settling time are relevant for voltage regulation and clock-



(a) (b)

Figure 4.10: Measured voltage response to a load-step (a) without loop-delay modeling and (b) with loop delay modeling. Settling times are measured from the clock edge when the droop is first detected till V_{dd} recovers to within 1% of V_{ref}

ing architectures like *UniCaP* [45, 46], the relevant metric for DLDOs is the worst-case droop and settling time.

Analyzing the voltage response to an I_{load} step is sufficient to determine the voltage response of a regulator that is a linear, time-invariant system (LTI). However, because the LTI model for DLDO architectures is an approximation, characterizing worst case regulator response using a step current is not sufficient: evaluating and validating DLDO operation under random current loading is critical. Figure 4.13 shows the voltage response of the LDO resulting from random loading provided by the programmable synthetic load. The processor was programmed to generate a pseudo-random data sequence which was periodically written to the memory-mapped synthetic load registers. Peak I_{load} amplitude was limited to 12mA. Measured waveforms demonstrate the ability of the CLDO to regulate V_{dd} under random loading conditions.

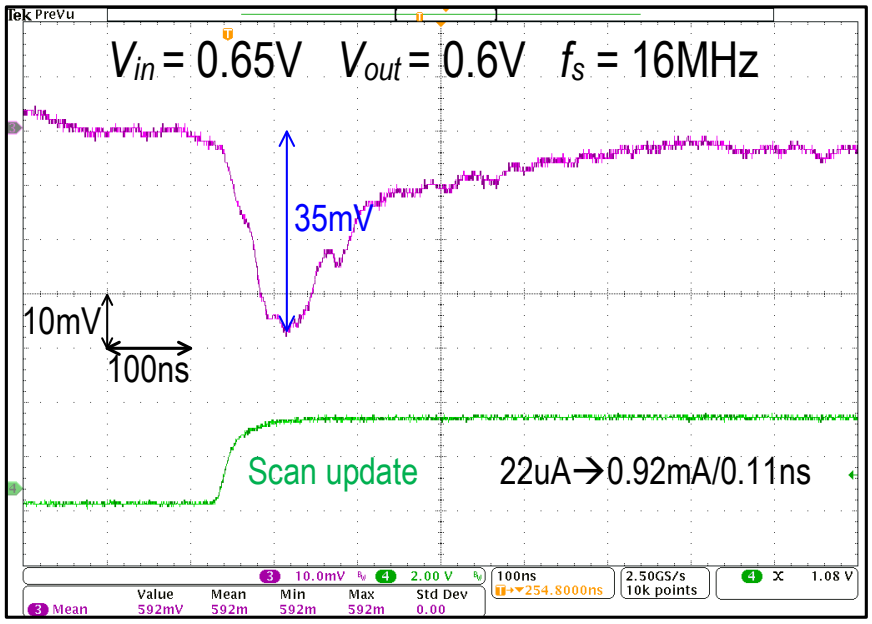


Figure 4.11: Measured voltage response to a load-step under low V_{in}.

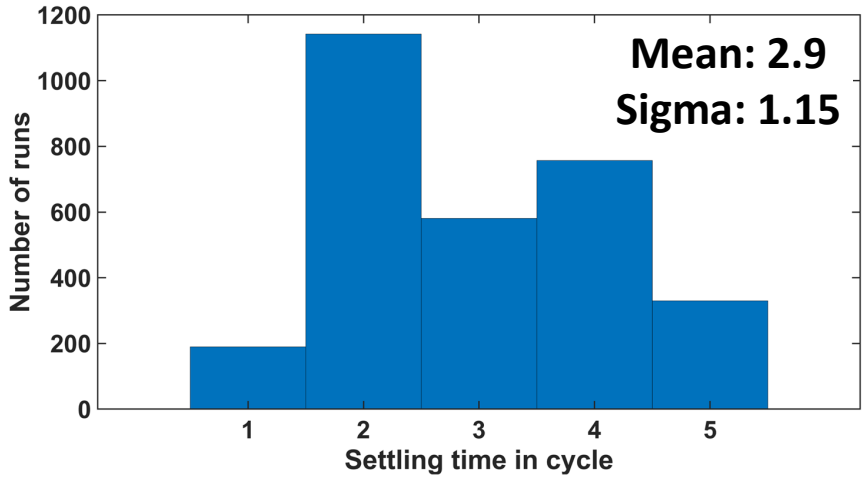


Figure 4.12: Histogram of LDO recovery time. Measurements were obtained at V_{in}=1.1V, V_{dd}=1.0V.

Measured load regulation and current-efficiency curves for the CLDO test-chip are reported in Figure 4.14. More significant computation relative to a traditional DLDO, and the design for low-latency same-cycle LDO header update in an older, low-power CMOS process leads to higher energy dissipation associated with CLDO control. However, switch-

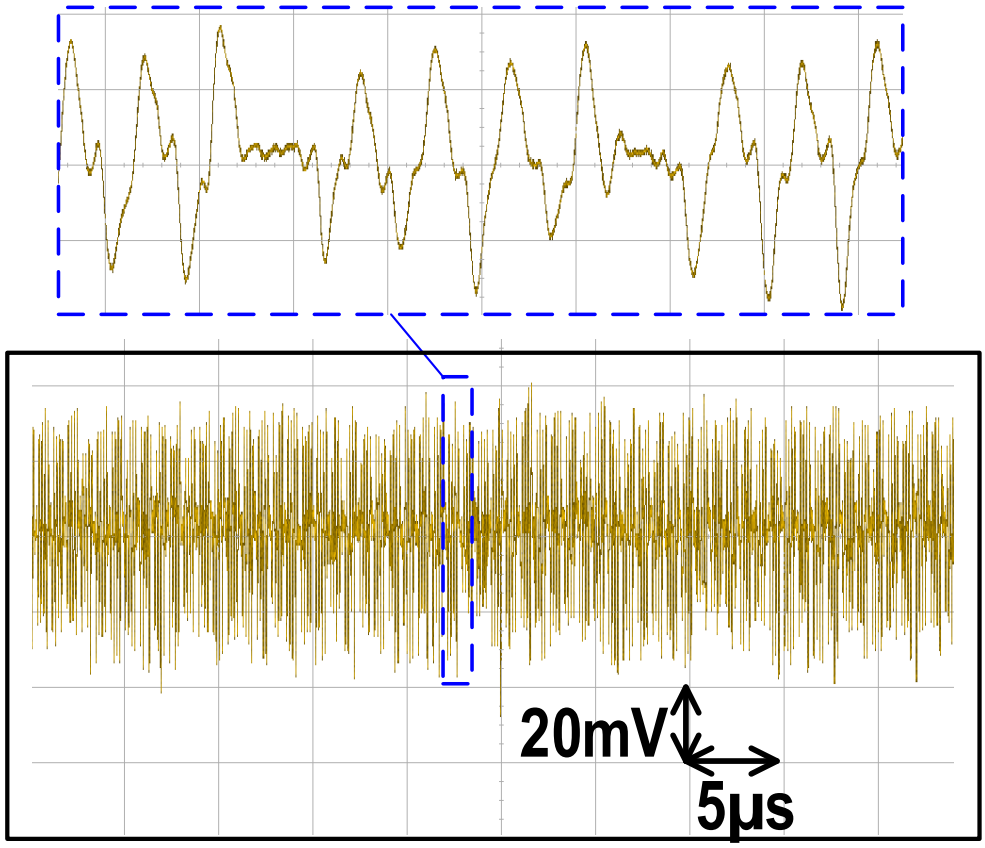
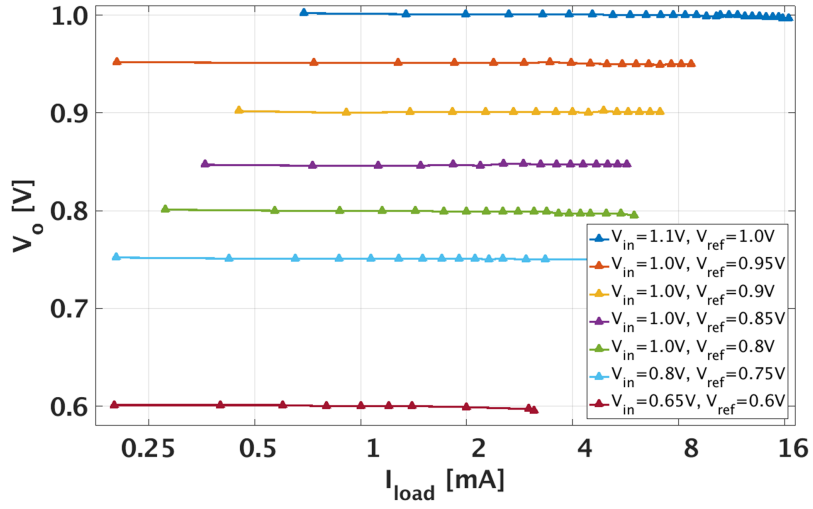


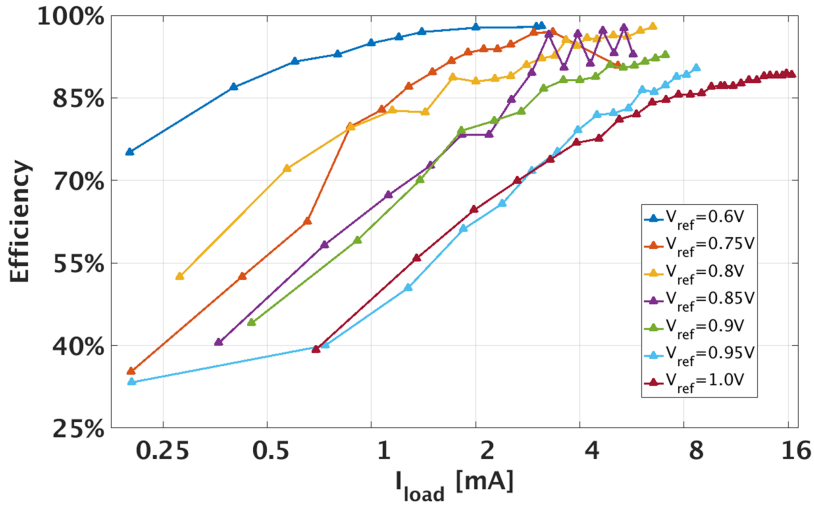
Figure 4.13: Measured voltage regulator response under random load conditions at V_{in}=1.1V, V_{dd}=1.0V, f_s=100MHz.

ing losses in the controller can be reduced in faster process nodes, and amortized in the case of larger and more dissipative voltage domains. CLDO efficiency is therefore expected to improve with larger voltage domains and in advanced process technologies. The proposed architecture may be well suited for smaller voltage domains if the system-level efficiency benefits from reduced supply droop guardbands outweigh regulator efficiency losses.

Figure 4.15 compares the proposed CLDO with related efforts. Leveraging computing allows for V_{dd} restoration in the lowest number of cycles. Although total settling time in [36] is lower, we note that the lower response-time results from a significantly higher switching frequency. CLDO implementation using faster, more efficient process technologies will afford improvements in quantizer resolution and enhanced time-domain modeling, including the



(a)



(b)

Figure 4.14: Measured load regulation and current-efficiency curves for the CLDO.

effect of $V_{dropout}$ on I_{lsb} which is currently not captured in the model. While discrete sampling remains a fundamental limiter to DLDO settling time, these enhancements can be expected to achieve maximal reduction in droop and settling time.

Paper	[5]	[3]	[9]	[8]	[20]	[21]	[22]	[23]	**This work**
Technology	14nm	65nm	40nm	130nm	65nm	65nm	65nm	65nm	**65nm**
Type	Digital	Digital	Digital	Digital	Digital	Digital	Digital	Analog assisted DLDO	**Digital**
Control topology	Time-driven	Event-driven	Time-driven	Time-driven	Time-driven	Event-driven	Event-driven	Time-driven	**Time-driven**
Control type	Linear (type-2) + non-linear	Linear PI	Linear PID	Linear PI + RDS	Linear PI (Beat frequency)	Linea PI	Linear PI + residue tracking	Linea PI	**Computational (digital)**
Autonomous PVT tracking	No	No	Yes	No	No	No	No	No	**Yes**
V_{in} [V]	1 – 1.15	0.45 – 1	0.6 – 1.1	0.5 – 1.2	0.6 – 1.2	0.5 – 1.0	1.2	0.6 – 1.2	**0.65 – 1.15**
V_{out} [V]	0.5 – 1.12	0.4 – 0.95	0.5 – 1	0.45 – 1.14	0.4 – 1.1	0.45 – 0.95	0.5 – 1	0.55 – 1.15	**0.6 – 1.1**
Max I_{LOAD} [mA]	2200	1.44	210	4.6	100	5.6	700	500	**16.3**
C_L [nF]	18	0.1	20	1	0.04	0.1	0.1	0.9	**0.25**
I_Q [µA]	14960	8.1 – 258	22.6 – 98.5	24 – 221	100 – 1070	18.1 (min)	254	500	**80 – 1200**
F_{sample} [MHz]	400 (linear) 800 (nonlinear)	N/A	N.R.	5 - 75	N/A	N/A	N/A	100	**16 – 100**
ΔV_{out} [mV] @ ΔI_{LOAD} / T_{EDGE}	100mV @ 1A/10ns	34mV @ 1.44mA	36mV @ 200mA	40mV @ 0.7mA	108mV @ 50mA	49.8mV @ 2.3mA/0.1ns	130mV @ 700mA	125mV @ 450mA/20ns	**46mV @ 5.6mA/0.1ns**
Droop settling: # of cycles (absolute time)	11.4 (14.6ns)	N/A (11200ns)	N.R. (13000ns)	33 (1100ns)	N/A (1240ns)	N/A (26ns)	N/A (>7us)	25 (250ns)	**2.9 (29ns)**

Figure 4.15: Comparison with related works.

4.6 Conclusion

This chapter describes a digital, computationally enabled architecture for rapid, stable voltage regulation across $PV_{in}V_{dd}T$ conditions. Computation is leveraged to evaluate accurate time-domain models of regulator behaviour, and to enable AGT, a technique that performs low-precision statistical analysis of quantized V_{dd} measurements to classify regulator response and autonomously tune regulator loop gain for rapid, stable response. A 65nm CMOS test-chip demonstration of the proposed CLDO regulating V_{dd} for a Cortex-M0 processor and a linear-algebra co-processor is presented.

Chapter 5

MODEL PREDICTIVE CONTROL OF INTEGRATED BUCK CONVERTER FOR DIGITAL SOC DOMAINS

5.1 Introduction

Minimizing peak V_{dd} droop (V_{droop}) margins in response to I_{load} transients is increasingly critical to SoC efficiency. Toward this goal, the increased adoption of adaptive clocking [18, 17] motivates the additional objective of τ_{settle} minimization: the time taken to restore V_{dd} to a reference (V_{ref}). Several techniques for aggressive τ_{settle} reduction including Time-Optimal Control (TOC) [47] have been proposed [48]. However, these techniques either pose stability analysis challenges, or work only for load-step current waveforms – not random loading. For this reason, PID control continues to dominate. Model Predictive Control (MPC) is a modern, widely adopted control technique capable of providing stable and optimal response under random disturbances. However, its computational complexity—requiring a constrained optimization problem to be solved at each time step—has thus far restricted its use to applications that can tolerate tens of microseconds of computational latency [49]: achieving sub-10ns loop delay latencies, frequently required of IVR control, remains elusive. This dissertation describes how control and datapath optimizations, combined with the computational speed in modern CMOS, realize an IVR buck regulator with optimal transient response.

5.2 MPC Buck Architecture

Figure 5.1 shows the proposed MPC buck regulator test-chip architecture. A Solver module incorporates the sampled V_{dd} at cycle n ($V_{dd}[n]$) into an accurate (non-linear) predictive model that tracks and updates the impact of the applied duty cycle ($d[n]$) on inter-dependent buck state variables – cycle-averaged inductor current ($\overline{I_l}$) and capacitor voltage (V_{dd}). Traditional MPC uses this predictive model to plan an optimal buck duty-cycle trajectory for

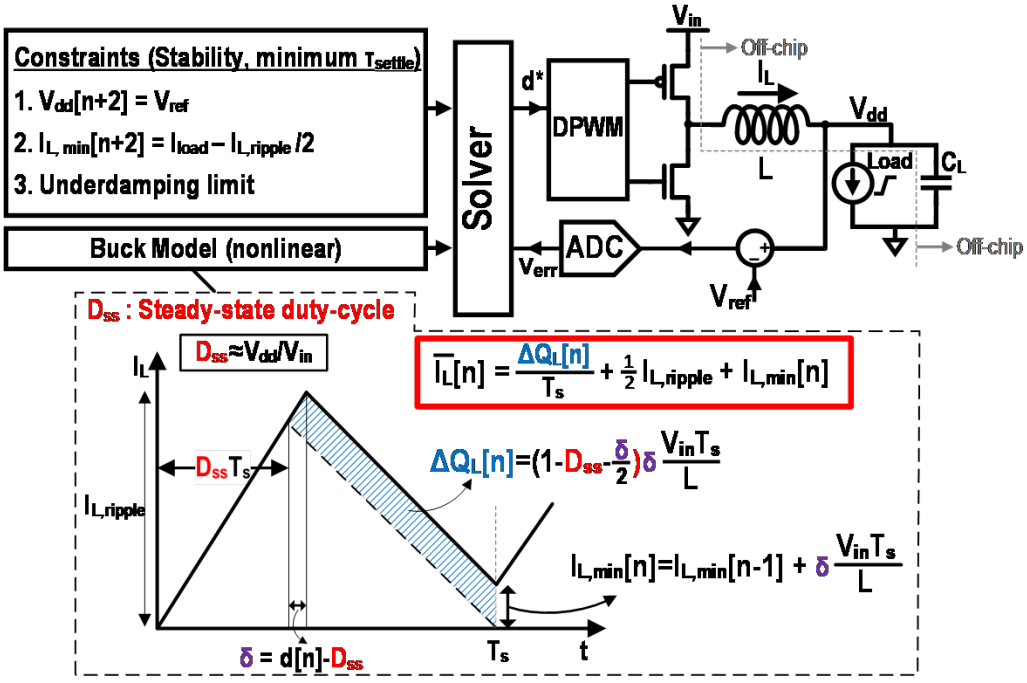


Figure 5.1: MPC Architecture and non-linea buck model used by Solver.

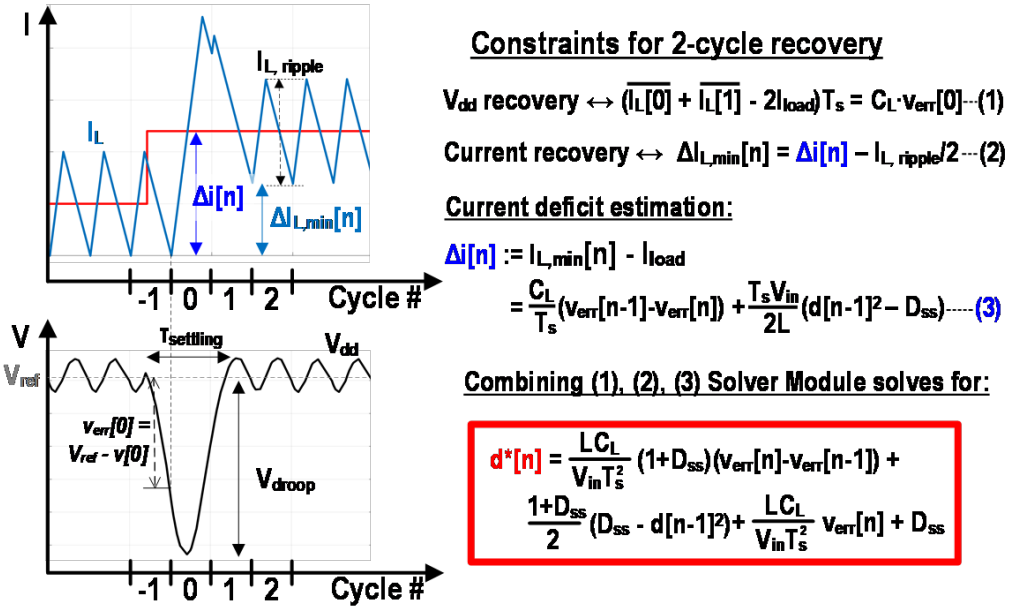


Figure 5.2: MPC Constraint Equations of $d^*[n]$ for 2-cycle V_{dd} restoration.

the next k-cycles ($d^*[n+i]$, i=0,1...k) which minimizes any desired objective function subject to constraints. However, to realize IVR-feasible low-latency control of an (approximately) 2nd order buck, the proposed formulation sets k=2, and avoids constrained optimization instead solving a much simpler problem of constraint satisfaction: the Solver must plan a 2-cycle trajectory ($d^*[n]$, $d^*[n+1]$) so that $V_{dd}[n+2] = V_{ref}$ and $I_L[n+2] = I_{load}$, to satisfy the minimum τ_{settle} criterion under non duty-cycle saturated conditions ($d[n] \neq 0$ or 1) to achieve V_{dd} recovery (Fig. 5.2).

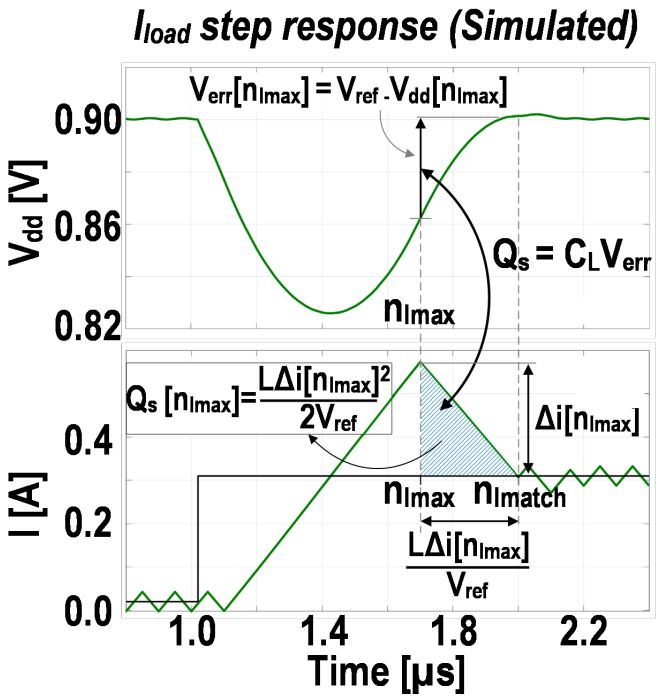


Figure 5.3: Stable MPC response under $d[n]$n saturation through Q_s tracking.

MPC also achieves optimal response under practical worst-case droop conditions involving duty-cycle saturation. For a large I_{load} step-up, such a response involves (1) maximally slewing I_L ($d^*[n]$=1) beyond I_{load} to stop and reverse V_{dd} droop recovery before (2) slewing I_L back toward I_{load} ($d^*[n]$=0) at time n_{Imax}. Eventually, (3) I_L=I_{load} at n_{Imatch}, after which I_L tracks I_{load} (Fig. 5.3). Timing n_{Imax} precisely is critical for a stable, minimal τ_{settle} response and is achieved by tracking the surplus charge $Q_s[n]$, delivered to the load

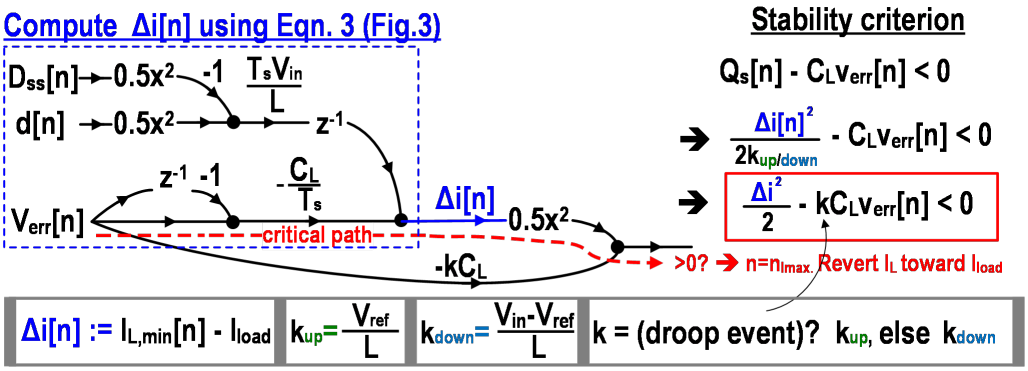


Figure 5.4: Dataflow evaluating stability criterion with saturated d_n.

if I_L were maximally slewed back to I_{load} at cycle n. Avoiding an under-damped response requires that $Q_s[n_{Imax}] - C_L(V_{ref} - V_{dd}[n]) = 0$, so that the surplus current provides just the right quantity of charge to restore V_{dd} without overshoot. Continuously evaluating the stability constraint (Dataflow graph shown in Fig. 5.4) to identify n_{Imax} allows stable, rapid MPC response even with random I_{load} disturbances. In contrast, traditional TOC techniques sense the I_{load} step magnitude and commit to a determined n_{Imax}, unable to adjust to further I_{load} changes until V_{dd} is achieved.

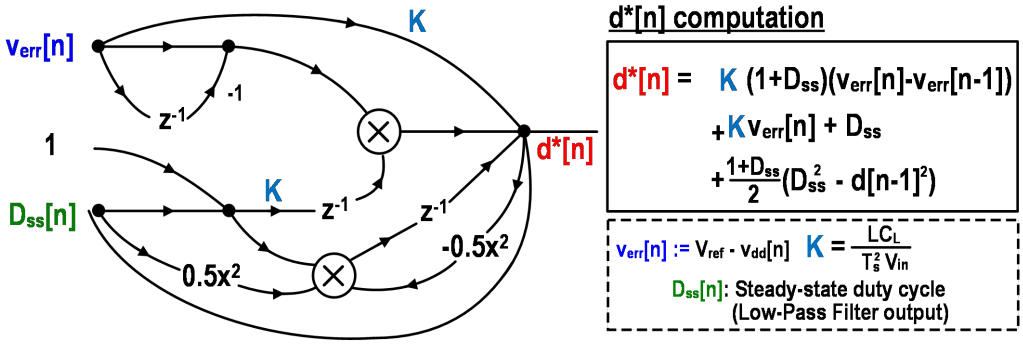


Figure 5.5: Dataflow graph for $d^*[n]$ computation.

Although the proposed MPC architecture avoids computationally intensive constrained optimization, datapath optimization remains necessary to lower computational latency and power dissipation. First, only $d^*[n]$ is computed in cycle n, because the computation of $d^*[n+1]$ is speculative under random loading conditions. The dataflow graph of the $d^*[n]$

computation (Fig. 5.5) is optimized by migrating computations away from the critical path between $V_{dd}[n]$ and $d^*[n]$ to reduce both latency and power dissipation (Fig. 5.3).

5.3 Test Chip and Measurement Results

Fig. 5.6 details the MPC-buck architecture implemented in 65nm CMOS (Die photograph in Fig. 5.7). An off-chip 1μH inductor and 1μF capacitor were used by the converter which was clocked at 10MHz. The error between V_{dd} and V_{ref}, V_{err} is first quantized by a 5mV resolution comparator bank before being forwarded to the MPC controller. A Digital Pulse Width Modulator (DPWM) converts the controller output code ($d[8:0]$) to pulses p and n. Delayed-clocked registers enable glitch-free same-cycle feedback while providing sufficient time for the comparator banks (2ns) and the MPC controller (4.5ns) to evaluate. The *Solver* consumes 0.7mW at 10MHz.

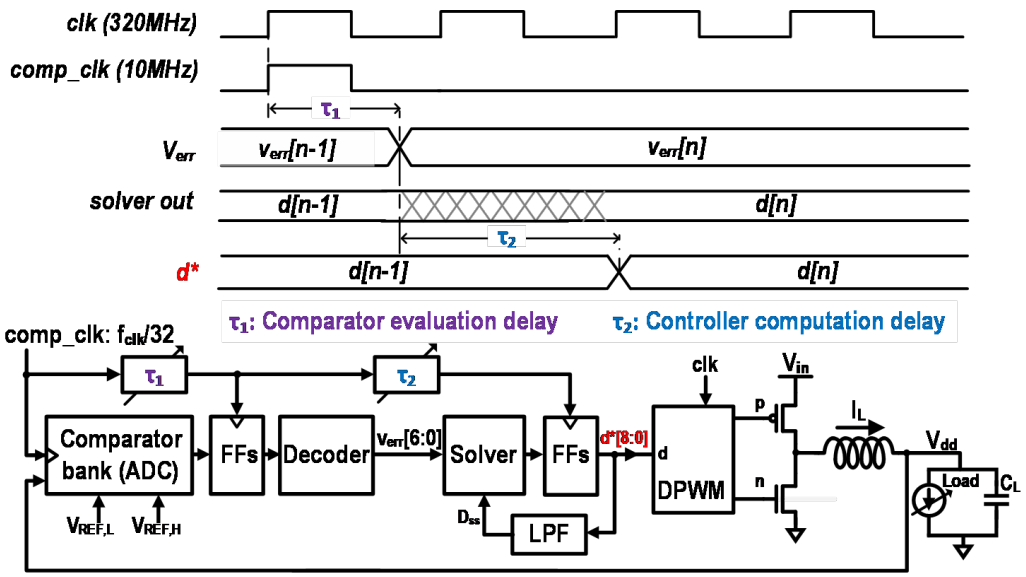


Figure 5.6: MPC Test chip architecture. Delayed clocks provide time for Comparator and Solver evaluation with glitch-free operation.

Fig. 5.8 shows measured MPC-buck step-up and -down responses to a 305mA/0.12ns I_{load} step at V_{dd}=1V and a 191mA/ 0.12ns step at 0.7V. The test chip includes a tunable PID controller for a baseline comparison. V_{droop} and τ_{settle} parameters corresponding to an optimally tuned PID controller are also annotated. Despite non-idealities resulting from

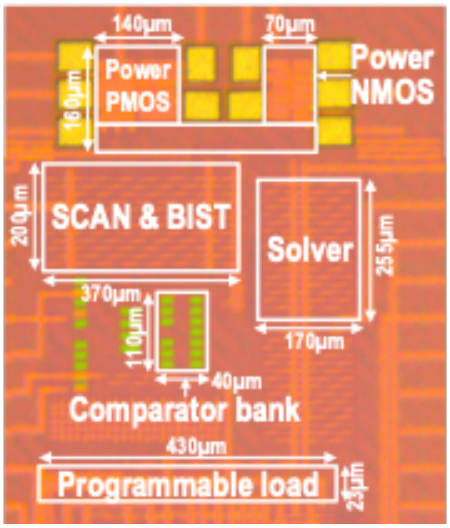


Figure 5.7: Die Photograph.

non-zero d_{min} and non-uniform quantization, the system achieves near time-optimal transient response. Compared to the baseline, MPC achieves marginal Vdroop improvement ($d[n]$= 0 or 1 during droop/surge) in both designs but a 2.49X τ_{settle} reduction. MPC can be applied to other VR modalities (LDOs, SC-regulators) and further leverage area and speed benefits offered by advanced CMOS nodes. These improvements, obtained through enhanced control, are compatible and expected to be additive with other circuit and package technology advances [50].

MPC stability, whose analysis is well documented in the literature [51], is also experimentally validated in this work. The synthetic load was configured with a pseudo-random I_{load} sequence in the 0-200mA range and the system transient response was measured (Fig. 5.9). Over 50ms (corresponding to over 500,000 I_{load} transitions), stable V_{dd} regulation was observed from the MPC-buck. Also shown is the stable transient V_{dd} response to an I_{load} pulse, where the step-down transition occurs during transient recovery of the preceding I_{load} step-up (Fig. 5.10). Fig.5.11 presents a comparison with related work.

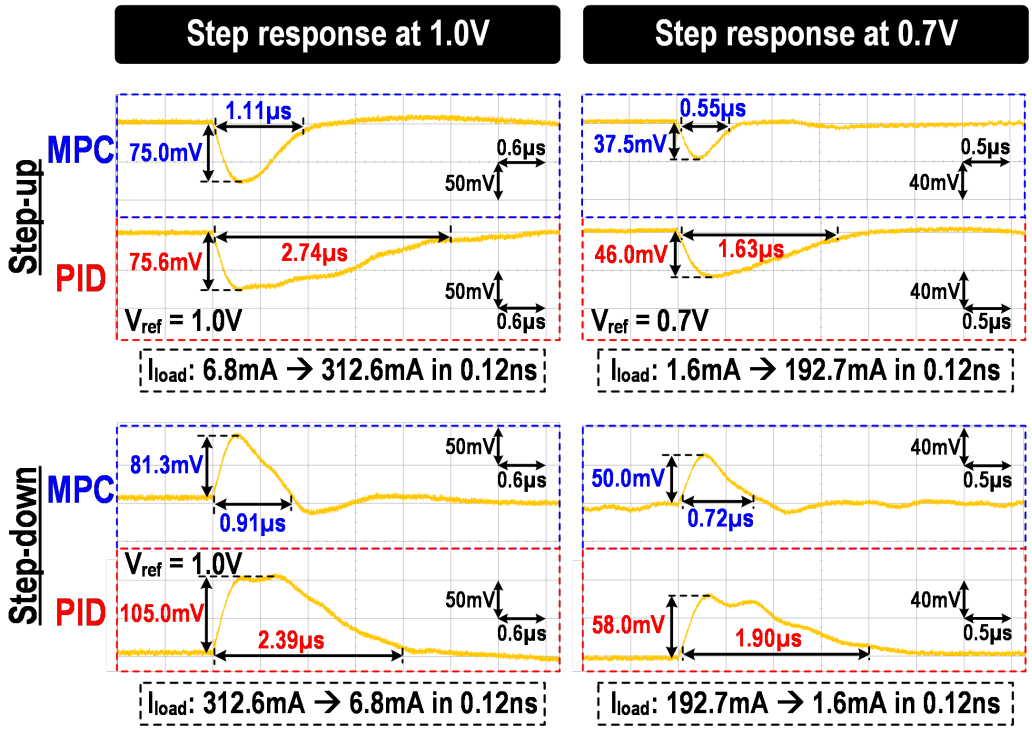


Figure 5.8: Measured V_{dd} response to a load-step for MPC and optimally tuned PID controllers. Both incur $d[n]$ saturation resulting in similar V_{droop}.

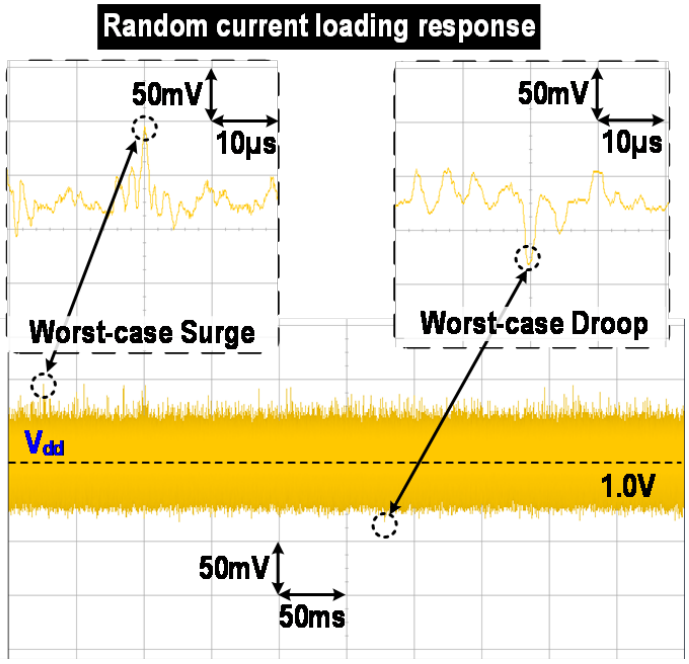


Figure 5.9: Measured MPC V_{dd} waveforms under random loading conditions.

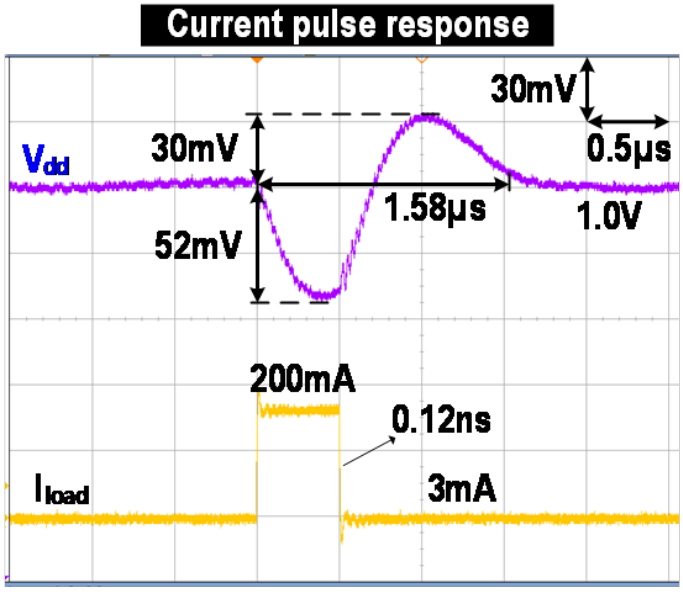


Figure 5.10: Measured MPC V_{dd} responses to I_{load} pulse with step-down occurring during τ_{settle}.

Paper	[4]	[5]	[9]	[10]	**This work**
Process	180nm	350nm	65nm	180nm	**65nm**
Type	Analog	Analog	Analog	Digital	**Digital**
Controller type	Time-optimal	Hysteresis (non-linear)	Time-domain current-mode	Time-based voltage-mode	**Model-predictive**
V_{in} [V]	3.3	2.7–4.2	1.8	1.8	**1.8**
V_{out} [V]	0.6–1.2	1.2–1.8	0.15 – 1.69	0.6–1.5	**0.6 – 1.0**
Max I_{LOAD} [mA]	1250	600	600	600	**400**
L [µH]	1	2.2	0.22	0.22	**1**
C_L [µF]	4.7	4.7	4.7	4.7	**1**
F_{sw} [MHz]	1.5	2.5–3.1	N.A.	11–25	**10**
Peak efficiency [%]	90.2	92	94.9	94	**91.9**
ΔV_{out} [mV] @ ΔI_{LOAD} / T_{EDGE}	10 @ 840mA/3ns	47 @ 500mA/N.A.	100 @ 480mA/0.1us	60 @ 500mA/N.A.	**37.5 @ 191mA/0.12ns**
Settling time [ns]	800	4700	3500	3000	**550**
Measured random I_{load} response	No	No	No	No	**Yes**

Figure 5.11: Comparison with related work.

Chapter 6

AUTO-TUNING OF DLDO DIGITAL CONTROLLERS

6.1 Introduction

As has been explained in Chapter 4, DLDOs face the challenge of high loop gain sensitivity to process, regulator input voltage (V_{in}), V_{dd} and temperature ($PV_{in}V_{dd}T$) which forces stability margins and degrades response times. Adaptive techniques have also been proposed in the literature to compensate for current-loading effects on loop dynamics [41], and to reduce V_{dd} ripple magnitude [52]. However, techniques capable of adapting to the broad $PV_{in}V_{dd}T$ range experienced by SoCs today remain critically important.

This chapter presents the implementation of autonomous gain-tuning (AGT), a technique that relies on computing the auto-correlation of the sampled V_{dd} with single-bit precision in order to classify regulator response and enable loop gain auto-tuning across $PV_{in}V_{dd}T$ variations on the computational *Solver*. Measurement results from the test chip in Chapter 4 indicate that the AGT mechanism performs broad loop-gain adjustments to track a wide range of $PV_{in}V_{dd}T$ configurations: even individual variation in V_{in} (0.9V–1.1V), V_{dd} (0.8V–1.0V) and temperature (-15°C–105°C) requires loop-gain modulation over a range of approximately 2×. This chapter also outlines the limitations of the proposed AGT technique and presents a more general method for adjusting loop gain parameters of DLDOs automatically.

6.2 AGT *Architecture*

The *Solver* in the Computational LDO discussed in Chapter 4 relies on accurate knowledge of G_S (Equation 4.5), which must be modeled with sufficient accuracy. From the (approximate) perspective of a discrete-time linear system, situating the poles at origin at design-time leaves them susceptible to drifting into the left half plane in the z-domain under $PV_{in}V_{dd}T$ variation during run time, leading to regulator instability. Indeed, G_S is

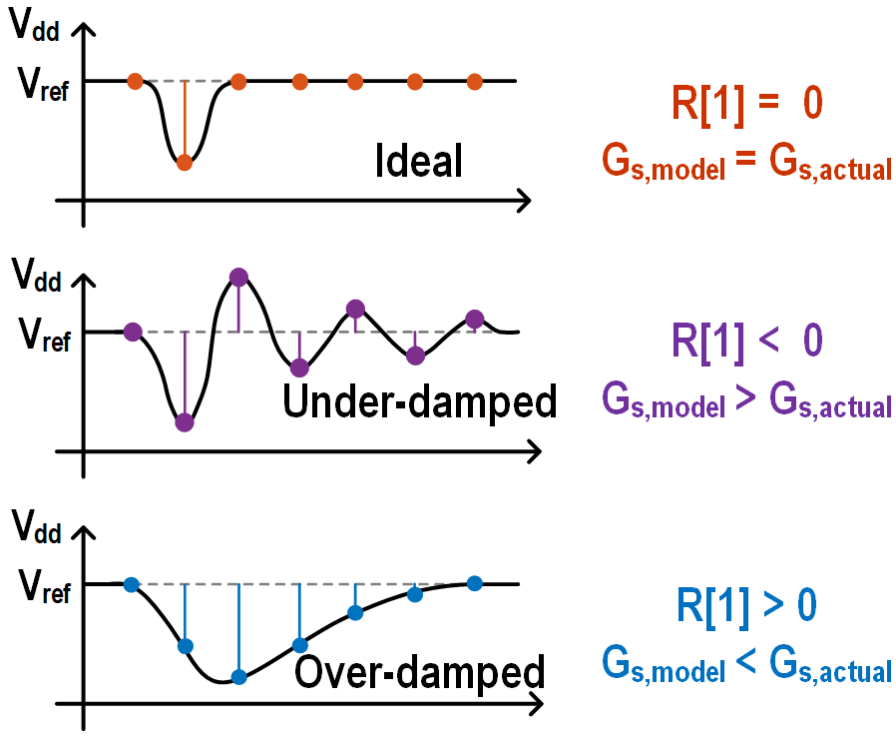


Figure 6.1: Motivation for using the sign of $R_{vo}[1]$ to guide gain-tuning. Varied polarities of $R_{vo}[1]$ are indicative of varied conditions of regulator transient response.

susceptible to variation in C_L (especially in the case of off-chip capacitors) and in I_{lsb} which is highly sensitive to $PV_{in}V_{dd}T$. Thus, the need to adapt loop-gain to variation, already important for DLDOs, is critical to achieve rapid, $PV_{in}V_{dd}T$-robust transient response in the CLDO. This need for adaptation motivates the idea of autonomous gain tuning (AGT), an all-digital technique for tuning the internal loop-gain parameter of the controller ($G_{S,model}$). AGT relies on the lag-1 auto-correlation of discrete-time V_{dd} samples ($R_{vo}[1]$) defined as:

$$\begin{aligned} R_{vo}[1] &:= E(v_o[n] - \overline{v_o[n]})(v_o[n-1] - \overline{v_o[n-1]}) \\ &= E(\Delta v_o[n])(\Delta v_o[n-1]). \end{aligned} \tag{6.1}$$

Simplification of the expression is enabled by observing that $\overline{v_o}$ is regulated to V_{ref} by the CLDO. In this work, we propose a heuristic: using $R_{vo}[1]$ to detect and correct $PV_{in}V_{dd}T$-induced loop gain variation. We proceed by first describing the rationale for the AGT

approach by examining the case of a CLDO step-response. We subsequently generalize the approach to the case of random current loading, and finally demonstrate that using only the sign-bit of $R_{vo}[1]$ is sufficient for AGT computation, vastly simplifying implementation.

Figure 6.1 illustrates the I_{load} step response of the CLDO under ideal, under-damped and over-damped conditions. We observe that for a design targeted for a single-cycle response time, $R_{vo}[1] = 0$ under ideal conditions. Similarly, $R_{vo}[1] < 0$ in the case of an under-damped response and $R_{vo}[1] > 0$ for an over-damped response. This observation points to the usefulness of $sign(R_{vo}[1])$, the polarity of $R_{vo}[1]$, in classifying whether the LDO is tuned for an ideal response under current $PV_{in}V_{dd}T$ conditions. We note that more realistic load

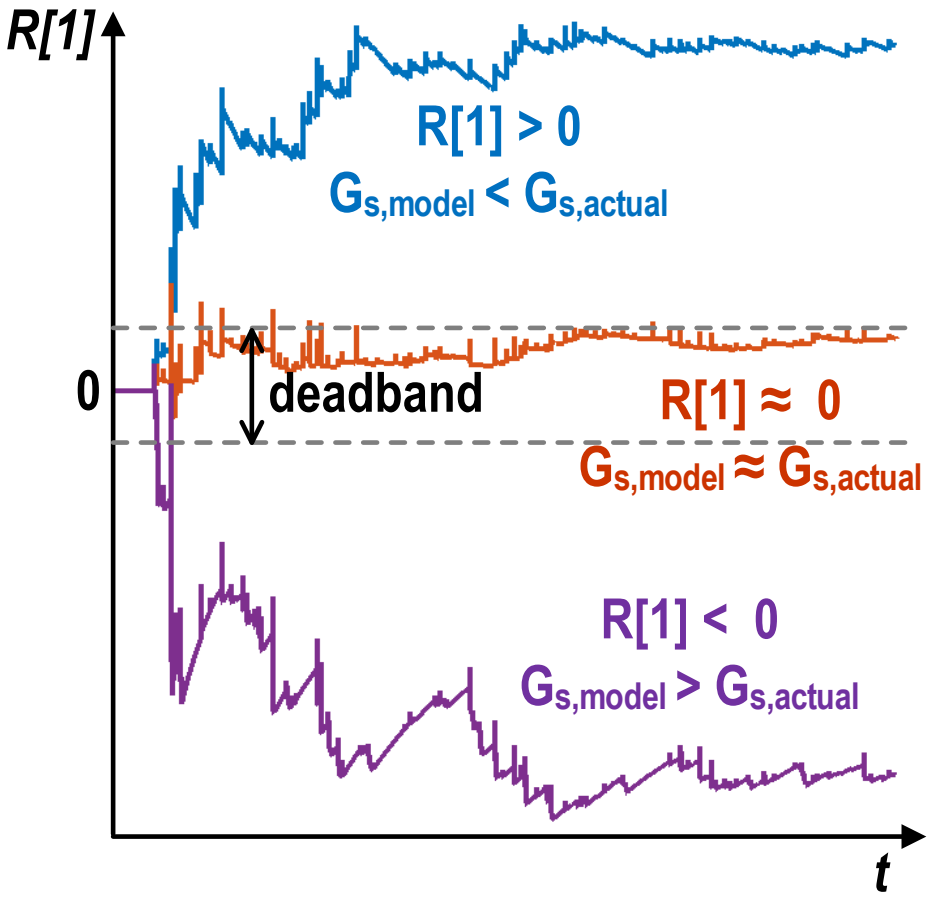


Figure 6.2: MATLAB simulation waveforms of calculated $R_{vo}[1]$ under ideal conditions, and under over-damped and under-damped conditions. $R_{vo}[1]$ is observed to be close to 0 under ideal tuning conditions.

currents can be viewed as a combination of time-shifted and scaled unit step load current events. Scaling and time-shifting does not affect the above observations about $R_{vo}[1]$ and consequently, simulations confirm that this heuristic extends well to the general case of current loading. This assertion is validated through silicon measurements in Section 6.4 .

Figure 6.2 shows waveforms of $R_{vo}[1]$ obtained in the presence of random current loading in MATLAB simulations under three different conditions: (1) ideal-tuned response where $G_{s,model}$–the loop-gain modeled by the controller–equals $G_{s,actual}$, the real $PV_{in}V_{dd}T$ dependent loop-gain; (2) over-damped response ($G_{s,model} > G_{s,actual}$); and (3) under-damped response ($G_{s,model} < G_{s,actual}$). The figure shows that the $R_{vo}[1]$ calculation converting to stable values whose polarity is consistent with our expectation, motivating its use to classify regulator response.

6.3 AGT *Implementation*

Section 6.2 presented the idea of using the statistics of the quantized v_o signal to determine nature of the regulator response, and autonomously tune the loop gain across $PV_{in}V_{dd}T$ conditions.

Precision requirements for the evaluation of $R_{vo}[1]$ can be significantly relaxed for AGT. The key simplifying observation is that the lower tracking bandwidth requirements of the AGT loop allow $R_{vo}[1]$ computations to occur over a long duration. Recall that evaluating $R_{vo}[1]$ involves accumulating the product of successive Δv_o samples over time. Equation 6.1 can be re-written in terms of a quantized sample of Δv_o ($\Delta v_{o,q}$), and its corresponding quantization error σ_{vo}:

$$R_{vo}[1] = E(\Delta v_{o,q}[n] + \sigma_{vo}[n])(\Delta v_{o,q}[n-1] + \sigma_{vo}[n-1]) \tag{6.2}$$

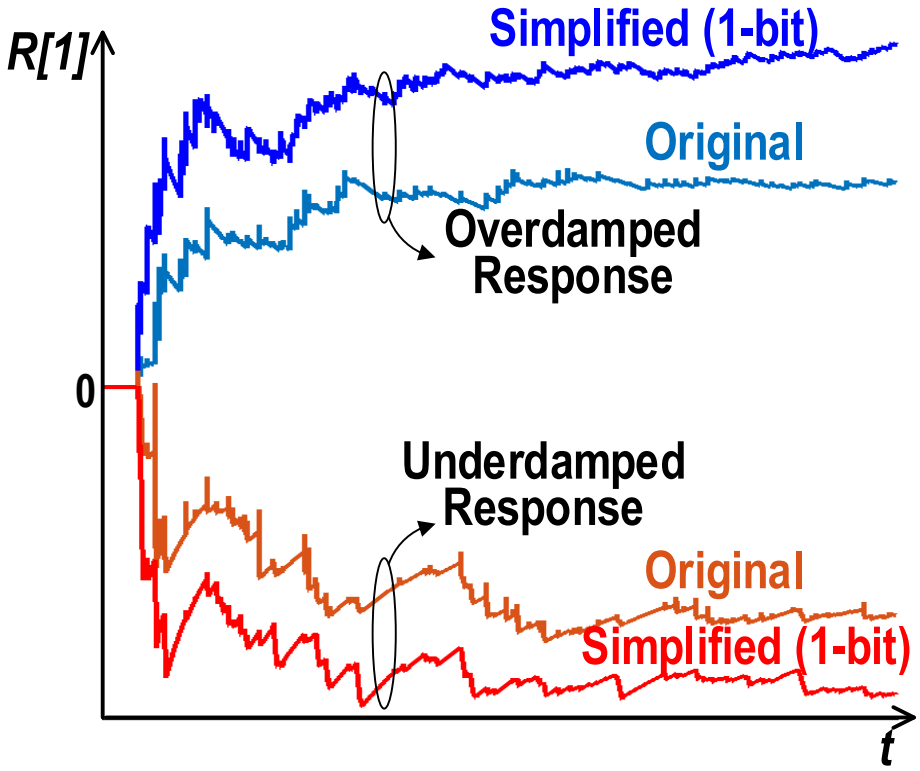


Figure 6.3: Matlab simulation waveforms corresponding to full-precision and 1-bit precision tracking of $R_{vo}[1]$. The offset term corresponding to Equation 6.3 is verified to be sufficiently small to justify 1-bit characterization of sign($R_{vo}[1]$).

Further simplification yields:

$$\begin{aligned} R_{vo}[1] &= E(\Delta v_{o,q}[n]\Delta v_{o,q}[n-1]) + E(\sigma_{vo}[n] \cdot \Delta v_{o,q}[n-1]) \\ &\quad + E(\sigma_{vo}[n-1] \cdot \Delta v_{o,q}[n])) + E(\sigma_{vo}[n]\sigma_{vo}[n-1]) \\ &= E(\Delta v_{o,q}[n]\Delta v_{o,q}[n-1]) \\ &\quad + 2cov(\sigma_{vo}, \Delta v_{o,q}) + E(\sigma_{vo}[n]\sigma_{vo}[n-1]) \end{aligned} \tag{6.3}$$

where $cov(\sigma_{vo}, \Delta v_{o,q})$ is the co-variance between the quantization noise and the quantized signal. Thus, using a quantized version of Δv_o to compute $R_{vo}[1]$ introduces an offset error. This offset error is low under typical loading conditions because quantization noise in cycle n is significantly uncorrelated with the sampled value in cycle n and with the quantization

error in the previous cycle. In such cases, the use of only the sign bit to estimate $R_{vo}[1]$ may be sufficient to determine its polarity. Notably, this offset vanishes in the case of perfect de-correlation.

Figure 6.3 shows the results of MATLAB simulations that evaluate the use of the sign bit of Δv_o for $R_{vo}[1]$ calculation. Simulations indicate relatively small error and good fidelity with the full-precision calculation. In this work, we therefore used only the sign-bit of v_o to evaluate $R_{vo}[1]$, relying on long-term averaging to suppress quantization errors.

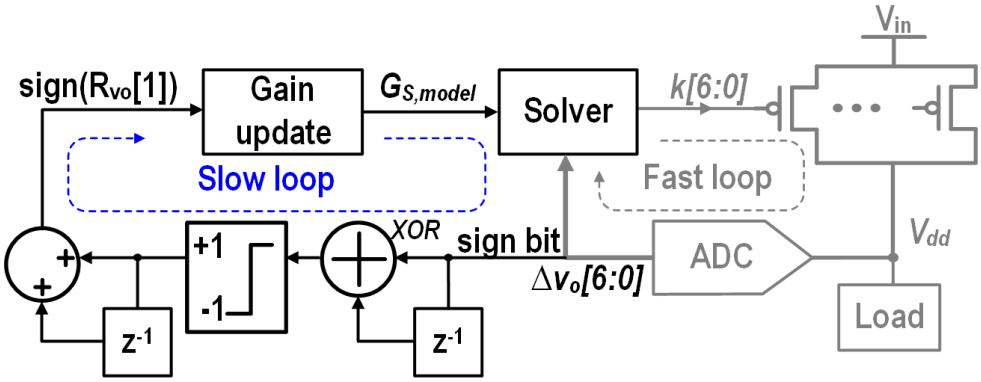
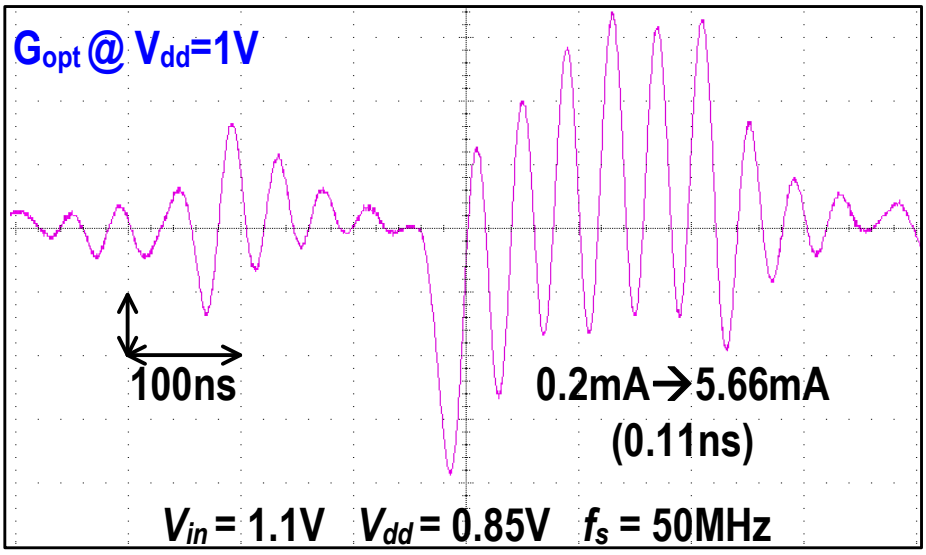
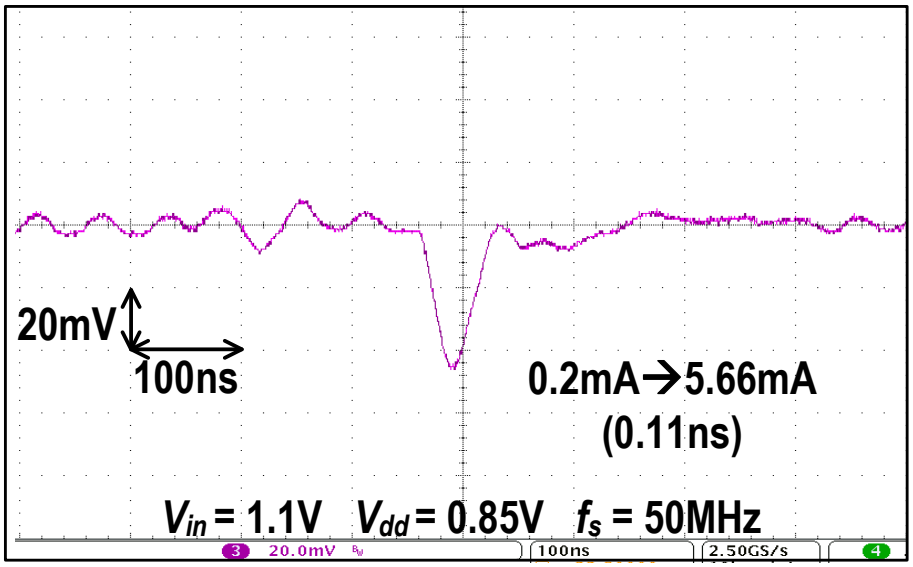


Figure 6.4: Block diagram of the AGT implementation relying on the sign-bit of the quantized v_o signal to auto-tune $G_{S,model}$.

Figure 6.4 shows the block diagram of the AGT implementation in our design. Note that this work focuses on tuning $G_{S,model}$ alone: an investigation into low-complexity techniques for jointly tuning $G_{S,model}$ and α is outside the scope of this effort. The product of successive sign-bits of Δv_o is performed using an XOR gate, and accumulated over a user-defined number of cycles. The number of accumulation cycles is user-programmable, although 40 cycles were found to provide sufficient accuracy. Because only the polarity of $R_{vo}[1]$ is of interest, the sign of the accumulated result can be directly used, avoiding the scaling required to compute $R_{vo}[1]$ itself . In this manner, efficient, low-precision statistical analysis of v_o enables auto-tuning $G_{s,model}$ in order to maintain rapid, stable response across the range of $PV_{in}V_{dd}T$ conditions.



(a)



(b)

Figure 6.5: CLDO voltage response to I_{load} step (a) with AGT disabled and (b) with AGT enabled.

6.4 AGT *Measurement Results*

Silicon measurements demonstrate the effectiveness of AGT relying on single-bit precision computations. Figure 6.5 demonstrates the effect that AGT has on stable operation when

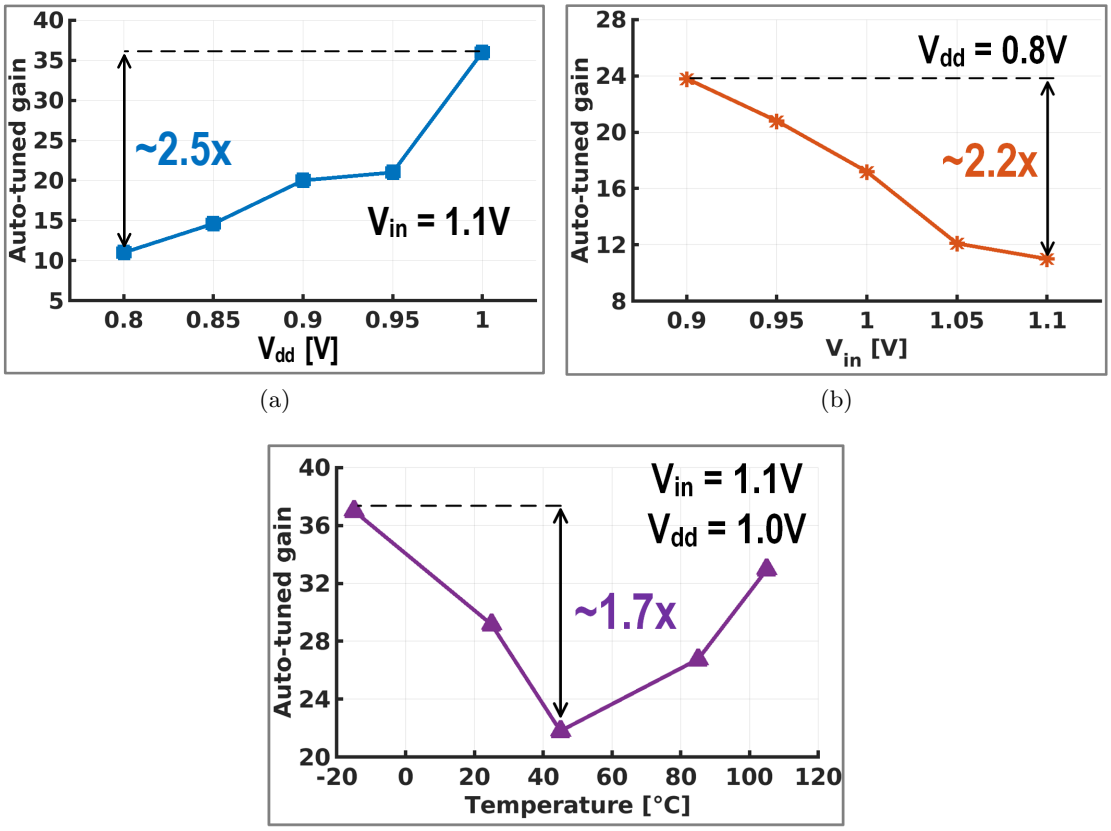


Figure 6.6: Recorded values of $G_{S,model}$ with AGT enabled across target operating range of V_{dd}, V_{in} and temperature.

the CLDO operates before and after a V_{dd} scaling event. In this experiment, the CLDO was first operated with V_{in}, V_{dd} at 1.1V and 1.0V respectively. At this time, AGT was enabled, allowing the system to auto-tune the loop-gain based on the prevalent $PV_{in}V_{dd}T$ conditions. For the first experiment, AGT was next disabled, and V_{dd} was scaled to 0.85V without any loop-gain update. The increased $V_{dropout}$ significantly increases I_{lsb}, decreasing $G_{S,actual}$ and causing instability (Figure 6.5a). Margining for this instability would require the loop-gain to be conservatively set at the maximum possible $V_{dropout}$ setting, severely degrading performance for minimum $V_{dropout}$ configurations, when supply noise is worst. To measure the effect of AGT, the experiment was repeated, but AGT was enabled throughout the voltage scaling process. The resulting measured CLDO voltage response is shown in

Figure 6.5b. Thus, AGT effectively tunes regulator loop gain for effective operation at $V_{dd} = 0.85$, maintaining rapid, stable response.

AGT was implemented to enable tuning across a broad range of $PV_{in}V_{dd}T$ variation. To demonstrate auto-tuning capability over a wide V_{in}, V_{dd} and temperature range, we performed three experiments, sweeping each of these variables while maintaining the remaining variables constant across a targeted range. In each experiment, stable response in the CLDO was validated at every point, and the $G_{S,model}$ determined by the AGT was captured. Figure 6.6 summarizes the results and the recorded $G_{S,model}$ values demonstrate that regulator loop-gain varies by approximately $2\times$ due to variation in each of these operating parameters.

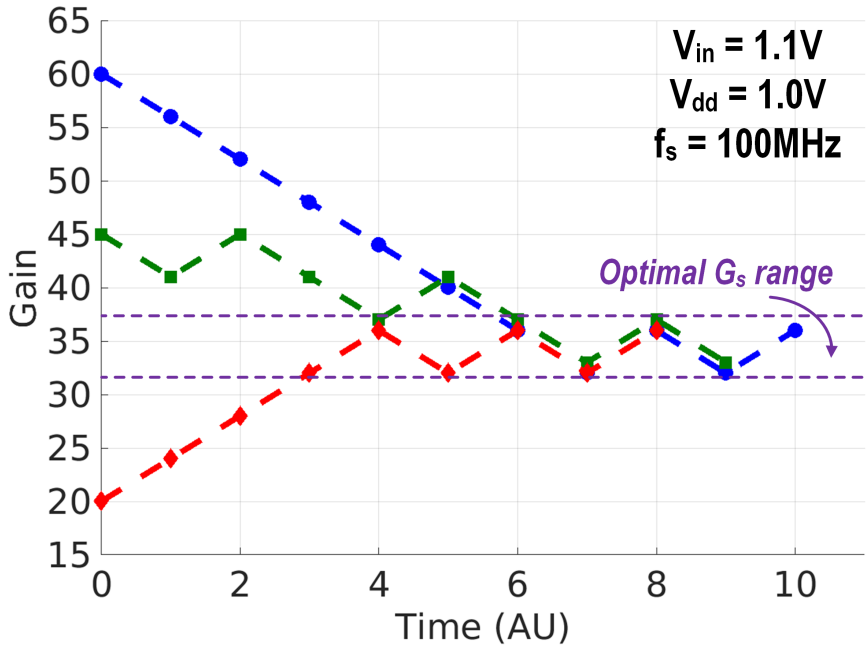


Figure 6.7: Run-time evolution of $G_{S,model}$ by AGT starting from three distinct and incorrect initial values. $G_{S,model}$ is found to converge in all three cases.

To test the capability of the AGT to approach the optimal loop gain under random I_{load} conditions, we ran three experiments using distinct initial values for $G_{S,model}$. Random load current was provided by executing the linear algebra co-processor to perform a sufficiently long sequence of matrix multiplications. $R_{vo}[1]$ statistics were gathered over 40 cycles in

each iteration and the AGT-tuned value of $G_{S,model}$ was registered. Figure 6.7 shows the trajectory of the recorded $G_{S,model}$ values, which experience constant-value corrections over successive tuning intervals of the AGT. The loop gain was found to converge well in all three experiments, regardless of the initial $G_{S,model}$ setting.

6.5 Limitations of AGT *Technique*

The proposed AGT technique, which performs low-precision statistical analysis of quantized V_{dd} measurements to classify regulator response and autonomously tune regulator loop gain for rapid, stable response, has been successfully implemented on a 65nm CMOS test chip. And silicon measurements demonstrate the significance and effectiveness of the mechanism in tuning loop-gain across $PV_{in}V_{dd}T$ variations. However, there are several key challenges associated with this technique which prevent it being applicable to a wide variety of designs.

First, the motivation of AGT is from the observation that v_o autocorrelation provides a low-complexity approach to determining transient LDO performance in Figure 6.1. Using the polarity of $R_{vo}[1]$ to indicate the relationship between $G_{s,model}$ and $G_{s,actual}$ is specific to the Computational LDO *Solver* design, targeting for single-cycle voltage recovery, and cannot be directly applied to other compensator designs. Second, in order to make the AGT technique valid, some assumptions have been made for the load current condition.

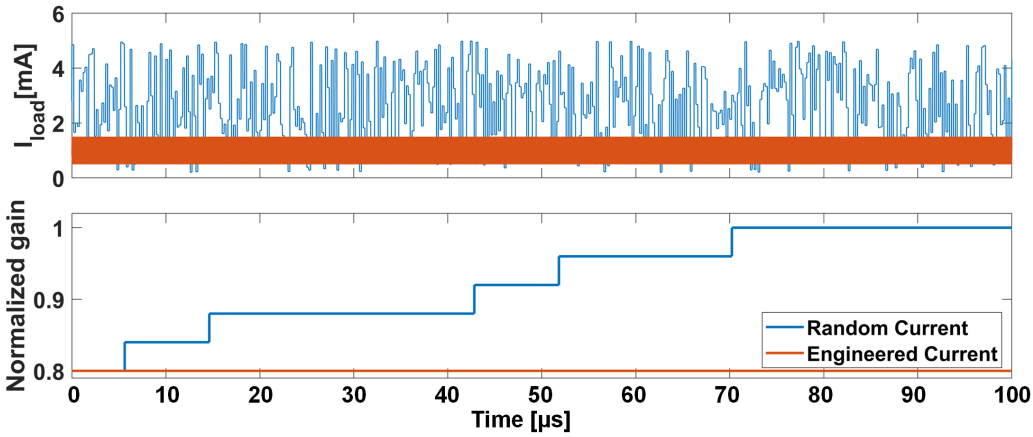


Figure 6.8: Matlab simulation of engineered current waveform to break the AGT loop and track solver gain $G_{S,model}$ incorrectly.

Figure 6.1, 6.2 show the classification of loop gain based on the sign of R[1] is a valid approach for current steps and random loading conditions. However, engineered current waveforms can hinder this approach and make the AGT loop mistrack solver gain.

An example of manipulated I_{load} pattern breaking AGT operation is shown in Figure 6.8. Blue curve shows the I_{load} waveform and run-time normalized gain ratio ($G_{s,model}/G_{s,actual}$) values under random loading condition while the orange one represents the case for a special designed I_{load} pattern. From this waveform, the AGT loop successfully tracks the actual solver gain and $G_{s,model}$ converges to $G_{s,actual}$ eventually under random current. In contrast, due to the specific I_{load} pattern, $R_{vo}[1]$ does not exhibit the expected result as for current steps or random current and AGT is unable to properly track $G_{s,actual}$. Instead, $G_{s,model}$ stays at the incorrect initial value.

6.6 PRBS-based Cross-correlation Method

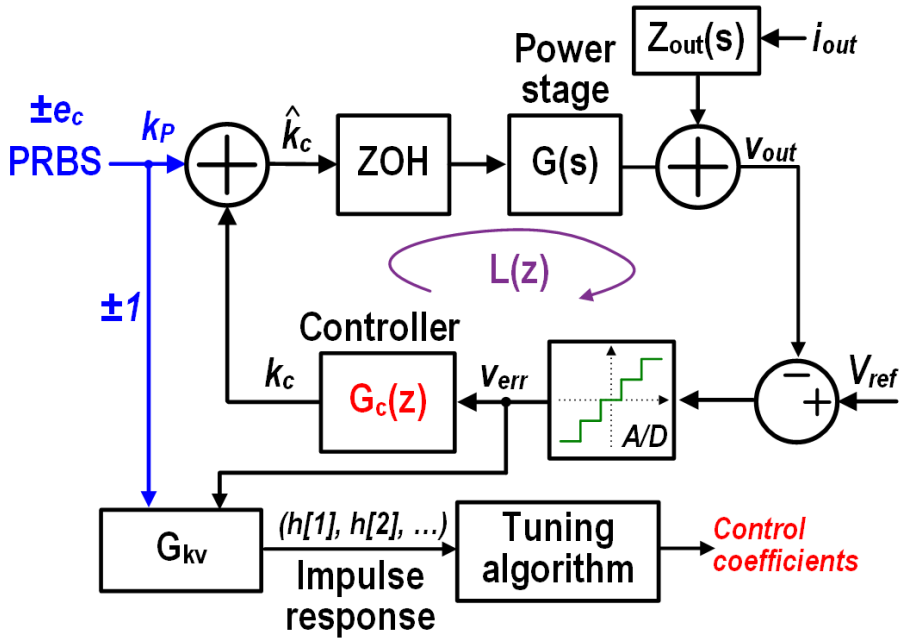


Figure 6.9: Architecture overview of PRBS-based cross-correlation technique.

To address the limitations of AGT technique, we propose a pseudo-random binary sequence (PRBS) based method which injects a PRBS signal (white-noise approximation)

at the controller output (k_C) and evaluates the loop impulse response measured via the cross-correlation of injected noise signal (k_u) and controller output after noise injection ($\hat{k_C}$).

6.6.1 Theory Background

The cross-correlation approach [53] is utilized to measure the impulse response. In Figure 6.9, considering $\hat{k_C}$ as a system output, its value is affected by two system inputs, k_u and i_{out}. Introducing the regulator (power-stage) transfer function $G(z)$, digital compensator $G_c(z)$ and the DLDO open loop output impedance $Z_{out}(z)$, the transfer functions of k_u to $\hat{k_C}$ ($H_u(z)$) and i_{out} to $\hat{k_C}$ ($H_i(z)$) can be represented as

$$H_u(z) = \frac{\hat{k_C}(z)}{k_u(z)} = \frac{1}{1 + G_c(z)G(z)} \tag{6.4}$$

$$H_u(z) = \frac{\hat{k_C}(z)}{i_{out}(z)} = \frac{Z_{out}(z)G_c(z)}{1 + G_c(z)G(z)} \tag{6.5}$$

For small-signal disturbances, the regulator can be approximated as a linear time-invariant discrete-time system in steady state so the time-domain representation of $\hat{k_C}$(k) becomes the sum of the convolution of $k_u$with the associated impulse response $h_u[k]$ of $H_u(z)$ and the convolution of i_{out} with impulse response $h_i[k]$ of $H_i(z)$.

$$\begin{aligned} \hat{k_C}[n] &= (h_u * k_u)[n] + (h_i * i_{out})[n] \\ &= \sum_{k=-\infty}^{\infty} h_u[k] \cdot k_u[n-k] + \sum_{k=-\infty}^{\infty} h_i[k] \cdot i_{out}[n-k] \end{aligned} \tag{6.6}$$

The lag-m cross-correlation of the input noise signal $k_u[k]$ and the output signal $\hat{k_c}[k]$ is

$$\begin{aligned} R_{k_u\hat{k_c}}[m] &= \sum_{n=-\infty}^{\infty} k_u[n-m] \\ &= \sum_{n=-\infty}^{\infty} h_u[n] \cdot R_{uu}[m-n] + \sum_{n=-\infty}^{\infty} h_i[n] \cdot R_{ui_{out}}[m-n] \end{aligned} \tag{6.7}$$

Now if the input signal $k_u[k]$ is selected to be white noise, then the following equation holds:

$$R_{uu}[m] = \delta[m] \qquad R_{ui_{out}}[m] = 0 \tag{6.8}$$

Under the condition in Equation 6.8, 6.7 is reduced to

$$R_{k_u\hat{k_c}}[m] = h_u[m] \tag{6.9}$$

Therefore, the loop impulse response can be measured by the cross-correlation between k_u and $\hat{k_C}$.

6.6.2 Implementation Details

For DLDO designs, PI controller is one of the most commonly used compensator types. If we assume the controller $G_c(z)$ is denoted as $\frac{cz+d}{z-1}$, then the open-loop transfer function of the system becomes $L(z) = \frac{K(cz+d)}{(z-1)(z-\alpha)}$, where K is the power stage DC gain. From the transfer function of $H_u(z) = \frac{1}{1+L(z)}$, we can derive the impulse response $h[0] = 1, h[1] = K \cdot c$. As $h[1]$ changes monotonically with respect to controller gain c (given a fixed K value), $h[1]_{measured}$ can be used to compare against the target value $h[1]_{targ}$ and based on the comparison result, the auto-tuning mechanism adjusts controller gain accordingly to achieve expected impulse response. The block-level implementation of the cross-correlation method for auto-tuning is shown in Figure 6.10

8-b PRBS signal is generated through a feedback shift register at the rate of the DLDO switching frequency. After every N (N=1, 2, ...) PRBS period, the accumulating cross-

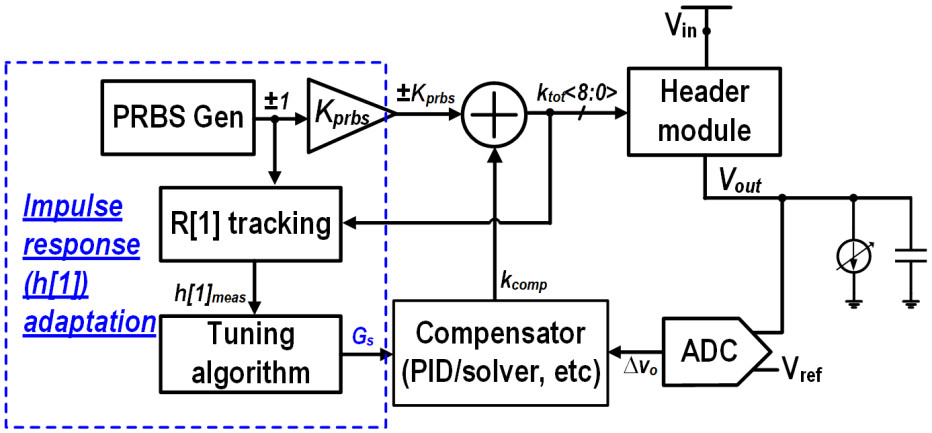


Figure 6.10: Block-level description of the cross-correlation based auto-tuning method.

correlation result $h[1]_{measured}$ is captured and compared against target impulse response/cross-correlation $h[1]_{targ}$. If the difference between $h[1]_{measured}$ and $h[1]_{targ}$ exceeds pre-defined threshold value, it indicates that the feedback loop are not operating with proper gain value and the auto-tuning loop will update controller gain accordingly. Figure 6.11 shows the co-simulation results of the proposed cross-correlation method with analog circuit simulated by *Hspice* and digital logic in verilog description by *VCS*. The simulation was repeated using three distinct initial controller gain value (G_s). All three curves converge to a common level representing optimal G_s.

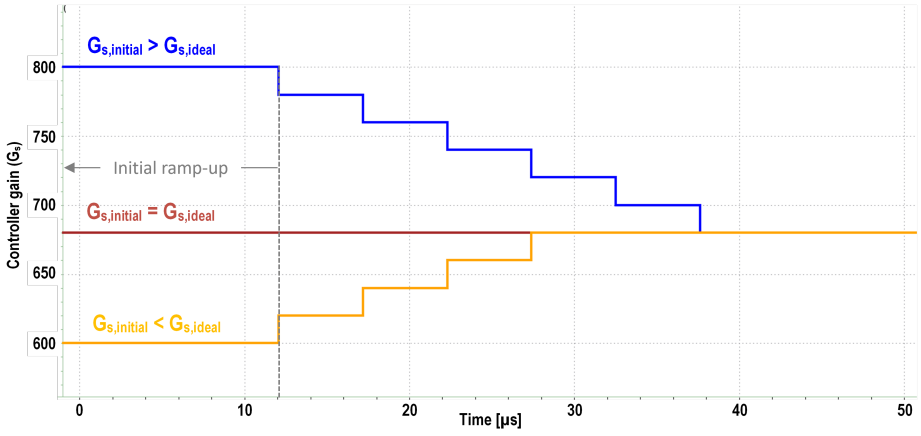


Figure 6.11: Simulation result of cross-correlation based impulse response adaption.

Chapter 7

CONCLUSION AND FUTURE WORK

The dissertation presents several techniques to minimize voltage guardbands and improve digital SoC system efficiency. In Chapter 2 we presented a unified clock and power architecture that combines the separate V_{dd} and clock regulation loops in conventional systems into a single loop. With the only control knob being V_{dd}, it avoids any possible timing violations in adaptive clocking systems using conventional PLLs. *UniCaP* architecture achieves aggressive V_{dd} margin reduction while providing performance guarantee at the same time by tracking the *REFCLK*. The buck-converter based prototype in 65nm CMOS demonstrated 10-13.5% V_{dd} reduction by reducing supply noise and temperature variation induced voltage margins. Over a wide range of voltages from 0.6-1.0V, a V_{dd} margin recovery of 95% was achieved with the regulator bandwidth well below 1MHz.

Chapter 3 identified outstanding issues within existing adaptive clocking architectures and presented a new, improved *UniCaP* architecture. The phase-locked adaptive clocking architecture relies on dual path feedback for rapid V_{dd}/f_{clk} restoration and phase recovery. A telescopic TRO design was also introduced to improve the matching between TRO and the critical path, further improving V_{dd} margin reduction. The 65nm CMOS test chip demonstrated 96.2% average V_{dd} margin recovery across 0.6 to 1.0V and achieved 38X $T_{recovery}$ improvement over prior work [7]. Measurements that quantify the impact of clock distribution on V_{dd} margin reduction on adaptive clocking systems were also reported in this chapter.

Chapter 4 presented computational regulation, a technique for fast and stable transient response. The new approach which exploits advances in computing performance and efficiency was first applied to realize DLDOs with stable and fast transient responses. The key insight in computational regulation is to develop a time-domain regulator model with enhanced accuracy over a simplified linear discrete-time model and keep evaluating the system

of equations in run time for stable and rapid response. With the advances in computing speed and efficiency in advanced technology nodes, this relatively more computing-intensive regulation technique has become possible.

With increased adoption of adaptive clocking, τ_{settle} minimization becomes an additional objective in voltage regulation besides peak V_{dd} droop minimization. Inspired by the success of computational regulation applied on a DLDO, Chapter 5 described the application of Model Predictive Control on an integrated buck converter. Continuously evaluating the *Solver* module every cycle to plan for an optimal two-cycle recovery trajectory, MPC-buck is capable of realizing 2.49X settling-time improvement over optimally tuned Proportional Integrated Differential (PID) control and achieving stable and near-optimal transient response under random disturbances.

Chapter 6 discussed the issue of worst-case stability margining for loop gain variations across $PV_{in}V_{dd}T$ conditions, which impacts VR response and overall SoC system efficiency. AGT, a technique that performs low-precision statistical analysis of quantized V_{dd} measurements to classify regulator response and autonomously tune regulator loop gain for rapid, stable response, was introduced in this chapter. Silicon measurements demonstrate the significance and effectiveness of the implemented AGT mechanism in tuning loop-gain across $PV_{in}V_{dd}T$ variation. Additionally, limitations of the AGT technique were discussed, as well as a cross-correlation-based method for automatically adjusting controller gains to match target impulse response.

All the techniques presented in this dissertation are mostly limited to power management blocks in a digital system and are constrained to reactive approaches in response to load disturbances. The on-going research in the group is exploring potential opportunities to incorporate information from hardware performance counters in microprocessors and possibly predict supply droop based on this information and take countermeasures to prevent droop events from happening. This technique is expected to be more computationally intensive compared to existing approaches but be more effective due to its proactive nature.

As an increasing proportion of world-wide energy resources are consumed for computing, energy efficient design is a critical component to addressing the imminent scarcity challenges that ensue. This dissertation has explored some of the impactful areas of inquiry into

energy efficient computation. It is hoped that the work will inspire further inquiry into the potential for effectively leveraging computation and co-design to address important problems in energy efficient computing which have proved to be stubborn to traditional circuit-dominated techniques.

BIBLIOGRAPHY

[1] "Increasing Data Center Efficiency with Server Power Measurements," p. 8.

[2] Q. Wu, M. Pedram, and X. Wu, "Clock-gating and its application to low power design of sequential circuits," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 3, pp. 415–420, Mar. 2000, conference Name: IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications.

[3] Z. Hu, A. Buyuktosunoglu, V. Srinivasan, V. Zyuban, H. Jacobson, and P. Bose, "Microarchitectural techniques for power gating of execution units," in *Proceedings of the 2004 international symposium on Low power electronics and design*, ser. ISLPED '04. New York, NY, USA: Association for Computing Machinery, Aug. 2004, pp. 32–37. [Online]. Available: https://doi.org/10.1145/1013235.1013249

[4] "Multi-Voltage Design," in *Low Power Methodology Manual: For System-on-Chip Design*, M. Keating, D. Flynn, R. Aitken, A. Gibbons, and K. Shi, Eds. Boston, MA: Springer US, 2007, pp. 21–31. [Online]. Available: https://doi.org/10.1007/978-0-387-71819-4_3

[5] H. Kaul, M. Anders, S. Hsu, A. Agarwal, R. Krishnamurthy, and S. Borkar, "Near-threshold voltage (NTV) design: opportunities and challenges," p. 6.

[6] X. Sun, S. Kim *et al.*, "A combined all-digital pll-buck slack regulation system with autonomous ccm/dcm transition control and 82% average voltage-margin reduction in a 0.6-to-1.0v cortex-m0 processor," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb 2018, pp. 302–304.

[7] X. Sun, F. u. Rahman, V. R. Pamula, S. Kim, X. Li, N. John, and V. S. Sathe, "An All-Digital Fused PLL-Buck Architecture for 82% Average Vdd-Margin Reduction in a 0.6-to-1.0-V Cortex-M0 Processor," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 11, Nov. 2019.

[8] X. Sun, A. Boora, R. Pamula, C.-H. Huang, D. Peña-Colaiocco, and V. S. Sathe, "Unicap-2: Phase-locked adaptive clocking with rapid clock cycle recovery in 65nm cmos," in *2020 IEEE Symposium on VLSI Circuits*, 2020, pp. 1–2.

[9] X. Sun, A. Boora, W. Zhang, V. R. Pamula, and V. Sathe, "A 0.6-to-1.1v Computationally Regulated Digital LDO with 2.79-Cycle Mean Settling Time and Autonomous Runtime Gain Tracking in 65nm CMOS," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, Feb. 2019, pp. 230–232.

[10] X. Sun *et al.*, "Model predictive control of an integrated buck converter for digital soc domains in 65nm cmos," in *2020 IEEE Symposium on VLSI Circuits*, June 2020.

[11] E. A. Burton, G. Schrom *et al.*, "FIVR — Fully integrated voltage regulators on 4th generation Intel® Core™ SoCs," in *2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014*, March 2014, pp. 432–439.

[12] X. Zhang, P. Chen *et al.*, "A 0.6 V Input CCM/DCM Operating Digital Buck Converter in 40 nm CMOS," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 11, pp. 2377–2386, Nov 2014.

[13] G.-Y. Wei and M. Horowitz, "A fully digital, energy-efficient, adaptive power-supply regulator," *IEEE Journal of Solid-State Circuits*, vol. 34, no. 4, pp. 520–528, April 1999.

[14] S. R. Sridhara, M. DiRenzo *et al.*, "Microwatt Embedded Processor Platform for Medical System-on-Chip Applications," *IEEE Journal of Solid-State Circuits*, vol. 46, no. 4, pp. 721–730, April 2011.

[15] T. Fischer, F. Anderson, B. Patella, and S. Naffziger, "A 90nm variable-frequency clock system for a power-managed itanium/sup/spl reg//-family processor," in *ISSCC. 2005 IEEE International Digest of Technical Papers. Solid-State Circuits Conference, 2005.* IEEE, 2005, pp. 294–599.

[16] N. Kurd, P. Mosalikanti *et al.*, "Next Generation Intel¯ Core™ Micro-Architecture (Nehalem) Clocking," *IEEE Journal of Solid-State Circuits*, vol. 44, no. 4, pp. 1121–1129, April 2009.

[17] K. Wilcox, R. Cole *et al.*, "Steamroller module and adaptive clocking system in 28 nm cmos," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 1, pp. 24–34, Jan 2015.

[18] K. A. Bowman, S. Raina *et al.*, "A 16 nm All-Digital Auto-Calibrating Adaptive Clock Distribution for Supply Voltage Droop Tolerance Across a Wide Operating Range," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 1, pp. 8–17, Jan 2016.

[19] D. Bol, J. D. Vos *et al.*, "SleepWalker: A 25-MHz 0.4-V Sub-mm^{2}7-μW/MHzMicrocontroller in 65-nm LP/GP CMOS for Low-Carbon Wireless Sensor Nodes," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 1, pp. 20–32, Jan 2013.

[20] S. Gangopadhyay, S. B. Nasir *et al.*, "UVFR: A Unified Voltage and Frequency Regulator with 500MHz/0.84V to 100KHz/0.27V operating range, 99.4% current efficiency and 27% supply guardband reduction," in *ESSCIRC Conference 2016: 42nd European Solid-State Circuits Conference*, Sept 2016, pp. 321–324.

[21] F. U. Rahman, S. Kim *et al.*, "An all-digital unified clock frequency and switched-capacitor voltage regulator for variation tolerance in a sub-threshold arm cortex m0 processor," in *2018 IEEE Symposium on VLSI Circuits*, June 2018, pp. 65–66.

[22] B. Kim, W. Xu, and C. H. Kim, "A supply-noise sensitivity tracking pll in 32 nm soi featuring a deep trench capacitor based loop filter," *IEEE Journal of Solid-State Circuits*, vol. 49, no. 4, pp. 1017–1026, 2014.

[23] D. Jiao and C. H. Kim, "A programmable adaptive phase-shifting pll for clock data compensation under resonant supply noise," in *2011 IEEE International Solid-State Circuits Conference*. IEEE, 2011, pp. 272–274.

[24] B. Litkouhi and H. Khalil, "Multirate and composite control of two-time-scale discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 7, pp. 645–651, July 1985.

[25] Trescases, W. T. Ng, and S. Chen, "Precision gate drive timing in a zero-voltage-switching dc-dc converter," in *2004 Proceedings of the 16th International Symposium on Power Semiconductor Devices and ICs*, May 2004, pp. 55–58.

[26] Y. K. Ramadass and A. P. Chandrakasan, "Minimum Energy Tracking Loop with Embedded DC-DC Converter Delivering Voltages down to 250mV in 65nm CMOS," in *2007 IEEE International Solid-State Circuits Conference. Digest of Technical Papers*, Feb 2007, pp. 64–587.

[27] T. Xanthopoulos, *Clocking in Modern VLSI Systems*, 1st ed. Springer Publishing Company, Incorporated, 2009.

[28] P. Chen, C. Wu, and K. Lin, "A 50 nW-to-10 mW Output Power Tri-Mode Digital Buck Converter With Self-Tracking Zero Current Detection for Photovoltaic Energy Harvesting," *IEEE Journal of Solid-State Circuits*, vol. 51, no. 2, pp. 523–532, Feb 2016.

[29] S. Angkititrakul and H. Hu, "Design and analysis of buck converter with pulse-skipping modulation," in *2008 IEEE Power Electronics Specialists Conference*, June 2008, pp. 1151–1156.

[30] F. u. Rahman, S. Kim *et al.*, "A unified clock and switched-capacitor-based power delivery architecture for variation tolerance in low-voltage soc domains," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1173–1184, April 2019.

[31] J. Wang, S. Member, and B. H. Calhoun, "Techniques to extend canary-based standby VDD scaling for SRAMs to 45 nm and beyond," Solid-State Circuits," *IEEE Journal of*, pp. 2514–2523, 2008.

[32] T. D. Burd, T. A. Pering *et al.*, "A dynamic voltage scaled microprocessor system," *IEEE Journal of solid-state circuits*, vol. 35, no. 11, pp. 1571–1580, 2000.

[33] D. Kim, J. Kim, H. Ham, and M. Seok, "A 0.5v-VIN1.44ma-class event-driven digital LDO with a fully integrated 100pf output capacitor," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017, pp. 346–347.

[34] D. Kim and M. Seok, "A Fully Integrated Digital Low-Dropout Regulator Based on Event-Driven Explicit Time-Coding Architecture," *IEEE Journal of Solid-State Circuits*, vol. 52, no. 11, pp. 3071–3080, Nov. 2017.

[35] S. Li and B. H. Calhoun, "A 745pa Hybrid Asynchronous Binary-Searching and Synchronous Linear-Searching Digital LDO with 3.8×105 Dynamic Load Range, 99.99% Current Efficiency, and 2mv Output Voltage Ripple," in *2019 IEEE International Solid-State Circuits Conference - (ISSCC)*, Feb. 2019, pp. 232–234.

[36] T. Mahajan, R. Muthukaruppan, D. M. Shetty, S. Mangal, and H. K. Krishnamurthy, "Digitally controlled voltage regulator using oscillator-based ADC with fast-transient-response and wide dropout range in 14nm CMOS," in *2017 IEEE Custom Integrated Circuits Conference (CICC)*, Apr. 2017, pp. 1–4.

[37] Y. Li, X. Zhang, Z. Zhang, and Y. Lian, "A 0.45-to-1.2-V Fully Digital Low-Dropout Voltage Regulator With Fast-Transient Controller for Near/Subthreshold Circuits," *IEEE Transactions on Power Electronics*, vol. 31, no. 9, pp. 6341–6350, Sep. 2016.

[38] L. G. Salem, J. Warchall, and P. P. Mercier, "A Successive Approximation Recursive Digital Low-Dropout Voltage Regulator With PD Compensation and Sub-LSB Duty Control," *IEEE Journal of Solid-State Circuits*, vol. 53, no. 1, pp. 35–49, Jan. 2018.

[39] Y. Okuma, K. Ishida *et al.*, "0.5-V input digital LDO with 98.7% current efficiency and 2.7-µA quiescent current in 65nm CMOS," in *IEEE Custom Integrated Circuits Conference 2010*, Sep. 2010, pp. 1–4.

[40] S. B. Nasir and A. Raychowdhury, "On limit cycle oscillations in discrete-time digital linear regulators," in *2015 IEEE Applied Power Electronics Conference and Exposition (APEC)*, Mar. 2015, pp. 371–376.

[41] S. B. Nasir, S. Gangopadhyay, and A. Raychowdhury, "A 0.13µm fully digital low-dropout regulator with adaptive control and reduced dynamic stability for ultra-wide dynamic range," in *2015 IEEE International Solid-State Circuits Conference - (ISSCC) Digest of Technical Papers*, Feb. 2015, pp. 1–3.

[42] N. H. Weste and D. Harris, *CMOS VLSI design: a circuits and systems perspective*. Pearson Education, 2015.

[43] V. Kucera, "Deadbeat control, pole placement, and LQ regulation," *Kybernetika*, vol. 35, pp. 681–692, 1999.

[44] H. Hu, V. Yousefzadeh, and D. Maksimovic, "Nonuniform A/D Quantization for ImprovedDynamic Responses of Digitally ControlledDC–DC Converters," *IEEE Transactions on Power Electronics*, vol. 23, no. 4, pp. 1998–2005, Jul. 2008.

[45] X. Sun, S. Kim, F. u. Rahman, V. R. Pamula, X. Li, N. John, and V. S. Sathe, "A combined all-digital PLL-buck slack regulation system with autonomous CCM/DCM transition control and 82% average voltage-margin reduction in a 0.6-to-1.0v cortex-M0 processor," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 302–304.

[46] F. u. Rahman, S. Kim, N. John, R. Kumar, X. Li, R. Pamula, K. A. Bowman, and V. S. Sathe, "A Unified Clock and Switched-Capacitor-Based Power Delivery Architecture for Variation Tolerance in Low-Voltage SoC Domains," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 4, pp. 1173–1184, Apr. 2019.

[47] J. P. LaSalle, "Time Optimal Control Systems," *Proceedings of the National Academy of Sciences*, vol. 45, no. 4, pp. 573–577, Apr. 1959. [Online]. Available: https://www.pnas.org/content/45/4/573

[48] Y.-C. Li, C.-J. Chen, and C.-J. Tsai, "A Constant On-Time Buck Converter With Analog Time-Optimized On-Time Control," *IEEE Transactions on Power Electronics*, vol. 35, no. 4, pp. 3754–3765, Apr. 2020.

[49] S. Vazquez, J. Rodriguez, M. Rivera, L. G. Franquelo, and M. Norambuena, "Model Predictive Control for Power Converters and Drives: Advances and Trends," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 2, pp. 935–947, Feb. 2017.

[50] E. A. Burton, G. Schrom, F. Paillet, J. Douglas, W. J. Lambert, K. Radhakrishnan, and M. J. Hill, "FIVR — Fully integrated voltage regulators on 4th generation Intel® Core™ SoCs," in *2014 IEEE Applied Power Electronics Conference and Exposition - APEC 2014*, Mar. 2014, pp. 432–439, iSSN: 1048-2334.

[51] C. E. García, D. M. Prett, and M. Morari, "Model predictive control: Theory and practice—A survey," *Automatica*, vol. 25, no. 3, pp. 335–348, May 1989. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0005109889900022

[52] W. Tsou, W. Yang *et al.*, "Digital low-dropout regulator with anti PVT-variation technique for dynamic voltage scaling and adaptive voltage scaling multicore processor," in *2017 IEEE International Solid-State Circuits Conference (ISSCC)*, Feb. 2017, pp. 338–339.

[53] R. J. Polge and E. M. Mitchell, "Impulse response determination by cross correlation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-6, no. 1, pp. 91–97, 1970.

[54] A. Emami-Naeini and G. Franklin, "Deadbeat control and tracking of discrete-time systems," *IEEE Transactions on Automatic Control*, vol. 27, no. 1, pp. 176–181, February 1982.

[55] S. Kundu, M. Liu, S. Wen, R. Wong, and C. H. Kim, "A Fully Integrated Digital LDO With Built-In Adaptive Sampling and Active Voltage Positioning Using a Beat-Frequency Quantizer," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 1, pp. 109–120, Jan. 2019.

[56] D. Kim, S. Kim, M. Seok, H. Ham, and J. Kim, "0.5v-VIN, 165-mA/mm2 Fully-Integrated Digital LDO Based on Event-Driven Self-Triggering Control," in *2018 IEEE Symposium on VLSI Circuits*, Jun. 2018.

[57] J. Park and D. Jeong, "A Fully Integrated 700ma Event-Driven Digital Low-Dropout Regulator with Residue-Tracking Loop for Fine-Grained Power Management Unit," in *2018 IEEE Symposium on VLSI Circuits*, Jun. 2018, pp. 231–232.

[58] Y. Lu, F. Yang, F. Chen, and P. K. T. Mok, "A 500ma analog-assisted digital-LDO-based on-chip distributed power delivery grid with cooperative regulation and IR-drop reduction in 65nm CMOS," in *2018 IEEE International Solid - State Circuits Conference - (ISSCC)*, Feb. 2018, pp. 310–312.

[59] E. F. Camacho, *Model predictive control (Advanced Textbooks in Control and Signal Processing)*. Springer, 2007.

[60] J.-I. Seo, B.-M. Lim, and S.-G. Lee, "A 96.5% Efficiency Current Mode Hysteretic Buck Converter With 1.2% Error Auto-Selectable Frequency Locking," *IEEE Transactions on Power Electronics*, vol. 33, no. 9, pp. 7733–7743, Sep. 2018.

[61] J.-G. Kang, J. Park, M.-G. Jeong, and C. Yoo, "A Time-Domain-Controlled Current-Mode Buck Converter With Wide Output Voltage Range," *IEEE Journal of Solid-State Circuits*, vol. 54, no. 3, Mar. 2019.

[62] S. J. Kim, Q. Khan, M. Talegaonkar, A. Elshazly, A. Rao, N. Griesert, G. Winter, W. McIntyre, and P. K. Hanumolu, "High Frequency Buck Converter Design Using Time-Based Control Techniques," *IEEE Journal of Solid-State Circuits*, vol. 50, no. 4, pp. 990–1001, Apr. 2015.

[63] F. Atallah, K. Bowman, H. Nguyen, J. Jeong, D. Yingling, Y. Sun, B. Appel, A. Polomik, M. Harinath, J. Morelli, T. Moore, N. Reeves, A. Cassier, and A. Raychowdhury, "19.3 a 7nm all-digital unified voltage and frequency regulator based on a high-bandwidth 2-phase buck converter with package inductors," in *2019 IEEE International Solid- State Circuits Conference - (ISSCC)*, 2019, pp. 316–318.