

MODELING, DESIGN AND OPTIMIZATION OF IC POWER DELIVERY
WITH ON-CHIP REGULATION

A Dissertation

by

SUMING LAI

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Peng Li
Committee Members,	Gwan S. Choi
	Edgar Sánchez-Sinencio
	Eun J. Kim
Head of Department,	Chanan Singh

December 2014

Major Subject: Computer Engineering

Copyright 2014 Suming Lai

ABSTRACT

As IC technology continues to follow the Moore's Law, IC designers have been constantly challenged with power delivery issues. While useful power must be reliably delivered to the on-die functional circuits to fulfill the desired functionality and performance, additional power overheads arise due to the loss associated with voltage conversion and parasitic resistance in the metal wires. Hence, one of the key IC power delivery design challenges is to develop voltage conversion/regulation circuits and the corresponding design strategies to provide a guaranteed level of power integrity while achieving high power efficiency and low area overhead.

On-chip voltage regulation, a significant ongoing design trend, offers appealing active supply noise suppression close to the loads and is well positioned to address many power delivery challenges. However, to realize the full potential of on-chip voltage regulation requires systemic optimization of and tradeoffs among settling time, steady-state error, power supply noise, power efficiency, stability and area overhead, which are the key focuses of this dissertation. First, we develop new low-dropout voltage regulators (LDOs) that are well optimized for low power applications. To this end, dropout voltage, bias current and speed are important competing design objectives. This dissertation presents new flipped voltage follower (FVF) based topologies of on-chip voltage regulators that handle ultra-fast load transients in nanoseconds while achieving significant improvement on bias current consumption. An active frequency compensation is embedded to achieve high area efficiency by employing a smaller amount of compensation capacitors, the major silicon area contributor. Furthermore, in one of the proposed topologies an auxiliary digital feedback loop is employed in order to lower quiescent power consumption further.

Second, coping with supply noise is becoming increasingly more difficult as design complexity grows, which leads to increased spatial and temporal load heterogeneity, and hence larger voltage variations in a given power domain. Addressing this challenge through a distributed methodology wherein multiple voltage regulators are placed across the same voltage domain is particularly promising. This distributive nature allows for even faster suppression of multiple hot spots by the nearby regulators within the power domain and can significantly boost power integrity. Nevertheless, reasoning about the stability of such distributively regulated power networks becomes rather complicated as a result of complex interactions between multiple active regulators and the large passive subnetwork. Coping with this stability challenge requires new theory and stability-ensuring design practice, as targeted by this dissertation. For the first time, we adopt and develop a hybrid stability framework for large power delivery networks with distributed voltage regulation. This framework is local in the sense that both the checking and assurance of network stability can be dealt with on the basis of each individual voltage regulator, leading to feasible design of large power delivery networks that would be computationally impossible otherwise. Accordingly, we propose a new hybrid stability margin concept, examine its tradeoffs with power efficiency, supply noise and silicon area, and demonstrate the resulted key design implications pertaining to new stability-ensuring LDO circuit design techniques and circuit topologies. Finally, we develop an automated hybrid stability design flow that is computationally efficient and provides a practical guarantee of network stability.

ACKNOWLEDGEMENTS

Through the past four and a half years of my PhD life, I am very grateful for all I have received.

I would first like to express my profound gratitude to my advisor, Prof. Peng Li. His guidance and support are most essential for the work presented in this dissertation to be accomplished. With a rather comprehensive set of knowledge and expertise in areas from circuit simulation to circuit design, he constantly provides enlightening advice and valuable contribution to my research work. Prof. Li is also a good source of encouragement that has not only helped me get through disheartening failures and, more often, strengthened my core to move on for a higher goal that I could not have thought about. Moreover, Prof. Li's diligence and ambition has also silently spurred me to try to keep abreast.

I would like to thank Prof. Edgar Sánchez-Sinencio. From his teaching I harvested a good deal of analog circuit design related knowledge and skills that are essential to my research work on voltage regulation. I also owe a lot to Prof. Gwan S. Choi, and Prof. Eun J. Kim for serving on my PhD committee and providing insightful feedbacks on my preliminary exam.

I would also like to thank many other colleagues in the Computer Engineering Group who contributed a lot on my dissertation by providing inspiring technical discussions. Dr. Boyuan Yan and Dr. Zhiyu Zeng had been great partners and friends to me before they left Texas A&M. I was fortunate to have Dr. Yan as my roommate for one year. His incandescent enthusiasm to research and unusual diligence has burned a model into my mind that continuously reminds me a correct attitude to my life. Dr. Yan's expertise in power grid analysis provided very helpful

discussions that were important pieces to the foundation of my work on stability of the distributed on-chip regulation. Dr. Zeng, who initiated a good momentum of research on distributed on-chip voltage regulators, literally introduced me to power delivery problems that modern IC's suffer from. With his expertise in power grid simulation and optimization, Dr. Zeng also provided great technical help. I am also thankful to Tong Xu, Leyi Yin, Yong Zhang, and Yongtae Kim for helpful discussions and conversations on related topics that broadened my sight of research. I also thank my former officemate, Navid Abedini, for his peaceful yet warm personality that was essential to the nice office environment we shared.

Last but not the least, I owe my family's unselfish support without which it certainly would be very hard for me to complete my PhD degree. Special thank goes to my wife, Xin. Three-year separated marriage life was by no means easy for either of us. And through all these years I only received and continue received her tremendous emotional and substantial support without a word of complaint. The hardest time there we walked through together is and will forever be one of the most beautiful memories we share.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	xiii
1. INTRODUCTION	1
2. An Area-Efficient On-Chip LDO with Ultra-Fast Transient Regulation	7
2.1 Background	10
2.1.1 Load- and Switching-Induced Noise on the Power Grids	10
2.1.2 FVF-Based LDO Topologies	11
2.2 The Proposed LDO Topology and Circuit Implementation	13
2.2.1 The Proposed Multi-Loop LDO Topology	13
2.2.2 Output Impedance	16
2.2.3 Stability Analysis and Frequency Compensation	20
2.2.4 Notch in the Output Impedance and PSRR	26
2.2.5 Dynamic Biasing	28
2.3 Simulation Results	31
2.3.1 Load Regulation	31
2.3.2 Robustness to Process, Temperature and Mismatches	36
2.3.3 Line Regulation	39
2.3.4 Comparison with the Antetypes	39
2.3.5 Benefits of the Notch	43
2.4 Performance Comparisons	48
2.5 Summary	52
3. A Power-Efficient On-Chip LDO Assisted by Switched Capacitors for Fast Transient Regulation	53
3.1 Concepts of the Proposed Techniques	55
3.1.1 The Switched Decoupling Capacitor	55
3.1.2 The Switched Positioning Capacitor	58
3.1.3 Discussions of the Two Techniques	59
3.2 Circuit Analysis and Implementation	60

3.2.1	Design of the Switched Capacitor Circuits	61
3.2.2	Design of the LDO	67
3.3	Simulation Results	68
3.3.1	LDO with Switched Decoupling Capacitors	69
3.3.2	LDO with Switched Positioning Capacitors	70
3.3.3	Performance Comparisons	73
3.4	Summary	74
4.	Design of Distributed On-Chip Regulators with Ensured Stability	76
4.1	The First Glance on Distributed Regulator Design	77
4.2	Investigation of PDN Stability	80
4.3	PDN Partitioning and Modeling	85
4.3.1	Concepts of Proposed Partition and Modeling	85
4.3.2	The Proposed Network Partition and Modeling	86
4.4	The Theoretical Framework	90
4.4.1	Preliminaries	91
4.4.2	Two Classical Stability Theorems	94
4.4.3	Hybrid Stability Theorem	96
4.4.4	Hybrid Stability Framework for PDNs	96
4.5	New Hybrid Stability Margin Concept And Efficient Stability Check- ing of the PDN	101
4.5.1	Stability Checking of the PDN	101
4.5.2	Hybrid Stability Margin (HSM)	103
4.6	Practical PDN Network Design	104
4.6.1	Design Flow	105
4.6.2	LDO Design Insights and Performance Tradeoffs	107
4.6.3	Illustrative Design Optimization	116
4.7	Experimental Study	118
4.7.1	Multiple LDOs in a Small Network	118
4.7.2	Multiple LDOs in a Large Network	120
4.7.3	Performance Trade-Offs	124
4.8	Summary	126
5.	CONCLUSION AND FUTURE WORK	127
5.1	Conclusion of the Dissertation	127
5.2	Future Work	128
	REFERENCES	131

LIST OF FIGURES

FIGURE	Page
1.1 Power supply trends by ITRS [1].	1
1.2 A power delivery system with a switching DC-DC converter followed by on-chip linear regulators.	3
1.3 Global feedback loops present in a PDN with distributed on-chip regulation.	5
2.1 Diagram of the power delivery system with switching DC-DC converter followed by on-chip linear regulators driving the on-chip power grids.	8
2.2 Spectrum density of the load current generated by clocked circuitry with pseudo-random inputs.	11
2.3 Topologies of the reported FVF-based LDOs [4, 8, 9].	12
2.4 Topology of the proposed LDO.	14
2.5 The full schematic of the proposed LDO.	16
2.6 Small-signal model of the proposed LDO. (g_{mp} , g_{mc} , g_{mB} , g_{ma} , and g_{ma2} are the transconductances of M_p , M_c , M_{db} , the EA's input device, and the IA, respectively; g_{dsp} , g_{dsc} , and g_{dsB} represent the drain-source conductance of M_p , M_c , M_{db} with g_a and g_{a2} representing the equivalent output resistances of EA and IA, respectively; C_{gsx} and C_{gdx} are respectively the gate-to-source capacitance and gate-to-drain capacitance of the devices of the same device-name subscript x (e.g., C_{gsc} corresponds to M_c); C_{an} is the output capacitance at the negative output terminal of EA while C_{ci} is the i -th compensation capacitor ($i = 1 \dots 3$); R_1 , R_2 , C_1 and C_L are the same as those in Fig. 2.4.) . . .	18
2.7 Illustrative Bode plot of the loop gain of the proposed LDO. (a) Before and after compensation. (b) With different load currents.	21
2.8 Simulated Bode plots of the proposed LDO. (a) Before and after compensation. (b) Different load currents ($C_L = 100\text{pF}$). (c) Different C_L ($I_L = 0\text{A}$).	22

2.9	Relationship of quiescent current and input voltage of class AB amplifier. (a) Schematic of a simplified class AB amplifier. (b) I-V curve. (c) The simulation results of the total bias current for the LDO vs. load current.	30
2.10	Transient responses to load current stepping between 0mA and 100mA. (a) $t_r = 10\text{ns}$, $C_L=100\text{pF}$. (b) $t_r = 1\text{ns}$, $C_L=600\text{pF}$. (c) $t_r = 100\text{ps}$, $C_L=600\text{pF}$	32
2.11	Transient responses to load current stepping between 1mA and 101mA with the transition time of 10ns and $C_L=100\text{pF}$	33
2.12	Transient responses to load current stepping between 1mA and 101mA with the transition time of 1ns and $C_L=600\text{pF}$	34
2.13	Transient responses to load current stepping between 1mA and 101mA with the transition time of 0.1ns and $C_L=600\text{pF}$	35
2.14	Monte Carlo simulation results (1000 samples) at $C_L = 600\text{pF}$. (a) Steady-state output voltage at $I_L = 0\text{A}$. (b) Steady-state output voltage at $I_L = 0.1\text{A}$. (c) The average biasing current of the LDO. (d) The maximum voltage drop when load current is switching between 1mA and 101mA within 100ps with $C_L = 600\text{pF}$. (e) The voltage overshoot when load current is switching between 1mA and 101mA within 100ps with $C_L = 600\text{pF}$	37
2.15	Temperature-sweep simulation results. (a) Steady-state output voltage. (b) Transient load regulation with load current switching between 1mA and 101mA within 100ps and $C_L = 600\text{pF}$	38
2.16	Transient responses to input voltage steps. (a) 1-ns V_{in} transition time with $C_L = 600\text{pF}$. (b) 1- μs V_{in} transition time with $C_L = 100\text{pF}$	40
2.17	Comparisons of the three LDOs' output impedance. (a) At $I_L=1\text{mA}$, $C_L=1\text{pF}$. (b) At $I_L=100\text{mA}$, $C_L=1\text{pF}$	41
2.18	Comparison of the three LDOs' load transient responses at 1- μs transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.	42
2.19	Comparison of the three LDOs' load transient responses at 0.1- μs transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.	44

2.20	Comparison of the three LDOs' load transient responses at 10-ns transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.	45
2.21	Load regulation performance comparison of the implementations with and without the notch ($I_L = 100\text{mA}$ and $C_L = 1\text{nF}$). (a) Output impedance. (b) Comparisons of transient responses.	46
2.22	Line regulation performance comparison of the implementations with and without the notch ($I_L = 1\text{mA}$ and $C_L = 600\text{pF}$). (a) The PSRRs. (b) Comparisons of transient responses.	47
2.23	Influences on the impedance notch frequency of V_{in} and I_L . (a) Varying V_{in} . (b) Varying I_L	49
3.1	An illustration of the benefit on power saving from a well-designed regulator.	54
3.2	Conceptual schematics of the switched capacitor techniques. (a) The existing technique. (b) The proposed switched decap technique. (c) The proposed switched positioning cap technique.	56
3.3	Schematics of the switched capacitor circuits. (a) The push-up circuit. (b) The pull-down circuit. (c) The schematic of the comparator. . . .	60
3.4	Illustrations of I_L and V_{out} with the switched decap technique. (a) When $t_r \gg t_{resp}$. (b) When $t_r \ll t_{resp}$	62
3.5	Illustrations of I_L and V_{out} with the switched positioning capacitor technique.	66
3.6	The schematic of the LDO.	68
3.7	Comparison of transient load regulation.	69
3.8	Transient load regulation with $t_r=5\text{ns}$	71
3.9	Transient line regulation with $t_r=5\text{ns}$	71
3.10	Monte Carlo simulation results.	72
3.11	Temperature dependence of the performances.	72
4.1	Illustration of the power delivery network with distributed on-chip regulators.	77
4.2	(a) The generic LDO structure. (b) The two-port Y-parameter model of the generic LDO.	80

4.3	Illustration of the problem when applying the open-loop methods to the stability problem under discussion. (a) Traditional stability checking in the design of a single LDO. (b) Problem illustration when applying open-loop method to the PDN.	82
4.4	Illustration of the inter-LDO loops in the PDN.	83
4.5	Pole analysis results that demonstrate a instability-arousing pole movement (each cross represents a pole location).	84
4.6	Transient analysis results that demonstrate the stability problem. . .	85
4.7	Partition of the PDN model.	87
4.8	PDN modeling with the system-wide feedback loop. (a) The complete PDN model with system inputs and outputs. (b) The PDN model reduced to contain only signals pertaining to the stability issue. . . .	89
4.9	The signal-flow graph of the system. ($i_{1,G}, i_{2,G}, V_{1,G}, V_{2,G}$ and $i_{1,Z}, i_{2,Z}, V_{1,Z}, V_{2,Z}$ are the same as in Fig. 4.7.)	90
4.10	The illustration of the serial resistance at each port of the H block. .	97
4.11	Hybrid stability margin at a frequency point.	104
4.12	Stability-ensuring design flow.	106
4.13	Demonstrations of exemplary stability-enhancing schemes for LDO output stage design. (a) Scheme I: simple circuit modification on the output stage. (b) Scheme II: topology change for the output stage. .	109
4.14	The illustrations of two VDD grid structures for distributed regulators. (a) The common global VDD grid structure. (b) The proposed global VDD grid structure.	112
4.15	The illustration of a practical way to weaken the inter-LDO interactions.	113
4.16	The illustration of splitting self-admittances in LDO's Y-parameter model.	114
4.17	The LDO topology used in the practical implementations [25].	117
4.18	Pole analysis showing the stability of the PDN designed with the proposed approach.	119
4.19	Transient analysis confirming the stability of the PDN with the stability-ensured LDOs.	120

4.20	The loop gain and λ_{min} of the initial design.	121
4.21	The loop gain and λ_{min} of the stability-ensured design.	122
4.22	The transient simulation results showing the instability of the PDN with the LDOs designed in a standard manner.	122
4.23	The transient simulation results confirming the stability of the PDN with the stability-ensured LDOs.	123

LIST OF TABLES

TABLE	Page
2.1 List of Parameters of Devices in Fig. 2.5	16
2.2 Performance Comparisons of the Proposed LDO with Some Existing Works	50
3.1 Performance Comparisons of the Proposed LDOs with Prior Art	73
4.1 The Performance Trade-offs	124
4.2 The Comparison of Trade-offs in Different Global VDD Grid Structures	126

1. INTRODUCTION

Almost all the electronic devices that operate at a different supply voltage from their external voltage supplies need a kind of circuit called voltage converter/regulator to generate the proper voltages. A good-quality power delivery network (PDN) is essential for providing such stable and correct supply voltages to the on-chip functional circuits. A significant trend, shown in Fig. 1.1, indicated by the International Technology Roadmap for Semiconductors (ITRS) [1], reveals two important challenges for the design of IC power delivery networks. First of all, as the supply voltage keeps scaling down towards the sub-threshold regime to further save chip power consumption, circuit delay will become much more vulnerable to supply noise, which indicates that PDN design will have to face an even more stringent constraint on power integrity. Second, given the fact that more and more modules or IP's are

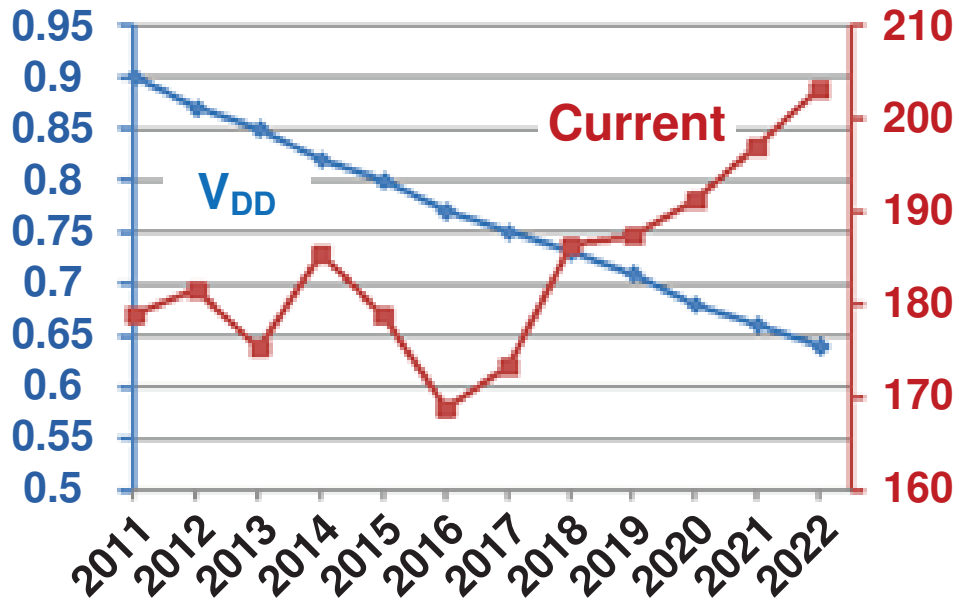


Figure 1.1: Power supply trends by ITRS [1].

being integrated into a single chip to offer more functionality and better performance, the current demand of the chip is increasing in general. The boosted current demand in turn will further deteriorate the already-severe IR drop and dynamic voltage drop caused by parasitic package and on-chip trace inductance (known as di/dt supply noise), thus further degrading the on-chip power integrity.

Due to the foregoing drive, in the past decade, there has been an intensive research effort dedicated to complete system-on-chip design solutions that include on-chip voltage regulation modules [2–4, 10, 21]. By integrating regulators with other analog and/or digital functional circuits, on the one hand, the end products can be of smaller volumes and costs. On the other hand, as each regulator is placed local to its loading circuits and blocks the undesirable power noise by the package and off-chip components, the on-chip regulator is able to provide stronger voltage regulation and offers remarkable improvement on on-chip power integrity. Other benefits of on-chip voltage regulation include facilitation of various on-chip voltage domains, suppression of package resonance [6], reduction of product footprint.

On-chip low-dropout voltage regulators (LDO) are popular as the post-regulator in the hybrid regulation scheme illustrated in Fig. 1.2, owing to the fact that they have better transient regulation performance per unit silicon area compared with other types of on-chip voltage regulators. Unlike off-chip counterparts, without help of the huge output capacitor an on-chip LDO has to combat on its own challenges including ultra-fast load transients, stability, power/current efficiency, area efficiency, and high-frequency power supply ripple rejection [7, 9, 16, 25]. Especially in this gigahertz era, sub-nanosecond response time is usually required for the LDO to output a good-quality supply voltage [21].

The improved supply integrity is also good for achieving higher power efficiency because this reduces overdesign margin on the nominal supply voltage level. On the

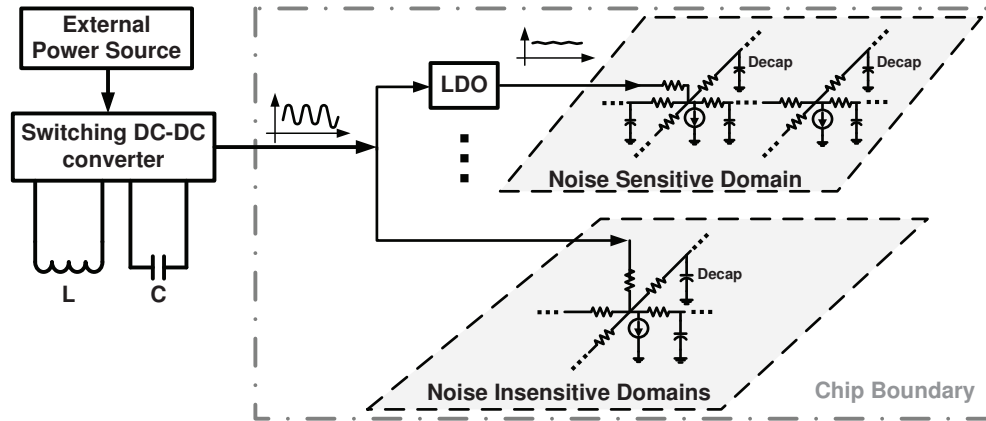


Figure 1.2: A power delivery system with a switching DC-DC converter followed by on-chip linear regulators.

other hand, the major efficiency hindrance in LDO-regulated PDNs is the LDO's dropout voltage. The smaller the dropout voltage, the better the power efficiency. Nevertheless, lowering dropout voltage usually results in degraded transient response. Sustaining a high power efficiency is fighting with achieving good transient response. Thus, a harsh trade-off problem between transient response time and power efficiency is left for designers to solve through innovation.

In the state-of-the-art high-performance chips, the increasing complexity and size of the chip pose great difficulty in pulling off timing closure in the existence of the exacerbated spacial variation of supply voltage which logic path delays are sensitive to. One promising solution is to mitigate the spacial variation by distributing an array of on-chip LDOs across the power domain [21]. As such, the effective distance from the loading circuit to the regulator is further reduced and so is the voltage droop. Supply voltage difference across that domain is, therefore, better confined.

While a single on-chip LDO per power domain has been researched for about a decade, it is only until recently that the PDN architecture that incorporates dis-

tributed on-chip voltage regulation has emerged and gained increasing attention. As always, new techniques imply new challenges and problems to tackle.

Stability is the first problem of all. On the one hand, the frequency compensation scheme for each LDO needs to be devised to emphasize area efficiency of the LDO which becomes important in this distributed regulation design. Schemes that need smaller compensation capacitors will be preferred over those require large ones. On the other hand, with each LDO being stable, is it always true that the whole PDN is stable when we construct the distributed regulation architecture using a multiplicity of stable LDOs? The question has been first answered recently as “no” in our work presented in [29,30], according to which, the global feedback loops (as illustrated in Fig. 1.3) such as the inter-LDO feedback paths and the feedback paths formed by parasitic capacitive coupling in the PDN are the main factors to be blamed for the network instability. Since the stability of the network comprised of stable LDOs designed in a conventional way is not guaranteed, it calls for a new LDO design technique that not only ensures the stability of each individual LDO but also guarantees the network stability when distributing these LDOs across the chip.

This dissertation is motivated by the foregoing problems in the design of PDNs with on-chip voltage regulation. In view of the lack of on-chip LDO designs that are capable of handling sub-nanosecond load transients while still consuming a relatively small amount of quiescent current, Chapter II of this dissertation introduces the design and analysis of a proposed new on-chip LDO topology that fits to the vacancy. More specifically, the specific contributions include:

1. A multi-feedback loop LDO topology that is capable of handling ultra-fast load transients as well as achieving ultra-high DC regulation accuracy and relatively low quiescent current;

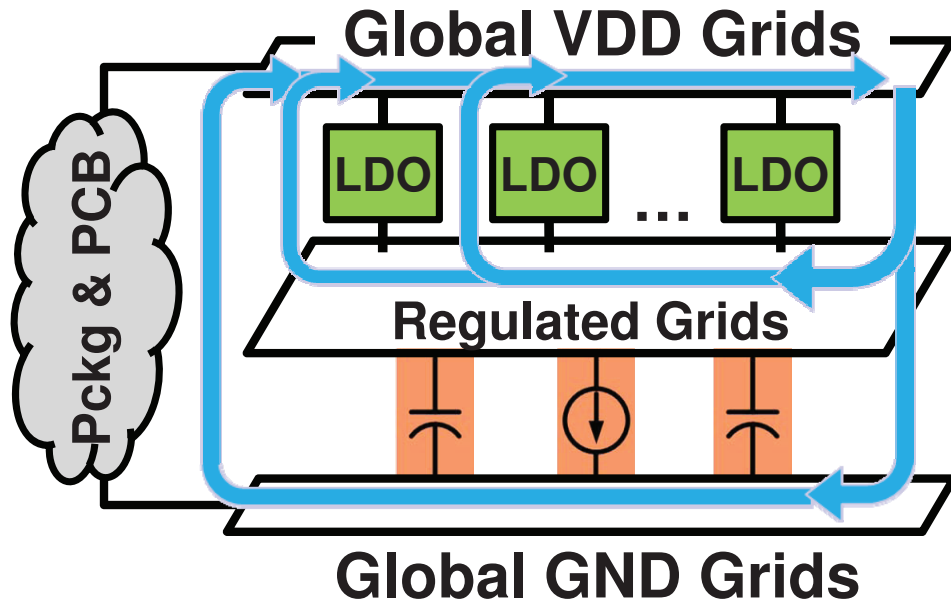


Figure 1.3: Global feedback loops present in a PDN with distributed on-chip regulation.

2. An active frequency compensation scheme that allows a significant reduction of compensation capacitors and hence good for achieving high area efficiency;
3. A notch filtering mechanism at the preceding switching converter's switching frequency for additional power supply ripple rejection.

Based upon this LDO topology, Chapter III of this dissertation presents a switched-capacitor based LDO topology which further improves the LDO's power efficiency while retaining its fast transient regulation ability. The contribution of this work is the proposal of employing a capacitor to assist the LDO to regulate the voltage by storing charges during the idle periods of load and promptly offering charges when fast load transient occurs. This allows the LDO to be biased with less quiescent current while still providing good quality of supply voltage.

Chapter IV of this dissertation is dedicated to solving the stability problem in the design of power delivery systems with distributed on-chip regulators. This work

for the first time exposes the potential instability issue in distributed voltage regulation and provides an efficient LDO design methodology to allow LDO designers to guarantee the stability of the entire system. The major contributions include:

1. Identification of potential catastrophic network stability failures resulted from the application of the conventional LDO design methodology;
2. A theoretically rigorous hybrid stability analysis framework that checks the power delivery stability with computational efficiency, which serves as the foundation for the proposed design and optimization methodology with assurance of PDN stability;
3. An efficient LDO design methodology that ensures the power delivery stability and optimizes the trade-offs between stability and other performance metrics;
4. Novel LDO and passive power grid design insights.

2. AN AREA-EFFICIENT ON-CHIP LDO WITH ULTRA-FAST TRANSIENT REGULATION*

In this gigahertz era, on-chip LDOs have to face fast load-current variations with the rise/fall time on the order of nanosecond or even less [4]. Conventional off-chip LDOs count on a large off-chip capacitor or the package parasitic capacitance of several micro-farads to tackle fast load transients, whereas implementation of the on-chip LDO cannot afford such a huge capacitor. Alternatively, on-chip LDOs have to respond faster to load transients so as to compensate the absence of the external capacitor. Furthermore, implementations of fine-grain on-chip power domains and/or dynamic-voltage-scaling (DVS) techniques embrace an aggressive integration of multiple on-chip regulators [5,6], where area efficiency is one of the key constraints. Generally, the area cost of LDO mainly comes from the frequency compensation capacitors, which are usually several to tens of picofarads [7,9], and the pass transistor. It is important to reduce those two area consumptions to enable the aforementioned implementations.

In addition, since inductor-based switching DC-DC converters are on-average superior in power efficiency and the maximum load current, LDOs are usually in the power delivery system illustrated in Fig. 2.1 to achieve good overall power efficiency and stable output voltage simultaneously. In the system, LDOs are supplied by the switching DC-DC converter and loaded by power grids. On the input interface, for effective isolation of its output from the input voltage ripples the LDO needs to have good power supply ripple rejection (PSRR) over a wide range of frequencies,

*Reprinted with permission from “A fully on-chip area-efficient CMOS low-dropout regulator with fast load regulation” by S. Lai, and P. Li, 2012. *Analog Integrated Circuits and Signal Processing*, 72(2): 433–450, Copyright [2012] by Springer.

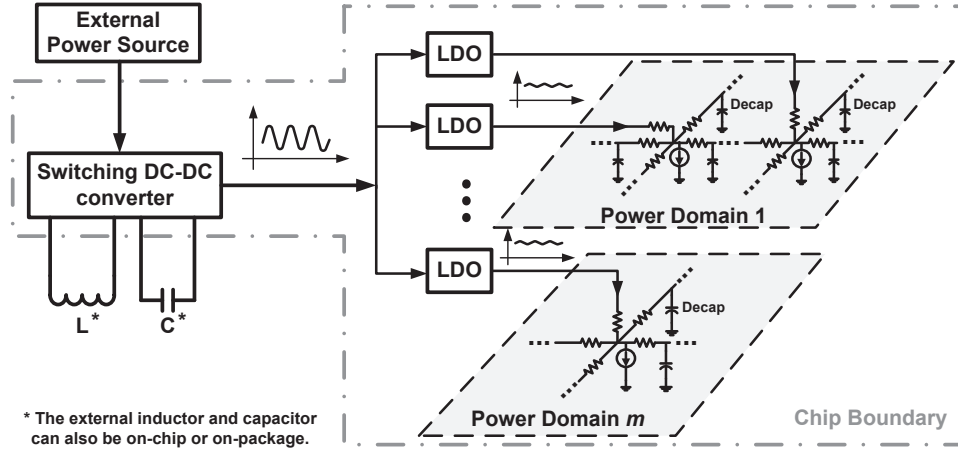


Figure 2.1: Diagram of the power delivery system with switching DC-DC converter followed by on-chip linear regulators driving the on-chip power grids.

especially at the switching frequency of the preceding converter. On the output side, the LDO should be able to drive a wide range of load capacitance due to the uncertainty of the actual capacitance rendered by the power grids.

Recently, a class of on-chip LDO topologies adopting flipped voltage follower (FVF) as the LDOs' output stages emerged to tackle fast load-current transients [4, 16]. The shunt feedback connection in FVF results in a lower output impedance [15], which is helpful to achieve fast load regulation. However, the improvements were limited by the relatively weak feedback loop gain at DC and low frequencies compared with the conventional error-amplifier-based LDO topologies as the one in [7]. As a result, the steady-state regulation performance is traded off for good transient response. To achieve even better transient response without sacrificing the steady-state performance any further, a variety of capacitive coupling techniques from the output of LDO to the biasing circuits were adopted [8, 9]; furthermore, an additional amplifier was also inserted into the feedback path in an attempt to improve the steady-state performance of the FVF-based LDO [9]. Effectively as the capacitive

coupling technique works, it is at the cost of considerable silicon area occupied by those additional capacitors; the stability of LDO may be also compromised by the coupling, which was not addressed in either of the two works. Likewise, the insertion of the additional amplifier inevitably endangered stability and the Miller frequency compensation accordingly applied in [9] involved a considerably large compensation capacitor.

To address the above issues, an additional feedback path constructed with a fully differential error amplifier is introduced to the FVF-based topology in the way that has both the steady-state regulation performance and the transient response enhanced. Furthermore, a novel active frequency compensation scheme is conceived that allows the load capacitance of the LDO to vary over a wide range and needs smaller compensation capacitors in total. And hence it achieves higher area efficiency than Miller compensation scheme. In addition, an output-impedance-oriented dynamic biasing scheme is proposed that boosts the LDO's bias current when it is most needed to lower the LDO's output impedance and reduces it, otherwise, to save power. The LDO also features a magnitude notch in both its PSRR and output impedance that provides further suppression of the supply voltage ripple and the load-induced output voltage fluctuation. The rest of this chapter is organized as follows. Section 2.1 gives discussions on the on-chip LDO design background. Detailed circuit analysis and discussions are presented in Section 2.2, followed by thorough simulation results in Section 2.3. Finally, a comprehensive performance comparison with some recent works is conducted in Section 2.4.

2.1 Background

2.1.1 Load- and Switching-Induced Noise on the Power Grids

The supply voltage ripple caused by the switching converter and the load current variations are two major sources of noise for the power grids. The resultant LDO output voltage variations are referred to as switching-induced and load-induced noise, respectively, in the following context. For the switching-induced noise, assuming the DC-DC converter adopts the popular pulse-width modulation (PWM) control mode, it is well-understood that the power spectrum of the noise source must have a high peak around the switching frequency. As a result, for less switching-induced noise, good LDO PSRR around the switching frequency should be achieved.

For the load-induced noise, the spectrum density is much less predictable and time-varying, but it also has a general pattern to follow. First and foremost, the LDOs' load circuits need to consume a certain amount of average power to fulfill their functions. Therefore, in the power spectrum density of the load current, there is a significant component at DC. Besides, the digital circuit driven by LDOs are mostly synchronized by clocks. The activities of these circuits could be triggered by both of the clock edges. For example, the buffers in the clock trees dissipate peak power at both edges of clock, so does the master-slave type of circuits. Therefore, it can be speculated that the load current of LDO has a considerable power component at the clock frequency (f_{clk}) and even at $2f_{\text{clk}}$. To support the speculation, post-layout simulations are done for a hundred clock cycles on a digital block including three 8-bit pipelined adders as well as the linear feedback shift registers for random input-signal generation. The time-domain simulation result as well as the normalized spectrum density are shown in Fig. 2.2. Because the inputs are pseudo-random this result is expected to reflect the generality.

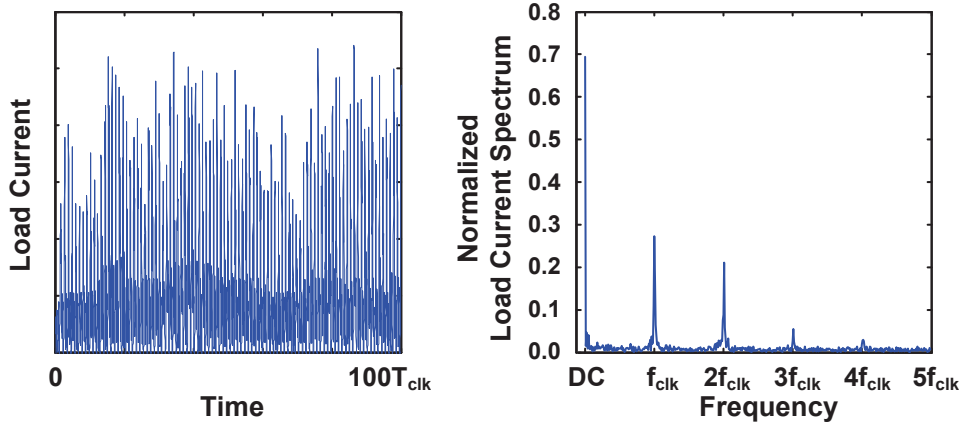


Figure 2.2: Spectrum density of the load current generated by clocked circuitry with pseudo-random inputs.

Accordingly, the LDO should provide high PSRR and low output impedance especially at those peak frequencies to achieve low total output voltage noise.

2.1.2 FVF-Based LDO Topologies

Flipped voltage follower, an enhanced source follower, is adopted as the output stage of the LDO in [16], showing a successful way to unite both low dropout voltage and fast load regulation response in one LDO structure. The FVF topology is shown in Fig. 2.3(a). The incoming line voltage and the output voltage supplied to the loading circuits are referred to as V_{in} and V_{out} , respectively; M_p is the pass transistor; the bias current source, I_{bias} , fixes the gate-source voltage (V_{gsc}) of M_c , the gate potential of which is set by a voltage source, V_{set} , making the source potential (i.e., V_{out}) fixed at $V_{set} + V_{gsc}$. On the aspect of the steady-state behavior, ideally speaking, V_{out} will not change as the load current (I_L) and only ‘follows’ the change of V_{set} . In reality, ΔI_L renders a considerable shift of V_{out} due to the finite feedback loop gain. As shown in Fig. 2.3(a), there are two signal loops in FVF LDO. The first one is well-known in source follower as ‘local’ feedback loop [15]. Comprised by

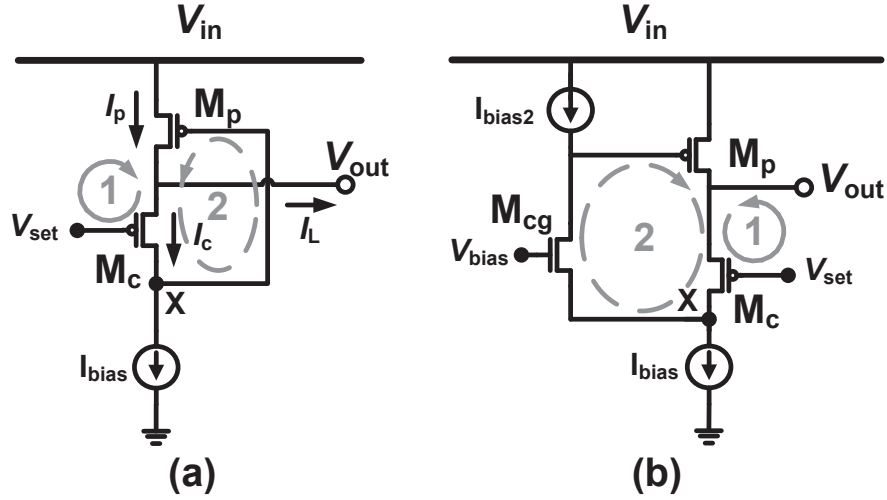


Figure 2.3: Topologies of the reported FVF-based LDOs [4, 8, 9].

merely one transistor, M_c , this loop instantly transforms ΔI_L -induced ΔV_{out} into in-phase variation of I_c so that ΔI_L is immediately compensated to some degree. Afterwards, the second loop takes over: ΔI_c is sensed by the current sensor comprised of M_c and I_{bias} that converts ΔI_c into variation of V_X , the voltage at node X; then ΔI_p is generated by ΔV_X through M_p , compensating for the rest major part of ΔI_L . Therefore, it is intuitive that ΔV_{out} is determined by DC gain of the second loop. Through this insight, the improved flipped voltage follower is developed as shown in Fig. 2.3(b). An additional amplifier in common-gate configuration is inserted in the second loop. Although the voltage gain at node X is not as much as in the original topology due to the smaller impedance introduced to node X by M_{cg} , the overall loop gain can still be enhanced by the additional amplifier.

On the aspect of the transient behavior, after ΔI_L occurs V_{out} often oscillates a bit before settling to the preset voltage. During oscillation, it is very likely that the voltage drop (or overshoot) of V_{out} is much larger than the steady-state ΔV_{out} if the load current increases (or decreases) abruptly. These transient behaviors are

not only dependent on the LDO’s steady-state characteristics, but the loop gain at high frequencies (or the bandwidth of the loop) is also a key factor. In this sense, the topology in Fig. 2.3(b) more or less lowers down the bandwidth of the second loop as it inserts into the loop one more node where the change of voltage takes some time. As a result, the transient behavior of this topology is not as good as the original one if the load is varying fast.

The basic concept of the proposed LDO is by building up multiple feedback loops in such a way that each loop is in charge of lowering the output impedance for a particular frequency range, the output impedance of the LDO at DC, low and high frequencies can all be taken care of with good synergy of these loops.

2.2 The Proposed LDO Topology and Circuit Implementation

In this section, the proposed LDO topology as well as its circuit implementation is presented. The advantages of the LDO, including fast load regulation, the area-efficient frequency compensation scheme, the high-frequency notches in the output impedance and the PSRR, and the impedance-oriented dynamic biasing scheme are discussed in detail.

2.2.1 The Proposed Multi-Loop LDO Topology

The proposed LDO topology is shown in Fig. 2.4 with the load circuits represented by a load capacitor, C_L and a current source, I_L . The FVF LDO shown in Fig. 2.3(a) is chosen as the output stage. M_c and M_{db} are together working as a current sensor. Note that the gate potential of M_p (V_X) can be as low as the saturation voltage ($V_{ds,sat}$) of M_{db} , allowing a smaller aspect ratio of M_p (w_p/l_p , equal to $1350\mu\text{m}/80\text{nm}$ in this implementation). Furthermore, it is desired but not necessary for the pass transistor to work in saturation region thanks to the wide dynamic range of V_X . As a result, the dropout voltage can be reduced so as to

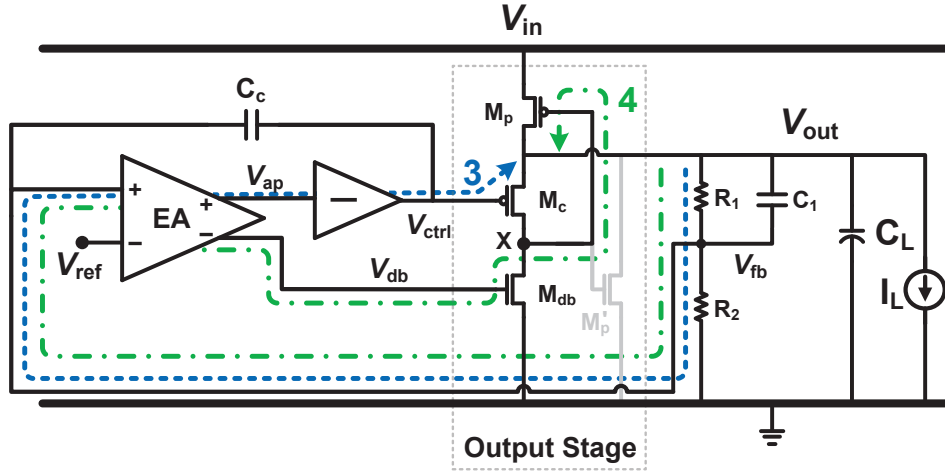


Figure 2.4: Topology of the proposed LDO.

improve power efficiency. A drawback for this type of FVF LDO is the limited input voltage range. When V_{in} jumps too high, the gate potential of M_p is forced to rise by the loop, but the output voltage is the upper limit. As a result, M_p cannot be turned off sufficiently and the LDO will lose ability to regulate the output. However, in the applications of the proposed LDO, as aforementioned, the switching converter preceding to LDOs can typically reduce the variations of the supply voltage of LDOs to tens of mV, which the FVF LDO is competent to handle.

The transistor, M'_p , is optional to the circuit. Its function is to further suppress the overshoot of V_{out} when a sudden drop of the load current happens. The trade-off for this spike-suppression circuit is when the load current is small it causes a considerable amount of extra quiescent current since its gate-source voltage is high in this case, whereas it consumes negligible extra power with large load. Regarding to this trade-off, in this implementation, the aspect ratio of M'_p is set to as low as 450nm/90nm.

The speeds of the two feedback loops in the output stage (as discussed in Section

2.1.2) are high whereas the loop gains of them are low. And hence these two loops are useful for providing timely response to fast variations of load current while the accuracy of the finally settled V_{out} (steady-state characteristics) is left for the high gain loop to handle, which is introduced as follows. The proposed LDO constructs a high gain feedback path (depicted as a dashed curve labeled with ‘3’ in Fig. 2.4) using a voltage sensor (comprised of two resistors, R_1 , R_2 and a capacitor, C_1), an error amplifier (EA) and an inverting amplifier (IA) to generate a controlling signal, V_{ctrl} , that dynamically adjusts the gate potential of M_c , instead of fixating it to V_{set} as is shown in Fig. 2.3. Due to concerns of quiescent power and stability, this feedback path is only supposed to provide high gain at DC and low frequencies, and hence its bandwidth is confined to a value much lower than that of the second path in the output stage.

Because this feedback path is in parallel to (instead of cascading) the second path, the loop gain of the whole LDO is, roughly speaking, the sum of the gain of each individual loop. With the help of the two-stage amplification, the whole loop gain at DC and lower frequencies is boosted compared with that of the FVF LDO. As a result, both steady-state line and load regulations are enhanced. And the whole loop gain at high frequencies is taken care of by the second loop thanks to its wide bandwidth. Detailed analysis will be given in the following sub-sections.

Another loop, the signal path of which is illustrated as the dash-dot curve labeled with ‘4’ in Fig. 2.4, is introduced for the active frequency compensation and dynamic biasing, which will be discussed in detail in the following sub-sections.

The full schematic of the proposed LDO is shown in Fig. 2.5 and the design parameters are listed in Table 2.1. The implementation of EA adopts complementary input devices, which not only has the EA’s transconductance enhanced but also makes it symmetric around the differential pairs’ equilibrium point so as to better

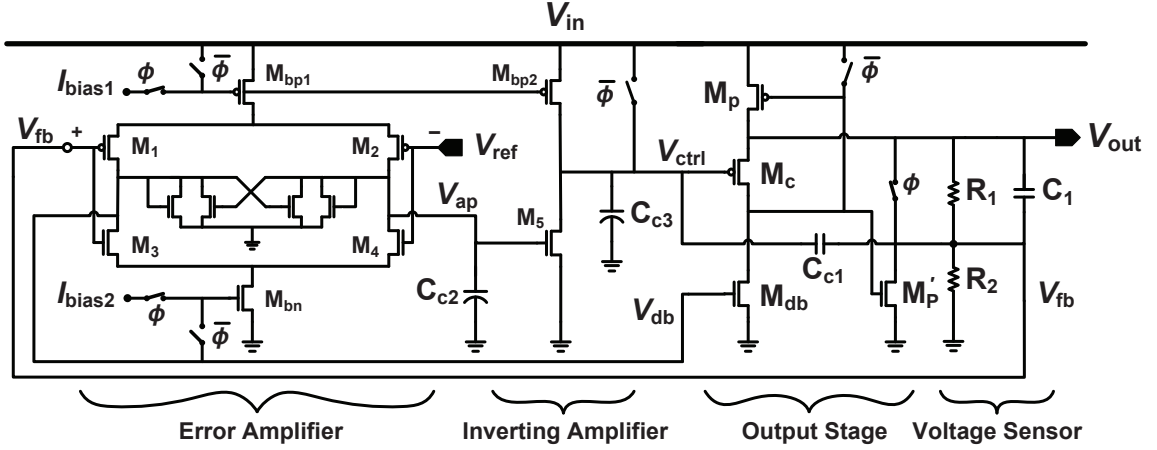


Figure 2.5: The full schematic of the proposed LDO.

Table 2.1: List of Parameters of Devices in Fig. 2.5

M_p :1350 μ m/80nm	M'_p : 450nm/90nm	M_c :80 μ m/80nm
M_{db} :27 μ m/80nm	$M_{1,2}$:12 μ m/160nm	$M_{3,4}$:960nm/160nm
M_5 :2.2 μ m/120nm	M_{dio} :740nm/160nm	M_{crs} :660nm/160nm
M_{bp1} :96 μ m/480nm	M_{bp2} :14.4 μ m/480nm	M_{bn} :6 μ m/480nm
C_{c1} :502fF	C_{c2} :651fF	C_{c3} :405fF
C_1 :207fF	R_1 :78K Ω	R_2 :122K Ω

* M_{dio} and M_{crs} represent the two identical diode-connected and cross-coupled load transistors in the EA, respectively.

suppress both the output voltage overshoots and droops. A weak positive feedback is employed in the EA's load to enhance the slew rate, while the IA is realized by a simple single-ended common-source amplifier. The seven switches are optional and their function is discussed in later sub-section.

2.2.2 Output Impedance

The output impedance of LDO determines the load regulation performance. The aforementioned loops contribute to output impedance in different frequency ranges and together can achieve a wide-range low output impedance, the analytical discus-

sion of which is given as follows.

The output impedance, $Z_{\text{out}}(s)$, of the proposed LDO including the load capacitance, C_L , is derived upon the small-signal circuit model shown in Fig. 2.6 with the meanings of symbols explained in the figure caption. The model construction detail is not imperative to the discussion of this sub-section and will be covered in the next sub-section. By closing the outermost loop (i.e., shorting $V_{\text{ol.in}}$ and V_{out} in Fig. 2.6 together), $Z_{\text{out}}(s)$ can be derived as

$$Z_{\text{out}}(s) \approx \frac{1}{[g_o(s) + sC_L][1 + H_{\text{ol}}(s)]}, \quad (2.1)$$

where the term $H_{\text{ol}}(s)$ is the open-loop transfer function from $V_{\text{ol.in}}(s)$ to $V_{\text{out}}(s)$ (or the loop gain of the proposed LDO); the term $[g_o(s) + sC_L]$ is the open-loop output impedance of the LDO (i.e., the equivalent output impedance of only the output stage loaded with decoupling capacitors) with $g_o(s)$ being the part contributed by the LDO's output stage (i.e., the conventional FVF LDO shown in Fig. 2.3(a)) and sC_L being the decoupling capacitor's contribution. The expression of $g_o(s)$ is given by

$$g_o(s) \approx g_{\text{mc}} + (g_{\text{mp}} + sC_{\text{gsp}}) \cdot A_{\text{CS}}(s) + g_{\text{dsp}}, \quad (2.2)$$

where $A_{\text{CS}}(s)$ is the gain of the current sensor in the output stage and is approximated by

$$A_{\text{CS}}(s) \approx \frac{g_{\text{mc}} + sC_{\text{gdp}}}{g_{\text{dsc}} + g_{\text{dsB}} + s(C_{\text{gsp}} + C_{\text{gdp}})}. \quad (2.3)$$

The first and second parts of $g_o(s)$ respectively reflect the helps of loop '1' and '2' in FVF, with the third part being the intrinsic output resistance. And the help of loop '3' is reflected from (2.1) which indicates that, compared with that of the FVF LDO, the closed-loop output impedance of the proposed LDO is improved by $[1 + H_{\text{ol}}(s)]$,

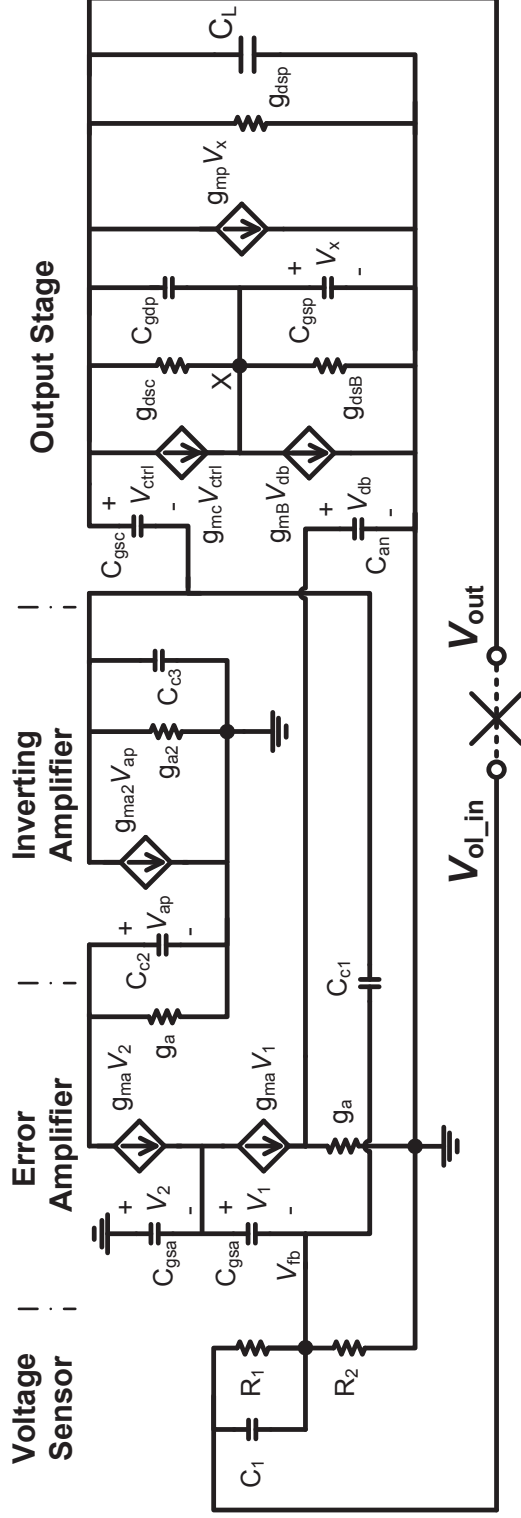


Figure 2.6: Small-signal model of the proposed LDO. (g_{mp} , g_{mc} , g_{mB} , g_{ma} , and g_{ma2} are the transconductances of M_p , M_c , M_{db} , the EA's input device, and the IA, respectively; g_{dsp} , g_{dsc} , and g_{dSB} represent the drain-source conductance of M_p , M_c , M_{db} with g_a and g_{a2} representing the equivalent output resistances of EA and IA, respectively; C_{gsx} and $C_{gd:x}$ are respectively the gate-to-source capacitance and gate-to-drain capacitance of the devices of the same device-name subscript x (e.g., C_{gsC} corresponds to M_c); C_{an} is the output capacitance at the negative output terminal of EA while C_{ci} is the i -th compensation capacitor ($i = 1 \dots 3$); R_1 , R_2 , C_1 and C_L are the same as those in Fig. 2.4.)

with $H_{ol}(s)$ given by

$$H_{ol}(s) = \frac{V_{out}(s)}{V_{ol,lin}(s)} \quad (2.4)$$

$$\approx \frac{K(s+z_1)\left(s^2+s\frac{\omega_{LCZ}}{Q_{LCZ}}+\omega_{LCZ}^2\right)}{(s+p_1)\left(s^2+s\frac{\omega_{LCP}}{Q_{LCP}}+\omega_{LCP}^2\right)\left(s^2+s\frac{\omega_{OS}}{Q_{OS}}+\omega_{OS}^2\right)}, \quad (2.5)$$

where the zeros and poles are to be elaborated in Section 2.2.3 and the factor K represents those ultra-high-frequency poles and zeros that are out of concern. In (2.5), the pole, p_1 , is the dominant pole, within which the amplitude of $H_{ol}(s)$ is dominated by the high low-frequency gain of the loop ‘3’, i.e., approximately

$$A_{EA_DC}A_{IA_DC}R_2/(R_1 + R_2). \quad (2.6)$$

Beyond p_1 , $|H_{ol}(s)|$ begins to roll off and even goes below 1 after the unity-gain frequency. Therefore, beyond the unity-gain frequency the improvement factor, $[1+H_{ol}(s)]$, approximately degrades to 1 and $Z_{out}(s)$ is roughly the same as that of FVF LDO. Now that the loop ‘4’ extends the bandwidth of $H_{ol}(s)$ by introducing a pair of left-half-plane complex zeros (LCZ), it also offers contribution in the frequency range from ω_{LCZ} (the frequency where the LCZ locates) to the unity-gain frequency, more discussion of which is presented in the following sub-section.

Therefore, each loop of the LDO takes part in improving $Z_{out}(s)$ within a certain frequency range and the first three loops can be independently tuned to achieve a specific output impedance while the fourth loop, along with the compensation capacitors, tackles the stability problem of the whole LDO.

2.2.3 Stability Analysis and Frequency Compensation

The proposed LDO has four extra poles to be concerned in addition to the two poles in the original FVF topology. Three of the extra poles are introduced by the high-gain loop: one in the voltage sensor and the other two in the two-stage amplification circuits (i.e., EA and IA). And the last one is related to the fourth loop. Therefore, its stability is not automatically guaranteed and careful analysis as well as proper frequency compensation is needed to make sure the synergy among those loops is stable. The stability analysis is conducted upon the aforementioned small-signal model as well through inspecting the zeros and poles of the open-loop transfer function (or the loop gain), H_{ol} . Derived from the circuit shown in Fig. 2.4, the model applies the first-level MOSFET AC model to those transistors in Fig. 2.4 and adopts well-known small-signal models [15] of the typical fully differential amplifier and the common-source amplifier for the EA and IA, respectively. Note that $V_{ol,in}$ here is not the supply voltage of the LDO which is treated as AC ground, but is a virtual input voltage for deriving the open-loop transfer function. In the state-of-the-art technologies, the parasitic capacitances of the devices are small, thus the poles are somewhat closed to each other before compensation. As a result, the frequency response of the loop gain without compensation would be very likely to have negative phase margin that indicates closed-loop instability, which is illustrated by the dashed line in Fig. 2.7(a) and by the simulation result in Fig. 2.8(a).

To solve this problem the pole-splitting frequency compensation schemes are conventionally used such as the Miller compensation techniques. However, with a large load capacitance, depending merely on Miller compensation techniques may result in multiple large compensation capacitors that take up much silicon area; also, doing so tends to achieve small loop bandwidth (i.e., low unity-gain frequency). This work

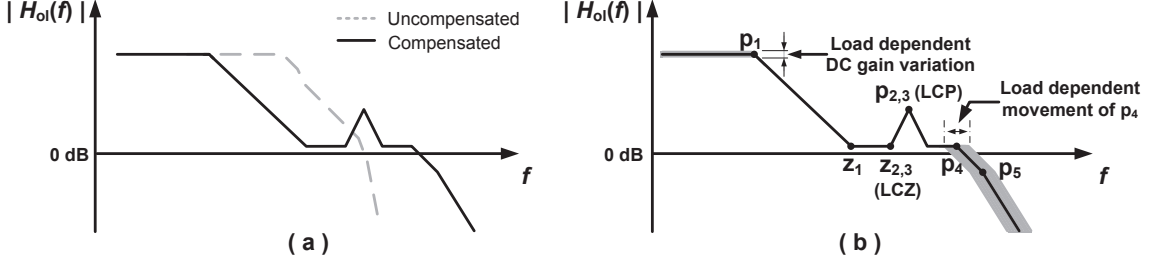


Figure 2.7: Illustrative Bode plot of the loop gain of the proposed LDO. (a) Before and after compensation. (b) With different load currents.

introduces an active bypass across the IA and M_c , which starts from the negative output of EA through the active element M_{db} and then ends at node X. Through a complicated derivation, it is proven that this bypass can generate a pair of left-half-plane complex zeros (LCZ) that improves the phase margin of H_{ol} . In addition, a pole-splitting technique is also adopted by connecting the input of EA and the output of IA with the capacitor C_{c1} , which makes the dominant pole far from the rest of the poles. Also, by using C_{c1} , the two poles related to the EA and IA are transformed from real poles before compensation into a pair of complex poles (referred to as LCP in the following discussion), the function of which is discussed later. Another two grounded capacitors, namely C_{c2} and C_{c3} , are simply connected to the outputs of EA and IA, respectively, to affect the positions of the LCZ and LCP. Relative locations of major poles and zeros are illustrated in Fig. 2.7(b). The following pole/zero analysis is unraveled in the sequence of frequency. The lowest-frequency pole (i.e., the dominant pole), p_1 , lowered by the pole-splitting technique, is given by

$$p_1 \approx -\frac{2g_a g_{a2}}{g_{ma} g_{ma2} R_p C_{c1}} = -\frac{1}{R_p (A_{EA,DC} A_{IA,DC} C_{c1})}, \quad (2.7)$$

where R_p , the equivalent resistance at the input node of EA, is equal to $R_1 R_2 / (R_1 +$

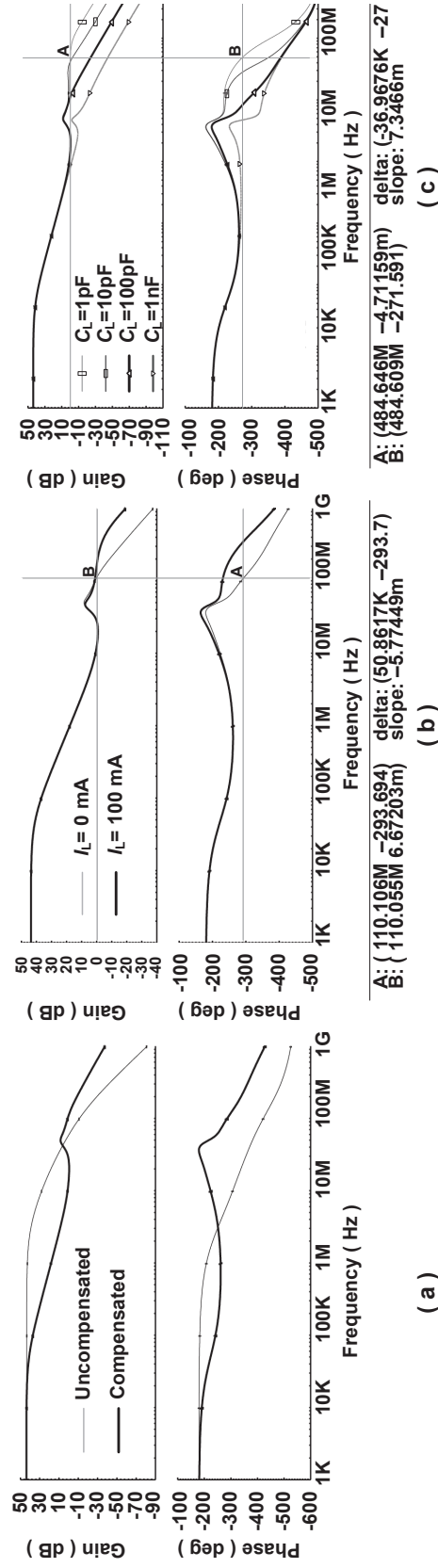


Figure 2.8: Simulated Bode plots of the proposed LDO. (a) Before and after compensation. (b) Different load currents ($C_L = 100$ pF). (c) Different C_L ($I_L = 0$ A).

R_2) and $(A_{EA_DC}A_{IA_DC}C_{c1})$ is the equivalent capacitance at that node given by the Miller effect on C_{c1} . After p_1 , the magnitude of H_{ol} starts to roll off with the slope of -20dB/dec; the phase is also dropping.

Next to p_1 , the lowest zero is from the capacitive bypass of R_1 and can be expressed as

$$z_1 = -\frac{1}{R_1C_1}. \quad (2.8)$$

This zero counteracts the influence of p_1 on the Bode plot, leveling off the magnitude of H_{ol} and tending to pull the phase back.

The LCZ, namely z_2 and z_3 in Fig. 2.7(b), are designed to the frequencies higher than z_1 , which not only uplift the magnitude of H_{ol} but also dramatically pull the phase up as shown in the simulation results in Fig. 2.8. It is good for the Q-factor of this LCZ to be high because the higher it is, the more the phase gets pulled up and hence the better the phase margin. The location of the high-Q LCZ on the Bode plot, as well as the Q-factor, is given approximately by

$$\begin{aligned} \omega_{LCZ} &\approx \sqrt{\frac{g_{mc}g_{ma2}g_a}{g_{mB}C_{c2}(C_{c1} + C_{c3})}}, \\ Q_{LCZ} &\approx \frac{\sqrt{g_{mc}g_{mB}g_{ma2}g_aC_{c2}(C_{c1} + C_{c3})}}{g_{mB}[C_{c2}g_{a2} + (C_{c1} + C_{c3})g_a]}. \end{aligned} \quad (2.9)$$

It is inferred from (2.9) that this pair of zeros are introduced by the active bypass in the fourth loop because if without the loop, i.e., connecting the gate of M_{db} to a fixed voltage bias instead of the negative output port of the EA, then it is equivalent to setting g_{mB} to zero. Then according to (2.9), the LCZ should locate at infinitely high frequency. Thanks to this active bypass, the resultant LCZ not only pulls up the phase but also extends the loop bandwidth as it elevates the magnitude too. However, these two aspects alone cannot guarantee a good phase margin because

over-extending the bandwidth can possibly lead to pushing the unity-gain frequency up to a certain frequency where the phase is already severely deteriorated by high-frequency poles, such as p_4 and p_5 . As a matter of fact, designing this LCZ should coordinate with the design of z_1 in the manner that z_1 should not be too close to p_1 in order to make $|H_{oi}|$ at LCZ low enough to keep the extension of bandwidth from exceeding the ‘stability-safe’ region.

The LCP introduced by C_{c1} is designed right higher than the LCZ, so that a magnitude peak happens in the Bode plot at the frequency approximately where the LCP locates. If the distance between the LCZ and the LCP is far enough, this peak can be high and cause an obvious notch in the magnitude of the output impedance of the LDO. As is motivated by the scenarios discussed in Section 2.1.1, this notch is helpful for better reduction of the LDO’s output voltage variation, and will be discussed and demonstrated with simulation results later. The frequency of the peak can be approximated by

$$\omega_{\text{peak}} \approx \omega_{\text{LCP}} \approx \sqrt{\frac{g_{ma}g_{ma2}}{2C_{c2}(C_1 + C_{c3} + C_1C_{c3}/C_{c1})}}, \quad (2.10)$$

with the Q-factor given by

$$Q_{\text{LCP}} \approx \frac{R_p \sqrt{0.5g_{ma}g_{ma2}C_{c1}C_{c2}(C_1C_{c1} + C_1C_{c3} + C_{c1}C_{c3})}}{C_{c2}(C_{c1} + C_{c3})}. \quad (2.11)$$

Regarding to (2.8)–(2.11), the passive parameters involved in the frequency compensation are C_{c1} , C_{c2} , C_{c3} as well as C_1 . By tuning them, the stability problem of the LDO can be solved without changing the loops’ DC gains as well as the quiescent current which are designed prior to the compensation. Note that the leveling-off of the magnitude after the pair of LCP is not caused by a zero since, at frequencies below p_4 , the number of poles is three and so is that of the zeros. The illustrative

magnitude curve in Fig. 2.7(b) should have leveled off right after $p_{2,3}$. However, the authors deliberately drew a peak at $p_{2,3}$ in an attempt to demonstrate the magnitude peaking caused by relatively high Q of the complex poles.

The two poles, p_4 and p_5 , contributed by the output stage, are the roots of the factor $(s^2 + s\omega_{OS}/Q_{OS} + \omega_{OS}^2)$ in the denominator of H_{ol} , which are apparently load-dependent. Under different load current conditions as well as different amounts of load capacitance, these two poles vary over a wide range, as illustrated by Fig. 2.7(b). The expressions for ω_{OS} and Q_{OS} , given by (2.12), are derived under the condition that $C_L \gg C_{gsp}, C_{gdp}$ and C_{gsc} . As (2.12) indicates, when C_L increases, p_4 and p_5 , moving down towards the low frequency region, will decrease the bandwidth, and vice versa. On the other hand, when I_L increases, g_{mp} and g_{dsp} get dominantly large, thus p_4 and p_5 will rise to higher frequencies. When I_L is as low as close to zero, in which case both g_{mp} and g_{dsp} will dramatically decrease, the bias current of the output stage will be boosted by the dynamic biasing technique (discussed in Section 2.2.5) to keep g_{mp} sustaining a relatively large value. Additionally, g_{mc} is also helpful to alleviate the down-move of the poles.

$$\begin{aligned} \omega_{OS} &\approx \sqrt{\frac{g_{mp}(g_{mc} + g_{dsc}) + g_{dsp}(g_{dsB} + g_{dsc})}{C_L(C_{gsp} + C_{gdp})}}, \\ Q_{OS} &\approx \frac{\sqrt{[g_{mp}(g_{mc} + g_{dsc}) + g_{dsp}(g_{dsc} + g_{dsB})]C_L(C_{gsp} + C_{gdp})}}{g_{mp}C_{gdp} + (g_{dsc} + g_{dsB})C_L + g_{dsp}(C_{gsp} + C_{gdp})}. \end{aligned} \quad (2.12)$$

The analysis above are supported by the simulated Bode plots in Fig. 2.8(b) and (c). It can be read from the cursors that when I_L is 0mA the phase of H_{ol} at the unity-gain frequency is about -293.7 degrees. Considering the initial phase shift of 180 degrees at DC, the phase margin is 66.3 degrees and is 116.9 degrees at 100-mA load condition. In Fig. 2.8(c), with the amounts of load capacitance ranging from 1pF to 1nF at the step of one decade, the phase margins are accordingly 88.8, 32,

66.3 and 94.3 degrees, respectively. It has to be mentioned that although when C_L is 10pF the phase margin is not so good as the other cases, yet the LDO is still stable in this case. Furthermore, those good phase margins when C_L is equal to the other values indicate that once C_L is somehow defined for sure to be around 10pF, the LDO can be re-designed specifically to achieve better phase margin.

2.2.4 Notch in the Output Impedance and PSRR

As aforementioned, the magnitude of H_{ol} has a high peak within the unity-gain frequency. Regarding to (2.1), it can be inferred that Z_{out} has a corresponding notch at the peak frequency. The width and depth of the notch are related to the distance between LCZ and LCP as well as the Q-factors of them. As indicated by (2.9) and (2.11), increasing C_{c1} can enlarge the distance between LCZ and LCP and hence make the notch deeper, so that the LDO can achieve better suppression of load-induced noise around ω_{peak} .

It is also worth to mention that, indicated by (2.9)–(2.11), this notch is almost immune to load current conditions because the parameters in those equations vary little with I_L . And hence, although those equations are derived from the small-signal model, this notch works even when the load current variation is large.

However, there is a trade-off that larger C_{c1} , which is good for the notch, can lead to lower p_1 according to (2.7), which degrades Z_{out} within the frequency range from p_1 to the frequency where the notch starts. As a result, for those load current variations with its power spectrum density dominantly clustering around a certain frequency besides DC, say, the local clock frequency as discussed in Section 2.1.1, C_{c1} should be designed to create a deep impedance notch at that frequency; for those with its power spectrum density widely spreading out (e.g., a step waveform), C_{c1} should be as small as possible on the premise of achieving good enough phase

margin. Therefore, for regulation of a local block, the circuits in which are probably well-synchronized with little skew, larger C_{c1} is preferable; for regulation of those unsynchronized blocks or those tolerant of large clock skews, smaller C_{c1} is better.

To increase the flexibility for this notch, programmable capacitors can even be applied to C_{c1} and the other two compensation capacitors to enable digital tunability of Z_{out} . So that if the prediction of the characteristics of upcoming load currents is available [18] (which is not in the scope of this work though), the predictor can send digital tuning signals to set the LDOs with the most suitable output impedances.

The PSRR of the LDO is also possible to possess this notch depending on the forward path from the power supply to the output of the LDO. There are four such paths in the proposed topology: through the bandgap reference input, through the bias circuits of the EA and IA, and through the pass transistor. Since the noise through the former three paths can be effectively suppressed by the PSRRs of the bandgap, the EA and IA, respectively, and further get filtered by C_{c2} and C_{c3} before it gets to the output node, the major part of supply noise comes from the pass transistor path. In this sense, the PSRR can be approximated as

$$PSRR \approx \frac{g_{mp} + g_{dsp}}{[g_o(s) + sC_L][1 + H_{o1}(s)]}. \quad (2.13)$$

Since the nominator in (2.13) has no additional zero, the H_{o1} can also introduce the same notch into PSRR as in Z_{out} (in fact, there is an zero introduced by C_{gsp} which is omitted because it lies far higher than the notch frequency and hence has little impact on the appearance of the notch). Similarly, by aligning the PSRR notch to the switching frequency of the preceding switching converter, the LDO can achieve better suppression of the supply voltage ripple. The benefits of the notch are to be demonstrated by simulation results in Section 2.3.

2.2.5 Dynamic Biasing

Dynamic biasing techniques are beneficial in power saving because the bias current of LDO is adaptively adjusted so as to let quiescent power be consumed only when it is needed. Traditional dynamic biasing techniques lay more emphasize on power-efficiency. In [11], the quiescent current is adjusted in phase with I_L variation. In this case, when I_L is large, the bandwidth of LDO is extended by consuming additional power to reduce the output impedance at high frequencies; on the other extreme, I_q is largely reduced in order to maintain good power efficiency, but degrades Z_{out} . Different from this scheme, the proposed dynamic biasing scheme is carried out on two levels. The first level, termed LDO local dynamic biasing, gives priority to improving Z_{out} when the load circuits operate in normal mode.

As indicated by (2.1) and (2.2), the major change of Z_{out} due to the change of I_L comes from Δg_{mp} and Δg_{dsp} . When on the light-load condition, little current flows through M_p and M_p works in the sub-threshold region with very small g_{mp} and g_{dsp} ; when on heavy-load condition, M_p works in saturation region or even in linear region and hence g_{mp} and g_{dsp} will increase by several orders of magnitude. Although $A_{CS}(s)$ and H_{ol} can be somewhat higher at light-load conditions, its impact is overwhelmed by that of g_{mp} and g_{dsp} variations. Consequently, Z_{out} will increase significantly as I_L decreases. Since the worst Z_{out} happens on light-load conditions, when I_L is low the proposed scheme allocates larger bias current to the output stage to increase g_{mp} and g_{dsp} and eventually lower Z_{out} . Also, it is well-known that Z_{out} is worse at high frequencies than at low frequencies. Therefore, this dynamic biasing scheme, which only changes bias current of the output stage, targets on improvement of Z_{out} at high frequencies instead of that at low frequencies which is already good enough.

Based on this concept, the circuit for the local dynamic biasing is realized with

transistors, namely M_{db} and M_c in Fig. 2.4. That is, only the quiescent current of the output stage is dynamically adjusted since the output stage has a significant impact on high-frequency output impedance. Assume the output voltage has settled to the preset value. Then M_c and M_{db} are working like a typical class AB amplifier as illustrated in Fig. 2.9(a). The relationship between the quiescent current (I_{bias}) and the input voltage (V_{set}) of a class AB amplifier is illustrated in Fig. 2.9(b), which is analogous to the relationship between the bias current of M_{db} and V_{db} . Then the dynamic biasing of the output stage is as follows. When I_L is 0A, the output stage is biased at some point near and on the right-hand side of the peak (illustrated as the solid line in Fig. 2.9(b)). As I_L goes up, V_{out} drops which triggers the loop ‘4’ as well as loop ‘3’ to pull up V_{db} and V_{ctrl} . Thus, the bias point of M_c and M_{db} is moving down along the direction shown as the dashed arrowed line in Fig. 2.9(b). Fig. 2.9(c) gives the simulation result of the total bias current (I_q) of the proposed LDO versus I_L . By dynamic biasing, I_q is about 408 μ A on average.

By now this scheme only works when the load circuits operate in normal mode, i.e., the activity of load circuits is high. In this scenario, the load current of the LDO is switching frequently and for most of time stays above, say, 30% of the rated maximum load current. Hence, the moment when I_L drops to zero lasts very shortly and the negative impact of the proposed ”reverse” dynamic biasing on light-load power efficiency is negligible, however, the benefit on reducing undershoot of V_{out} is large as to be shown in Section 2.3.1.

For the case that the load circuits are idle or in ”sleep” mode, to improve the power efficiency, the power gating concept is borrowed as the second level dynamic biasing. Consider the system shown in Fig. 2.1 where several LDOs are together regulating one power domain, several power switches are introduced into the topology as shown in Fig. 2.5 for turning on and off the LDO. When the load circuits enter

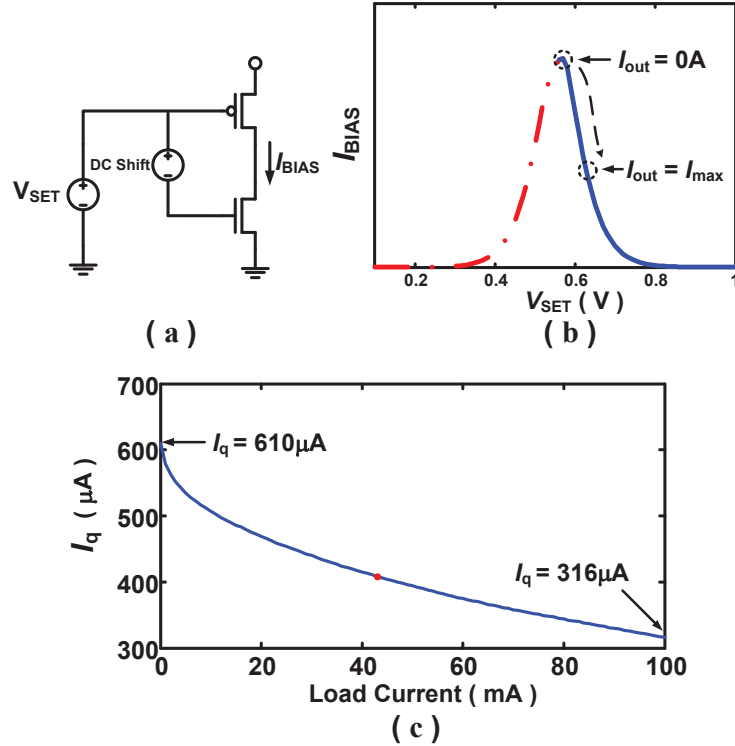


Figure 2.9: Relationship of quiescent current and input voltage of class AB amplifier. (a) Schematic of a simplified class AB amplifier. (b) I-V curve. (c) The simulation results of the total bias current for the LDO vs. load current.

the "sleep" mode, most of the LDOs can be turned off while leaving only one or two LDOs on to maintain the rated output voltage so as to enhance power efficiency. Note that in the 90nm CMOS technology we used, the decap leakage current is about $1.9\mu A/pF$. As a result, if for example a power domain has decap of 10nF, the leakage current by decap will be about 19mA, let alone to say the subthreshold leakage of MOSFETs. Therefore, even in sleep mode, due to leakage current from load circuits or decaps, the regulators can still achieve good power efficiency. And to the authors knowledge, the state-of-the art CPUs, like the Intel Nehalem series, are already supporting on-chip power gating. we expect no daunting difficulty in implementing load activity monitoring and digital control of those switches.

2.3 Simulation Results

The LDO shown in Fig. 2.5 is designed in a commercial 90-nm CMOS technology. The circuit simulations are done using BSIM4 model obtained from the foundry. The nominal input and output voltages are 1.2V and 1V, respectively, and the rated maximum load current, I_{\max} , is 100mA. For the decoupling capacitance (decap) up to 1nF, the total amount of compensation capacitors, namely C_{c1} , C_{c2} and C_{c3} , is 1.6pF; if the bypassing capacitor, C_1 , in the voltage sensor is counted in, it is up to 1.8pF in total, occupying an area of about 0.003mm².

2.3.1 Load Regulation

Fig. 2.10 shows the transient responses to load currents jumping between 0mA and 100mA with different rise times. For steady-state characteristics, the settled output voltage of 1.00026V at zero load drops to 999.92mV at the load of 100mA, achieving an ultra-high load regulation accuracy of 0.003mV/mA. The maximum voltage drop for 10-ns rise time of I_L with 100-pF decap is about 43mV; it increases to 113mV and 122mV for 1-ns and 100-ps rise time, respectively, with 600-pF decap. Furthermore, the transient responses to load current stepping between 1mA and 101mA are also simulated. This is in an attempt to emulate more realistic situations in which the decap leakage current as well as sub-threshold leakage of MOSFETs in the load circuits contributes a considerable amount of load current, say, 1mA in this case. The step size remains 100mA for comparison purpose. The results are shown in Fig. 2.11, Fig. 2.12 and Fig. 2.13 for the above three types of load transients respectively. It is shown that the maximum voltage drops are 28mV, 88mV and 95mV, respectively. The improvements are brought by the smaller output impedance at the starting point of the load current step. For example, at the beginning of the I_L step in the first case (i.e., $I_L=0$ mA), Z_{out} at DC is about 109m Ω , while it is 32m Ω

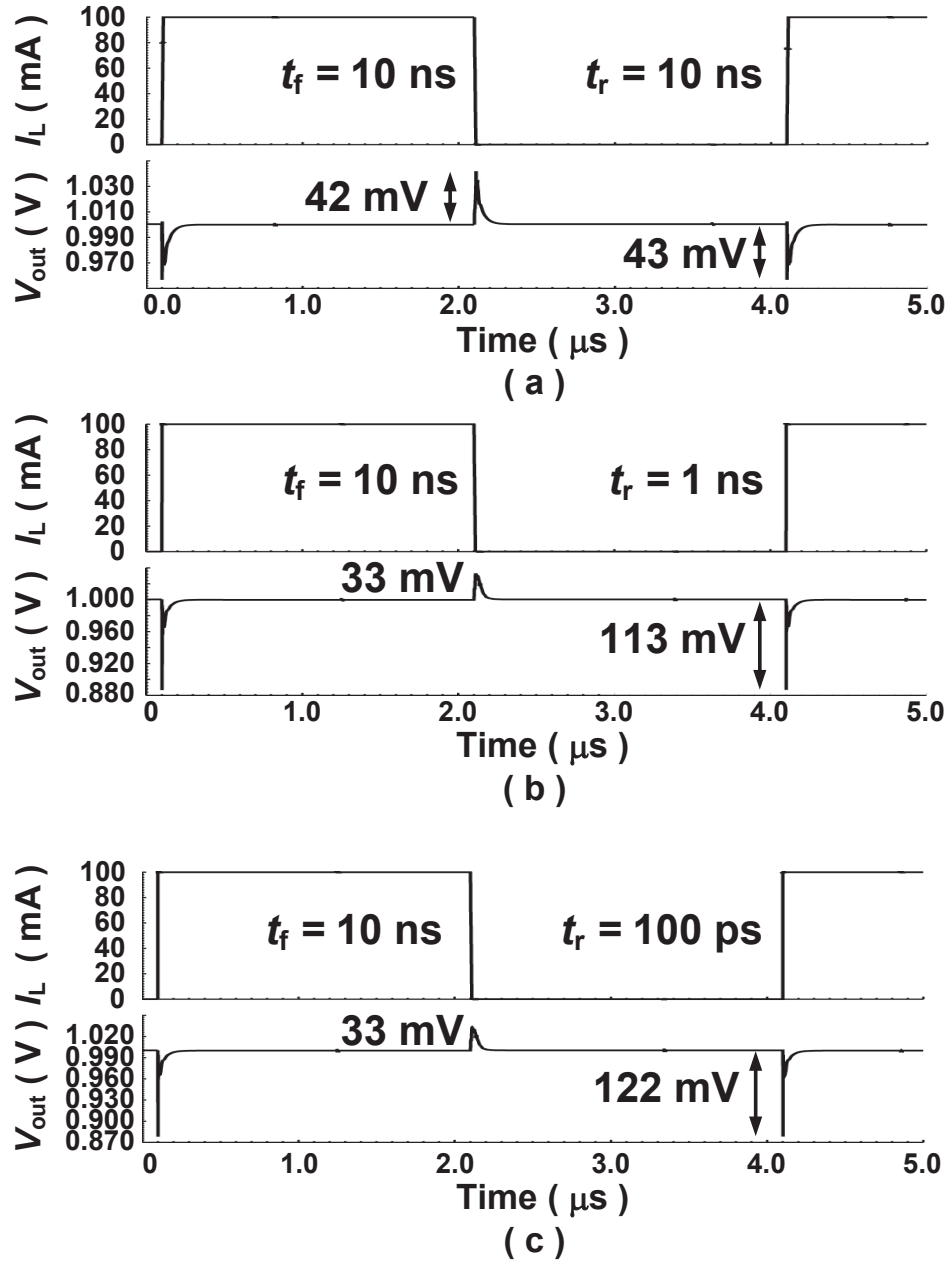
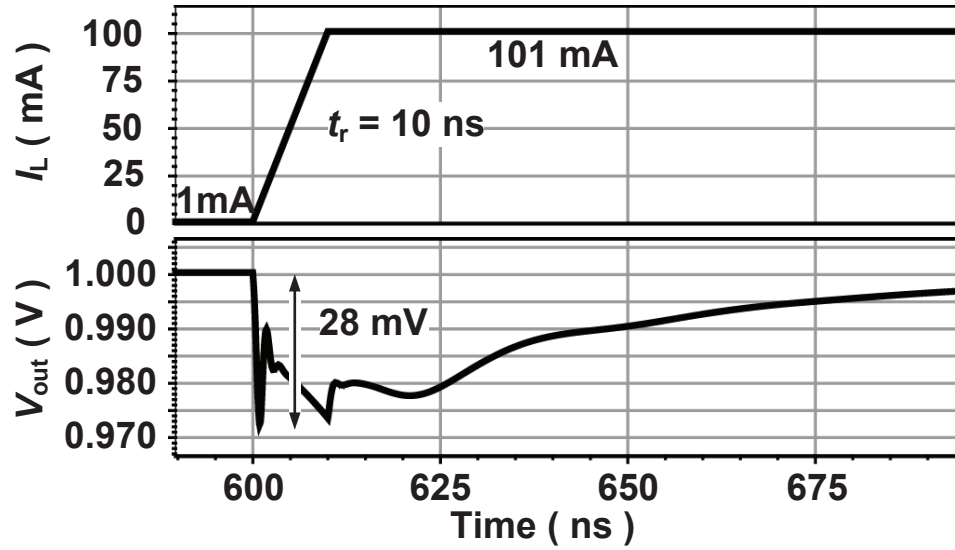
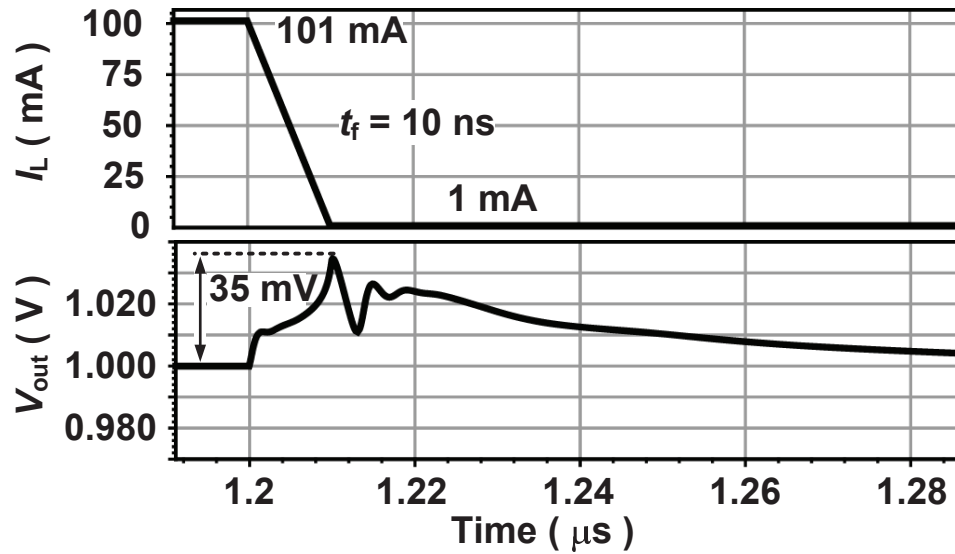


Figure 2.10: Transient responses to load current stepping between 0mA and 100mA. (a) $t_r = 10\text{ns}$, $C_L=100\text{pF}$. (b) $t_r = 1\text{ns}$, $C_L=600\text{pF}$. (c) $t_r = 100\text{ps}$, $C_L=600\text{pF}$.

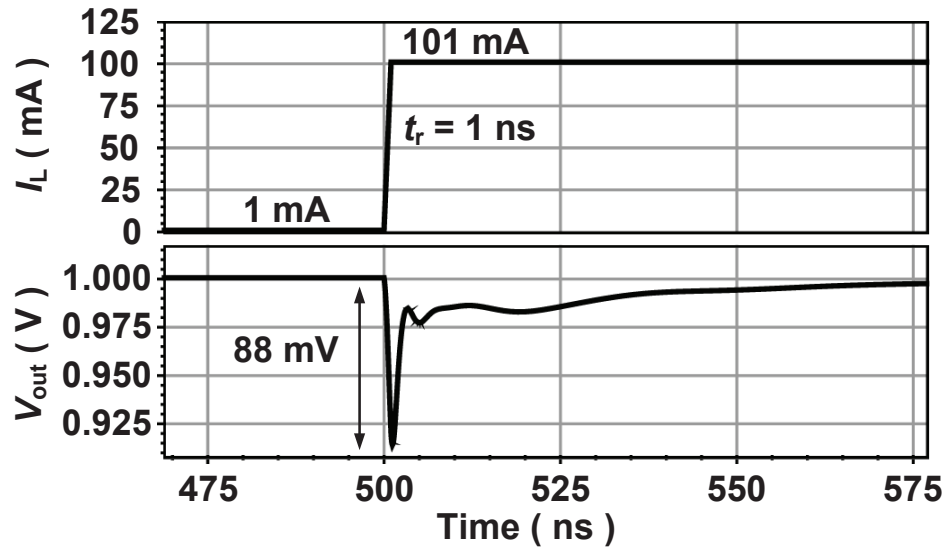


(a)

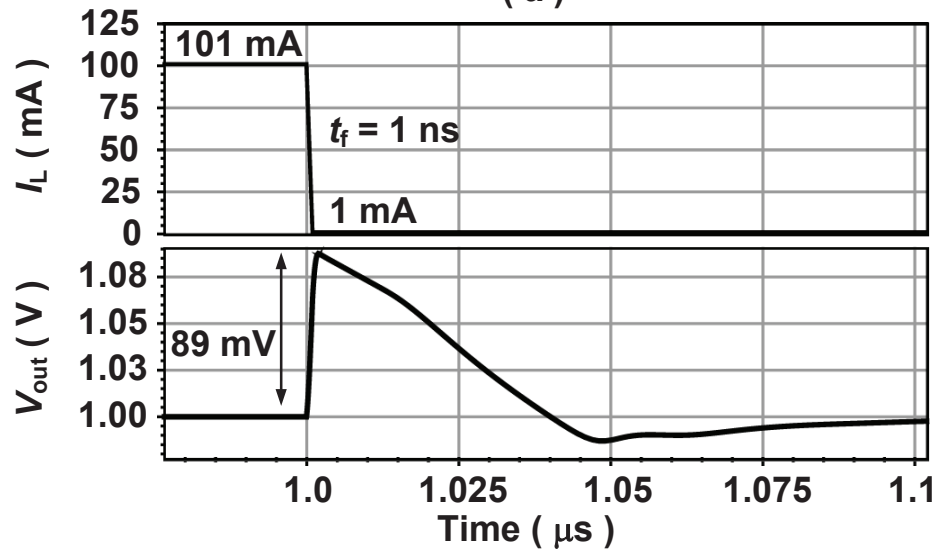


(b)

Figure 2.11: Transient responses to load current stepping between 1mA and 101mA with the transition time of 10ns and $C_L=100$ pF.

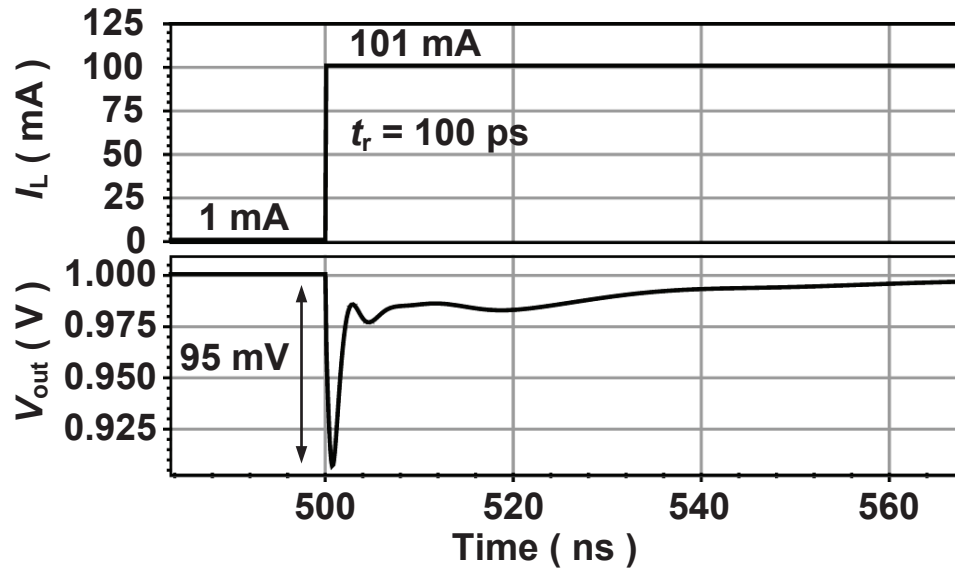


(a)

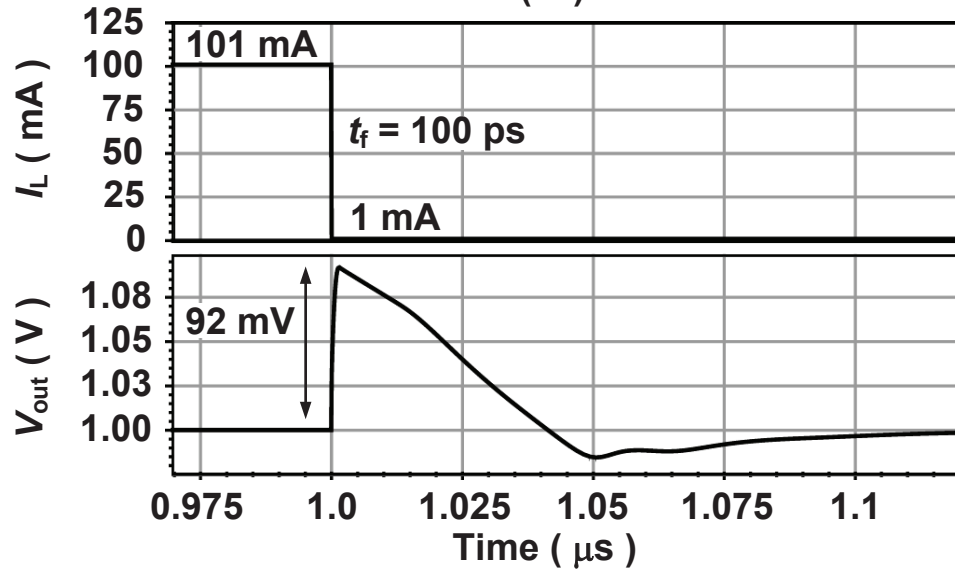


(b)

Figure 2.12: Transient responses to load current stepping between 1mA and 101mA with the transition time of 1ns and $C_L=600$ pF.



(a)



(b)

Figure 2.13: Transient responses to load current stepping between 1mA and 101mA with the transition time of 0.1ns and $C_L=600$ pF.

in the second case (i.e., $I_L=1\text{mA}$). This also verifies the concept of the proposed dynamic biasing scheme that Z_{out} on the minimum-load condition is to be reduced in order to reduce the maximum voltage drop.

2.3.2 Robustness to Process, Temperature and Mismatches

To capture the sensitivity of the proposed LDO to the process variations as well as device mismatches, a 1000-sample Monte Carlo simulation is conducted using process and mismatch models the foundry provides. The steady-state characteristics at $I_L=0\text{A}$ and $I_L=100\text{mA}$ are shown in Fig. 2.14(a) and (b), respectively. The mean and standard deviation of the steady-state output voltage at $I_L=0\text{A}$ are 1.0005V and 2.5mV respectively, with those at $I_L=100\text{mA}$ being 1.00005V and 2.4mV respectively. The average bias currents of the 1000 samples are also calculated from the simulation data with the mean of $424\mu\text{A}$ and the standard deviation of $72.6\mu\text{A}$. The binned result is shown in Fig. 2.14(c) with the number of samples in each bin labeled right above the bin. The transient load characteristics is simulated with the load current stepping between 1mA and 101mA within 100ps and with 600-pF decap. The results are shown in Fig. 2.14(d) and (e). The mean and standard deviation of the maximum voltage drop are 92.5mV and 7.4mV respectively, with those of the output voltage overshoot being 90.9mV and 8.3mV respectively. Of the 1000 samples there are only 50 samples, either voltage drops or overshoots of which exceed 10% (i.e., a conventional specification for LDO's transient load regulation performance) of the rated output voltage.

In reality, the LDO is supposed to work under various temperatures. A sweep simulation on temperature is also conducted to verify the LDO's performance under the temperature ranging from -40°C to 85°C . The result shown in Fig. 2.15(a) shows that the variations of the steady-state output voltage for both $I_L=0\text{A}$ and $I_L=100\text{mA}$

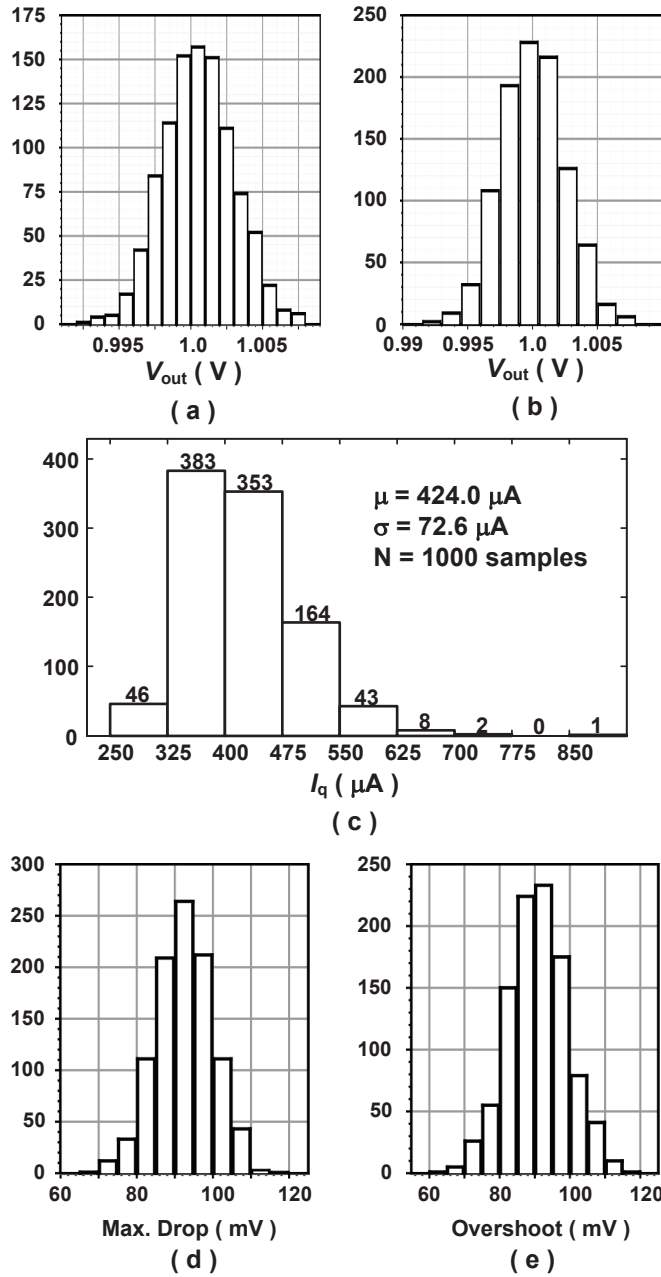


Figure 2.14: Monte Carlo simulation results (1000 samples) at $C_L = 600\text{pF}$. (a) Steady-state output voltage at $I_L = 0\text{A}$. (b) Steady-state output voltage at $I_L = 0.1\text{A}$. (c) The average biasing current of the LDO. (d) The maximum voltage drop when load current is switching between 1mA and 101mA within 100ps with $C_L = 600\text{pF}$. (e) The voltage overshoot when load current is switching between 1mA and 101mA within 100ps with $C_L = 600\text{pF}$.

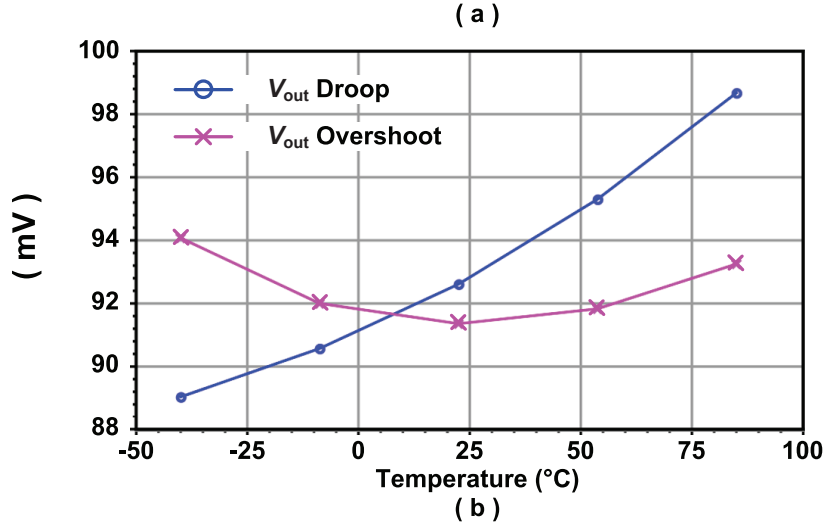
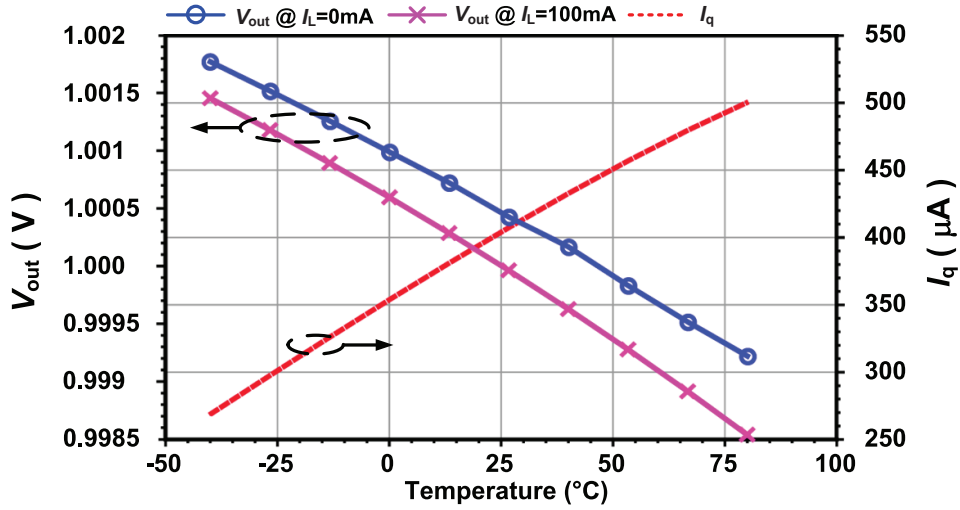


Figure 2.15: Temperature-sweep simulation results. (a) Steady-state output voltage. (b) Transient load regulation with load current switching between 1mA and 101mA within 100ps and $C_L = 600pF$.

are confined within 3mV; it also demonstrates that the quiescent current will increase with temperature at the slope of roughly $1.9\mu A/^\circ C$. And the maximum voltage drop and overshoot vary within the ranges from 89mV to 98.6mV and from 91.5mV to 94mV, respectively. And hence, the LDO can still meet the specifications under a wide range of temperature.

2.3.3 Line Regulation

The transient line responses are also simulated for a 1.15–1.25-V input voltage step with both 1-ns and 1- μ s transition times under load conditions of $I_L=0$ mA and $I_L=100$ mA. For steady-state characteristics, the settled output voltage is jumping between 1.00028V ($V_{in}=1.25$ V) and 1.00071V ($V_{in}=1.15$ V), achieving line regulation accuracy of 4.3mV/V. The transient output voltage variation with 1-ns V_{in} transition time is less than 13mV when $I_L=0$ mA and less than 55mV when $I_L=100$ mA as shown in Fig. 2.16(a), while the variation with 1- μ s V_{in} transition time is less than 0.7mV when $I_L=0$ mA and less than 1.3mV when $I_L = 100$ mA as shown in Fig. 2.16(b).

2.3.4 Comparison with the Antetypes

To better demonstrate the evolution of the proposed LDO from the basic FVF topologies, the original FVF-based LDO as shown in Fig. 2.3(a) and an improved one shown in Fig. 2.3(b) (which is the base of the LDO in [8]) are re-designed in 90-nm technology with their quiescent currents approximately the same as the proposed one in an attempt to perform a valid comparison. Only 1-pF load capacitance mimicking power line (without decoupling) parasitic capacitance is attached to each output node of the three LDOs, so as to compare these LDOs' load regulation performances without the help of decaps.

On the aspect of AC characteristics, Fig. 2.17 shows the output impedances of these three LDOs on the load conditions of both $I_L=1$ mA and $I_L=100$ mA. Z_{out} of the LDO in [8] (LDO2) at frequencies from DC to 1MHz is lower than that of the original topology (LDO1) as discussed in Section 2.1.2, and the proposed LDO (LDO3) achieves the lowest Z_{out} in this frequency range, about 15–20-dB lower than that of LDO2. At frequencies from 10MHz to hundreds of MHz, Z_{out} of LDO2 degrades more quickly with frequency than LDO1 because of LDO1's wider bandwidth.

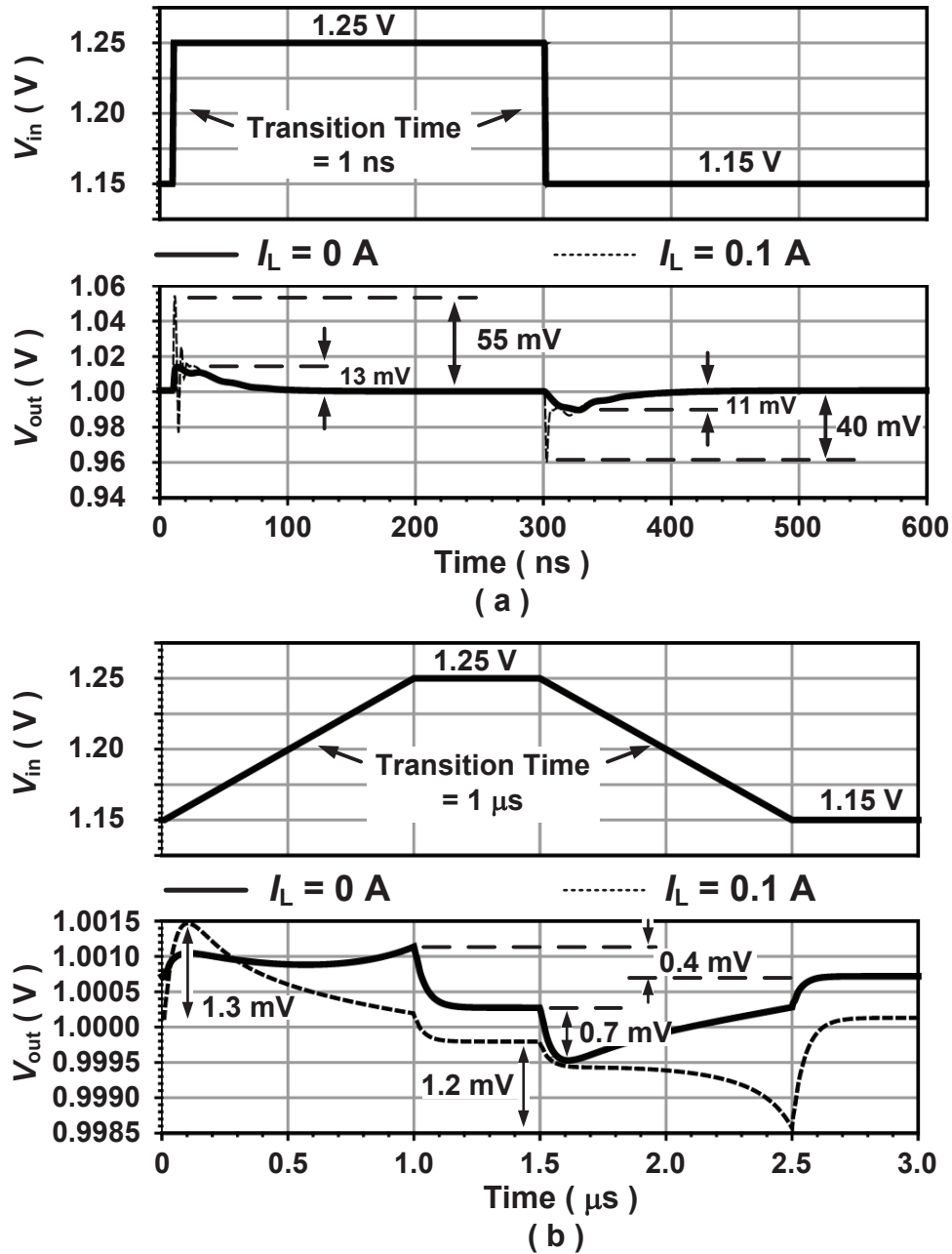


Figure 2.16: Transient responses to input voltage steps. (a) 1-ns V_{in} transition time with $C_L = 600\text{pF}$. (b) 1- μs V_{in} transition time with $C_L = 100\text{pF}$.

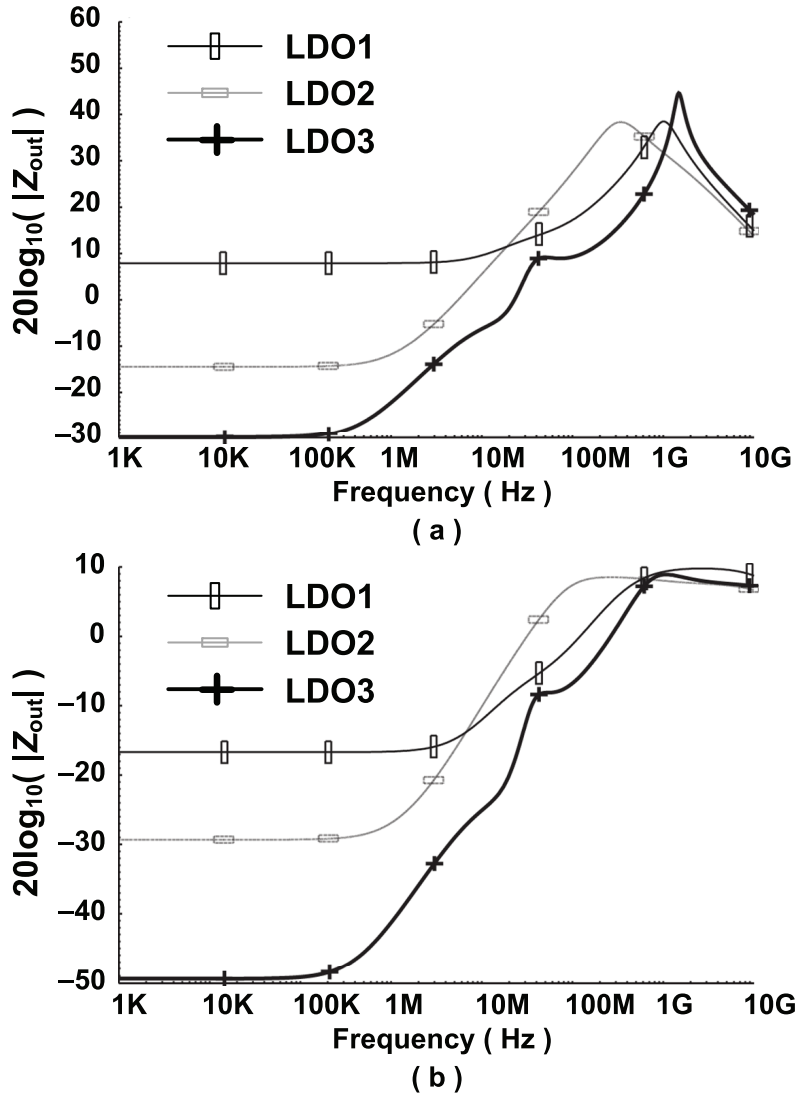


Figure 2.17: Comparisons of the three LDOs' output impedance. (a) At $I_L=1\text{mA}$, $C_L=1\text{pF}$. (b) At $I_L=100\text{mA}$, $C_L=1\text{pF}$.

Better than these two, LDO3 achieves the lowest Z_{out} , though close to that of LDO1. Therefore, the proposed LDO has the best output impedance in the frequency range from DC to hundreds of MHz. Although for even higher frequencies Z_{out} of the proposed LDO appears worse than the other two, the actual Z_{out} in this frequency band is mostly lowered by using large amount of decaps.

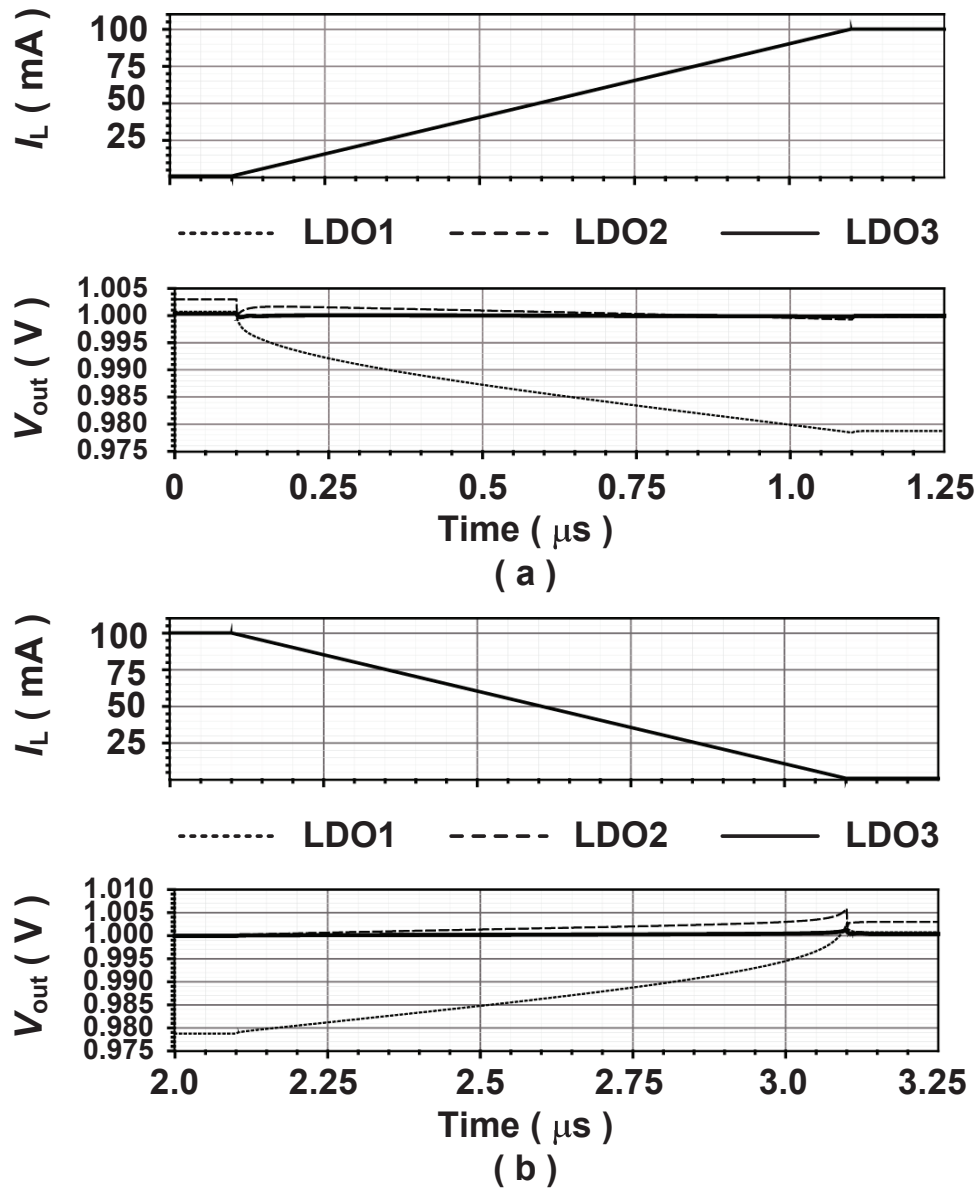
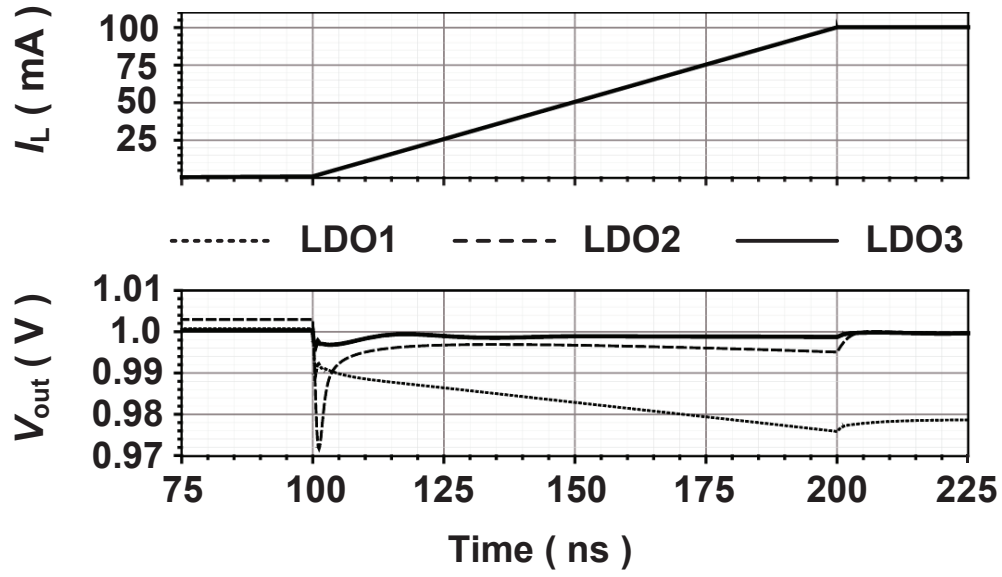


Figure 2.18: Comparison of the three LDOs' load transient responses at 1- μs transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.

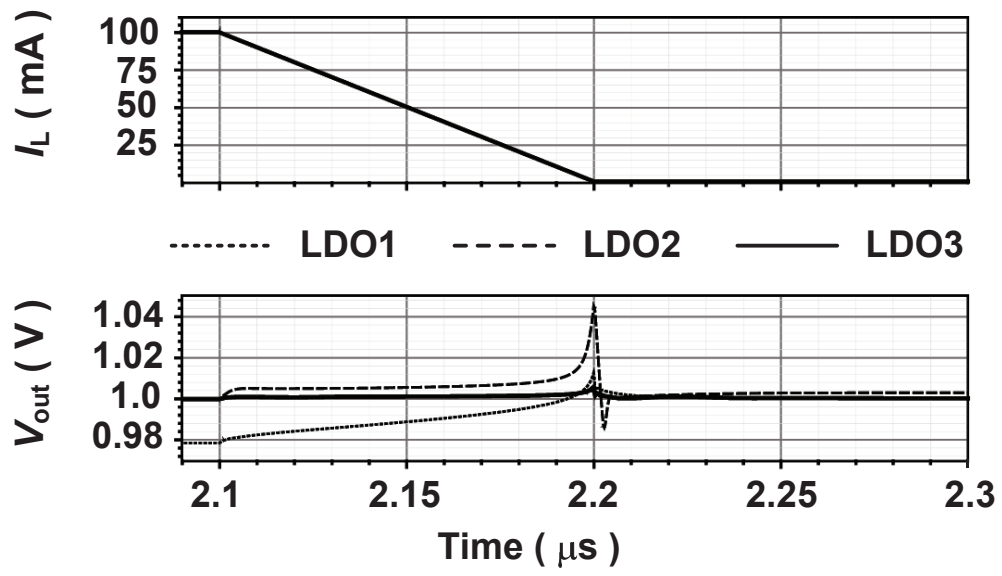
The load transient responses of the three LDOs are also compared shown in Fig. 2.18, Fig. 2.19 and Fig. 2.20. For load current ramping within $1\mu\text{s}$ (shown in Fig. 2.18), the voltage drop and overshoot of the proposed LDO are 0.7mV and 1.2mV , respectively. There is at least 95% improvement over LDO1 with 22.5-mV voltage drop and 25-mV overshoot, and at least 80% improvement over LDO2 whose voltage drop and overshoot are 3.6mV and 6mV respectively. For load current jumping within 100ns (shown in Fig. 2.19), the voltage drop and overshoot of the proposed LDO are 3.8mV and 5.9mV , respectively. Compared with 24.8-mV voltage drop and 33.7-mV overshoot by LDO1 and 31.2-mV drop and 45.5-mV overshoot by LDO2, the improvement is still more than 80% of the better of the two. For the load transient of 10-ns transition times (shown in Fig. 2.20), the voltage drop and overshoot of LDO3 are 37mV and 35.4mV respectively, while those of LDO1 are 82.9mV and 79.4mV and those of LDO2 are 198.5mV and 187.8mV , respectively. In the comparison, LDO1 achieves better transient load regulation than LDO2 when the transients of I_L are fast, while LDO2 wins over LDO1 when I_L transients are slow. But in both cases, the proposed LDO performs the best of the three.

2.3.5 Benefits of the Notch

Fig. 2.21 demonstrates the effect of the impedance notch on load regulation. The non-notched implementation is realized by simply changing C_{c1} from 500fF to 85fF and C_{c2} from 650fF to 1pF . Fig. 2.21(a) shows the output impedance difference with and without a notch. Although the implementation with the notch exhibits worsened Z_{out} in the band of $10\text{K}\text{--}30\text{M}\text{-Hz}$, it gives a 25-dB improvement at about 66MHz . Because the power spectrum density of the load transients can peak around a certain frequency which the notch can be aligned to, the notched LDO is still able to achieve better noise suppression in the scenarios discussed in Section 2.2.4. Transient simu-



(a)



(b)

Figure 2.19: Comparison of the three LDOs' load transient responses at 0.1- μ s transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.

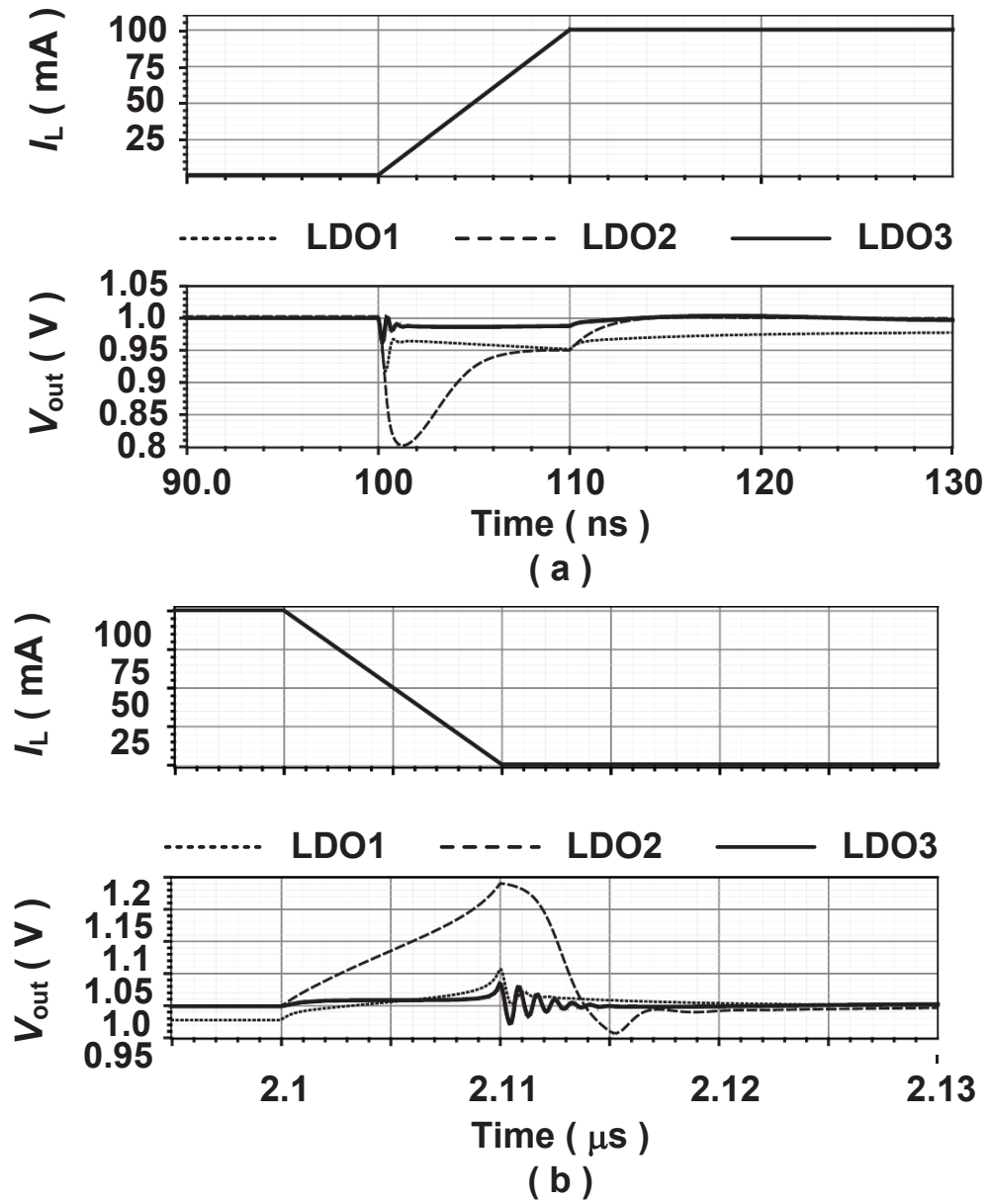
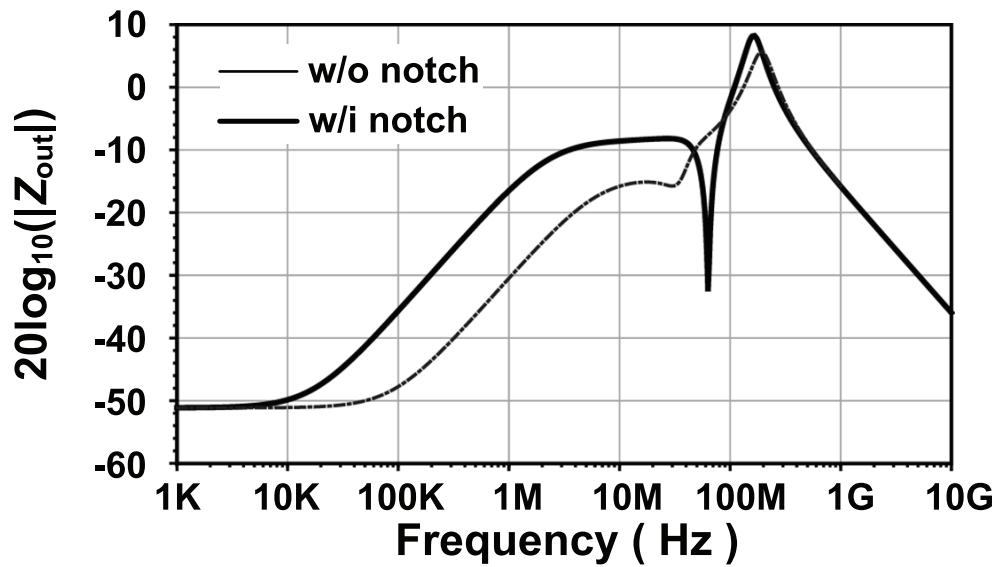
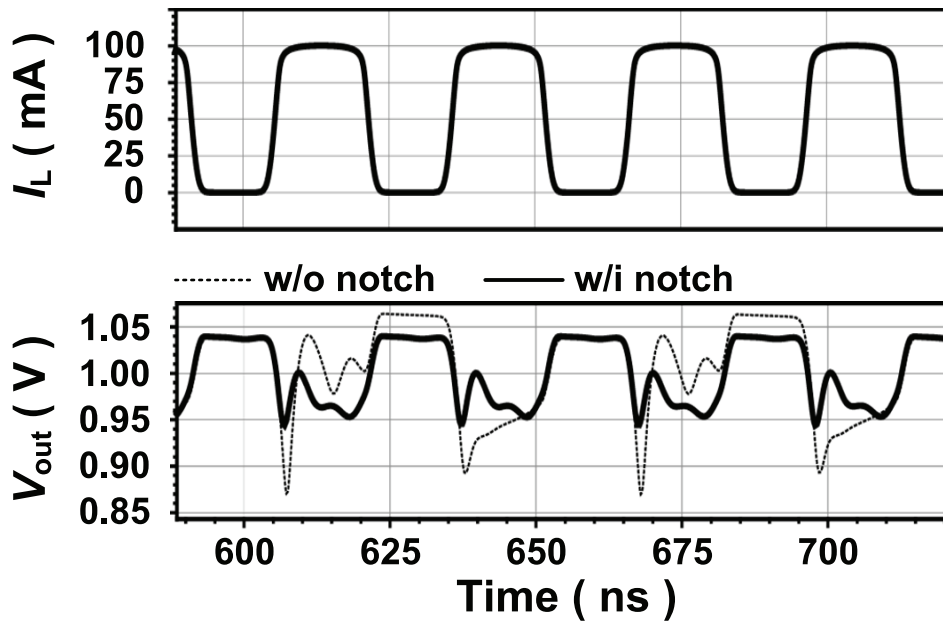


Figure 2.20: Comparison of the three LDOs' load transient responses at 10-ns transition time of the load current with C_L being 1pF. (a) Voltage drops. (b) Overshoots.

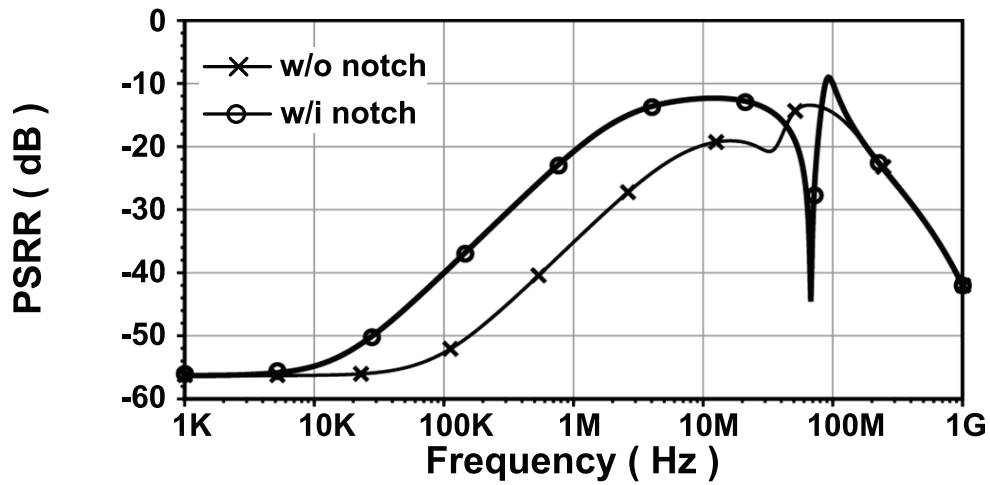


(a)

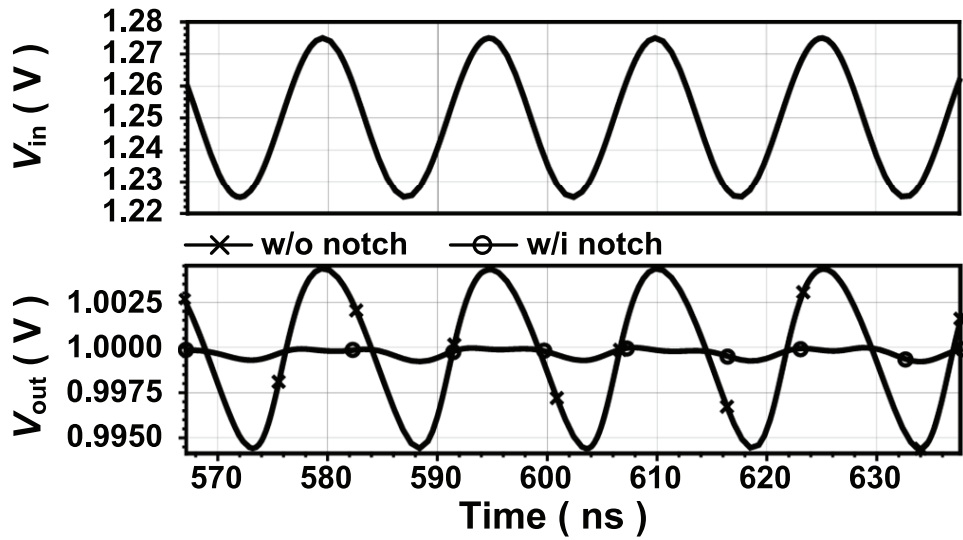


(b)

Figure 2.21: Load regulation performance comparison of the implementations with and without the notch ($I_L = 100\text{mA}$ and $C_L = 1\text{nF}$). (a) Output impedance. (b) Comparisons of transient responses.



(a)



(b)

Figure 2.22: Line regulation performance comparison of the implementations with and without the notch ($I_L = 1\text{mA}$ and $C_L = 600\text{pF}$). (a) The PSRRs. (b) Comparisons of transient responses.

lation result shown in Fig. 2.21(b) verifies the benefit of the notch. A peak-to-peak output voltage variation of 95mV_{pp} of the "notched" LDO caused by a periodic load current ramping up and down within 7–8 ns (corresponding to frequencies around 66MHz, i.e., the notch frequency in this implementation) is compared with that of 193mV_{pp} by the one without the notch, achieving an improvement of over 50%.

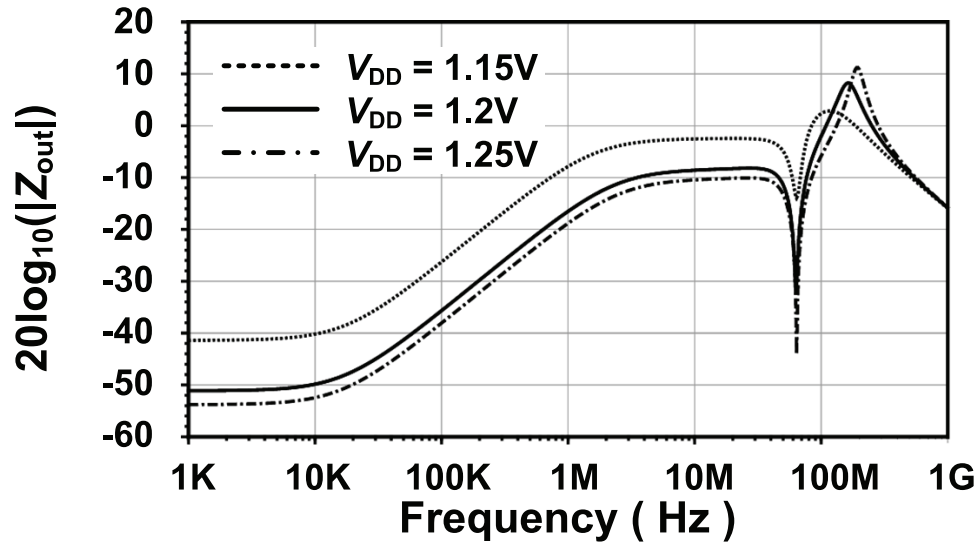
Fig. 2.22 demonstrates the effect of the PSRR notch on line regulation. The PSRRs of the two implementations are compared in Fig. 2.22(a). It is shown that the PSRR notch frequency is almost the same as that of output impedance notch which verifies what is discussed in Section 2.2.4. The transient simulation is set up in the following way: the input voltage of LDO is provided by a sine voltage source with amplitude of 25mV at the exact frequency of 66MHz, mimicking the output ripple of the preceding switching DC-DC converter. And the result in Fig. 2.22(b) shows that the LDO with the notch suppresses supply ripple to as small as about $750\mu\text{V}_{\text{pp}}$ compared with that of about 10mV_{pp} by the one without the notch.

Note that the notch frequency, 66MHz, is for demonstration and can be tuned, according to (2.11), to the actual switching frequency in a specific application.

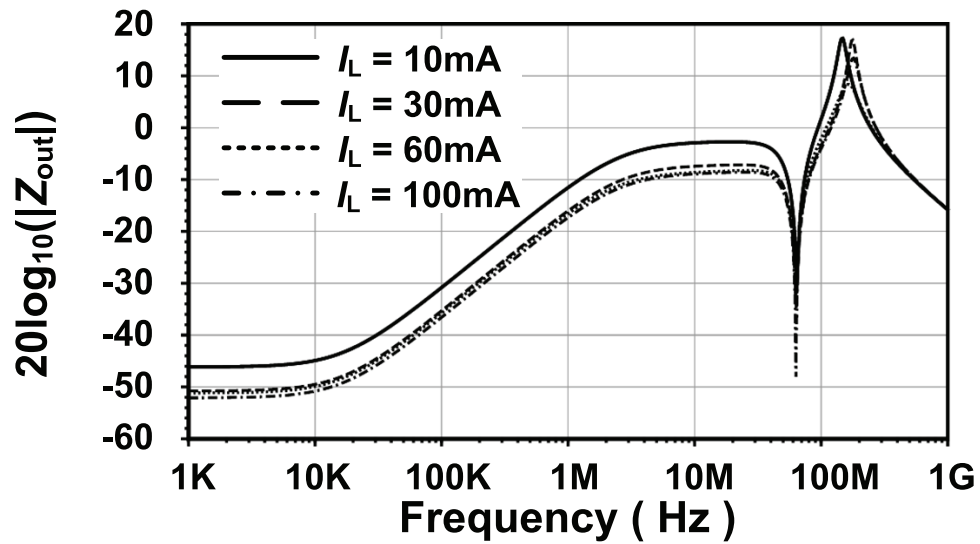
Fig. 2.23 also shows a good immunity of the notch frequency to variations of both the input voltage and load current, which, from another perspective other than time-domain simulation results, verifies that the notch can work under large-signal condition, although it is demonstrated by small-signal analysis results.

2.4 Performance Comparisons

The performance comparisons of the propose LDO with some recently-published on-chip LDOs are summarized in Table 3.1. Since the maximum voltage drop due to load variation is closely related to the transition time (t_r) of the load current, the comparison chooses the kind of figure of merits (FOM) that takes t_r into account.



(a)



(b)

Figure 2.23: Influences on the impedance notch frequency of V_{in} and I_L . (a) Varying V_{in} . (b) Varying I_L .

Table 2.2: Performance Comparisons of the Proposed LDO with Some Existing Works

	[8]	[9]	[14]	[13]	[12]	This work
Year	2010	2010	2010	2011	2011	2011
Technology	350-nm	90-nm	350-nm	65-nm	350-nm	90-nm
V_{in} (V)	0.95-1.4	0.75-1.2	1.8-4.5	1.65-1.95	≥ 1.4	1.2
V_{out} (V)	0.75-1.2	0.5-1	1.6	1.2	1.2	1
Dropout Voltage (mV)	200	200	200	450	200	200
I_q (μ A)	43	8	20	132	34.6	408
I_{max} (mA)	100	100	100	200	50	100
Load capacitor range for stability (pF)	0,100,1000	0-50	100	150	0-200	0-1000
On-Chip capacitor (pF)	6	7	7	0	26	1.8
Chip area (mm^2)	0.155	0.019	0.145	0.08	0.08	0.005*
Stead-State load regulation (mV/mA)	0.4	0.1	0.109	0.078	0.003	0.003
Transient load regulation ($V_{in} = 1.2V$)	ΔV_{out} (mV)	70 ($t_r=1\mu s$)	114 ($t_r=100ns$)	97 ($t_r=100ns$)	16 ($t_r=1\mu s$)	75 ($t_r=0.3\mu s$)
	ΔI_L (mA)	99	97	100	149	49.9
Stead-State line regulation (mV/V)	N/A	3.78	57.4	0.43	8.8	4.3
Transient line regulation (@ I_{max})	ΔV_{out} (mV)	N/A	($t_r=10\mu s$)	N/A	N/A	27
	ΔV_{in} (mV)	N/A	420	N/A	N/A	200
Transition time ratio, K	10,000	1,000	1,000	10,000	3,000	1/1,000/10,000
FOM ₁ (V)	0.304 ($t_r=1\mu s$)	0.009 ($t_r=100ns$)	0.0194 ($t_r=100ns$)	0.142 ($t_r=1\mu s$)	0.156 ($t_r=0.3\mu s$)	0.0004/0.069/0.204 ($t_r=0.1ns/100ns/1\mu s$)
FOM ₂ (pico-second)	1.8e-3 ($C_o=6pF$)	4.8e-3 ($C_o=50pF$)	0.0194 ($C_o=100pF$)	0.0143 ($C_o=150pF$)	0.104 ($C_o=100pF$)	2.33/2.4e-4/7e-5 ($C_o=600pF/0.35pF/0.35pF$)
FOM ₃ (pico-Coulomb)	1.82 ($C_o=6pF$)	0.45 ($C_o=50pF$)	1.94 ($C_o=100pF$)	21.3 ($C_o=150pF$)	15.6 ($C_o=100pF$)	0.23/0.024/0.071 ($C_o=600pF/0.35pF/0.35pF$)

* The value is estimated according to the design rules for the process.

FOM₁ in [9] is given by

$$\text{FOM}_1 = K \frac{\Delta V_{\text{out}} I_{\text{q}}}{\Delta I_{\text{L}}}, \quad (2.14)$$

where K is load current transition time ratio that is defined by

$$K = \frac{t_{\text{r}} \text{ used in the work being compared}}{\text{The smallest } t_{\text{r}} \text{ among all compared works}}. \quad (2.15)$$

The unit is Volt. However, this FOM does not reflect the fact that the output capacitor also has a significant impact on the voltage drop. And hence, we also perform the comparisons of the FOM defined as [4]

$$\text{FOM}_2 = C_{\text{o}} \frac{\Delta V_{\text{out}} I_{\text{q}}}{\Delta I_{\text{L}}^2}, \quad (2.16)$$

where C_{o} is the total capacitance of all extrinsic capacitors connected to the output of the LDO. The unit of this FOM is second. Finally, we define another figure of merit (FOM₃) that combines considerations in the above two as

$$\text{FOM}_3 = C_{\text{o}} \cdot \text{FOM}_1. \quad (2.17)$$

The unit of this FOM is Coulomb. All the three FOMs are encouraged to be small.

The proposed LDO achieves comparably fast load transient response with respect to the design in [4] while the quiescent power it consumes is about 7% of that in [4]. And the total amount of on-chip capacitors inside the LDO is less than 2pF, compared with 6pF in [8] or 7pF in [9] or 26pF in [12], and hence it occupies smaller chip area, achieving higher area efficiency. Whereas the LDO in [13] adopts N-type source follower as the output stage which has tightly-constrained gate-to-source voltage,

plus it is implemented by using the thick-oxide devices, resulting in a relatively large chip area. Regarding to FOM₁, the proposed LDO is more than ten times better than the best among the others for the applications where there are ultra-fast load variations, while it is also improved significantly regarding to the other two FOM's for the applications where the load variations are relatively slow. Note that in calculation of the three FOM's, the output capacitor of the proposed LDO is set to be 600pF for the 100-ps load transients, and be less than 1pF for the 100-ns/1- μ s load transients.

2.5 Summary

This chapter demonstrates an on-chip LDO topology with multiple feedback loops that enhances both steady-state and transient load regulation performances as well as the suppression of the input voltage ripple. An active frequency compensation scheme is also presented to improve the LDO's area efficiency while ensuring stability. Designed in 90-nm CMOS technology, the LDO shows robustness to process and temperature variations as well as device mismatches by thorough simulations. Performance comparisons with recently reported works manifest that the LDO achieves better load regulation by more than ten times than the best of the compared if consuming the same amount of quiescent current, and only occupies chip area of about 60 percents of the smallest among its peers. And hence, it is advantageous to employ the proposed LDO for voltage regulation in modern high-performance ASICs.

3. A POWER-EFFICIENT ON-CHIP LDO ASSISTED BY SWITCHED CAPACITORS FOR FAST TRANSIENT REGULATION*

The fully on-chip linear regulator (LDO), compared with its counterparts (e.g., fully integrated switching-mode DC-DC converters) is usually more area-efficient, and hence can be placed closer to the loading circuits for better supply noise suppression [4, 9]. Specifically, it better shields the load circuits from the static and dynamic voltage drop caused by parasitic inductance and resistance of the package and off-chip components. And it can also provide isolation of the on-chip decoupling capacitors (decap) from the package [6], rendering improved power supply integrity which is otherwise endangered by fast load current variations and package-decap resonance.

As mentioned in Chapter 1, improved power integrity is helpful to improve power efficiency. The tangible benefit in terms of power saving by a well-regulated power delivery system is illustrated in Fig. 3.1. For specified functioning and performance of the load circuits, there is a lower bound of supply voltage ($V_{sup.min}$). To accommodate dynamic voltage drop, the actual supply voltage (V_{DD}) in an unregulated system should be elevated considerably to allow for some margin, which causes excessive power consumption. In contrast, in a regulated system, the significantly reduced transient noise on the output of the LDO (V_{reg}) allows smaller margin above $V_{sup.min}$. On the other hand, the noise on the supply voltage of the regulator (V'_{DD}) can also be smaller than that on V_{DD} of an unregulated system, because the IR drop from the external supply to the regulator is less than that to the load circuits and

*Reprinted with permission from “A power-efficient on-chip linear regulator assisted by switched capacitors for fast transient regulation” by S. Lai, & P. Li, 2013. *Proceedings of 14th International Symposium on Quality Electronic Design*, 682–688, Copyright [2013] by IEEE.

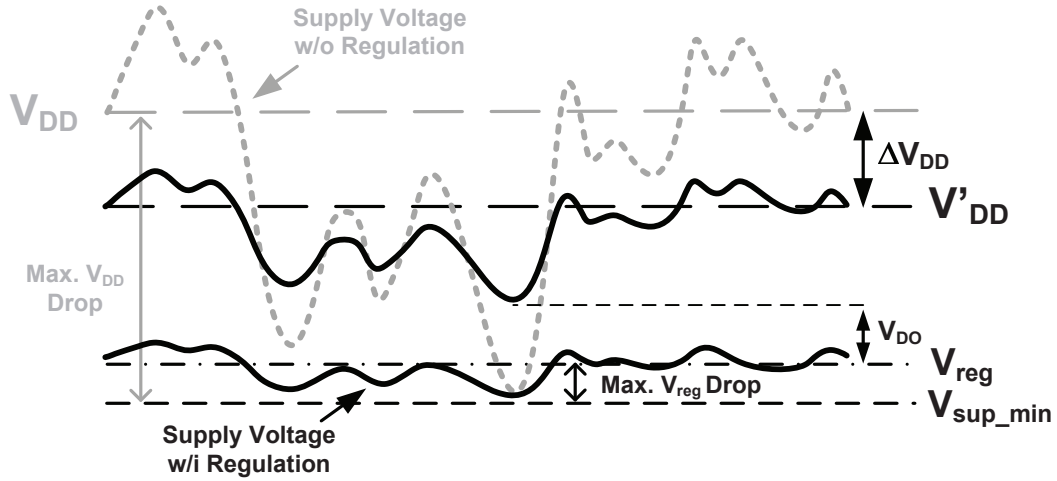


Figure 3.1: An illustration of the benefit on power saving from a well-designed regulator.

also, the package-induced resonance is suppressed. Therefore, if the dropout voltage (V_{DO}), i.e., the desired least voltage difference between the input and output of LDO, is sufficiently small, then V'_{DD} can be actually lower than V_{DD} , reducing the aforementioned excessive power consumption.

Unfortunately, the downside of traditional on-chip LDOs is the considerably large V_{DO} that is needed for good transient regulation performance. Since the power efficiency of LDOs is upper bounded by $V_{reg}/(V_{reg} + V_{DO})$, the high dropout voltage is the major hindrance to achieving high power efficiency of the system.

To reduce V_{DO} by a certain factor, the width of the LDO's pass transistor needs to be increased roughly by the same factor, which in turn can severely retard the feedback control loop and deteriorate transient performance. To combat this problem, the traditional method prescribes a proportional increase of the bias current of the pass transistor's driving circuit, which could result in an uproar of quiescent power consumption. Therefore, the trade-off among the dropout voltage, transient response and current efficiency of traditional on-chip LDOs makes it difficult to achieve high

power efficiency and good transient response simultaneously.

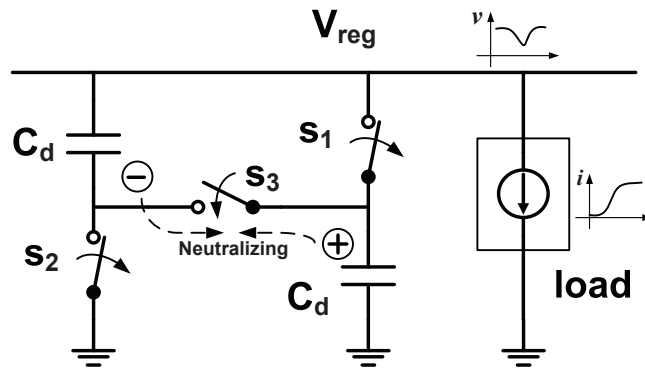
To this end, we propose an implementation of a power-efficient on-chip LDO with significantly reduced the dropout voltage, and two effective techniques that employ switched capacitors to overcome the degradation of the transient performances. The first technique switches a portion of on-chip decaps between the LDO's input and output to exploit the voltage difference. As charged up to a higher voltage, these decaps are able to provide more charges to the load than non-switched ones. The second technique utilizes the "spare" time (i.e., when V_{reg} is relatively stable) to store charges on the switched capacitors. Once an abrupt transient load variation occurs, the switched capacitors are hooked up to the gate of the pass transistor to speed up the process of charging/discharging of the gate capacitance of the pass transistor. And thus it can save a considerable amount of adjusting time for the LDO. With the second technique implemented, the proposed LDO achieves about 70mV dropout voltage while maintaining within 10% output voltage fluctuation under load current step of 5ns rise/fall time. With load capacity of 100mA, the LDO, including the switched capacitor circuit, only consumes quiescent current of only $38\mu A$ in total. While the power overhead brought by the switched capacitor circuit is about 30% of the original consumption, the dynamic voltage drop is improved by about 80%.

3.1 Concepts of the Proposed Techniques

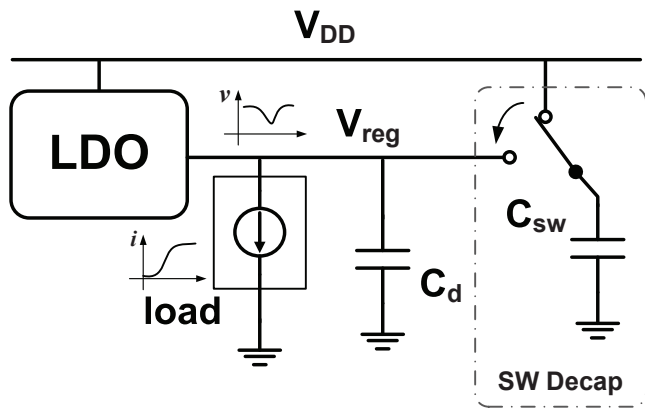
In this section, the conceptual illustrations of the two proposed techniques are presented. And the advantages of the techniques on power saving are demonstrated as well.

3.1.1 *The Switched Decoupling Capacitor*

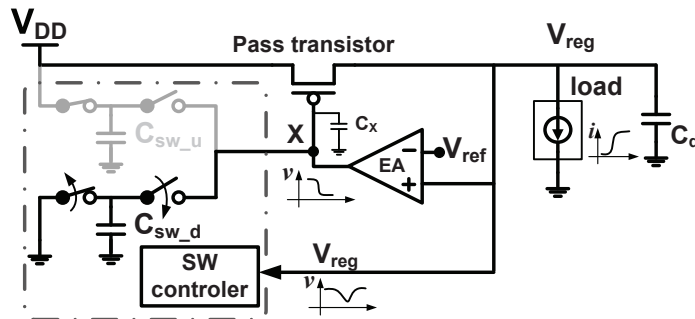
In the past decade, a few works on active decoupling capacitor (decap) design were proposed to conquer the transient noise in power delivery systems (PDN) without on-



(a)



(b)



(c)

Figure 3.2: Conceptual schematics of the switched capacitor techniques. (a) The existing technique. (b) The proposed switched decap technique. (c) The proposed switched positioning cap technique.

chip regulators [35,36], which the authors believe can also be adopted in the on-chip regulated PDNs. Suppose we apply the active decap to the regulated power grids. The conceptual schematic is drawn in Fig. 3.2(a). At the beginning, the switches, s_1 and s_2 , are closed with s_3 opened, so that the two decaps, C_d , are connected to the power line in parallel. Once a sufficiently large voltage drop, ΔV_{reg} happens due to a sudden demand of load current, a sensing circuit will disconnect s_1 and s_2 and connect s_3 . Thereby, the two decaps are connected in series and can deliver extra charge, in the amount of $C_d(V_{reg} - \Delta V_{reg})/2$, to the load to prevent further decrease of the supply voltage.

Similar to this decap topology switching, we first propose the switched decap concept in a regulated power delivery network that exploits the difference between the supply voltage and the regulated voltage, as illustrated in Fig. 3.2(b). When a voltage drop of ΔV_{reg} occurs, C_{sw} will be switched from the V_{DD} rail to the V_{reg} rail. The amount of extra charge by the switched decap is $C_{sw}(V_{DD} - V_{reg} + \Delta V_{reg})$. Compared with the active decap techniques, the proposed technique has the following three key advantages. 1), all the charge stored on C_{sw} is utilized, i.e., there is no waste of charge during the whole process, whereas in the scheme shown in Fig. 3.2(a) a neutralization of charges happens when the two capacitors are switched into serial connection. The evidence of the neutralization is in the difference of available charges of the two configurations. While in the parallel configuration the charges on decaps available for the load is $2C_dV_{ref}$, in serial configuration the available charge is only $\frac{C_d}{2} \times 2V_{ref} = C_dV_{ref}$, which is halved, indicating that the other half is not available for the load any more. The amount of the neutralized charge is equal to the amount of charge provided to the load. 2), when switching back to the parallel connection, C_{sw} in the proposed technique will have little influence on V_{reg} , while the two C_d 's in the active decap technique will draw up to C_dV_{ref} charge from V_{reg} , which can

be large enough to cause considerable secondary drop on the V_{reg} . 3), in the active decap technique, there are more switches to operate, making control circuit more complicate and potentially consuming more power; furthermore, s_1, s_2 and s_3 should not be turned on at the same time to prevent short-circuit path from V_{reg} to ground, thus a specific delay circuit is needed to generate two non-overlapping sets of switch control signals.

In summary, the proposed switched decap technique is more power-efficient and can be a better fit in the regulated power delivery system.

3.1.2 The Switched Positioning Capacitor

While the previous technique exploits the LDO's input-to-output voltage difference, we further propose a voltage positioning technique which is also based on switched capacitors, but can work in the existence of small input-to-output voltage difference. The conceptual schematic is shown in Fig. 3.2(c). On top of a typical LDO circuit, the auxiliary circuit adopting the switched positioning capacitor technique is shown in the dash-dotted box.

The principle of the technique is as follows. Initially the regulated power line (V_{reg}) is above a preset threshold, the pull-down capacitor (C_{sw-d}) is connected to ground; once V_{reg} drops below the threshold, the switch controller block will connect C_{sw-d} onto the gate of the pass transistor. Then the gate potential of the pass transistor (V_X) will be pulled down due to charge sharing between C_{sw-d} and C_X (mainly the gate capacitance of the pass transistor). The time constant (τ_{sh}) of the change of V_X is then $R_{sw}C'_X$, where $C'_X = C_X C_{sw-d} / (C_X + C_{sw-d})$. Similarly, if a certain amount of V_{reg} overshoot happens, C_{sw-u} is switched to push up V_X . If by the LDO alone, V_X can only be adjusted with the time constant (τ_{EA}) of the error amplifier (EA) which can be much larger than τ_{sh} . Since V_X can be more quickly

positioned to the proximity of the “right” voltage level by the proposed technique, the LDO can react faster to the transient variations.

As for the quiescent power consumption, because the auxiliary circuits more resembles digital circuits, as will be shown in the next section, they only bring about a limited power overhead.

3.1.3 Discussions of the Two Techniques

Essentially, the fundamental reason for the above two proposed techniques to work is that: the required positive (or negative) charges are pre-stored in the switched capacitors when V_{reg} is relatively stable (i.e., no burst of variations), and when the fast transient occurs these capacitors can quickly share charges, saving some amount of charging/discharging time at critical nodes. Further, the storage of those pre-obtained charges consumes limited power, and hence rendering small quiescent power overhead.

Putting aside the common effectiveness on suppression of transient noise, the two techniques both have their own application scopes.

The switched decap technique allows the switched decap to work independent of the LDO main circuit. Therefore, the switched decaps can be widely distributed to the places closer to the “hot spots”, i.e., the heavy load noise sources, maximizing the benefit of local regulation.

On the other hand, the switched positioning capacitor technique is advantageous in regulators of very low dropout voltage. With a fast response time, this technique can be a competitive candidate in addressing the confliction between high power efficiency and good transient regulation.

3.2 Circuit Analysis and Implementation

In this section, detailed analysis on the design of the circuits are presented. The symbolic analysis is conducted based on the assumption that the load current variation is a step with a certain rise time (t_r). And we also assume that the LDO does not response to the V_{reg} variation earlier than its intrinsic response time (t_{resp}); and after t_{resp} the LDO can quickly adjust itself to the “right” state. By this means, the LDO is underestimated and the later conclusions based on these assumptions tend to be conservative. For simplicity of words, the analysis of transient behaviors at the rising edge of load current is performed; behaviors at the falling edge is similar and omitted.

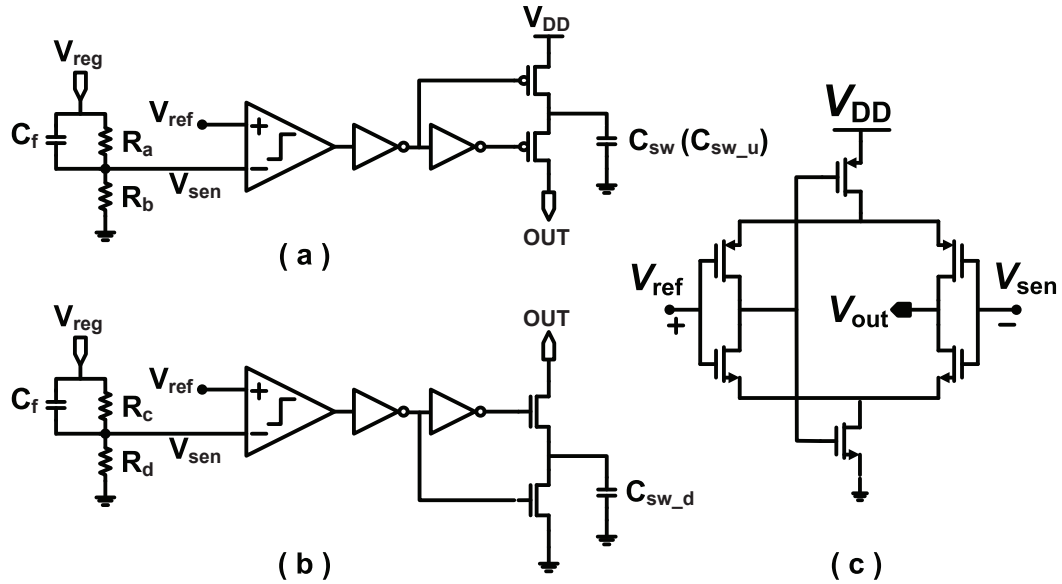


Figure 3.3: Schematics of the switched capacitor circuits. (a) The push-up circuit. (b) The pull-down circuit. (c) The schematic of the comparator.

3.2.1 Design of the Switched Capacitor Circuits

The schematics of the switch controllers are shown in Fig. 3.3. In order to cut down the number of voltage references required, all V_{ref} are the same as that used in the LDO. Different resistive voltage dividers (R_a , R_b and R_c , R_d) are used to have different thresholds for the push-up and pull-down switching. A feed-forward capacitor (C_f) is added for better sensing of high frequency variations on V_{reg} . A high slew-rate comparator [34] shown in Fig. 3.3(c) is adopted; two stages of inverters are appended to shape the output of the comparator and to enhance driving strength as well. The switches are implemented with PMOS's for push-up circuit as shown in Fig. 3.3(a) and with NMOS's for pull-down circuit as shown in Fig. 3.3(b).

3.2.1.1 Design of the Switched Decoupling Capacitor

The design of the switched decap is dependent on the relationship between the LDO's response time (t_{resp}) and the fastest transition time (or t_{r_min}) of load current according to the design specifications. There are two key types of the relationship as demonstrated in Fig. 3.4: the scenario when $t_r \gg t_{resp}$ and the one when $t_r \ll t_{resp}$. If $t_r \approx t_{resp}$, it is fair to say that this scenario is less stringent than the $t_r \ll t_{resp}$ scenario in terms of transient voltage drop of V_{reg} . During analysis on each scenario, we will develop design requirements for the amount of the switched capacitor, and the speed for the comparators in the switch controller.

In both scenarios, we first specify the maximum tolerable voltage droop of the LDO's output voltage as ΔV_{reg} and denote the threshold voltage as $V_{sw.th}$ and the reaction time of the sensor as t_p which includes the portion of time spent on charge sharing process. And again we assume the voltage will not change until the charge sharing is finished, which is conservative.

Since the load current is ramping at the slope of I_{max}/t_r , V_{reg} will drop on a

parabolic track, and the first V_{reg} dip is derived as

$$V_{dip1} = I_{max}(t_0 + t_p)^2 / (2t_r C_d), \quad (3.1)$$

where t_0 is the time for V_{reg} to drop from the steady-state value (denoted as V_{reg0}) down to V_{sw_th} (i.e., $V_{reg0} - V_{sw_th} = I_{max}t_0^2 / (2t_r C_d)$); C_d is the fixed decap as shown in Fig. 3.2(b).

As indicated by (3.1), if the specification for ΔV_{reg} is tight, then V_{dip1} is tightly constrained. According to (3.1), a large fixed decap, and/or quick reaction time of the controller, and perhaps higher V_{sw_th} are desired. The trade-offs are: large decap means larger silicon area needed; small t_p indicates higher quiescent and dynamic power by the switching circuits; higher V_{sw_th} can cause frequent switching

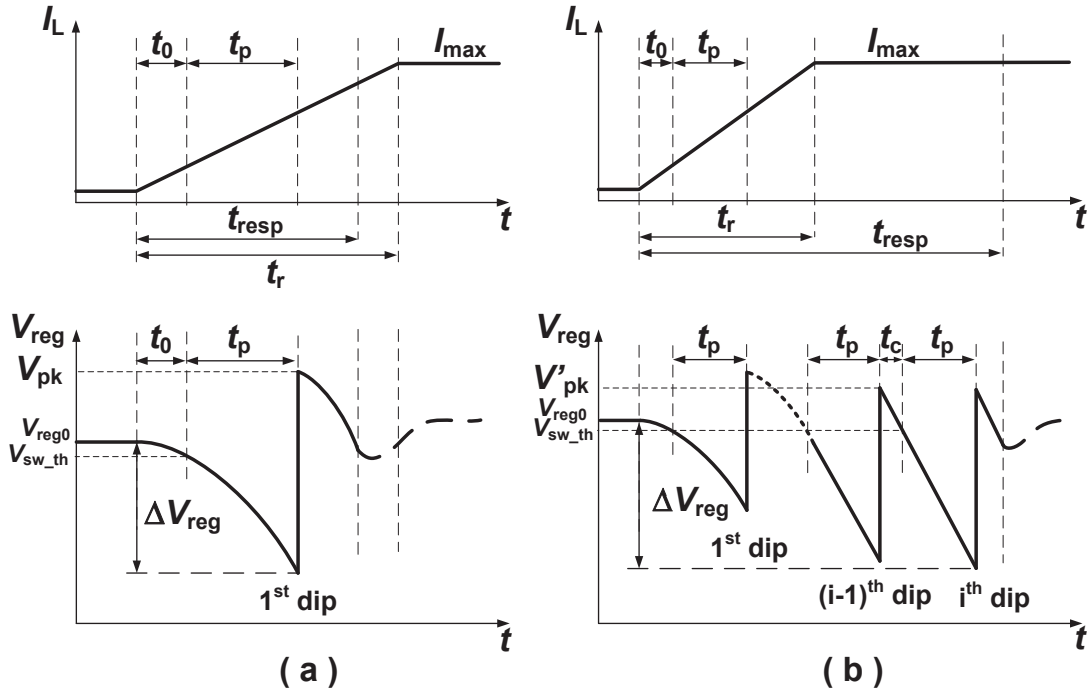


Figure 3.4: Illustrations of I_L and V_{out} with the switched decap technique. (a) When $t_r \gg t_{resp}$. (b) When $t_r \ll t_{resp}$.

and potentially can increase dynamic power.

The first voltage dip is common for all the two scenarios and what is next are different and discussed separately as follows.

The $t_r \gg t_{resp}$ Scenario In this scenario, the load current (I_L) versus the regulated voltage (V_{reg}) is illustrated in Fig. 3.4(a). Since the LDO can track the change of I_L before the I_L ramping is finished, the design goal for this scenario should be made to having the V_{reg} drops after the first dip no more than V_{dip1} , i.e., $V_{dip1} = \Delta V_{reg}$; otherwise, the speed advantage of the LDO is not well utilized. Thereby, we have

$$t_p \leq \sqrt{2C_d \Delta V_{reg} t_r / I_{max}} - t_0, \quad (3.2)$$

where $t_0 = \sqrt{2C_d(V_{reg0} - V_{sw.th})t_r / I_{max}}$. (3.2) implies the speed requirement for the whole sensing, controlling and charge sharing process. From the schematic shown in Fig. 3.3(a), there are four stages of delay: the delay of comparator, the two inverters and the delay of charge sharing between C_{sw} and C_X . As well established in digital circuit design, the optimal total delay of a chain is reached when the delay of each stage is equal. Thus we can obtain the time constant (τ_{sh}) for the charge sharing as $t_p/4$ divided by 0.69.

According the assumption mentioned in the beginning of this section, after the first dip and before t_{resp} , neither the switched capacitor nor the LDO provides charge to the load. The charge demanded by the load can only be provided by the fixed decap, C_d , and can be expressed as

$$\Delta Q_t = I_{max} t_{resp}^2 [1 - (\frac{t_0 + t_p}{t_{resp}})^2] / (2t_r). \quad (3.3)$$

On the other hand, the amount of charge provided by the fixed decap during the time

window that is after charge sharing but before V_{reg} falling again down to $V_{reg0} - \Delta V_{reg}$, is

$$\Delta Q_d = C_d[\Delta V_{reg} + (C_{sw}V_{DO} - C_d\Delta V_{reg})/(C_d + C_{sw})]. \quad (3.4)$$

Then C_{sw} can be calculated by solving $\Delta Q_d \geq \Delta Q_t$ on the condition that $\Delta Q_t \leq C_d(\Delta V_{reg} + V_{DO})$:

$$C_{sw} \geq \frac{C_d}{\frac{C_d(\Delta V_{reg} + V_{DO})}{\Delta Q_t} - 1}. \quad (3.5)$$

With C_{sw} and the time constant τ_{sh} obtained, the PMOS switch resistance can be determined, so is the size. Then, the load capacitance of the sensor (the comparator and inverters) which is the total gate capacitance of PMOS switches, is known, and then the rest of the circuit design is a well-established procedure in digital design that leverages logical effort to minimize the path delay.

The $t_r \ll t_{resp}$ Scenario With reasonable amount of C_d , there can be multiple dips after the first dip in this scenario. The worst dip (shown as the i -th dip in Fig. 3.4(b)) in this scenario happens after I_L has reached I_{max} for a while longer than $(t_0 + t_p)$ and the LDO still has not taken action. Then the charge demanded by the load is completely provided by C_d and C_{sw} . Therefore, we have

$$\Delta V_{reg} - (V_{reg0} - V_{sw.th}) = I_{max}t_p/C_d, \quad (3.6)$$

which determines t_p .

It can be inferred that the peak voltage, V'_{pk} , satisfies not only the charge sharing equation

$$V'_{pk}(C_d + C_{sw}) = C_d(V_{reg} - \Delta V_{reg}) + C_{sw}V_{DD}, \quad (3.7)$$

but also the equality:

$$V'_{pk} = V_{sw.th} + I_{max}t_c/C_d, \quad (3.8)$$

where t_c is defined as drawn in Fig. 3.4(b). Conservatively speaking, after each switching, C_{sw} should be charged up to V_{DD} within t_c . Since C_{sw} is switched out of the main circuit, this process is also referred to as the switch-out in the later content. To make it simple, let the switch-in and switch-out time constants equal, i.e., t_c should be at least $t_p/4$. As a result, through the above mentioned two equalities, the low bound of C_{sw} can be derived as

$$C_{sw} = \frac{\frac{5}{4}C_d[\Delta V_{reg} - (V_{reg0} - V_{sw.th})]}{V_{DO} + \frac{5}{4}(V_{reg0} - V_{sw.th}) - \Delta V_{reg}/4}. \quad (3.9)$$

Then the rest of the design is similar to that in the previous scenario. Lastly, when applying this technique, choose the most stringent scenario since designers cannot control t_r of I_L .

3.2.1.2 Design of the Switched Positioning Capacitor

For the analysis of the switched positioning capacitor, consider the scenario when $t_r < t_{resp}$. The I_L - V_{reg} plot is shown in Fig. 3.5. Again, the goal is to make the maximum V_{reg} drop happen at the first voltage dip, i.e.,

$$\Delta V_{reg} = I_{max}(t_0 + t_p)^2/(2t_r C_d). \quad (3.10)$$

In order to make sure the later voltage dip will not exceed this value, V_{reg} needs to be charged up to a sufficiently high level. Different from the previous technique, the output current (I_{out}) of the LDO, controlled by V_X , is increased after each switching of $C_{sw.d}$, because after $C_{sw.d}$ pulls down V_X and is switched back to ground, V_X

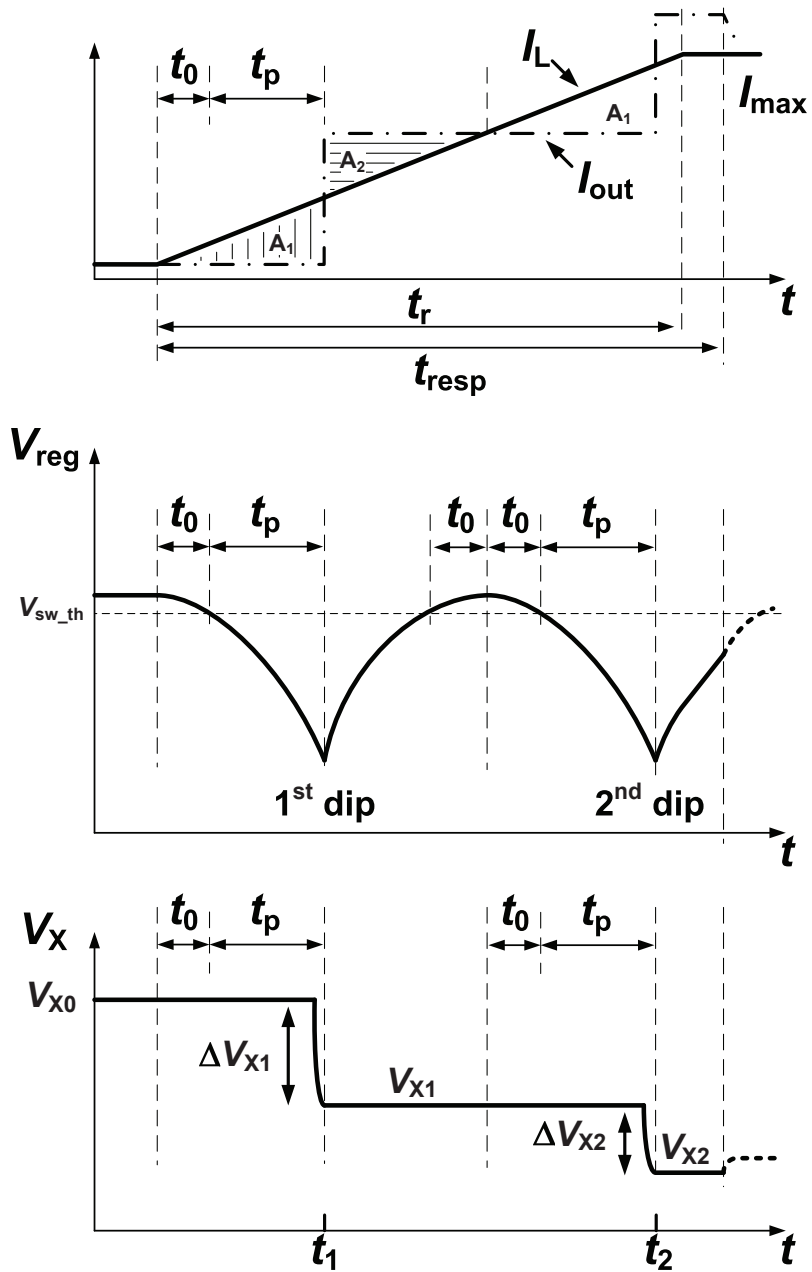


Figure 3.5: Illustrations of I_L and V_{out} with the switched positioning capacitor technique.

approximately stays at the level by the assumption that the LDO does not react within t_{resp} . Therefore, the goal can be reached by making sure that after switching,

I_{out} is momentarily large enough to have the supplemented charge (denoted by the area of A_2 in Fig. 3.5) is no less than the already drained charge (denoted by the area of A_1 in Fig. 3.5). Take the $A_2 = A_1$ scenario as an example. Then V_{reg} will be charged up to the original level by $I_{out} - I_L$ till $I_{out} = I_L$ and then be discharged by $I_L - I_{out}$ till V_{reg} drops by ΔV_{reg} . Afterwards, the switching happens again and all the process repeats till $t = t_r$. Then ΔI_{out} at t_1 should be no less than $2I_{max}(t_0 + t_p)/t_r$. Since the relationship between I_{out} and V_X is known with a fixed LDO design, ΔV_{X1} that generates ΔI_{out} can be calculated. Then $C_{sw.d}$ can be determined through

$$\Delta V_{X1} = V_{X0} * C_{sw.d} / (C_X + C_{sw.d}). \quad (3.11)$$

The time constant of the switch-in path is designed in the same way discussed in the previous subsection; the time constant of switch-out path is constrained by the narrowest switch-out time window, $2t_0$.

The design of the push-up capacitor, $C_{sw.u}$, is similar and it is worth to mention that the threshold for push-up should be distant enough from the threshold for pull-down in order to avoid a current path from V_{DD} through switches to ground.

3.2.2 Design of the LDO

The topology of the implemented on-chip LDO is shown in Fig. 3.6. Like a typical LDO topology, this LDO is also comprised of the output voltage sensor (including R_1 , R_2 and C_1), the error amplifier, and the push-pull output stage (including M_p , M_1 and M_2) that drives C_X and hence typically consumes relatively large quiescent current. To operate at a low supply voltage, all the PMOS's in the P-type current mirror (i.e., M_{bp1} - M_{bp8}) are low- V_{th} transistors; the body of M_{p1} (M_{p2}) is tied to the source of M_{n1} (M_{n2}) to boost the transconductance of the transistors; M_{c1} - M_{c5} are self-cascode composite transistors suitable for low-voltage applications.

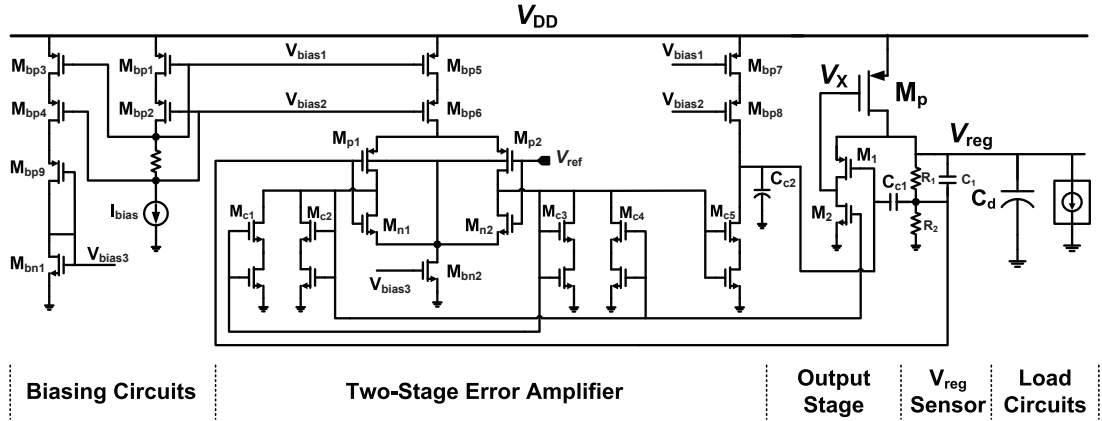


Figure 3.6: The schematic of the LDO.

When applying the switched positioning capacitors to this LDO, the outputs of the push-up and pull-down circuits shown in Fig. 3.3(a)(b) are both connected to V_X , the gate of M_p . All the references are the same; the ratios, R_a/R_b and R_c/R_d , satisfies the relationship:

$$V_{sw_d.th} = V_{ref}(R_a + R_b)/R_b < V_{reg0} < V_{ref}(R_a + R_b)/R_b = V_{sw_u.th} \quad (3.12)$$

. The two thresholds, $V_{sw_d.th}$ and $V_{sw_u.th}$, should be sufficiently away from each other as discussed.

3.3 Simulation Results

The proposed circuits are designed in a commercial 90nm CMOS technology. The simulation setup includes the package model to take into account the LdI/dt noise. Separate implementations of the switched decoupling capacitor and switched positioning capacitor techniques are done to show the common feature of them on improving transient noise suppression and to show the difference of suitable applications of them as well.

3.3.1 LDO with Switched Decoupling Capacitors

Since the switched decap technique utilizes the dropout voltage of the LDO, we first simulate the switched decap technique on the LDO which is designed with 150mV dropout voltage under 1.15V supply. The LDO itself consumes about $350\mu\text{A}$ quiescent current. The switched decap circuit adds about $160\mu\text{A}$, making the total quiescent current of $510\mu\text{A}$. The fixed decap is 500pF while the switched decap is 500pF.

The comparison on the transient load regulation ($t_r=100\text{ps}$) between the LDO with the switched decap circuit and the LDO redesigned with $510\mu\text{A}$ quiescent current and 1nF fixed decap is demonstrated in Fig. 3.7. About 40% reductions of both droop and overshoot of V_{reg} are observed.

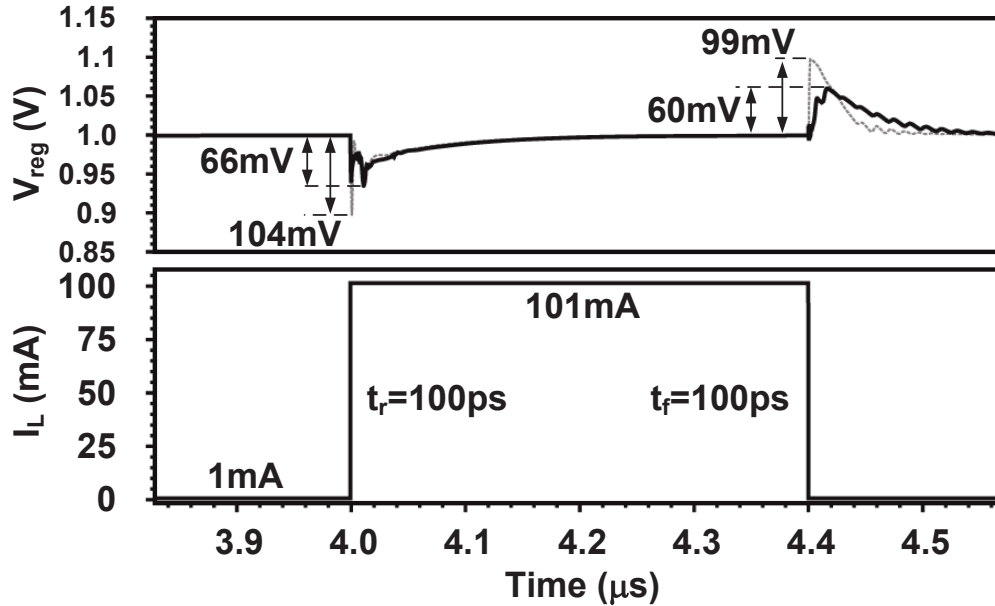


Figure 3.7: Comparison of transient load regulation.

3.3.2 LDO with Switched Positioning Capacitors

In this implementation, the nominal supply voltage of the LDO is 1V and the output voltage is 0.9V; the aspect ratio of the pass transistor is 2.4 μm /80nm. Only when V_{DD} drops below about 970mV, ΔV_{reg} exceeds 10% of 0.9V under load current step with 5ns rise/fall time. The total quiescent current of the LDO (including the auxiliary circuit) is 38 μA , giving the power efficiency of 89.97% and 86.79% respectively at 100mA and 1mA load conditions.

The detailed quiescent current consumption is listed as follows. The bias currents for the biasing circuit, error amplifier, output stage, and the voltage sensor are respectively 2 μA , 9 μA , 14 μA and 4 μA . For the switched positioning capacitor technique, the push-up and pull-down circuits consumes about 9 μA quiescent current, i.e., about 24% of that consumed by the whole LDO.

The fixed decap is about 200pF and the switched pull-down and push-up capacitors are 3pF and 2.5pF, respectively.

Fig. 3.8(a) and Fig. 3.9 are the simulation results on the transient load and line regulation with rise time of 5ns, respectively.

For transient load regulation, the maximum output voltage variation ΔV_{reg} is reduced from 486mV, given by the LDO without switched positioning capacitors, down to 80mV; similarly in line regulation, the maximum ΔV_{reg} is reduced from 92mV to 42mV. About 80% and 50% improvements on load- and line-induced transient noise, respectively, are observed. In addition, Fig. 3.8(b) zooms the signals within a small time window around a V_{reg} dip, which demonstrates the fast adjustment of V_X during the charge sharing process.

Fig. 3.10 shows the Monte Carlo simulation results to shown the robustness of the design with the presence of process variations and device mismatches. 445 out

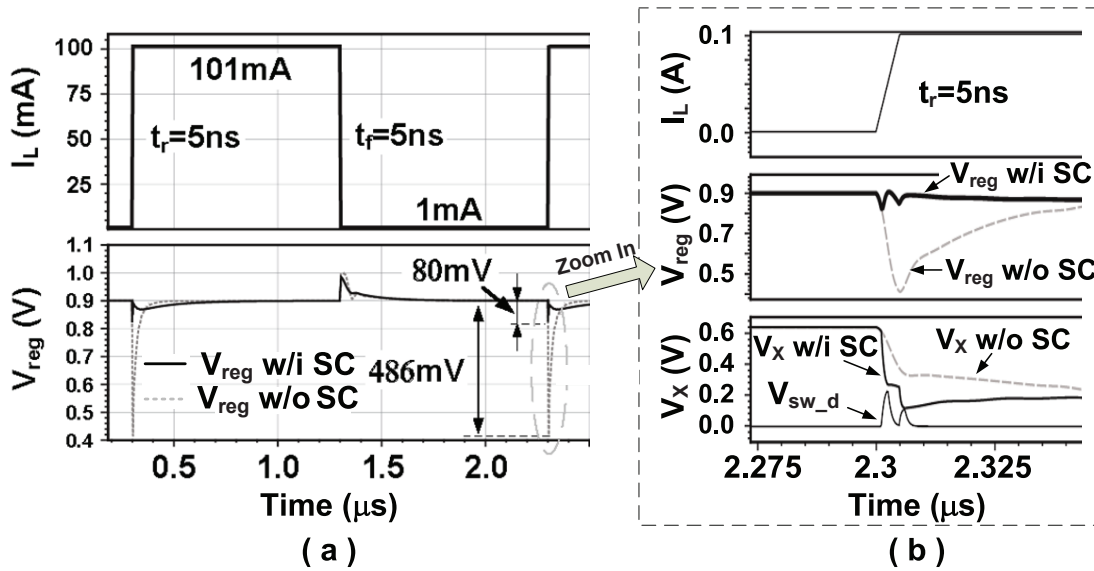


Figure 3.8: Transient load regulation with $t_r=5\text{ns}$.

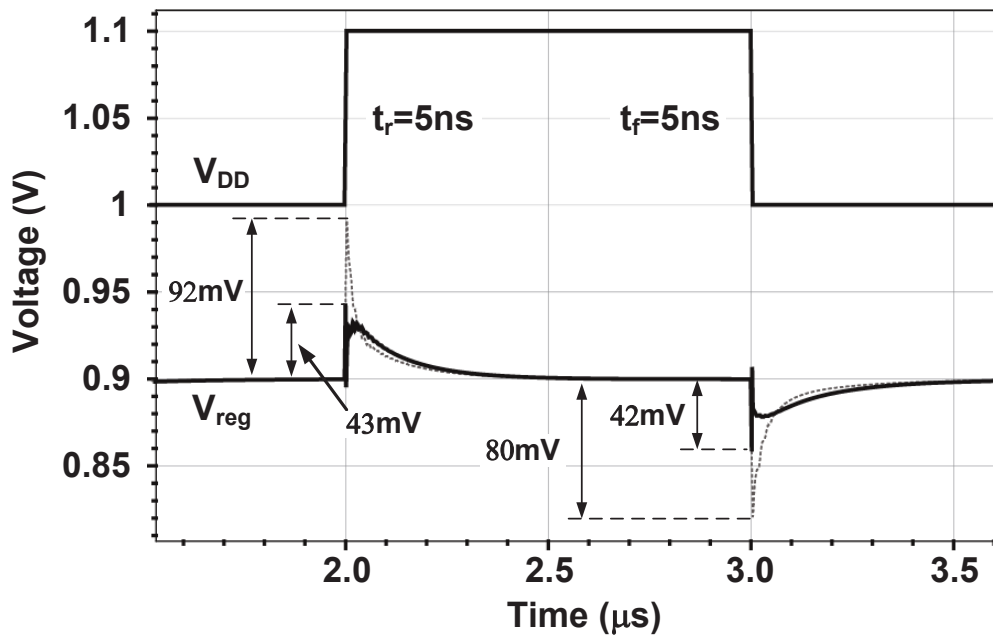


Figure 3.9: Transient line regulation with $t_r=5\text{ns}$.

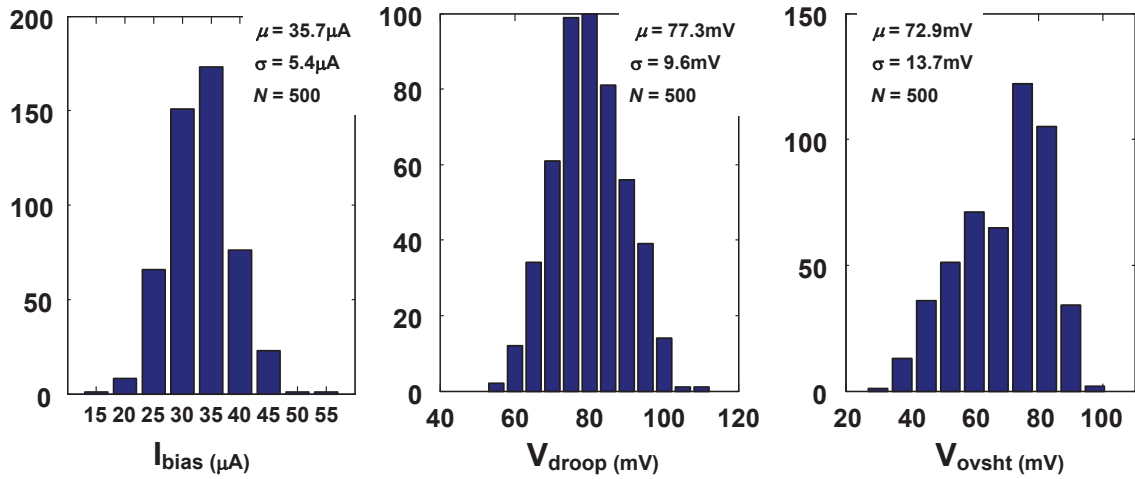


Figure 3.10: Monte Carlo simulation results.

of 500 samples have less than 10% voltage droop and 498 out of 500 have less than 10% overshoot. The mean value of I_{bias} is about $35.7 \mu A$.

A temperature sweep from -40 to 85 C° is conducted to show the performance independence on T . As shown in Fig. 3.11, over the swept range, the V_{reg} droop and overshoot are maintained 10% and the quiescent current monotonically increases with T and reaches about $48 \mu A$ at $85 C^\circ$.

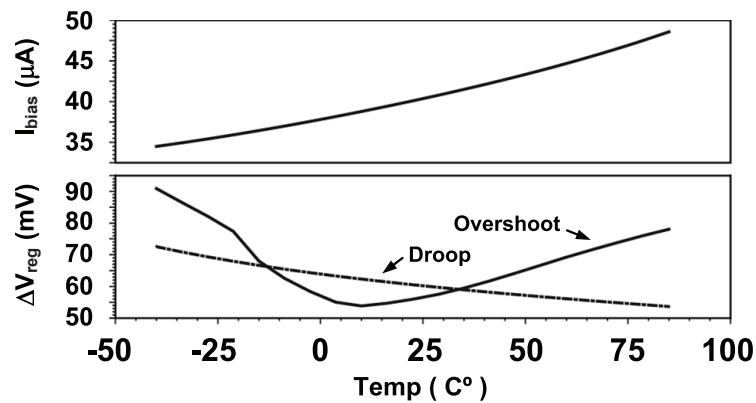


Figure 3.11: Temperature dependence of the performances.

Table 3.1: Performance Comparisons of the Proposed LDOs with Prior Art

	[9]	[13]	[25]	[37]	This work*
Year	2010	2011	2012	2012	2012
Technology	90-nm	65-nm	90-nm	.35 μ m	90-nm
V_{in} (V)	1.2	1.65	1.2	1.2	1
V_{out} (V)	1	1.2	1	1	0.9
Dropout Voltage (mV)	200	200	200	200	70
Power Efficiency @ $I_L=0.1A$	83.3%	72.7%	83.1%	83.3%	90%
Power Efficiency @ $I_L=1mA$	82.7%	64.2%	52.1%	81.1%	86.7%
I_q @ $I_L=1mA$ (μA)	8	132	601	28	38
I_{max} (mA)	100	200	100	100	100
Load regulation (mV/mA)	0.1	0.078	0.003	0.078	0.003
Transient Load Reg.					
ΔV_{out} (mV)	114	16	95	78	80
ΔI_L (mA)	97	149	100	99	100
C_d (pF)	50	150	600	100	200
Transition time ratio, K	20	200	0.02	200	1
FOM ₁ (V)	2e-4	0.0028	1e-5	0.0044	3e-5
FOM ₂ (pico-Coulomb)	0.0094	0.42	0.0069	0.44	0.0061

* The switched positioning design.

3.3.3 Performance Comparisons

The performance comparisons of the propose LDO with some recently-published on-chip LDOs are summarized in Table 3.1. Due to the fact that the major advantage of the switched decap technique is on distributed regulation scenario which is seldom seen from the existing LDO design literatures, the LDO design with the switched decap technique would not receive a fair treatment in this comparison. Therefore, only the LDO design with the switched positioning capacitors is considered in this comparison.

Since the maximum voltage drop due to load variation is closely related to the transition time (t_r) of the load current, the comparison chooses the kind of figure of

merits (FOM) that takes t_r into account. FOM₁ in [9] is given by

$$\text{FOM}_1 = K \frac{\Delta V_{\text{out}} I_{\text{q}}}{\Delta I_{\text{L}}}, \quad (3.13)$$

where K is load current transition time ratio that is defined by

$$K = \frac{t_r \text{ used in the work being compared}}{\text{The smallest } t_r \text{ among all compared works}}. \quad (3.14)$$

The unit is Volt. However, this FOM does not reflect the fact that the output capacitor also has a significant impact on the voltage drop. And hence, we adopt another figure of merit that combines considerations in the above two as [25]

$$\text{FOM}_2 = C_o \cdot \text{FOM}_1. \quad (3.15)$$

The unit of this FOM is Coulomb. All the two FOMs are encouraged to be small.

As observed from the table, in the comparisons on both FOM₁ and FOM₂, the proposed switched positioning technique exhibits evident advantages while achieving the highest power efficiency.

3.4 Summary

An on-chip linear regulator with switched capacitor circuit is proposed. By switching a capacitor that pre-stores desired amount of charges into the LDO main circuit, the LDO manifests fast response to the transient load change. Comprehensive simulation results show that the regulator achieves as low dropout voltage as near 50mV and hence has high power efficiency of close to 90%. The transient performance is significantly improved by the auxiliary switched capacitor circuits. The quiescent power overhead of the auxiliary circuit is small, making it suitable for

low-voltage high-performance applications.

4. DESIGN OF DISTRIBUTED ON-CHIP REGULATORS WITH ENSURED STABILITY*

As previously discussed, the on-chip integration of voltage regulators and converters has emerged as a promising means to address many IC power delivery challenges. Both static and dynamic supply voltage droops can be reduced and the package resonance also can be suppressed. A step further towards the on-chip regulation is placing multiple regulators, e.g. LDOs, close to heavy noise sources on the die in a distributive manner (as illustrated in Fig. 4.1) [6, 21, 28]. The development from the centralized on-chip LDO structure to the distributed on-chip regulation structure is driven by two factors. It is first of all intuitive that in distributed structure the longest distance from any load to the current suppliers (LDOs) will be largely reduced, so is the associated static and dynamic voltage drop. As the scale of integration is reluctant to stop following Moor's Law, a foreseeable problem is the spacial imbalance of power supply due to degraded IR drop incurred by increased geometrical distance and increased current demand of the circuit. The distributed LDO structure is promising in this sense. Secondly, a centralized LDO structure is also facing already fairly important electromigration (EM) problem. This is because the power strips supplying the central LDO are stressed to deliver the total current demanded by the power domain. Distributed LDO structure will have the stress to split to each LDO and will allow more freedom to optimize power routing.

*Most part of this chapter is reprinted with permission from "Localized stability checking and design of IC power delivery with distributed voltage regulators" by S. Lai, B. Yan, & P. Li, 2013. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(9):1321–1334, Copyright [2013] by IEEE.

Some of material in this chapter are reprinted with permission from "Stability assurance and design optimization of large power delivery networks with multiple on-chip voltage regulators" by S. Lai, B. Yan, & P. Li, 2012. *Proceedings of 2012 International Conference on Computer-Aided Design*, 247–254, Copyright [2012] by IEEE/ACM.

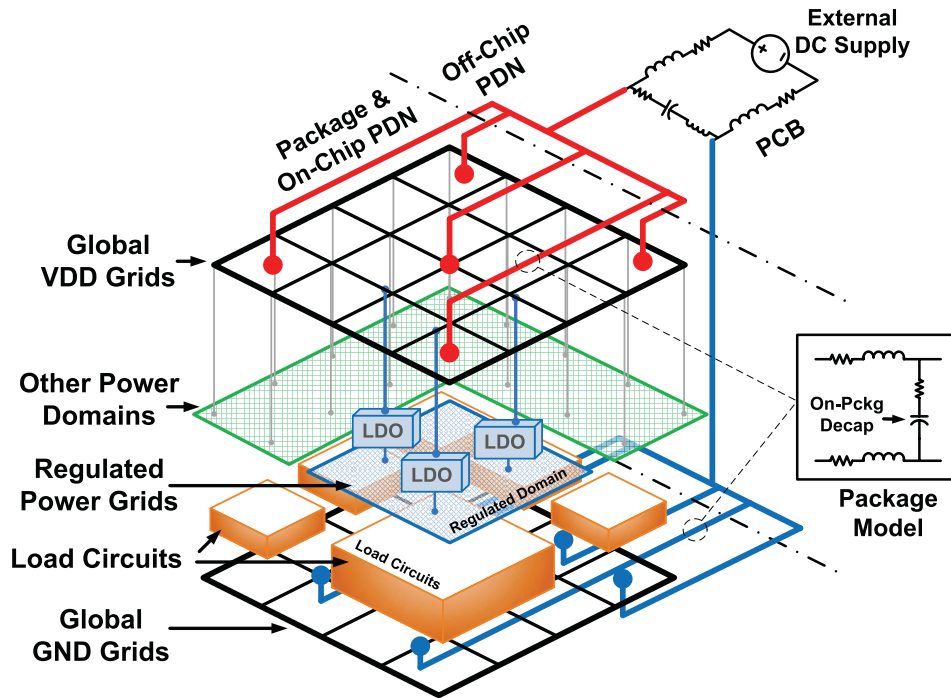


Figure 4.1: Illustration of the power delivery network with distributed on-chip regulators.

4.1 The First Glance on Distributed Regulator Design

While integrating multiple on-chip voltage regulators to facilitate distributed active regulation is appealing and represents a significant ongoing design trend, the design of such a distributed regulation system is not as easy and straightforward as it seems to be. As a starting point, one can always come up with a standalone regulator designed in the traditional way as discussed in previous sections, and deploy multiple such regulators in the distributed structure.

The first killjoy is instability of the distributed structure. While stability is a well solved problem in the traditional centralized regulation system, the autonomous nature of active voltage regulators placed closed to each other could indeed render the PDN unstable even though they are stable respectively on their own, which will

be further demonstrated in Section 4.2.

An unstable PDN can manifest itself with sustained supply voltage oscillations, which may cause severe degradation of circuit performance or even chip operation failure. But the understanding of the stability of the network is very challenging due to the complex interactions between multiple active regulators and the immense size of the passive RLC sub-network.

Traditional small-signal stability analysis methods, as commonly employed in the standard LDO design process, are incapable of addressing the above challenge; they are either unable to capture the effects of inter-regulator loops, a key characteristic of multi-LDO regulated PDNs, or computationally intractable for PDNs with a practical size. Phase/gain margins are commonly used by analog designers for checking the small-signal stability of analog circuits including LDOs. However, these methods are single-loop based, i.e., it is assumed that there exists only one dominant (outer) loop in the design and the stability analysis only pertains to this loop. In practice, phase or gain margins are computed mostly when the circuit is loaded with a simple lumped capacitor. In this chapter, it will be shown that the use of phase margin can lead to completely misleading prediction of the stability of PDNs regulated by distributed LDOs.

On the other hand, in theory, the small-signal stability of a PDN may be thoroughly detected by finding existence of any right half-plane poles (RHP) of the closed loop system. However, this has a computational cost that is cubic in the size of the PDN and is impractical for practical designs. The computational complexity exacerbates in an iterative design process in which LDOs may be tuned multiple times before it is finally pinned down.

It is pressing to develop a computationally tractable network stability-ensured method to facilitate the design of distributed regulation system.

In this chapter, a modeling and partitioning strategy of the PDN to describe the system-wide feedback loops in the PDN will be presented, making it possible to reason about stability while tracking the interactions between the LDOs and the passive RLC sub-network. Putting the proposed approach on a firm theoretical footing, this work then adopts and extends the recently emerged hybrid stability theory (HST) [22], developed originally for multi-variable robust control, to examine the stability of PDNs with multiple LDOs. It is rigorously proven that under a set of practical conditions a PDN is guaranteed to be stable. The use of HST allows us to combine the notions of small gain (of system-level loops) and passivity (of individual regulators) to impose more relaxed sufficient conditions for guaranteeing the network stability. Moving one step further, it is managed to leverage the proposed HST framework to achieve the goal of localized stability checking. That is, with one time AC simulation of the passive sub-network, the stability of the complete PDN can be determined by locally characterizing the gain and the passivity of individual LDOs. While the passivity of analog circuits and gains of system-level loops are unfamiliar concepts to typical analog designers, this chapter will show how these properties may be leveraged to render feasible stability checking of a given large PDN and empower practical iterative LDO design in a typical analog design flow.

After an effective and efficient stability checking method at hand, the next step will be to develop localized LDO design techniques that guarantee the stability of the PDN while achieving good power delivery/regulation performance. This work achieves that goal by first defining a hybrid stability margin (HSM) concept that numerically assesses the network stability and guides the trade-offs between stability and other design specifications for the optimization of LDOs. One key aspect of the proposed design methodology is the investigation of circuit level design techniques, e.g. proper choice of LDO topologies, and introduction of additional design

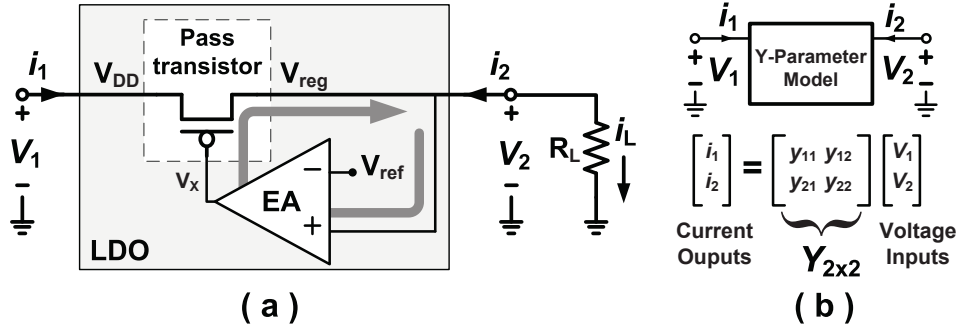


Figure 4.2: (a) The generic LDO structure. (b) The two-port Y-parameter model of the generic LDO.

freedoms, which may lead to the most efficient guarantee of the network stability and the best tradeoffs with other design specifications. Based upon these developments, transistor-level regulator design parameters are identified that are key to the system-wide stability and develop an automated localized LDO design flow that jointly optimizes several important design specifications pertaining to stability, voltage regulation and power efficiency.

4.2 Investigation of PDN Stability

While it is very attractive to apply the distributed on-chip voltage regulation in a power delivery network (PDN), the stability of the whole system has to be guaranteed in the first place.

Stability is a general concern for any feedback control systems. For example, Fig. 4.2(a) depicts a generic LDO circuit structure which includes a pass transistor whose pass resistance is dynamically tuned by a negative feedback loop (referred to as the ‘local loop’ in the rest of this chapter) to counteract the change of the output voltage (V_{reg}). As a result, V_{reg} can be maintained at a preset value regardless of either fluctuations of the global supply voltage, V_{DD} , or variations of the load current, i_L . Due to the feedback control, however, the circuit can be potentially unstable

and circuit designers need to perform stability checking to verify the LDO’s stability.

Unfortunately, it is particularly challenging in the LDO circuit design phase to guarantee the stability of the entire large-scale PDN in question, due primarily to the large network size and the complicated interactions among the on-chip regulators as well as the surrounding passive RLC sub-network. The classical stability-checking approaches traditionally used for regulator design can be categorized into two groups: the ones that check, via expensive pole analysis, the existence of right-half-plane (RHP) poles of the closed-loop transfer function of a system, and the ones (e.g., phase margin, or Nyquist plot) that leverage characteristics of the open-loop transfer function of a system.

The methods in the first group are not applicable to this multi-LDO PDN design for the following reasons. To search for right-half-plane (RHP) poles of the closed-loop transfer function of the system, an eigenvalue problem needs to be solved with a runtime cost of $O(N^3)$, where N is the number of nodes in the network. It is even daunting that every time the LDO design is modified, the eigenvalue problem has to be solved once again. Considering that the power delivery networks in practical designs can easily have millions of circuit nodes, the prohibitive cost involved will obviously disqualify this type of methods as a practical option. Another disadvantage is that, even if the system’s instability has finally been identified, designers are usually left with no clue on how to fix the problem.

The second group of approaches, while perfectly suitable for single-input and single-output systems (SISO) and widely adopted by regulator designers, can hardly be applied to the stability problem under discussion. For example, the classical phase margin method, inspecting narrowly at characteristics of the ‘local loop’ inside an individual LDO as illustrated in Fig. 4.3(a), cannot find in the PDN a major loop to open for stability analysis, as illustrated in Fig. 4.3(b). In this scenario, not only

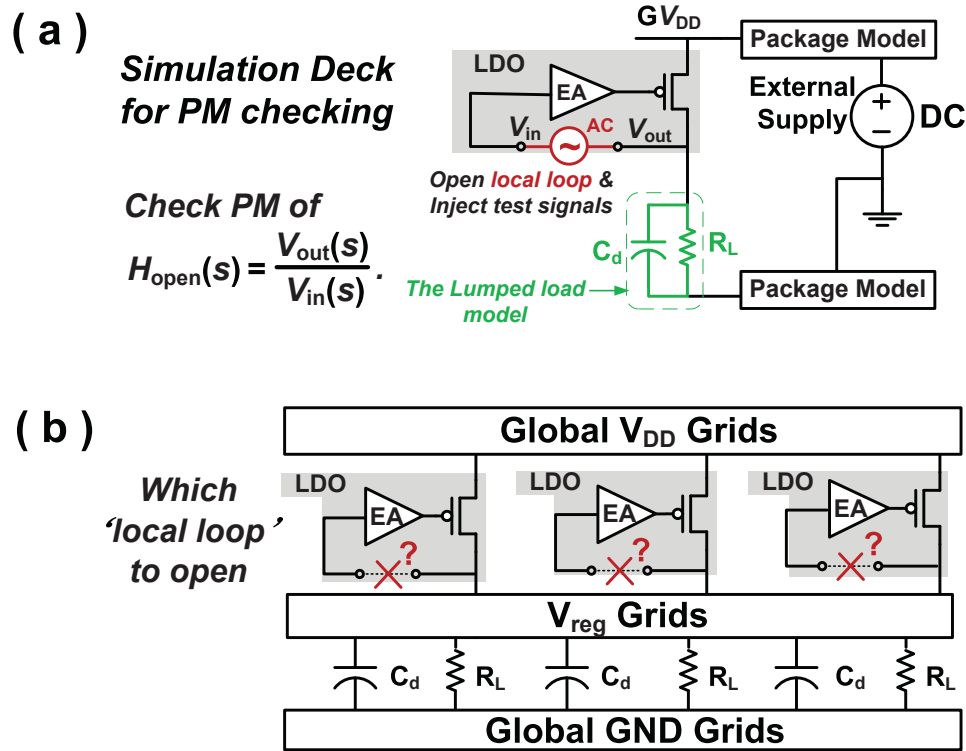


Figure 4.3: Illustration of the problem when applying the open-loop methods to the stability problem under discussion. (a) Traditional stability checking in the design of a single LDO. (b) Problem illustration when applying open-loop method to the PDN.

that the LDO under design is also a part of the load to itself, but there are also multiple inter-LDO feedback loops as depicted in Fig. 4.4 which may be accused of causing instability of the network but is invisible to the method. Therefore, it makes the stability conclusion given by this type of methods not reliable any more.

To further illustrate this point, a realistic LDO design [25] is adopted as an example. We first designed the LDO in the traditional manner, achieving a phase margin of about 110° under a typical load capacitor (decap) of about 100pF and above 40° under a wide range of decap from 1pF to 1nF, which is interpreted as a highly stable design by the conventional stability checking method. Interestingly, it

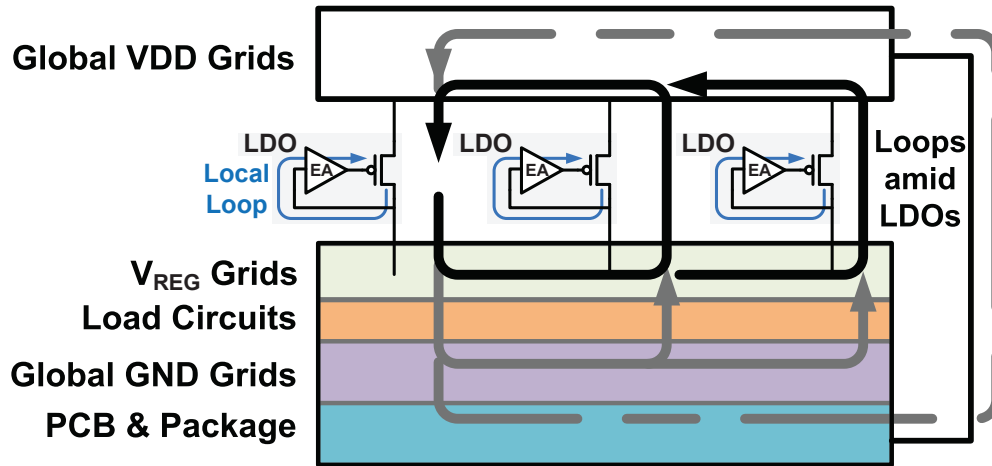


Figure 4.4: Illustration of the inter-LDO loops in the PDN.

was found that when multiple copies of this LDO design are integrated into a PDN, the entire network is possibly unstable. To gradually disclose how the stability of the PDN in this example is destroyed, the network stability was examined every time we added one more LDO into the PDN. To keep the loads to each LDO roughly constant as more LDOs were added, the total amounts of decap and load current in the PDN are increased proportionally with the number of LDOs. As the size of the power grids in this illustrative example is intentionally made small (about 20 nodes with the parasitic grid resistance being a few hundreds of $m\Omega$), the thorough pole analysis on the whole network can be applied to check the stability. The package model given in [33] is adopted in this example. Fig. 4.5 demonstrates the problematic pole movements extracted from the analysis results. It is observed that as the number of LDOs in the PDN increases there are a pair of complex poles moving from the left half of the s-plane toward the right half-plane (i.e., from stable region toward instability), which is further confirmed by the corresponding transient simulation results shown in Fig. 4.6 which demonstrates heavy oscillation of the local supply

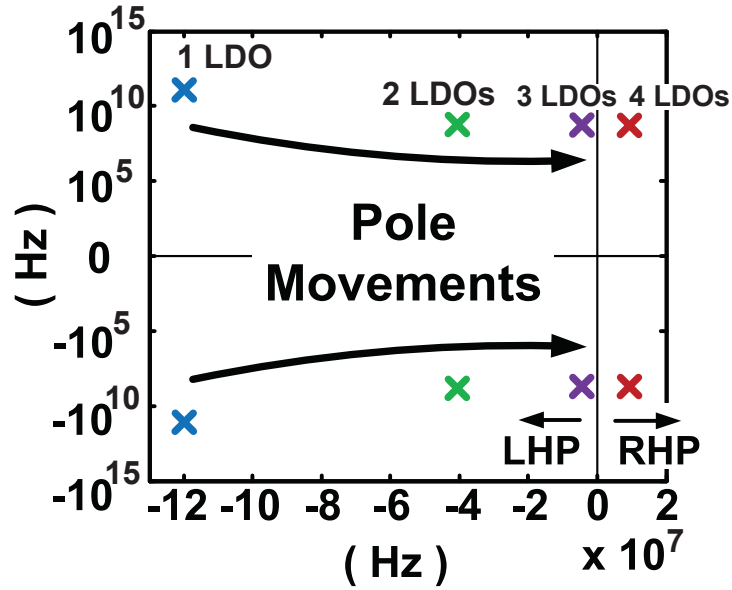


Figure 4.5: Pole analysis results that demonstrate a instability-arousing pole movement (each cross represents a pole location).

voltage (V_{reg}) occurs when there are four LDOs in the PDN.

The above example clearly shows that achieving a high phase margin for each stand-alone LDO does not provide any guarantee for the stability of the integrated network. One of the major reasons for the phase margin method to fail is the inappropriate handling of signal loops in the network. The phase-margin based LDO stability analysis is only positioned to capture the interaction between one LDO and the rest of the network. As already pointed out, this treatment is unable to take the interactions among the LDOs into account. Not surprisingly, inspecting one LDO at a time while assuming the rest of the circuit may be modeled as a simple passive load, as implied in the application of the phase margin method, can lead to erroneous conclusions about network stability. Therefore, building a sensible network model that captures all stability-endangering signal loops is the first critical step on tackling the problem.

4.3 PDN Partitioning and Modeling

Partitioning is a common practice in the divide-and-conquer paradigm for solving large complex problems. Towards the goal of establishing a theoretically rigorous and practically useful treatment of the stability challenge, an effective way of partitioning and modeling of the PDN is first presented, which facilitates the identification of a complete set of system-wide signal-flow loops responsible for stability of the entire network.

4.3.1 Concepts of Proposed Partition and Modeling

The PDN can be partitioned in a way to properly account for all key signal paths at the network level, which contribute in a significant way to stability. This

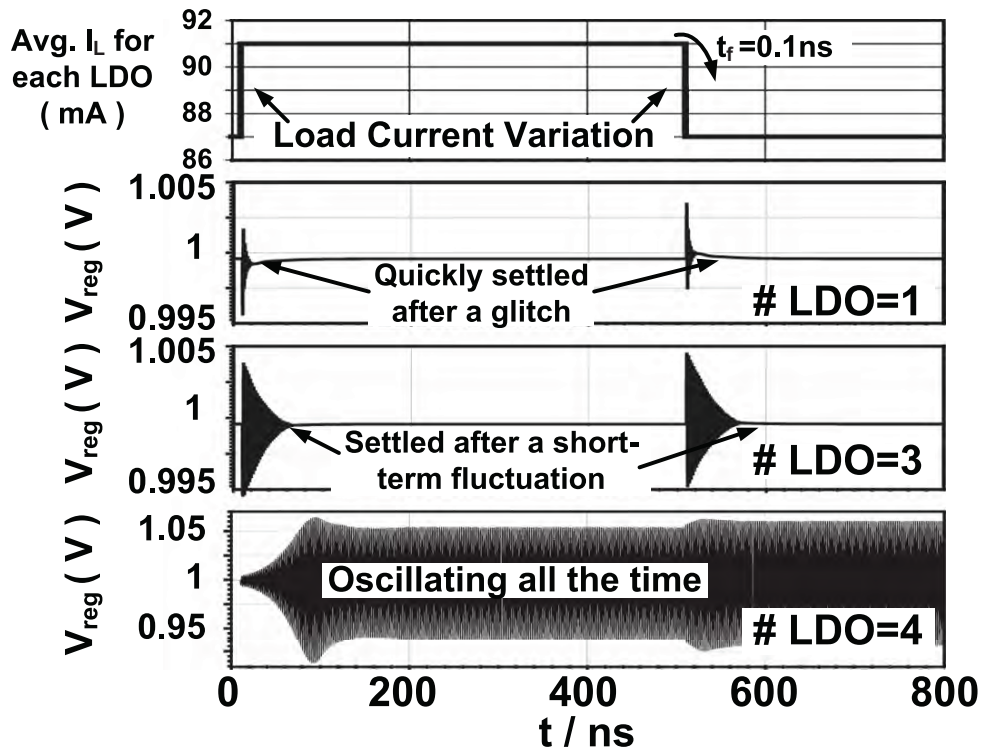


Figure 4.6: Transient analysis results that demonstrate the stability problem.

requires us to move away from SISO based approaches as typically adopted by analog designers and take a multi-port based modeling approach.

Furthermore, partitioning shall be done in a way to facilitate the iterative design process in which network stability may be checked multiple times as the LDOs are tuned. Thus, it is highly desirable to detach the bulky passive RLC sub-network, which requires a great effort to analyze, from this iterative design process. This leads us to considering a partition that separates the passive RLC sub-network from all the LDOs, resulting in two multi-port sub-systems: one that contains only the regulators, and the other comprised merely of the passive RLC sub-network serving as the load to the LDOs. This partitioning strategy has an appealing advantage. As will be shown later in this chapter, it allows us to spend only a one-time cost to characterize the passive sub-network using AC analysis, based on which stability constraints that are local to each individual LDOs are extracted prior to the iterative LDO design process. In the subsequent design process, these extracted local stability constraints are used to drive the optimization of each LDO while guaranteeing the stability of the complete network.

Note that in the proposed partitioning scheme, all the LDOs are grouped in a single multi-port sub-system despite the fact that their physical locations are spread out. In other words, the partitioning is done not based on physical vicinity, rather to electrically separate the LDOs from the passive sub-network.

4.3.2 The Proposed Network Partition and Modeling

The proposed partition of the PDN with n on-chip LDOs is illustrated in Fig. 4.7, where the dashed lines represent the partition boundaries, and the two subsystems are respectively represented by block \mathbf{G} that only contains the LDOs, and the passive sub-network \mathbf{Z} which is enclosed in the U-shaped dashed box. Between \mathbf{G} and \mathbf{Z}

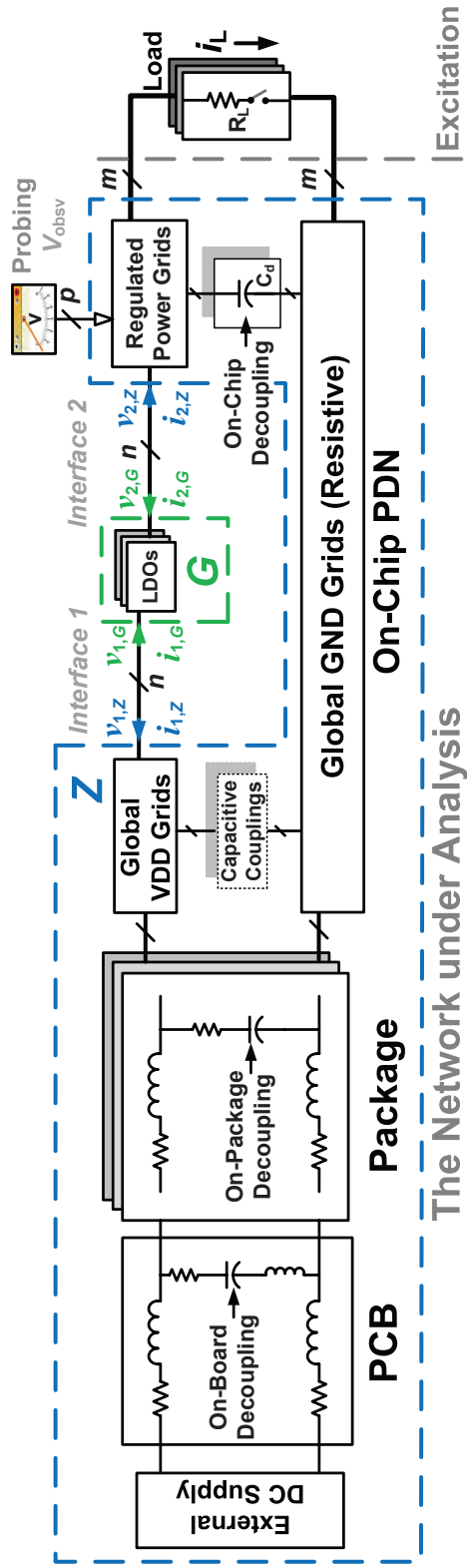


Fig. 4.7: Partition of the PDN model.

there are two types of interfaces corresponding to the V_{DD} ports and V_{reg} ports of the LDOs. Therefore, for n on-chip LDOs in the PDN, each subsystem has $2n$ interfacing ports. Besides the interfacing ports, block \mathbf{Z} is also connected to both the PDN's excitation inputs, which are the variations of the load currents \mathbf{i}_L , and the whole system's outputs, which can be any nodal voltages of interest on the power grids (\mathbf{V}_{obsv}).

As the LDOs are commonly linearized and, in order to utilize the signal-flow graph, this work models the LDO block by a $2n$ -port \mathbf{Y} -parameter model with each LDO described by the 2×2 Y-parameter matrix shown in Fig. 4.2(b). The transfer matrix of block \mathbf{G} is then given by

$$\underbrace{\begin{bmatrix} i_{1,1} \\ i_{2,1} \\ \vdots \\ i_{1,i} \\ i_{2,i} \\ \vdots \\ i_{1,n} \\ i_{2,n} \end{bmatrix}}_{2n \times 1} = \underbrace{\begin{bmatrix} \mathbf{Y}_{2 \times 2,1} & \dots & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{Y}_{2 \times 2,i} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{Y}_{2 \times 2,n} \end{bmatrix}}_{\mathbf{G}_{2n \times 2n}} \underbrace{\begin{bmatrix} v_{1,1} \\ v_{2,1} \\ \vdots \\ v_{1,i} \\ v_{2,i} \\ \vdots \\ v_{1,n} \\ v_{2,n} \end{bmatrix}}_{2n \times 1}, \quad (4.1)$$

where $i_{j,k}$ ($j = 1, 2; k = 1, 2, \dots, n$) represents the j -th port current of the k -th LDO, and similarly $v_{j,k}$ is its port voltage. It is worth to note that because of the way in which block \mathbf{G} is constructed, the LDOs are isolated to each other; accordingly, the matrix $\mathbf{G}_{2n \times 2n}$ is block diagonal with the i -th block being the 2×2 Y-parameter matrix of the i -th LDO, as can be observed from (4.1). The computational benefit

from this property will be discussed in Section 4.5.

The PDN then can be abstracted into a block diagram of a feedback control system shown in Fig. 4.8(a), where block \mathbf{G} interfaces with block \mathbf{Z} through $2n$ voltage signals and $2n$ current signals. Further, the excitation inputs and the outputs can be removed for stability analysis because for LTI systems, stability is an intrinsic property regardless of external system inputs or outputs. Thereby during stability analysis, block \mathbf{Z} can be reduced into block \mathbf{H} which only retains the interfacing ports with \mathbf{G} . By modeling \mathbf{H} with a $2n$ -port Z-parameter model whose inputs are $2n$ currents (\mathbf{i}_H) with the outputs being $2n$ voltages (\mathbf{v}_H), we simplify the system model into the one as shown in Fig. 4.8(b), to which stability theory can be readily applied.

By modeling the LDO block and the passive sub-network in the above way, the system's signal-flow graph can be built as shown in Fig. 4.9, where every electrical quantity (i.e., a current or voltage) or a "node" is only dependent on the upper-stream node. Therefore, when it is partitioned as illustrated by the dash-dotted line in Fig. 4.9, the output signals of the two partitions, namely \mathbf{i}_G and \mathbf{V}_H , are

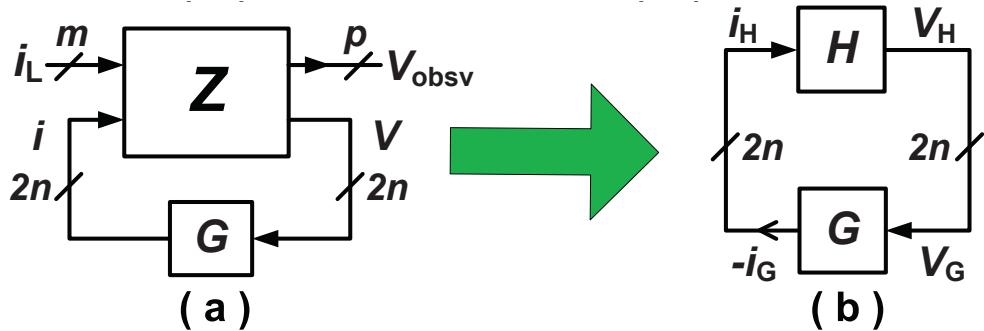


Figure 4.8: PDN modeling with the system-wide feedback loop. (a) The complete PDN model with system inputs and outputs. (b) The PDN model reduced to contain only signals pertaining to the stability issue.

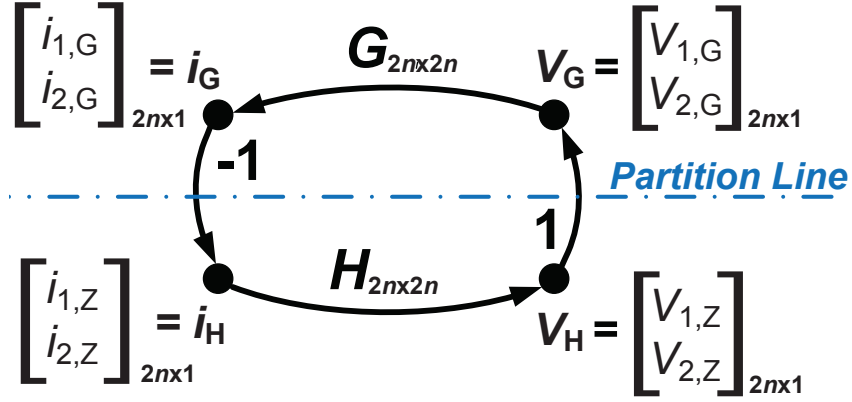


Figure 4.9: The signal-flow graph of the system. ($i_{1,G}, i_{2,G}, V_{1,G}, V_{2,G}$ and $i_{1,Z}, i_{2,Z}, V_{1,Z}, V_{2,Z}$ are the same as in Fig. 4.7.)

respectively determined only by the corresponding inputs (namely \mathbf{V}_G and \mathbf{i}_H) as well as the partition transfer matrices $\mathbf{G}_{2n \times 2n}$ and $\mathbf{H}_{2n \times 2n}$. In this way, the stability evaluation of the LDO block can be confined within the partition itself without any overlook of loading effect between the two partitions, which is important to the rigorousness of the proposed method.

From Fig. 4.9, the system-wide multi-variable feedback loop is identified starting from the inputs (\mathbf{i}_H) of block \mathbf{H} to its outputs (\mathbf{V}_H), which are directly fed to block \mathbf{G} , and the loop finally ends at the outputs (\mathbf{i}_G) of \mathbf{G} . As the positive directions of port currents are defined as flowing into the corresponding blocks, \mathbf{i}_G and \mathbf{i}_H are of the same magnitude but the opposite directions, i.e., the loop is a negative feedback.

4.4 The Theoretical Framework

Toward a rigorous theoretical guarantee rather than an empirical educated guess about the PDN stability, in this section the development of the theoretical framework that is not only suitable for effective and efficient stability checking, but offers more flexibility for achieving superior system performance, is laid out.

An ideal stability checking method shall have the following desirable properties: 1) it should be able to handle multi-input and multi-output (MIMO) feedback systems such as the one in Fig. 8 (b); and 2) it needs to avoid or at least greatly reduce the analysis cost associated with the large passive-network (block \mathbf{H}) in order to be computationally efficient; 3) the stability conditions adopted in the method shall not lead to poor regulation performance.

Based on the above discussion, the use of a combination of passivity and small gain principles offers an appealing solution to the stability problem at hand. This approach goes naturally with the network partitioning presented in the previous subsection and facilitates a localized checking methodology. Prior to delving into this theoretical framework, several key concepts and relevant mathematical backgrounds [20, 22, 27] are first introduced, followed by the theoretical framework specifically developed for the targeted PDNs.

4.4.1 Preliminaries

The stability concerned in this work is referred to as signal convergence in terms of the norm in L_2 -space. L_2 -space is the space of square integrable functions defined by $L_2 = \{ \mathbf{v} : \mathbb{R}^+ \mapsto \mathbb{R}^m \mid \int_0^\infty \mathbf{v}^T(t) \mathbf{v}(t) dt < \infty \}$ where \mathbf{v} is an arbitrary vector function of time and \mathbf{v}^T is its transpose. The L_2 -space is a Hilbert space, where the inner product defines the norm

$$\langle \mathbf{w}, \mathbf{v} \rangle = \int_0^\infty \mathbf{w}^T(t) \mathbf{v}(t) dt, \quad \|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} \quad (4.2)$$

where $\mathbf{v} \in L_2$, $\mathbf{w} \in L_2$ and $\langle \cdot, \cdot \rangle$ is the inner product.

Definition 1. (*System gain*) Consider a general square system with an input $\mathbf{w}(t) \in L_2$ and an output $\mathbf{y}(t) \in L_2$ mapped through an operator $\mathbf{M} : L_2 \rightarrow L_2$, the induced

L_2 -gain, or simply the system gain, is defined by

$$\gamma = \sup_{\forall \mathbf{w} \in L_2, \mathbf{w} \neq 0} \|\mathbf{y}\|_2 / \|\mathbf{w}\|_2. \quad (4.3)$$

A system possesses ‘finite gain’ if there exists $0 < \gamma < \infty$ such that

$$\gamma \langle \mathbf{w}, \mathbf{w} \rangle \geq \gamma^{-1} \langle \mathbf{y}, \mathbf{y} \rangle, \quad \forall \mathbf{w} \in L_2. \quad (4.4)$$

For any LTI system, the induced L_2 -gain is equivalent to the \mathcal{H}_∞ -norm of the system transfer matrix, \mathbf{M} , which is defined by $\|\mathbf{M}\|_\infty = \max_{0 \leq \omega < \infty} \|\mathbf{M}(j\omega)\|_2$, and

$$\|\mathbf{M}(j\omega)\|_2 = \max_i [\lambda_i(\mathbf{M}^H(j\omega)\mathbf{M}(j\omega))]^{\frac{1}{2}}, \quad (4.5)$$

where $\lambda_i(\mathbf{M})$ denotes the i -th eigenvalue of \mathbf{M} , and \mathbf{M}^H denotes the complex conjugate transpose of \mathbf{M} .

Definition 2. (*Passive systems*) A general square system with an input $\mathbf{w}(t) \in L_2$ and an output $\mathbf{y}(t) \in L_2$ mapped through the operator $\mathbf{M}: L_2 \rightarrow L_2$ is passive if there exist constants $\delta \geq 0$ and $\epsilon \geq 0$ such that $\forall \mathbf{w}$,

$$\langle \mathbf{w}, \mathbf{y} \rangle \geq \delta \langle \mathbf{w}, \mathbf{w} \rangle + \epsilon \langle \mathbf{y}, \mathbf{y} \rangle. \quad (4.6)$$

Further, if $\delta > 0$, then the system is called *input strictly passive*; if $\epsilon > 0$, then the system is *output strictly passive*; the system is *very strictly passive* if both $\delta > 0$ and $\epsilon > 0$. Based on (4.3) and (4.4), it can be easily derived that a system that is already ‘input strictly passive’ with finite gain is ‘output strictly passive’, and hence is ‘very strictly passive’.

The passivity of LTI systems can also be examined in the frequency domain. Consider that \mathbf{M} is an LTI system which has a minimal realization that is asymptotically stable; then we have [20]:

i) \mathbf{M} is passive if and only if its transfer matrix satisfies $\mathbf{M}(j\omega) + \mathbf{M}^T(-j\omega) \geq 0, \forall \omega \in \mathbb{R}$;

ii) \mathbf{M} is input strictly passive if and only if its transfer matrix satisfies that $\exists \delta > 0, \mathbf{M}(j\omega) + \mathbf{M}^T(-j\omega) \geq \delta \mathbf{I}, \forall \omega \in \mathbb{R}$, i.e., all eigenvalues of $\mathbf{M}(j\omega) + \mathbf{M}^T(-j\omega)$ are greater than or equal to δ .

Unfortunately, for many systems, a passive input-output map defined by (4.6) does not always exist. When a system's passive input-output relationship does not hold for a certain input case, we say that 'passivity violation' happens. In particular, for LTI systems, if there exists a frequency ω where the condition $\mathbf{M}(j\omega) + \mathbf{M}^T(-j\omega) \geq 0$ is not met, then passivity violation happens.

On the other hand, we also define the passiveness of the system with passivity violations as 'local passivity'. Before rigorously defining it, we first define a 'passivity filter' $\mathcal{A}: L_2 \rightarrow L_2$, which is a causal convolution operator; also we define $\mathbb{A} = \mathcal{A}\mathbf{I}$, where \mathbf{I} represents identity matrix.

Definition 3. (*Local Passivity*) A general square system with an input $\mathbf{w}(t) \in L_2$ and an output $\mathbf{y}(t) \in L_2$ mapped through the operator $\mathbf{M}: L_2 \rightarrow L_2$ is locally passive, if there exists a passivity filter \mathbb{A} and constants $\delta \geq 0$ and $\epsilon \geq 0$, such that

$$\langle \mathbb{A}\mathbf{w}, \mathbb{A}\mathbf{y} \rangle \geq \delta \langle \mathbb{A}\mathbf{w}, \mathbb{A}\mathbf{w} \rangle + \epsilon \langle \mathbb{A}\mathbf{y}, \mathbb{A}\mathbf{y} \rangle. \quad (4.7)$$

If $\exists \delta > 0$, and $\epsilon > 0$ that satisfy (4.7), the system is referred to as locally very strictly passive.

For LTI systems, denoting the frequency set where the system meets the passivity condition by $\Omega \triangleq \{\omega \in \mathbb{R} | \mathbf{M}(j\omega) + \mathbf{M}^T(-j\omega) \geq 0\}$, we define a frequency-dependent

function $\alpha(\omega): \mathbb{R} \rightarrow \{0, 1\}$ as [22]

$$\alpha(\omega) = \begin{cases} 1, & \omega \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (4.8)$$

Let $A(s)A(-s)$ be the spectral factorization of the Laplace transform of the inverse Fourier transform of $\alpha(\omega)$. Then we have

$$\alpha(\omega) = A(j\omega)A(-j\omega). \quad (4.9)$$

Further, the time domain equivalent of $A(s)$ is a causal convolution operator $\mathcal{A}: L_2 \rightarrow L_2$, referred to as the frequency selection operator in the rest of the chapter. Obviously, \mathcal{A} can be a passivity filter for LTI cases. Again, \mathbb{A} is also defined accordingly and has its Fourier transformation $\mathbf{A}(j\omega) = A(j\omega)\mathbf{I}$. Note that if an LTI system is passive, then it is locally passive with respect to any Ω including $\Omega = \{\omega | \omega \in \mathbb{R}\}$.

4.4.2 Two Classical Stability Theorems

Considering the Barkhausen oscillation conditions, it is intuitive that if the loop gain of a feedback system is less than one, then any oscillation through the loop will finally be attenuated and hence the system remains stable. The intuition leads us to the *small-gain theorem*, a classical stability theorem for general feedback systems.

Given the feedback system in Fig. 4.8(b) and the system gain defined by (4.3), the small-gain theorem states the following result [27]:

Theorem 1. (*Small-gain theorem*) *The negative feedback interconnection of the sub-systems $\mathbf{G}: L_2 \rightarrow L_2$ and $\mathbf{H}: L_2 \rightarrow L_2$ is L_2 -stable if the product of the gains of the two sub-systems is strictly less than one.*

That is, the whole system is L_2 -stable as long as $\gamma_{\mathbf{G}}\gamma_{\mathbf{H}} < 1$, where $\gamma_{\mathbf{G}}$ and $\gamma_{\mathbf{H}}$ are

respectively the gain of blocks \mathbf{G} and \mathbf{H} . As such, the theorem allows $\gamma_{\mathbf{G}}$ and $\gamma_{\mathbf{H}}$ to be separately evaluated through (4.5). Therefore, if one sub-system is fixed (as the passive sub-network) while the design of the other one is in process (as the LDOs), the gain evaluation on the fixed sub-system can be done once for all and be used to assist the iterative design of the other subsystem. Thus, the stability of the entire system can be checked locally on the other subsystem.

The small gain theorem, however, utilizing merely gain information of the sub-systems, tends to give a Pyrrhic victory for ensuring stability. This is because one of the sub-systems (e.g., the passive sub-network in the case) once has a very high gain at any operational frequencies of interest, the other one (e.g., the LDO block) would be mandated by the theorem to have a rather low gain, resulting in poor closed-loop system performance.

In addition to exploiting the characteristics of system gains of the sub-systems, another property that LDO designers may easily resort to is the phase information of the open-loop transfer function of an SISO system. For MIMO systems, *passivity* can be deemed, in some sense, as a quantity that correlates with the phase information of the system transfer matrix. Thus, as efforts are made trying to relax the harsh constraint on the gains (performance) imposed by the small-gain theorem, the *passivity* property is considered as another avenue to ensure stability.

The *passivity theorem* states the following useful result for the system in Fig. 4.8(b) [27]:

Theorem 2. (*Passivity theorem*) *The negative feedback interconnection of the sub-systems $\mathbf{G}:L_2 \rightarrow L_2$ and $\mathbf{H}:L_2 \rightarrow L_2$ is L_2 -stable if one system is passive while the other is very strictly passive.*

The theorem implies that the whole system is L_2 -stable if both $\epsilon_{\mathbf{G}} \geq 0, \delta_{\mathbf{G}} \geq 0$ and $\epsilon_{\mathbf{H}} > 0, \delta_{\mathbf{H}} > 0$, where $\epsilon_{\mathbf{G}}$ and $\delta_{\mathbf{G}}$ respectively represent the ϵ and δ of block \mathbf{G}

as defined in (4.6), so do $\epsilon_{\mathbf{H}}$ and $\delta_{\mathbf{H}}$ of block \mathbf{H} . Similar to the small-gain theorem, the passivity of each sub-system can also be checked separately.

While a system with only passive elements are necessarily passive, a system containing active elements cannot usually be passive. Therefore, the passivity theorem alone cannot be the silver bullet either. In fact, it is more often the case that analog circuits (such as regulators) behave like a passive system over a certain frequency range, suggesting potential good use of local passivity for ensuring stability.

4.4.3 Hybrid Stability Theorem

Recently, stability theorems that simultaneously exploits small gain and passivity properties of a general system, have emerged [22,23]. In particular, a *hybrid stability theorem* has been proposed to make use of the local passive behaviors. If a general system has passivity violations, the ‘finite gain’ property is instead exploited for stability by the theorem [22].

Theorem 3. (*Hybrid stability theorem*) *The negative feedback interconnection of the sub-systems $\mathbf{G}:L_2 \rightarrow L_2$ and $\mathbf{H}:L_2 \rightarrow L_2$ is L_2 -stable if the following three conditions are met: 1) $\exists \epsilon_{\mathbf{G}} \geq 0, \delta_{\mathbf{G}} \geq 0$ and $\exists \epsilon_{\mathbf{H}} \geq 0, \delta_{\mathbf{H}} \geq 0$, such that \mathbf{G} and \mathbf{H} are both locally passive with respect to a common passivity filter \mathbb{A} ; 2) $\epsilon_{\mathbf{G}} + \delta_{\mathbf{H}} > 0$ and $\epsilon_{\mathbf{H}} + \delta_{\mathbf{G}} > 0$; 3) when passivity violation happens, $\gamma_{\mathbf{G}}\gamma_{\mathbf{H}} < 1$ holds.*

While providing a sufficient condition for stability, Theorem 3 nevertheless offers much greater design freedom in achieving superior closed-loop performance by combining the two previous basic stability theorems.

4.4.4 Hybrid Stability Framework for PDNs

Based upon the above general stability theory, this work develops a specific hybrid stability framework for PDNs. The proposed framework is based on the following

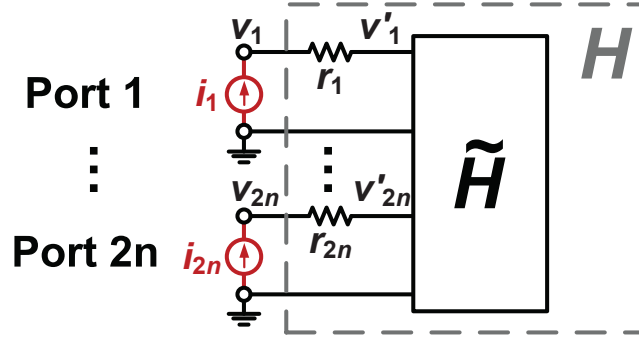


Figure 4.10: The illustration of the serial resistance at each port of the \mathbf{H} block.

two key observations of any realistic PDN of concern. LDOs are connected to the passive subnetwork (e.g., the global VDD grids and the regulated power grids in Fig. 4.7) through resistive metal wires and vias, which contribute to non-zero input serial resistance of the corresponding ports of the passive subnetwork as illustrated by the resistors $r_1 \dots r_{2n}$ in Fig. 4.10. Note that the impedance model of the passive subnetwork is denoted as block \mathbf{H} in the figure. Furthermore, the system gain of the passive sub-network in a realistic PDN, i.e., $\|\mathbf{H}(j\omega)\|_\infty$, cannot reach infinity, i.e., it is always upper bounded.

By virtue of the above observations, the following important property of the passive sub-network in such a PDN can be derived.

Property 1. *The passive sub-network of Fig. 4.10 is very strictly passive.*

Proof. According to Definition 2, it is to be proven that for the realistic passive sub-network $\mathbf{H}: \mathbf{i}_H(t) \in L_2 \rightarrow \mathbf{v}_H(t) \in L_2$, $\exists \epsilon_H > 0$ and $\delta_H > 0$, such that $\langle \mathbf{i}_H(t), \mathbf{v}_H(t) \rangle \geq \epsilon_H \langle \mathbf{i}_H(t), \mathbf{i}_H(t) \rangle + \delta_H \langle \mathbf{v}_H(t), \mathbf{v}_H(t) \rangle$.

To begin with, we know that $2\langle \mathbf{i}_H(t), \mathbf{v}_H(t) \rangle = \{\langle \mathbf{i}_H(t), \mathbf{v}_H(t) \rangle + \langle \mathbf{v}_H(t), \mathbf{i}_H(t) \rangle\}$,

which, by Parseval's theorem, is equivalent to the expression

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) [\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)] \mathbf{i}(j\omega) d\omega. \quad (4.10)$$

As well known that an LTI RLC network is passive [26], the matrix $\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)$ is therefore positive semi-definite. If we denote the passive network excluding those input resistors r_i ($i = 1, 2, \dots, 2n$) by $\widetilde{\mathbf{H}}$, then $\widetilde{\mathbf{H}}$ is also passive. From Fig. 4.10, it can be easily inspected that

$$\mathbf{v}_H = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_{2n} \end{bmatrix} = \begin{bmatrix} v'_1 \\ v'_2 \\ \vdots \\ v'_{2n} \end{bmatrix} + \mathbf{R} \begin{bmatrix} i_1 \\ i_2 \\ \vdots \\ i_{2n} \end{bmatrix} = (\widetilde{\mathbf{H}} + \mathbf{R}) \mathbf{i}_H, \quad (4.11)$$

where $\mathbf{R} = \text{diag}\{r_1, r_2, \dots, r_{2n}\}$, where $r_i \in \mathbb{R}_+$ ($i = 1, 2, \dots, 2n$). Then we have $\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega) = \widetilde{\mathbf{H}}(j\omega) + \widetilde{\mathbf{H}}^T(-j\omega) + 2\mathbf{R}$. Therefore, for $\forall \mathbf{X} \in \mathbb{R}^{2n}$ and $\mathbf{X} \neq \mathbf{0}$, we have

$$\begin{aligned} & \mathbf{X}^H [\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)] \mathbf{X} \\ &= \mathbf{X}^H [\widetilde{\mathbf{H}}(j\omega) + \widetilde{\mathbf{H}}^T(-j\omega) + 2\mathbf{R}] \mathbf{X} \\ &= \underbrace{\mathbf{X}^H [\widetilde{\mathbf{H}}(j\omega) + \widetilde{\mathbf{H}}^T(-j\omega)] \mathbf{X}}_{\geq 0} + \underbrace{\mathbf{X}^H (2\mathbf{R}) \mathbf{X}}_{> 0} \\ &\geq \min_{k=1,2,\dots,2n} \{2r_k\} \mathbf{X}^H \mathbf{X} \\ &> 0. \end{aligned} \quad (4.12)$$

Since $\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)$ is continuous with respect to ω and, according to (4.12), is positive definite, exists $l_{min} = \inf_{\omega \in \mathbb{R}} \underline{\lambda}(\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)) > 0$, where $\underline{\lambda}(\cdot)$ means the minimum eigenvalue. Also since $\|\mathbf{H}(j\omega)\|_\infty$ is upper bounded, $s_{max} = \sup_{\omega \in \mathbb{R}} \|\mathbf{H}(j\omega)\|_2$ exists. Hence by selecting $\epsilon > 0$ and $\delta > 0$ that meet the inequality

$$l_{min} \geq \epsilon + \delta s_{max}^2 > 0, \quad (4.13)$$

we have

$$\begin{aligned} & \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) [\mathbf{H}(j\omega) + \mathbf{H}^T(-j\omega)] \mathbf{i}(j\omega) d\omega \\ & \geq \frac{l_{min}}{2\pi} \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) \mathbf{i}(j\omega) d\omega \\ & \geq \frac{1}{2\pi} (\epsilon + \delta s_{max}^2) \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) \mathbf{i}(j\omega) d\omega \\ & \geq \frac{1}{2\pi} \epsilon \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) \mathbf{i}(j\omega) d\omega + \frac{1}{2\pi} \delta \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) \mathbf{H}^T(-j\omega) \mathbf{H}(j\omega) \mathbf{i}(j\omega) d\omega \\ & \geq \frac{1}{2\pi} \epsilon \int_{-\infty}^{\infty} \mathbf{i}^H(j\omega) \mathbf{i}(j\omega) d\omega + \frac{1}{2\pi} \delta \int_{-\infty}^{\infty} \mathbf{v}^H(j\omega) \mathbf{v}(j\omega) d\omega. \end{aligned} \quad (4.14)$$

That is, $\exists \epsilon_{\mathbf{H}} = \epsilon/2 > 0$ and $\delta_{\mathbf{H}} = \delta/2 > 0$, such that $\langle \mathbf{i}_{\mathbf{H}}(t), \mathbf{v}_{\mathbf{H}}(t) \rangle \geq \epsilon_{\mathbf{H}} \langle \mathbf{i}_{\mathbf{H}}(t), \mathbf{i}_{\mathbf{H}}(t) \rangle + \delta_{\mathbf{H}} \langle \mathbf{v}_{\mathbf{H}}(t), \mathbf{v}_{\mathbf{H}}(t) \rangle$. \square

Base on Theorem 3 and Property 1, the following corollary is developed that serves directly as the theoretical foundation for the proposed localized stability checking method as well as the automated stability-aware system optimization presented in later sections.

Corollary 1. *The feedback interconnection of a sub-system $\mathbf{G}:L_2 \rightarrow L_2$ and a very strictly passive sub-system \mathbf{H} is L_2 -stable if at $\forall \omega \in \mathbb{R}$, either one of the following two conditions is met: 1) $\gamma_{\mathbf{G}}(j\omega)\gamma_{\mathbf{H}}(j\omega) < 1$; 2) $\mathbf{G}(j\omega) + \mathbf{G}^T(-j\omega) \geq 0$.*

Proof. Apparently, when applying Corollary 1 to this design scenario, the passive sub-network would be the sub-system \mathbf{H} according to Property 1, and the LDO

block would be sub-system \mathbf{G} . If the first condition (i.e., $\gamma_{\mathbf{G}}\gamma_{\mathbf{H}} < 1$) is met, then there is no need for block \mathbf{G} to be locally passive as prescribed by Theorem 3. To prove this corollary one only needs to show that if the transfer matrix $\mathbf{G}(s)$ satisfies the second condition (i.e., $\mathbf{G}(j\omega) + \mathbf{G}^T(-j\omega) \geq 0$) over some frequency range Ω , then there exist $\epsilon_{\mathbf{G}} \geq 0$ and $\delta_{\mathbf{G}} \geq 0$, such that block \mathbf{G} is locally passive with respect to Ω . On the other hand, according to Property 1, block \mathbf{H} is locally very strictly passive with respect to Ω . Therefore, there exist $\epsilon_{\mathbf{H}} > 0$ and $\delta_{\mathbf{H}} > 0$ satisfying $\epsilon_{\mathbf{G}} + \delta_{\mathbf{H}} > 0$ and $\epsilon_{\mathbf{H}} + \delta_{\mathbf{G}} > 0$.

Given the transfer matrix $\mathbf{G}(s)$, define a frequency set $\Omega \triangleq \{\omega \in \mathbb{R} | \mathbf{G}(j\omega) + \mathbf{G}^T(-j\omega) \geq 0\}$ and the corresponding $\alpha(\omega)$ as well as the corresponding frequency selection operator \mathbb{A} . We define the convolution operator $\mathbb{G}: \mathbf{v}_{\mathbf{G}}(t) \in L_2 \rightarrow \mathbf{i}_{\mathbf{G}}(t) \in L_2$ that corresponds to $\mathbf{G}(s)$. Then according to the positive semi-definiteness of $\mathbf{G}(\omega) + \mathbf{G}^T(-\omega)$, for $\forall \mathbf{v}_{\mathbf{G}}(t) \in L_2$, we have

$$\frac{1}{2\pi} \int_{\Omega} \mathbf{v}_{\mathbf{G}}^H(j\omega) [\mathbf{G}(\omega) + \mathbf{G}^T(-\omega)] \mathbf{v}_{\mathbf{G}}(j\omega) d\omega \geq 0. \quad (4.15)$$

By introducing $\alpha(j\omega)$ into the integral to convert the integration range to be from $-\infty$ to $+\infty$, (4.15) is turned into

$$\begin{aligned} & \frac{1}{2\pi} \left[\int_{-\infty}^{\infty} \mathbf{v}_{\mathbf{G}}^H(j\omega) \mathbf{G}^H(\omega) (\alpha(\omega) \mathbf{I}) \mathbf{v}_{\mathbf{G}}(j\omega) d\omega + \right. \\ & \left. \int_{-\infty}^{\infty} \mathbf{v}_{\mathbf{G}}^H(j\omega) (\alpha(\omega) \mathbf{I}) \mathbf{G}(\omega) \mathbf{v}_{\mathbf{G}}(j\omega) d\omega \right] \geq 0. \end{aligned} \quad (4.16)$$

By substituting (4.9) for $\alpha(j\omega)$ in (4.16) and by Parseval's theorem, we get

$$\langle \mathbb{A} \mathbf{v}_{\mathbf{G}}(t), \mathbb{A} \mathbb{G} \mathbf{v}_{\mathbf{G}}(t) \rangle + \langle \mathbb{A} \mathbb{G} \mathbf{v}_{\mathbf{G}}(t), \mathbb{A} \mathbf{v}_{\mathbf{G}}(t) \rangle = 2 \langle \mathbb{A} \mathbf{v}_{\mathbf{G}}(t), \mathbb{A} \mathbb{G} \mathbf{v}_{\mathbf{G}}(t) \rangle \geq 0, \quad (4.17)$$

i.e., $\exists \epsilon_{\mathbf{G}} \geq 0$ and $\delta_{\mathbf{G}} \geq 0$, such that $\langle \mathbb{A}\mathbf{v}_{\mathbf{G}}(t), \mathbb{A}\mathbf{i}_{\mathbf{G}}(t) \rangle \geq \epsilon_{\mathbf{G}} \langle \mathbb{A}\mathbf{v}_{\mathbf{G}}(t), \mathbb{A}\mathbf{v}_{\mathbf{G}}(t) \rangle + \delta_{\mathbf{G}} \langle \mathbb{A}\mathbf{i}_{\mathbf{G}}(t), \mathbb{A}\mathbf{i}_{\mathbf{G}}(t) \rangle \geq 0$. \square

With the theoretical foundation built, in the next section, the way to perform stability checking for the PDN based on Corollary 1 is demonstrated.

4.5 New Hybrid Stability Margin Concept And Efficient Stability Checking of the PDN

Based on the hybrid stability theorem and the corollary, a rigorous and efficient stability checking method is first come up with, followed by a proposal of a new hybrid stability margin that assesses the system's stability. Such that the stability checking method can be incorporated into an *automated* optimization flow. The computational cost of the method is also analyzed.

4.5.1 Stability Checking of the PDN

The stability of the entire PDN is examined according to a frequency-sampling approach. Given a set of P points $\omega_k, k = 1, 2, \dots, P$ sampled in the frequency range of interest, the passivity and gain conditions are evaluated at each frequency ω_k . If for all frequencies at least one condition is satisfied, then the stability of the system is guaranteed.

4.5.1.1 Passivity Evaluation

Given a total number of n LDOs in the network, the passivity of the LDOs block at ω_k , is evaluated by finding the the smallest eigenvalues of the $2n \times 2n$ matrix $\mathbf{G}(j\omega_k) + \mathbf{G}^H(j\omega_k)$.

More efficiently, the evaluation can be performed on one LDO at a time, thanks to the fact that the transfer matrix \mathbf{G} is block diagonal, a feature of the LDO model mentioned in Section 4.3.2. The 2×2 admittance matrix of the j -th LDO is denoted

as $\mathbf{Y}_j(j = 1, 2, \dots, n)$. Therefore, the passivity of \mathbf{G} is evaluated by finding the value $\lambda_{min}(j\omega_k)$ given by:

$$\lambda_{min}(j\omega_k) = \min_{i=1,2;j=1,2,\dots,n} \{\lambda_i(\mathbf{Y}_j(j\omega_k) + \mathbf{Y}_j^H(j\omega_k))\}. \quad (4.18)$$

If $\lambda_{min}(j\omega_k) \geq 0$, the LDO exhibits passivity at ω_k , otherwise, passivity violation occurs.

Note that, there is no need to perform such passivity check for the large-scale passive load sub-network.

4.5.1.2 System Gain Evaluation

To decouple the design of LDO from the passive network, the evaluations of the L_2 -gain of the two subsystems are separately performed and inequality $\|\mathbf{G}\|\|\mathbf{H}\| < 1$ is targeted. At ω_k , $\|\mathbf{G}(j\omega_k)\|_2$ and $\|\mathbf{H}(j\omega_k)\|_2$ are first calculated using (4.5). Again, as \mathbf{G} is block diagonal, $\|\mathbf{G}(j\omega_k)\|_2$ can be obtained by

$$\|\mathbf{G}(j\omega_k)\|_2 = \max_{j=1,2,\dots,n} \|\mathbf{Y}_j(j\omega_k)\|_2, \quad (4.19)$$

where $\|\mathbf{Y}_j(j\omega_k)\|_2$ is the j -th block corresponding to the j -th LDO.

If $\|\mathbf{G}(j\omega_k)\|_2\|\mathbf{H}(j\omega_k)\|_2 < 1$, the system passes the stability checking at ω_k .

4.5.1.3 The Cost of Evaluation

Due to the small size of the LDO circuit, the cost of the passivity and gain evaluation for each LDO is very low. The overall cost of evaluation is dominated by the evaluation of the gain of the large passive load network $\|\mathbf{H}(j\omega_k)\|_2$, which involves an AC analysis to determine the transfer matrix $\mathbf{H}(j\omega)$ at ω_k .

Given that the total number and locations of the LDOs are predetermined and

the passive load sub-network is fixed, the evaluation only needs to be done once. Whenever the design of LDO is tuned, it is only needed to recompute (4.18) and (4.19) for the stability checking, which is very efficient because of the small size of the LDO.

If there are P sampling points, n LDOs, and N nodes in the passive sub-network, the cost of AC analysis for the passive sub-network is $O(PN^\alpha)$, given $n \ll N, P \ll N$ and typically α is somewhat greater than 1.0 depending on the sparsity of the circuit matrices. Note that the AC characterization of on-chip power grids including the package is routinely done in existing design flows even for PDNs without on-chip voltage regulation. In this sense, the proposed stability checking for regulated PDNs does not incur any significant additional analysis cost.

4.5.1.4 Other Considerations

The above checkings are made based on a linear modeling of the LDO circuit. In practice, nonlinearity of LDO circuits is traditionally handled by performing linearization at multiple operation points. For example, in the traditional phase margin method, analog designers need to plot the Bode plot and measure the phase margin at multiple different operation points of the circuit within its operation range. Similar to that, the proposed method can be applied while linearizing the LDO at different operation points and performing the checkings discussed previously at each operation point.

4.5.2 Hybrid Stability Margin (HSM)

An HSM that integrates the evaluations of passivity and gain into a single quantitative measure is further defined. HSM can be incorporated as a localized stability constraint into an automated stability-ensuring LDO design flow that will be described in the next section.

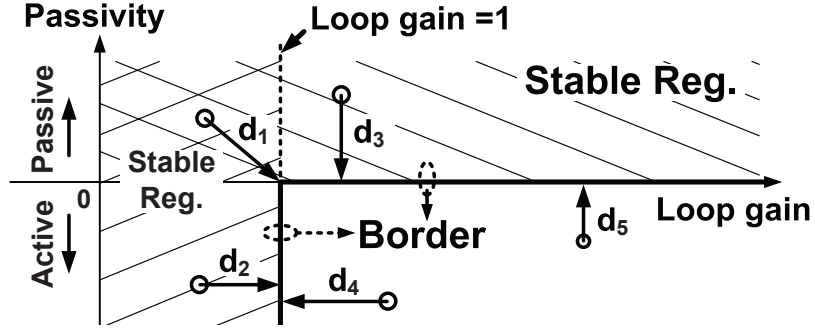


Figure 4.11: Hybrid stability margin at a frequency point.

First the HSM defined on an individual frequency basis is proposed. In Fig.4.11, the horizontal axis represents $\|\mathbf{G}(j\omega_k)\|_2\|\mathbf{H}(j\omega_k)\|_2$ and the vertical axis represents $\lambda_{min}(j\omega_k)$. Based on the evaluation of gain and passivity, an LDO design can be represented by a point in the plane. According to hybrid stability theorem, the stability-guaranteed region is the band area $0 < \|\mathbf{G}\|_2\|\mathbf{H}\|_2 < 1$ in union with the quadrant where $\lambda_{min} > 0$. The border of the region is depicted with the bold solid lines.

The hybrid stability metric is defined as a signed distance to the border of the stability-guaranteed region. In Fig. 4.11, there are five design cases evaluated at frequency ω_k , and each case is represented by a circle. The HSM(ω_k) for each case is the signed length of the corresponding arrowed line. The sign is positive if the circle is in the stability-guaranteed region, and is negative otherwise.

4.6 Practical PDN Network Design

The proposed stability checking approach provides a basis for evaluating the stability of a given PDN by means of a localized LDO HSM design constraint. This makes it possible to efficiently leverage this constraint to drive the LDO design optimization in an enhanced design flow. On the other hand, from a design perspective,

the introduction of HSM into the design process of LDOs does introduce new design issues. In many aspects, the techniques one may take to meet the proposed hybrid stability margin are with a flavor similar to ones that are commonly employed by the designers to meet conventional phase/gain margin targets. This similarity may help the adoption of the proposed design approach by typical designers. Nevertheless, an in-depth design analysis reveals unique design considerations pertaining to trade-offs between the new HSM and other LDO performances, choice of key transistor-level design parameters and LDO design topologies. In this section, a localized automated LDO design flow is demonstrated and key circuit-level design issues are discussed.

4.6.1 Design Flow

As elaborated in the previous section, all the information required by the proposed stability-checking approach can be obtained from AC simulations which circuit designers are well familiar with. Thus, the approach can be easily integrated into the conventional LDO design flow which the LDO designers are already accustomed to. As such, the stability-ensuring LDO design flow can be built upon the conventional flow with inclusion of one additional stability constraint.

The integration of the stability-checking approach is illustrated in Fig. 4.12. First of all, an initial LDO design with sufficient circuit performances is obtained using the conventional design methodology. The network stability evaluation over the specified frequency samples is then performed at each iteration until the stability is guaranteed and the performance requirements are satisfied. Note that, only the low-cost LDO circuit evaluations (as in the grey box in Fig. 4.12) are repeated in each design iteration, on the premises that the LDO sizing during optimization is well contained without affecting the passive sub-network structure, which may be achieved by measures like prescribing a fixed chip area large enough to accommodate

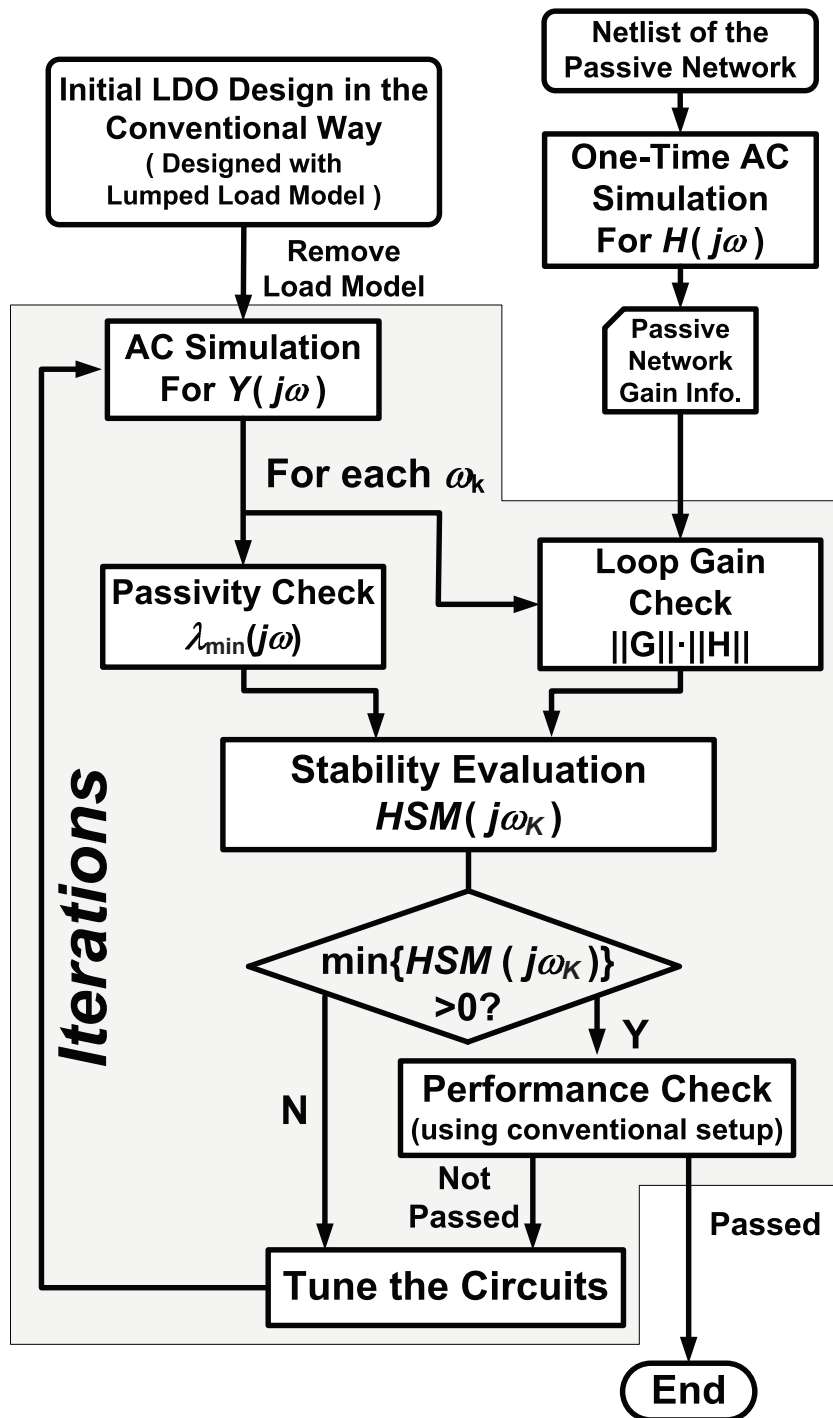


Figure 4.12: Stability-ensuring design flow.

sizing of the LDO within the optimization boundaries.

4.6.2 LDO Design Insights and Performance Tradeoffs

From a design point of view, the key issues in ensuring system stability are to properly control the gain, bandwidths, etc. These are in some sense no more than what are manually done in the standard LDO design process, including, but not limited to, reducing the 3-dB bandwidth (pole-splitting), increasing the quiescent current, and adjusting the gain of the “local loop”. Clearly, just like in the case of conventional phase or gain margin, there are trade-offs between stability and other performances. However, there exist several new design issues and opportunities for the case of hybrid stability, which are discussed below.

One powerful aspect of the proposed stability ensuring framework is that it leverages the notions of passivity and small gain in a complimentary way. This provides very useful degrees of design freedom for guaranteeing stability and trading off with other performances. In the following context, three types of design freedom are discussed, starting from the one that immediately exploits the frequency-dependent nature of the hybrid stability framework, and then the one that creates freedom through circuit or topology modifications, and the last one that explores freedom in the passive power grid design.

4.6.2.1 Exploiting Frequency Dependency

As described earlier, hybrid stability can be ensured by satisfying either the passivity or gain condition at each frequency. Optimal design of LDOs can be approached by choosing judiciously one of the two conditions to satisfy for each frequency in a way to minimize area and power overhead and influences on other performances. It is instructive to examine how such optimal designs may vary across different frequency ranges.

At *DC and low frequencies*, through investigations on the 2×2 Y -matrix of an LDO it is found that it is advantageous for the LDO designed to satisfy the gain condition ($\gamma_{\mathbf{G}}(j\omega)\gamma_{\mathbf{H}}(j\omega) < 1$). Specifically speaking, in this frequency band, the elements in the first column of the Y -matrix is smaller in magnitude than the corresponding elements in the second column roughly by a factor of A_{LL} , where A_{LL} represents the loop gain of the local loop of each individual LDO, a critical performance metric in LDO design. For good closed-loop regulation performances, a large A_{LL} is normally desired. On the other hand, by examining the property of the Y -matrix, it can be observed that the LDO can simply become not passive under large A_{LL} . Therefore, it is extremely hard, if not impossible, for an LDO to achieve good regulation performances while exhibiting passive characteristics in this frequency range. The conflict between passivity and regulation performance is somewhat intuitively straightforward since this is what active regulation is supposed to be as to differentiate from passive regulation. And in order to pass the HSM check while keeping good regulation performance, satisfying the gain condition shall be targeted.

On the other hand, it is critical to note that satisfying the gain condition does not necessarily imply lowering A_{LL} . A critical constraint-relaxing technique (hereafter referred to as the impedance splitting technique) is developed that allows to lower the gain of the system-wide loop without introducing much degradation of regulation performance (corresponding to a high A_{LL}). For continuity, the design implications resulted from the technique are discussed below while Section 4.6.2.4 will introduce a detailed discussion of the impedance splitting technique.

First take a look at a typical LDO structure illustrated in Fig. 4.13(a) as well as some important AC currents labeled as i_p , i_s , and i_{EA} . The small-signal currents i_p and i_{EA} are respectively the dynamic currents flowing in or out of the pass transistor

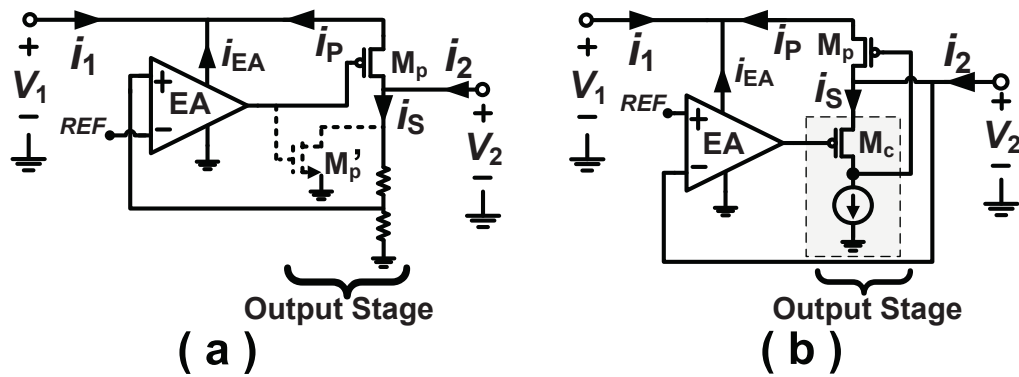


Figure 4.13: Demonstrations of exemplary stability-enhancing schemes for LDO output stage design. (a) Scheme I: simple circuit modification on the output stage. (b) Scheme II: topology change for the output stage.

and the error amplifier, while i_s is the dynamic ground current in the output stage. The impedance splitting technique then reveals that a generally effective way of satisfying the gain condition is to make $|i_s|$ larger than $|i_{EA}|$.

At *mid- and high-frequencies*, it is well known that the impedance peaking due to package parasitic inductance usually occurs, which is around the typical on-chip LDO's unity-gain bandwidth (GBW). Since the gain of \mathbf{Z} -parameter matrix of block \mathbf{H} is in a sense of impedance, the package resonance peakings are reflected in $\gamma_{\mathbf{H}}$ as similar peaks of value. While this research discovers that LDOs usually exhibit local passivity in a frequency band beyond its GBW, it is usually of less performance cost to force the LDO to meet passivity condition than the gain condition at those peaking frequencies. Tuning the LDO's GBW below the peaking frequencies can be one of the effective measures to meet passivity condition and it can be done by varying the value of LDO's internal capacitors (e.g., some compensation capacitors or some zero-generation capacitors) and/or reducing the LDO's bias current. It is worth to mention that analogous to the tradeoff made in the traditional phase margin method, there is apparently a tradeoff between bandwidth and stability. To achieve

good HSM, the LDO's GBW may need to be lowered as aforementioned.

In addition, it is also observed that the active devices in LDO can no longer react to fast signal changes beyond a certain high frequency, ω_h , and only their intrinsic and parasitic capacitors remain in play. For example, i_p in Fig. 4.13(a) is mostly conveyed through the path consisting of the gate-to-source capacitor and gate-to-drain capacitor of M_p ; i_s is through the grounded capacitors associated to the output port, including the drain diffusion capacitance of M_p . Because of the fact that the size of M_p is hundreds to thousands of times larger than transistors in EA, so are the capacitors associated, thus, $|i_s|$ can easily exceed $|i_{EA}|$, and the gain condition can be met in this frequency band with little design effort.

Summarily, the passivity condition is chosen in the package impedance peaking frequency range to relieve the efforts on handling the rugged impedance peaks, while the gain condition is selected at either DC and low-frequencies or the ultra-high frequencies.

4.6.2.2 Exploiting LDO Topology Modifications

Another important source of design freedom comes from LDO topology modifications. In particular, if the output stage is designed in such way that $|i_s|$ is greater than $|i_{EA}|$, the gain condition can be more easily met. Section 4.6.2.4 gives more detailed analysis on this claim.

According to this insight, a topological modification on the output stage is identified, i.e., by adding a pull-down pass transistor to the output stage (shown as the dashed NMOS M'_p in Fig. 4.13(a)) which is seldom seen in existing LDO topologies and is, to the best knowledge of the author, the first time acknowledged for its effectiveness in enhancing stability. Alternatively, in the same spirit, designers can choose another type of output stage topology, e.g., a source follower, as shown in

Fig. 4.13(b) to fulfill the same purpose. Since the selection of LDO topology should be made at the very beginning of the design process, this insight may help designers to make the right choice earlier, reducing possibility of design re-spinning.

4.6.2.3 *Exploiting Possible Structures of the Global VDD Grid*

Recalling that the proposed theoretic framework constrains the designs of two sub-systems for the stability of the whole network. While the design freedom of the LDO sub-system has been explored in the previous sections, this section explores the design freedom of the passive power grids, the other sub-system in the framework.

Apparently, among the two HST conditions designers can only work around the gain condition on block \mathbf{H} , which implies that the general design goal of block \mathbf{H} is to lower the gain of block \mathbf{H} . While it is easy to come to the idea of widening all the power routing (i.e., power grid wires) which may be done to the maximum extent given practical physical design constraints, another avenue of altering the structure of the global VDD grid is explored in this section.

Intuitively, considering that the stability problem is caused by inter-LDO interactions, one would argue that if the LDOs are placed far enough away from each other, the interactions can be made weak enough to be neglected. It can be mathematically proven that placing LDOs far apart can effectively lower the gain of block \mathbf{H} because the non-diagonal elements in the transfer matrix \mathbf{H} are decreased in so doing. However, this method contradicts with the fundamental idea of load sharing by distributing multiple LDOs in one power domain.

To keep load sharing while minimizing interactions among the LDOs, this work proposes to use an exclusive VDD grid for each LDO while the outputs of LDOs are still tied together by a shared Vreg grid, which is depicted in Fig. 4.14(b). Compared with the VDD grid structure in Fig. 4.14(a), the proposed structure

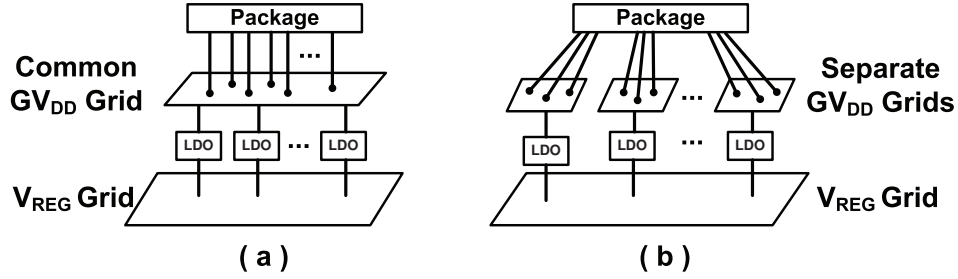


Figure 4.14: The illustrations of two VDD grid structures for distributed regulators. (a) The common global VDD grid structure. (b) The proposed global VDD grid structure.

effectively elongates the distance between the input ports of LDOs which weakens the interactions among them. On the other hand, the output ports of LDOs are placed closed enough to offer the benefit from distributed regulation.

In practice, if the number of LDOs is large, the proposed global VDD grid structure would cause difficulty in power and/or I/O plannings for physical designers as the number of VDD power islands increases. A compromise can be reached by retaining the top metal layer for the global VDD grid while separating the VDD grids on the lower metal layers as illustrated in Fig. 4.15.

The effectiveness on relaxing stability constraint by the proposed global VDD grid structure is verified in Section 4.7.3.

4.6.2.4 The Impedance Splitting Technique for Relaxing Stability Constraint

In Section 4.6.2.1 and 4.6.2.2, the inequality $|i_s| > |i_{EA}|$ is pointed out as a helpful design guide for meeting the gain condition prescribed by Corollary 1 without compromising regulation performance significantly. The detailed development of this insight is discussed as follows.

To begin with, reconsider the LDO's 2×2 Y-matrix in Fig. 4.2(b). To lower the gain of block \mathbf{G} , the element values of the Y-matrix are inevitably to be decreased,

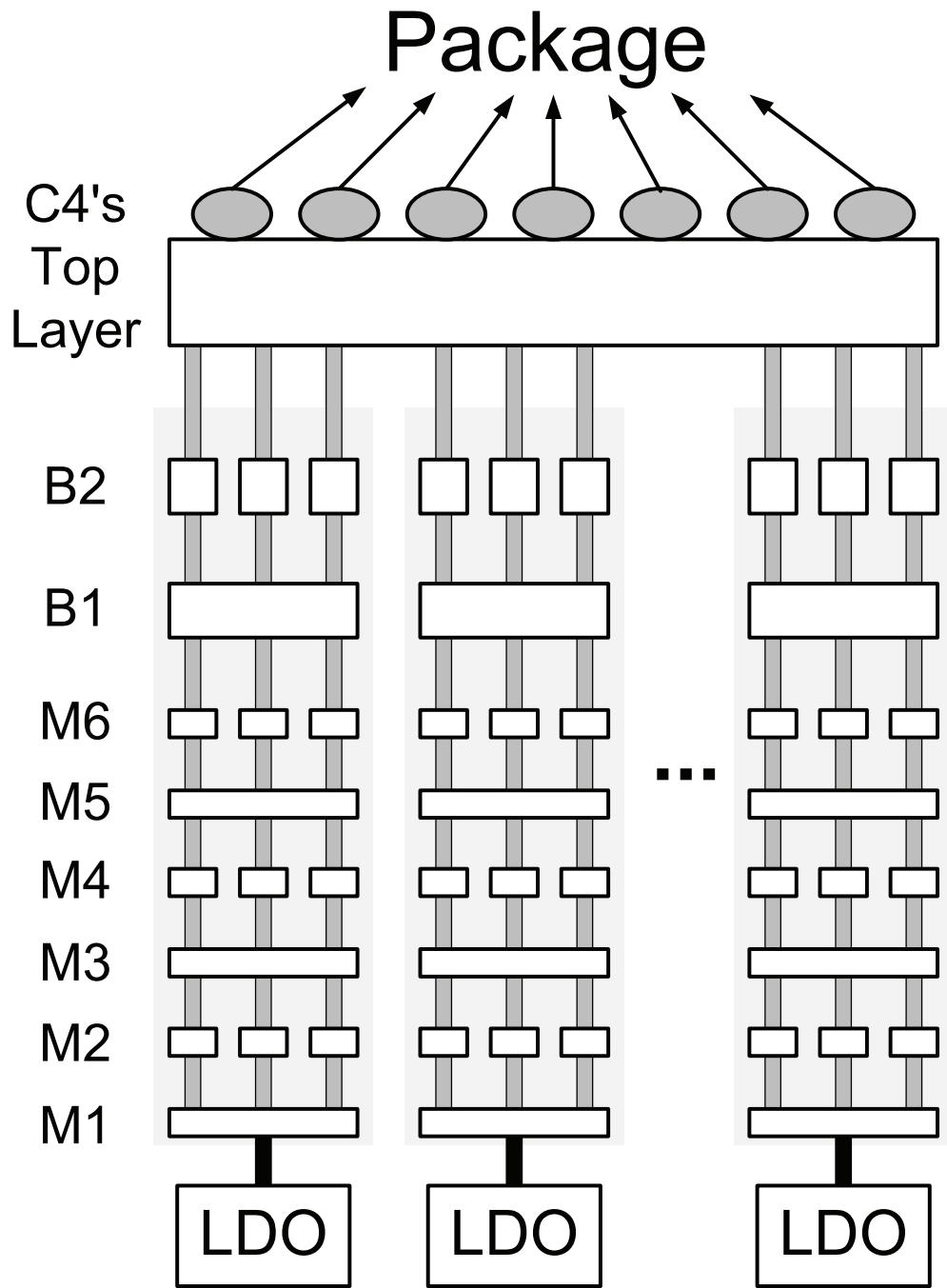


Figure 4.15: The illustration of a practical way to weaken the inter-LDO interactions.

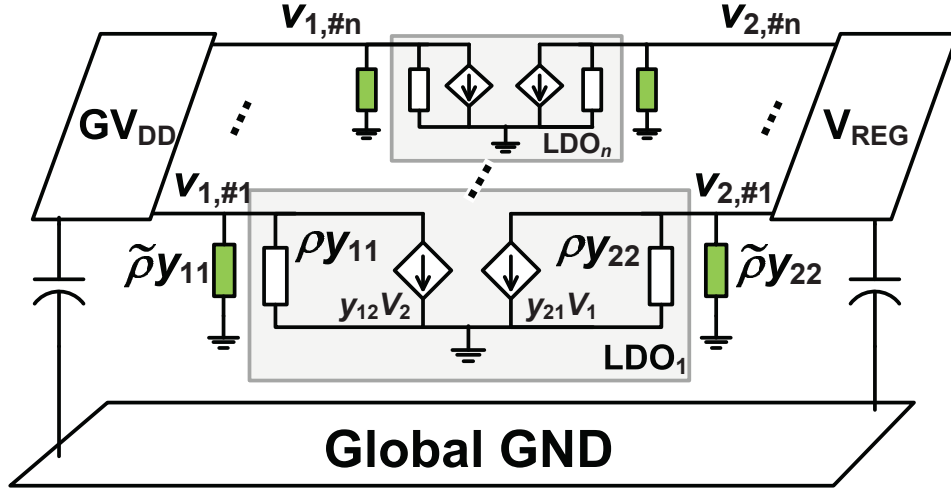


Figure 4.16: The illustration of splitting self-admittances in LDO's Y-parameter model.

especially the dominant elements. As mentioned in Section 4.6.2.1, y_{12} and y_{22} are the dominant ones at low frequencies. Given the fact that a large y_{22} is key to achieve good load regulation, the on-chip voltage regulation can be compromised if y_{22} is significantly reduced.

In order to solve this dilemma and further relax the stability versus performance trade-off, this work proposes to re-partition the system by splitting the self-admittances (y_{11} and y_{22}) into two parts with one part remaining in the LDO block and the other part pushed into the passive network, as illustrated in Fig. 4.16.

Note that the splitting is only performed to meet the gain condition as part of the stability checking process. In order to have a uniform mathematical description of the two sub-system before and after the re-partitioning, a frequency-dependent splitting coefficient, $\rho (= 1 - \tilde{\rho})$, defined in the same way as $\alpha(\omega)$ discussed in Section 4.4.1 is introduced:

$$\rho(\omega) = \begin{cases} 1, & \omega \in \Omega, \\ 0, & \text{otherwise,} \end{cases}$$

where Ω is the set of frequencies at which the LDO block satisfies the passivity condition. This splitting is automatically controlled by $\rho(\omega)$ as the stability checking is being performed along the frequency axis. At frequency bands where the gain condition is preferred to satisfy, the self-admittances of LDO block is deemed as elements in block \mathbf{H} (i.e., $\rho = 0$); otherwise, they are assigned back to LDO block (i.e., $\rho = 1$). Obviously, the splitting coefficient is the same as the frequency selection function $\alpha(\omega)$. Thus, the splitting is perfectly synchronized with the switching of the two hybrid stability conditions to meet. Specifically, when to meet the passivity condition, the splitting is not performed and block \mathbf{H} is still locally very strictly passive and the local passivity of block \mathbf{G} is to be examined; when to meet the gain condition, additional elements are hooked up to the passive sub-network. For the latter case, because block \mathbf{H} is changed, recalculation of its gain is needed, which can be easily done given that $\mathbf{H}(j\omega)$ is only a small $2n$ -port model.

The benefit from re-partitioning is that, when targeting at gain condition, block \mathbf{G} loses self-admittance elements in the matrix which results in lowered $\gamma_{\mathbf{G}}$ while on the other hand, the self-impedances of block \mathbf{H} are lowered too (due to additional impedance in parallel), resulting in lowered $\gamma_{\mathbf{H}}$. In this way meeting the gain condition is in fact helped by increasing $|y_{22}|$. For example, at DC and low frequencies, after moving self-admittances out of block \mathbf{G} , $\gamma_{\mathbf{G}}(\omega)$ is approximately $|y_{12}(\omega)|$, while $\gamma_{\mathbf{H}}(\omega)$ is roughly as large as $1/|y_{22}(\omega)|$. Therefore, $\gamma_{\mathbf{G}}(\omega)\gamma_{\mathbf{H}}(\omega) \approx |y_{12}(\omega)/y_{22}(\omega)|$, which shows that an increase of $|y_{22}|$ actually reduces the gain of the system-wide loop.

Since increasing $|y_{22}|$ with respect to $|y_{12}|$ is helpful to meet the gain condition, its design implications is further examined. By definition of admittance matrix, $y_{12} = \frac{i_1}{V_2}|_{V_1=0}$; $y_{22} = \frac{i_2}{V_2}|_{V_1=0}$. From Fig. 4.13(a), it is shown that $|i_1| = |i_p + i_{EA}|$ and $|i_2| = |i_p + i_s|$. Therefore, one way to increase $|y_{22}|$ with respect to $|y_{12}|$ is to make

$|i_s|$ greater than $|i_{EA}|$, which can be accomplished by designers through one of many means as discussed in Section 4.6.2.2.

In essence, the fundamental reason for this splitting to work is the conservativeness brought by meeting $\gamma_G \gamma_H < 1$ which is only a sufficient condition for stability. Different partition of the system can lead to different degrees of such conservativeness and hence potential benefits can be obtained by seeking a proper partition of the system. Since the re-partitioning does not add or remove any elements into or from the system, the entire system is physically unchanged; the only thing it changes is the way that the whole system is analyzed.

4.6.3 Illustrative Design Optimization

To illustrate the application of the proposed techniques, this work develops an optimization-based automated design flow using an optimizer to run the iterations shown in Fig. 4.12. The objective function for this optimization contains two classes of terms: one for penalizing performance degradations and the other for penalizing instability. In general, any performance metric can be considered in the optimization. For an illustration purpose, the LDO's performance metrics considered in this objective function include, but not limited to, the load regulation accuracy of the LDO (*ACC*) defined by $1 - \frac{|V_{reg} - V_{preset}|}{V_{preset}}$ which is an important DC characteristic that measures how close the actual output voltage V_{reg} is to the target voltage V_{preset} ; the gain-bandwidth product, GBW; the quiescent current, I_q , which emphasizes the quiescent current efficiency; the average output admittance, y_{avg} , which can largely reflect how good the dynamic regulation is. And y_{avg} is defined by $\frac{1}{\omega_n - \omega_0} \int_{\omega_0}^{\omega_n} |y_{22}(j\omega)| d\omega$, where ω_0 and ω_n are respectively the lowest and highest frequencies of interest. These terms

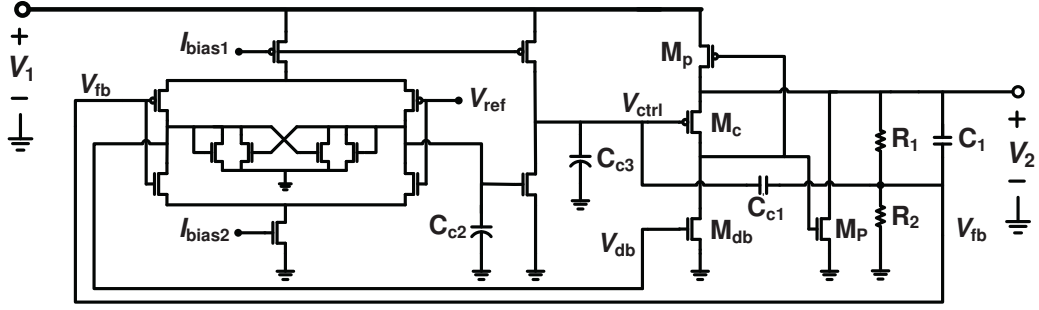


Figure 4.17: The LDO topology used in the practical implementations [25].

are properly normalized and included in the objective function to be minimized:

$$f = \alpha \left(\frac{ACC_n}{ACC} \right)^k + \beta \left(\frac{GBW_n}{GBW} \right)^t + \eta \left(\frac{I_q}{I_{q-n}} \right)^p + \gamma \lg \left(\frac{y_{avg-n}}{y_{avg}} \right) + \theta 10^{-HSM}, \quad (4.20)$$

where α , β , η , γ , and θ are the weights for respective performance penalty terms, which reflect optimization biases according to a specific practical set of design requirements; the exponential or logarithmic functions are used to prevent the optimizer from straying far away from the optimal point, and to deal with large differences in the orders of magnitude of those quantities. Specifically, the first four terms in (4.20) indicate that the greater ACC , GBW , $1/I_q$ and y_{avg} are with respect to the ones achieved by the initial design (i.e., ACC_n , GBW_n , I_{q-n} and y_{avg-n}), the smaller f is, and the closer the design will be to the optimum. Note that if in the situation where there are hard constraints on these performances, one can also change the penalty functions into the ones dealing with the differences between the actual values and the hard constraints. Since negative HSMs do not guarantee stability, an exponential function is chosen to heavily penalize any negative HSM so that stability will be enforced.

For the LDO proposed in Chapter 2 (redrawn in Fig. 4.17), due to their impor-

tance to hybrid stability and other performance specifications, several transistor-level design parameters are chosen: the widths of M_p , M_c and M_{db} , and the amounts of pole/zero-tuning capacitors C_{c1} , C_{c2} , C_{c3} , and C_1 . The width of the pass transistor (M_p) influences ACC , ω_{-3dB} , GBW and y_{avg} in a major way, whereas the widths of M_c and M_{db} are influential on the bias current, I_q , and the i_s in Fig. 4.13. The results of the proposed optimization are presented in detail in the following section.

4.7 Experimental Study

In this section, two experimental PDN designs are showcased to demonstrate the effectiveness and efficiency of the proposed approach. While the PDN sizes are different in the two cases, the adopted LDO topology is the same as shown in Fig. 4.17. And the same package model [33] is adopted. Both cases aim at an optimized PDN design with four on-chip LDOs. An LDO is initially designed in the traditional manner with sufficient circuit performances and a good phase margin (referred to as the ‘initial LDO design’ in the rest of the chapter), and then respectively the proposed approach is adopted to optimize the initial LDO design. The circuits are designed and optimized in a commercial 90nm CMOS technology. And the APPS optimizer [24] is adopted to tune the LDO.

4.7.1 Multiple LDOs in a Small Network

As discussed in Section 4.2, the brute-force method for stability checking is only feasible for small networks. To verify the effectiveness of the proposed stability-ensuring LDO design approach, return to the case with the small PDN (of about only 20 nodes) adopted in the example in Section 4.2, and apply the proposed approach to optimize the LDO for the PDN’s stability, such that the classical pole analysis method can be adopted to judge the effectiveness of the proposed approach. Comparisons are also made with the example in Section 4.2 which showed that an LDO designed

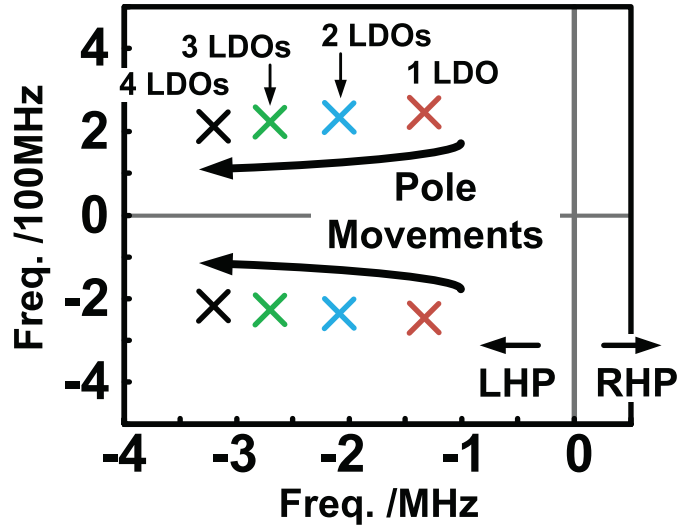


Figure 4.18: Pole analysis showing the stability of the PDN designed with the proposed approach.

in a traditional manner with a good phase margin cannot guarantee the network stability.

In the pole analysis, a pair of complex poles are revealed that move most evidently as the number of LDOs changes. Fig. 4.18 shows the movement of the poles on the s-plane as the number of LDOs integrated into the network is increased. In contrast to the rightward pole movement happened in the counterexample shown in Fig. 4.5, in this PDN with stability-enforced LDOs the movement is leftward and there are no RHP poles, meaning that the system is stable and the proposed approach is effective in ensuring the stability of the whole network. It is further confirmed by the transient simulation results shown in Fig. 4.19 which demonstrates the waveforms of the regulated voltage V_{reg} under load current variations. Compared with the heavy oscillation of V_{reg} in the counterexample shown in Fig. 4.6, V_{reg} in this case settles after the load current disturbance, reflecting the stability of the system.

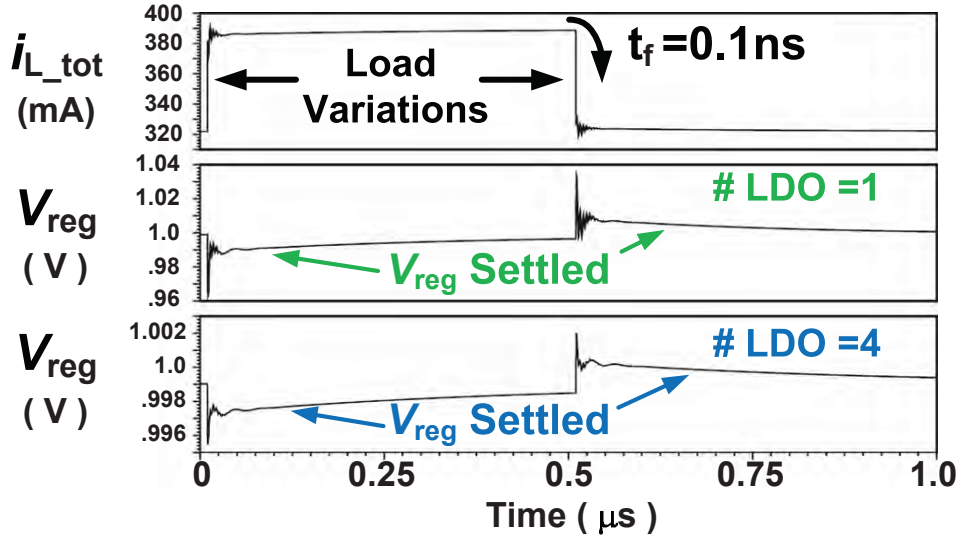


Figure 4.19: Transient analysis confirming the stability of the PDN with the stability-ensured LDOs.

4.7.2 Multiple LDOs in a Large Network

Further, the application of the proposed approach to the optimization of LDOs for a PDN of over 200K nodes is presented in this subsection, in an attempt to demonstrate the effectiveness and efficiency of the approach in large PDN design scenarios.

Stability Checking Along Frequency Axis The frequency-wise stability checking on the initial LDO design and the one designed in the stability-ensuring method are respectively illustrated in Fig. 4.20 and Fig. 4.21, with the loop gain and passivity metric, λ_{min} , being plotted in dashed lines and in dash-dotted lines, respectively. In both figures, the frequency ranges in which the gain condition is met, are labeled as “A”, the ranges where passivity condition is met are labeled as “C”, and the ranges where both conditions are met are “B”, while the potentially unstable range is “D”. As shown in Fig. 4.20, the initial design violates the hybrid stability criteria at the

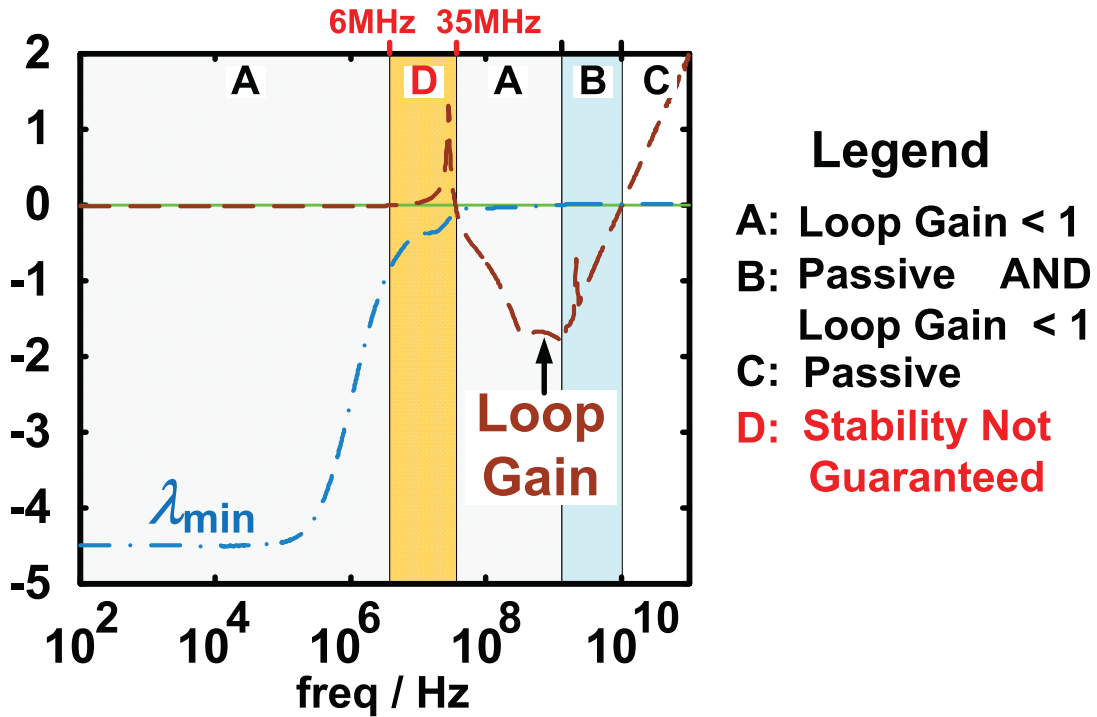


Figure 4.20: The loop gain and λ_{min} of the initial design.

frequency band from about 6 to 35 MHz where the loop gain exceeds unity while $\lambda_{min} < 0$. And it is shown in Fig. 4.21 that, by the proposed approach, the initial design can be successfully optimized into the design that satisfies the HSM criteria over all frequencies and thus guarantee the stability of the whole network.

Effectiveness of the Approach In this case, since the poles searching method is impractical, the transient simulation results are instead used to confirm the stability of the system. To examine the initial design, first plug four copies of the initial LDO into the PDN. As shown in Fig. 4.22, an arbitrarily selected nodal voltage on the regulated power grids (V_{reg}) as well as the one on the global VDD grids (GV_{DD}) renders continuing oscillations. In contrast, replace the initial LDOs with the ones given by the proposed approach and the result in Fig. 4.23 shows only slight

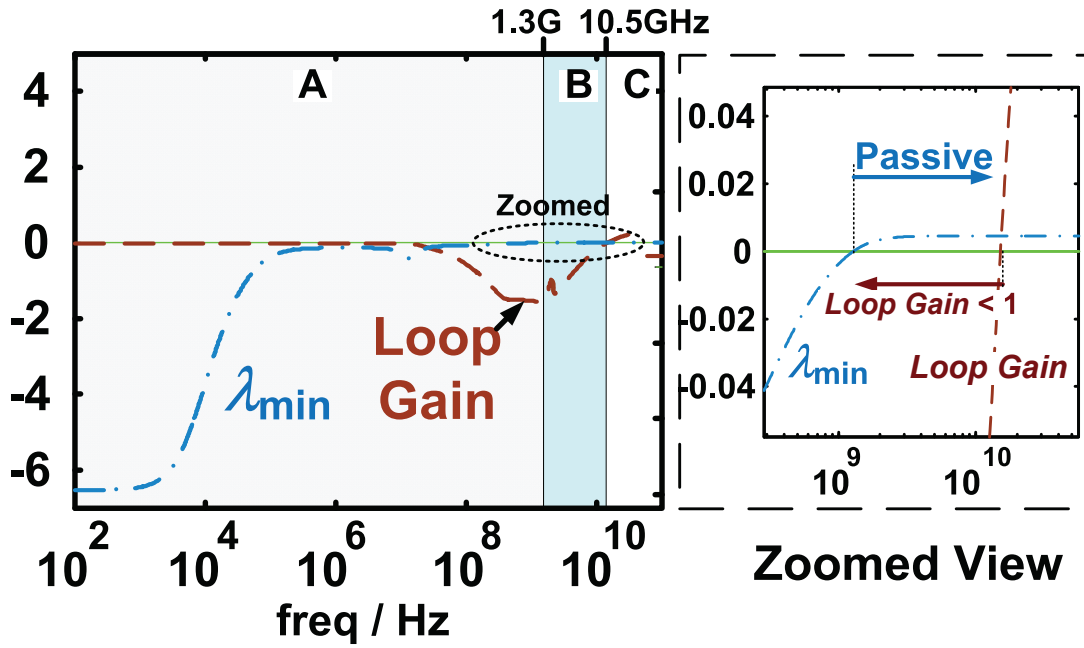


Figure 4.21: The loop gain and λ_{min} of the stability-ensured design.

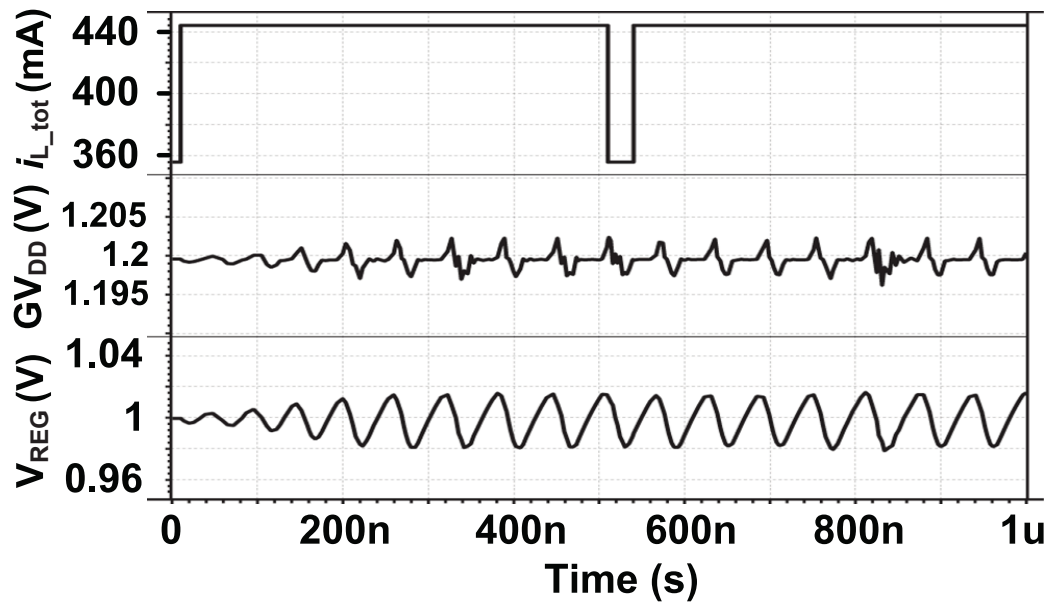


Figure 4.22: The transient simulation results showing the instability of the PDN with the LDOs designed in a standard manner.

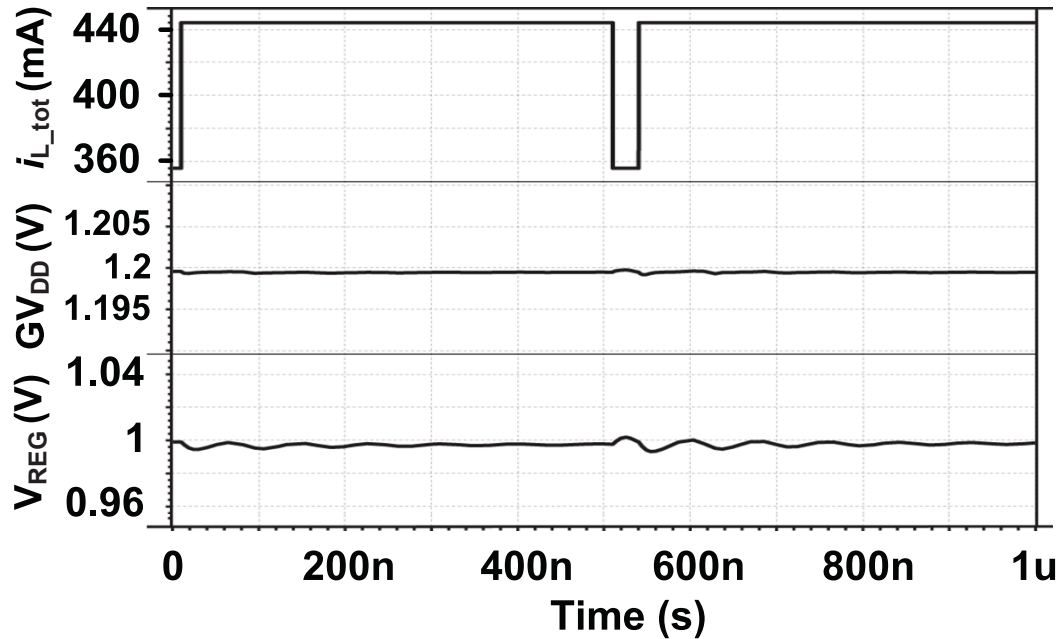


Figure 4.23: The transient simulation results confirming the stability of the PDN with the stability-ensured LDOs.

fluctuations when i_L variations occur, after which the voltages become settled.

Efficiency of the Approach As indicated by Fig. 4.12, there are two additional sources of design time cost: the AC simulations for the gain characterization of the passive network and the iterations of stability checking.

The former are performed at frequencies ranging from 1Hz up to 1THz with 200 samples per decade. There are four LDOs in this case and hence eight ports in the passive network, and the simulation by using an in-house simulator takes about 11 hours. Note that AC simulations are also a common practice in power grids analysis without regulators. So actually no additional cost is added by doing so.

The rest of the stability assurance procedure (the iterations) is taken over by the optimizer. The optimization takes about 116 minutes (including simulator invocation time) to reach the optimal performance trade-offs while ensuring stability.

Table 4.1: The Performance Trade-offs

	Initial LDO	Opt. LDO in the large PDN	Opt. LDO in the small PDN
<i>HSM</i>	-18.9	0.01	5e-3
Stability	Unstable	Stable	Stable
Load Reg. Acc.	99.96%	99.90%	99.91%
GBW (MHz)	511	422	380
I_q (μA)	469	518	340
y_{avg} (S)	5.17	7.18	4.26
y_{avg}/I_q (S/ μA)	0.011	0.0139	0.0125

In summary, the total design time in this case is about 13 hours with 11 hours being consumed on the one-time simulation of the passive sub-network.

4.7.3 Performance Trade-Offs

When designing an on-chip regulated PDN, stability is the primary design target. Without stability, the whole chip is easily subjected to power failure. Therefore, comparisons between the stability-ensured LDO designs in the above two cases with the unstable initial LDO design are, in this sense, not meaningful. However, to gain the insights, the comparisons on several performance metrics are performed in this sub-section. Also, in order to get a more complete picture of the tradeoffs, this work sets up the two optimization cases with different sets of performance weights (i.e., α , β , η , and γ) to represent different performance biases: in the small PDN case, the quiescent current consumption is particularly stressed, while in the large PDN case, the dynamic regulation performance is emphasized more than the other two performances.

Table 4.1 lists the comparisons among the three designs. The network stability metric *HSM*, negative in the initial design, is greatly optimized to be positive in both optimization cases indicating that the PDN stability is ensured. While it is obvious

that the unstable LDO design cannot be used in the PDN, this work first shows that, in the large PDN case, by consuming 10.4% more quiescent power, the network stability is ensured. In addition, an improvement of 37.8% on the dynamic regulation performance metric y_{avg} is also obtained, bringing forth a 26.3% improvement on y_{avg}/I_q , an efficiency quantity that measures the regulation performance gained per unit quiescent power consumed. By emphasizing low power consumption, in the small PDN case, the quiescent power is saved by 27.5% and the stability is ensured at a cost of 25.6% GBW reduction and a 18.2% degradation of y_{avg} . The resultant y_{avg}/I_q is nevertheless improved by 13.6%.

Table 4.2 performs the comparison of trade-offs for LDOs in two different global VDD grid structures as discussed in Section 4.6.2.3 (illustrated in Fig.4.14). For both cases the power grids are extracted from a realistic power grid design that consists of 9 metal layers and 4 LDOs. The four bottommost layers are assumed to be intensively used for local signal routings and hence there is no VDD sharing among the LDOs on these layers for both cases. The VDD sharing starts from M5 all the way up to the top layer for the case with a common global VDD grid, while for the case with separated global VDD grids the four LDOs have their own VDD Grids and only until the topmost layer those VDD grids are connected together as illustrated in Fig. 4.15. And the LDOs in the two cases are optimized with the same set of performance weights.

It is verified that by simply breaking up the global VDD grid, the trade-off between stability and performance can be loosened noticeably. Specifically, the LDOs optimized under a common global VDD grid would need to suffer from a slight degradation on the output impedance and consumes a little bit more quiescent current compared with the initial LDO. On the other hand, the LDOs optimized under a separate VDD grid structure achieves stability with output impedance and quiescent

Table 4.2: The Comparison of Trade-offs in Different Global VDD Grid Structures

	Initial LDO	Opt. LDO w/i a Comm. Global VDD Grid	Opt. LDO w/i Diff. Global VDD Grids
<i>HSM</i>	-0.31/-0.37*	0.0075	0.0162
Stability	Unstable	Stable	Stable
Load Reg. Acc.	99.96%	99.93%	99.93%
GBW (MHz)	511	638	684
I_q (μA)	469	473	465
y_{avg} (S)	5.17	5.08	5.18
y_{avg}/I_q (S/ μA)	0.011	0.0107	0.0111

* HSM for the initial LDO is evaluated in the two types of VDD grids respectively.

current consumption being even improved slightly. The beauty of the method of breaking up VDD grid is gaining performance with little cost.

4.8 Summary

A hybrid theoretical framework for addressing the stability challenges of large PDNs with integrated LDOs is presented in this chapter. A practical design methodology is developed to allow for the localized design of LDOs while ensuring the system-wide stability, leading to trackable stability-driven design optimization of large PDNs. By virtue of unique design freedoms in the framework, useful design insights on stability-ensuring LDO design are discussed. Experimental results demonstrate the effectiveness and efficiency of the proposed method and also show that the enforced PDN stability does not necessarily incur significant performance degradations.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion of the Dissertation

The dissertation presents circuit design and methodology solutions to two major problems in the design of modern IC power delivery with on-chip regulation:

- Two low-dropout voltage regulator topologies are proposed to improve power consumption and hence power efficiency of the regulator while maintaining the ability to handle fast transient regulation requirements. The first proposed regulator topology employs multiple feedback loops to achieve a frequency compensation without a need of big compensation capacitors and thereby occupies much less silicon area. With the improved area efficiency, the compensation scheme also accommodates a wide range of output capacitor values ranging from 0 to 1nF, offering flexibility for the power grid decap insertion. By reasonable allocation of quiescent power to the loops in the regulator, fast transient response is achieved with much less power consumed compared with its counterparts. The second proposed regulator topology employs a switched-capacitor based transient booster which only kicks into operation when an abrupt transient current demand occurs while remaining idle otherwise. During relative stable period of load, on the one hand, the main regulator circuit assumes the regulation responsibility with neither the need to be ultra-fast responsive to the worst-case transient of load, nor a need for large bias current; on the other hand, the switched-capacitor circuit stands by and consumes only a small amount of static power. In this way, low total quiescent current consumption associated with good suppression of fast load transients is achieved.

- Regarding to the lately emerged distributed on-chip voltage regulation technology, the dissertation first reveals by a realistic case study that instability could possibly happen if using the traditional regulation design methodology. A theoretically elegant stability checking framework is then proposed. The framework is built upon a partition of the power delivery network that splits active circuits from the passive sub-network, and developed by a complimentary combination of two classic stability theorems, namely the small-gain theorem and the passivity theorem, offering additional freedom for the circuit design to satisfy the stability conditions. In-depth analysis on the design tradeoffs in the distributed regulator design is afterward performed from which meaningful design insights for designs of both regulators and power grids are attained. An automatic optimization flow based on the framework is developed and experiment results verify the effectiveness and efficiency of the proposed methodology.

5.2 Future Work

As the lifeline of high-performance chips, power delivery network is and will continue to be an active research topic that is imbued with innovation opportunities on multiple levels, such as PDN architecture, design methodology, transistor-level circuit design, optimization algorithm, and so on.

In this section, we first discuss about future work on the distributed LDO design. As the stability is discussed under the context of linear time-invariant analysis in this work, with the development of mixed-signal types of regulators it is imperative to develop effective and efficient stability checking/ensuring methods for these types of circuits which requires not only linear system analysis, but also calls for insights from nonlinear analysis to explain problems and develop solutions. For example, it is still difficult to adopt the digitally assisted LDO topology discussed in Chapter 3 into

the distributed regulator architecture because the digital feedback loop behavior is not linear and cannot be incorporated into the hybrid stability framework discussed in Chapter 4. Therefore, an extension of the proposed hybrid stability framework or even a completely new framework is yet to explore in order to accommodate a broader types of regulators.

Besides stability, in the distributed LDO topic there are other issues from high level optimization to low level circuit design. For example, there is still a lack of scalable system-level modeling for power efficiency, load/line regulation performance, area, etc. Without appropriate modeling, it is hard to say how many LDOs are needed to achieve an optimal/near-optimal PDN. Also, the placement of distributed LDOs in 3-D integrated circuits has not been addressed yet. For example, it is not answered yet which one of the following two structures is better, placing all LDOs in one silicon layer or distribute them in each layer. On the other hand, on the circuit design level, there are problems with mismatches between LDOs, difficulty in delivering a common reference voltage to the LDOs, the question on using digital LDOs instead of analog ones, etc. Researchers have recently given an initial solution to the first two aforementioned problems (i.e., the mismatch and reference delivery problems) [21]. However, it is an incomplete solution as it only tabs one location of the chip and only the supply voltage at that chip will be forced to be exactly the reference voltage (or some ratio of it). The problem may be solved by introducing multiple-input and multiple-output control mechanisms in the reference feedback loop.

Looking further ahead, LDOs may not be the only choice for distributed regulation system. Currently LDO is the most area efficient one which makes it feasible to place quite a number of LDOs in one power domain. However, LDOs have major limitations on power efficiency and not so DVS friendly. In the future, if the quality of

on-chip capacitors and inductors or even memristors can be improved significantly, it is possible to replace LDOs with other switching-mode regulator topologies with generally superior efficiency, or more likely, a heterogeneous architecture that contains multiple kinds of regulators for different loads.

REFERENCES

- [1] International Technology Roadmap for Semiconductors. 2011 Overall Roadmap Technology Characteristics (ORTC) Tables. <http://www.itrs.net>
- [2] G. Patounakis, Y.W. Li, & K.L. Shepard. A fully integrated on-chip DC-DC conversion and power management system. *IEEE J. Solid-State Circuits*, 39(3):443–451, 2004.
- [3] Y.-H. Lee, Y.-Y. Yang, K.-H. Chen, Y.-H. Lin, S.-J. Wang, K.-L. Zheng, et al. A DVS embedded power management for high efficiency integrated SoC in UWB system. *IEEE J. Solid-State Circuits*, 45(11):2227–2238, 2010.
- [4] P. Hazucha, T. Karnik, B.A. Bloechel, C. Parsons, D. Finan, & S. Borkar. Area-efficient linear regulator with ultra-fast load regulation. *IEEE J. Solid-State Circuits*, 40(4):933–940, 2005.
- [5] W. Kim, M.S. Gupta, G.-Y. Wei, & D. Brooks. System level analysis of fast, per-core DVFS using on-chip switching regulators. *Proc. of IEEE Intl. Symp. High Performance Computer Architecture*, 123–134, 2008.
- [6] Z. Zeng, X. Ye, Z. Feng, & P. Li. Tradeoff analysis and optimization of power delivery networks with on-chip voltage regulation. *Proc. of Design Automation Conference (DAC)*, 831–836, 2010.
- [7] R.J. Milliken, J. Silva-Martínez, & E. Sánchez-Sinencio. Fully on-chip CMOS low-dropout voltage regulator. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 54(9):1879–1890, 2007.
- [8] P.Y. Or, & K.N. Leung. An output-capacitorless low-dropout regulator with direct voltage-spike detection. *IEEE J. Solid-State Circuits*, 45(2):458–466, 2010.

- [9] J. Guo, & K.N. Leung. A 6- μ W chip-area-efficient output-capacitorless LDO in 90-nm CMOS technology. *IEEE J. Solid-State Circuits*, 45(9):755–759, 2010.
- [10] E. Alon, & M. Horowitz. Integrated Regulation for Energy-Efficient Digital Circuits. *IEEE J. Solid-State Circuits*, 43(8):1795–1807, 2008.
- [11] Y.-H. Lam, & W.-H. Ki. A 0.9V 0.35 μ m adaptively biased CMOS LDO regulator with fast transient response. *IEEE Intl. Solid-State Circuits Conference (ISSCC)*, 442–443, 2008.
- [12] G. Blakiewicz. Output-Capacitorless low-dropout regulator using a cascoded flipped voltage follower. *IET Circuits, Devices and Systems*, 5(5):418–423, 2011.
- [13] T. Jackum, G. Maderbacher, W. Pribyl, & R. Riderer. Fast transient response capacitor-free linear voltage regulator in 65nm CMOS. *IEEE Intl. Symp. Circuits and Systems (ISCAS)*, 908–905, 2011.
- [14] E.N.Y. Ho & P.K.T. Mok. A capacitor-less CMOS active feedback low-dropout regulator with slew-rate enhancement for portable on-chip application. *IEEE Trans. Circuitis Syst. II: Express Briefs*. 57(2):80–84, 2010.
- [15] P.R. Gray, P.H. Hurst, S.H. Lewis, & R.G. Meyer. *Analysis and Design of Analog Integrated Circuits*. Wiley, New York, 2001.
- [16] T.Y. Man, K.N. Leung, C.Y. Leung, P.K.T. Mok, & M. Chan. Development of single-transistor-control LDO based on flipped voltage follower for SoC. *IEEE Trans. Circuits Syst. I, Reg. Papers*, 55(5):1392–1401, 2008.
- [17] M. El-Nozahi, A. Amer, J. Torres, K. Entesari, & E. Sánchez-Sinencio. High PSR low drop-out regulator with feed-forward ripple cancellation technique. *IEEE J. Solid-State Circuits*, 45(3):565–577, 2010.

- [18] V.J. Reddi, M.S. Gupta, G. Holloway, G.-Y. Wei, M.D. Smith, & D. Brooks. Voltage emergency prediction: using signatures to reduce operating margins. *Proc. of IEEE Int. Symp. High Performance Computer Architecture*, 18–29, 2009.
- [19] B. Anderson & S. Vongpanitlerd. *Network analysis and synthesis: a modern systems theory approach*. Network series. Prentice-Hall, 1973.
- [20] B. Brogliato, R. Lozano, B. Maschke, & O. Egeland. *Dissipative Systems Analysis and Control*. Springer, London, 2007.
- [21] J. F. Bulzacchelli, Z. Toprak-Deniz, T. M. Rasmus, J. A. Iadanza, W. L. Bucossi, S. Kim, et al. Dual-loop system of distributed microregulators with high DC accuracy, load response time below 500 ps, and 85-mV dropout voltage. *IEEE J. Solid-State Circuits*, 47(4):863–874, 2012.
- [22] J. R. Forbes & C. J. Damaren. Hybrid passivity and finite gain stability theorem: stability and control of systems possessing passivity violations. *IET Control Theory and Applications*, 4(9):1795–1806, 2010.
- [23] W. M. Griggs, B. D. O. Anderson & A. Lanzon. Interconnections of nonlinear systems with “mixed” smallgain and passivity properties and associated input-output stability results. *Systems and Control Letters*, 58(4):289–295, 2009.
- [24] G. A. Gray & T. G. Kolda. Algorithm 856: Appspack 4.0: asynchronous parallel pattern search for derivative-free optimization. *ACM Trans. Math. Softw.*, 32(3):485–507, 2006.
- [25] S. Lai & P. Li. A fully on-chip area-efficient cmos low-dropout regulator with fast load regulation. *Analog Integrated Circuits and Signal Processing*, 72(2):433–450, 2012.

- [26] A. Odabasioglu, M. Celik, & L. Pileggi. Prima: passive reduced-order interconnect macromodeling algorithm. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 17(8):645–654, 1998.
- [27] A. van der Schaft. *\mathcal{L}_2 -gain and passivity techniques in nonlinear control*. Springer, London, 2000.
- [28] P. Zhou, D. Jiao, C. Kim, & S. Sapatnekar. Exploration of on-chip switched-capacitor DC-DC converter for multicore processors using a distributed power delivery network. *IEEE Custom Integrated Circuits Conference (CICC)*, 1–4, 2011.
- [29] S. Lai, B. Yan, & P. Li. Stability assurance and design optimization of large PDNs with multiple on-chip LDOs. *IEEE/ACM Conf. Computer-Aided Design (ICCAD)*, 247–254, 2012.
- [30] S. Lai, B. Yan, & P. Li. Localized stability checking and design of IC power delivery with distributed voltage regulators. *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, 32(9):1321–1334, 2013.
- [31] J. Doyle, B. Francis, & A. Tannenbaum. *Feedback control theory*. Dover Books on Electrical Engineering Series, Dover Publications, 2009.
- [32] R. Riaza & C. Tischendorf. The hyperbolicity problem in electrical circuit theory. *Math. Meth. Appl. Sci.*, 33(17):2037–2049, 2010.
- [33] M. Gupta, J. Oatley, R. Joseph, G.-Y. Wei & D. Brooks. Understanding voltage variations in chip multiprocessors using a distributed power-delivery network. *Design, Automation and Test in Europe Conference and Exhibition*, 1–6, 2007.
- [34] P.E. Allen, & D.R. Holberg. *CMOS Analog Circuit Design*, Oxford University Press, New York, 2012.

- [35] M. Ang, R.Salem, & A. Taylor. A on-chip voltage regulator using switching decoupling capacitors, *IEEE Intl. Solid-State Circuits Conference (ISSCC)*, 438-439, 2000.
- [36] X. Meng, & R. Saleh. An improved active decoupling capacitor for “hot-spot” supply noise reduction, *IEEE J. Solid-State Circuits*, 44(2):584-593, 2009.
- [37] C. Zhan & W. Ki. An output-capacitor-free adaptively biased low-dropout regulator with subthreshold undershoot-reduction for SoC, *IEEE Trans. Circuits Syst. I: Regular Papers*, 59(5):1119-1131, 2012.