Power Delivery and Management in Nanoscale ICs

by

Inna P. Vaisband

Submitted in Partial Fulfillment

of the

Requirements for the Degree
Doctor of Philosophy

Supervised by
Professor Eby G. Friedman

Department of Electrical and Computer Engineering
Arts, Sciences and Engineering
Edmund J. Hajim School of Engineering and Applied Sciences

University of Rochester
Rochester, New York
2015

Dedication

This work is dedicated to my parents, Elga Levi and Alexander Vaisband.

Biographical Sketch



Inna Vaisband received the Bachelor of Science degree in computer engineering and the Master of Science degree in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in, respectively, 2006 and 2009, and the Ph.D. degree in electrical engineering from the University of Rochester, Rochester, New York, in 2015. She is currently a Researcher with the Department of Electrical Engineering, University of Rochester, Rochester, New York. Between 2003 and 2009, she held a variety of software and hardware R&D positions at Tower Semiconductor Ltd., G-Connect Ltd., and IBM Ltd., all in Israel. In summer 2012, Inna was employed as a Visiting Researcher in the research group of Prof. B. Murmann at Stanford University. Her current research interests include the analysis and design of high performance integrated circuits, analog design, and

on-chip power delivery and management. Dr. Vaisband is an Associate Editor of the Microelectronics Journal.

The following publications were a result of work conducted during her doctoral study:

Journal Papers

1. I. Vaisband, E. G. Friedman, R. Ginosar, and A. Kolodny, "Low Power Clock Network Design," *Journal of Low Power Electronics and Applications*, Volume 1, No. 1, pp. 219–246, June 2011.

2. I. Vaisband and E. G. Friedman, "Heterogeneous Methodology for Energy Efficient Distribution of On-Chip Power Supplies," *IEEE Transactions on Power Electronics*, Volume 28, No. 9, pp. 4267–4280, September 2013.

3. I. Vaisband, M. Azhar, E. G. Friedman, and S. Kose, "Digitally Controlled Pulse Width Modulator for On-Chip Power Management," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, Volume 22, No. 12, pp. 2527–2534, November 2014.

4. I. Vaisband and E. G. Friedman, "Energy Efficient Adaptive Clustering of On-Chip Power Delivery Systems," *Integration, the VLSI Journal*, Volume 48, pp. 1–9, January 2015.

5. I. Vaisband, B. Price, S. Kose, Y. Kolla, E. G. Friedman, and J. Fischer, "A 28 nm Ultra-Small LDO for Fast and Efficient Load Regulation in Portable Devices," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* (under review).

6. I. Vaisband and E. G. Friedman, "Stability in Distributed Power Delivery Networks," *IEEE Transactions on Circuits and Systems I: Regular Papers* (under review).

Conference Papers

7. I. Vaisband, E. Friedman, R. Ginosar, and A. Kolodny, "Energy Metrics for Power Efficient Crosslink and Mesh Topologies," *Proceedings of the IEEE Symposium on Circuits and Systems*, pp. 1656–1659, May 2012.

8. S. Kose, I. Vaisband, and E. G. Friedman, "Digitally Controlled Wide Range Pulse Width Modulator for On-Chip Power Supplies," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2251–2254, May 2013.

9. I. Vaisband and E. G. Friedman, "Computationally Efficient Clustering of Power Supplies in Heterogeneous Real Time Systems," *Proceedings of the IEEE Symposium on Circuits and Systems*, pp. 1628–1631, June 2014.

10. I. Vaisband and E. G. Friedman, "Power Network On-Chip for Scalable Power Delivery," *Proceedings of the ACM/IEEE International Workshop on System Level Interconnect Prediction*, June 2014.

11. I. Vaisband and E. G. Friedman, "Dynamic Power Management with Power Network-on-Chip," *Proceedings of the IEEE International NEWCAS Conference*, June 2014.

Patent Disclosures

12. I. Vaisband and E. G. Friedman, "Power Network On-Chip for Scalable Power Delivery," United States Patent (pending).

Acknowledgements

The years I spent at the University of Rochester have been a unique experience that I will treasure for life. I would like to take this opportunity to express my highest gratitude to all of my exceptional teachers, dear friends, and beloved family for their guidance and support throughout the years of my life.

Foremost, I would like to express special gratitude to my academic advisor, Professor Eby G. Friedman, for the opportunities he has given me to learn and grow as his student. His extraordinary dedication to mentoring and motivating each of his students has allowed me to evolve personally and professionally. With his guidance and consideration, I was able to gain valuable insights into the world of knowledge, and shape my own vision of research. Professor Friedman, I am fortunate, indeed, to have undertaken my doctoral studies under your supervision. Thank you.

I would like to thank the members of my committee, Professors Engin Ipek, Zeljko Ignjatovic, Oleg Prezhdo, and my committee chair, Professor Sandhya Dwarkadas,

for serving on my proposal and defense committees. Their valuable comments greatly improved my dissertation. I would also like to express my appreciation to the University of Rochester, and specifically, the Department of Electrical and Computer Engineering, for providing a friendly and encouraging environment with all the required tools to support high quality academic research.

Professor Avinoam Kolodny, your invaluable mentorship throughout all the years of my doctoral studies have been of great importance. Your sincere advice was always warm, timely, and helpful. Thank you, Professor Kolodny. I am grateful to Professor Boris Murmann for providing me with the opportunity to work as part of his research group at Stanford University, as well as his guidance and support during the collaborative project. I thank Mahmoud Saadat, a Ph.D. student at Stanford University, for sharing his experience during this collaborative project. I would also like to thank Jeff Fischer and Yesh Kolla for the opportunity to collaborate with Qualcomm. I'd like to express special gratitude to Burt Price of Qualcomm for his time, professional guidance, and all of the technical discussions that broadened and deepened my knowledge and research experiences.

I feel lucky to have spent these years at the University of Rochester with the members of the High Performance VLSI/IC Design and Analysis Laboratory: Professor Emre Salman, Doctor Renatas Jakushokas, Professor Selçuk Köse, Professor

Ioannis Savidis, Ravi Patel, Alex Shapiro, Boris Vaisband, Mohammad Kazemi, Shen Ge, Kan Xu, and Nathan Kistner. Thank you all for your sincere friendship. It has been my great pleasure working with you. I also thank RuthAnn Williams for her friendship and constant help on the administrative side.

I am grateful to all my friends in Israel and those in the United States for making the distance across the Atlantic Ocean so much shorter, and my life experience so much brighter. I thank Eugene Gurevich for his continuous support and motivation. I would like to express my warmest gratitude to my uncle, Doctor Mark Levy, for instilling in me a love of science through his exceptional patience and teaching talent. I express my deepest appreciation to my brother, Boris Vaisband, who has stood by my side through every step of my life, and my parents, Elga Levi and Alexander Vaisband, who have smoothed out every bump on my path with their unconditional love, wise advice, and invincible faith in me. Last, but not least, my encouraging and patient husband, and dearest friend, Alexander Partin, your kindness, keen intelligence, and deep sense of humor have been a source of great inspiration and strength. Thank you for lighting my way.

Abstract

The continued advance of society and emerging markets requires functionally diverse semiconductors. The future of heterogeneous, high performance systems is strongly dependent upon the power delivery system, and deeply affected by the quality of on-chip power, availability of fine grain dynamically controlled voltage levels, and the ability to manage power in real-time. To satisfy evolving power delivery requirements, an effective power delivery solution is required.

In this dissertation, a platform for scalable power delivery and management is proposed. The key concept of this platform is to manage the overall energy budget with fine grain distributed on-chip power networks, providing local feedback paths from the billions of loads to multiple, locally intelligent power routers. Essential information such as timing slacks, voltages, currents, and temperatures sensed within the individual power domains is used to locally manage power in real-time. The overall energy budget is also adjusted in a near real-time manner by communicating

local power management decisions among the individual power routers within the proposed platform.

A computationally efficient methodology to co-design different types of power supplies within different levels of hierarchy has been proposed, and key circuits for distributed voltage regulation and dynamic voltage scaling have been developed. A distributed power delivery system with six ultra-small fully integrated low-dropout regulators has been fabricated in a 28 nm CMOS process and exhibits a fast transient response with excellent load regulation. To evaluate stability in a distributed power delivery system, a passivity-based stability criterion has been proposed. To provide a circuit level means for dynamically scaling the voltage in adaptive systems, a digitally controlled pulse width modulator has been developed including closed-form duty cycle expressions.

The proposed power network on-chip provides a means to efficiently deliver and intelligently manage high quality power in complex heterogeneous ICs. The issues of design complexity and scalability are addressed by providing a modular architecture that supports integration of diverse functional blocks and adapts to specific power demands without requiring re-design of the power delivery system.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professor Eby G. Friedman (advisor), Professor Engin Ipek, and Professor Zeljko Ignjatovic of the Department of Electrical and Computer Engineering, and Professor Oleg Prezhdo of the Chemistry Department. All of the work described in the dissertation was completed independently by the student.

This research is supported in part by the Binational Science Foundation under Grant No. 2012139, the National Science Foundation under Grants No. CCF-0541206, CCF-0811317, CCF-0829915, and CCF-1329374, the New York State Office of Science, Technology and Academic Research to the Center for Advanced Technology in Electronic Imaging Systems, the Intelligence Advanced Research Projects Activity under Grant No. W911NF-14-C-0089, and by grants from Qualcomm, Cisco Systems, Samsung, and Intel.

Table of Contents

Dedication **ii**

Biographical Sketch **iii**

Acknowledgements **vii**

Abstract **x**

Contributors and Funding Sources **xii**

List of Tables **xix**

List of Figures **xxi**

1 Introduction **1**

1.1 Power delivery in integrated circuits 9

1.2 Outline 12

2 Power Converter/Regulator Circuits **18**

2.1 Switching mode power supplies 20

2.2 Switched-capacitor converter 25

2.3 Linear voltage regulator 29

2.4 Comparison of monolithic power supplies 32

2.5 Summary 35

3 Modeling, Analysis, and Optimization Techniques for Power Delivery **36**

3.1 Modeling power delivery systems 39

3.2 Analysis of power delivery system 45

3.2.1 Hierarchical power grid analysis 47

3.2.2 Statistical power grid analysis 47

3.2.3 Effective resistance-based power analysis 49

3.3 Optimization of power delivery systems 50

3.4 Summary 54

4 Heterogeneous Power Delivery **55**

4.1 Power converter topologies 58

4.1.1 Switching converters 58

4.1.2 Linear converters 66
4.1.3 Comparison of power supply topologies 71
4.2 Heterogeneous power delivery system 72
4.2.1 Number of on-chip power regulators 75
4.2.2 Number of off-chip power converters 76
4.2.3 Power supply clusters 81
4.3 Summary 84

5 Energy Efficient Adaptive Clustering of Power Supplies in Heterogeneous Dynamically Controlled Systems 86

5.1 Dynamic control in heterogeneous power delivery systems 89
5.2 Computationally efficient power supply clustering 92
5.2.1 Near-optimal power supply clustering 93
5.2.2 Power supply clustering with dynamic programming 103
5.3 Co-design of power delivery system in circuit benchmarks 109
5.3.1 Power supply clustering of IBM power grid benchmarks 110
5.3.2 Power supply clustering and existing power delivery solutions 113
5.4 Summary 115

6 Distributed Power Delivery with 28 nm Ultra-Small LDO Regulator 118

6.1 Power delivery system 121

6.1.1 Op amp based LDO 124

6.1.2 Adaptive bias 133

6.1.3 Adaptive compensation network 136

6.1.4 Distributed power delivery 139

6.2 Test Results 142

6.3 Summary 148

7 Digitally Controlled Pulse Width Modulator for On-Chip Power Management 150

7.1 Description of the proposed PWM architecture 153

7.1.1 Header circuitry 154

7.1.2 Duty cycle to voltage converter 157

7.1.3 Ring oscillator topology for pulse width modulation 159

7.1.4 Ring oscillator topology for pulse width modulation with constant frequency 164

7.2 Simulation results 167

7.2.1 Proposed pulse width modulator under PVT variations 168

7.2.2 Duty cycle controlled pulse width modulator 170

7.2.3 Duty cycle and frequency controlled pulse width modulator . . 172

7.3 Summary 173

8 Stability in Distributed Power Delivery Networks 175

8.1 Passivity-based stability of distributed power delivery systems 177

8.2 Distributed power delivery systems 180

8.2.1 Distributed power delivery model 180

8.2.2 Passivity analysis 188

8.3 Parametric circuit performance modeling technique 192

8.4 Experimental results 195

8.5 Summary 201

9 Power Network-on-Chip for Scalable Power Delivery 203

9.1 Power network-on-chip architecture 205

9.1.1 Power routers 208

9.1.2 Locally powered loads 209

9.1.3 Power grid 210

9.2 Case study 211

9.3 Summary 215

10 Future Work **217**

10.1 Power network on-chip for distributed power management 218

10.2 Secure power network on-chip 220

10.3 Summary 221

11 Conclusions **223**

Bibliography **229**

List of Tables

2.1 Comparison of SMPS, SC, and LDO topologies for power conversion. 33

5.1 Power supply clustering for a heterogeneous power delivery system with $S = 3$, $L = N = 4$, $V_{Drop} = 0.1$ volts and voltage domains (1 volt, 1 ampere), (1.49 volts, 1 ampere), (1.51 volts, 1 ampere), and (2 volts, 1 ampere). 95

5.2 Power grid specifications for 0.8 volt, 1.0 volt, 1.2 volt, 1.4 volt, and 1.6 volt domains based on ibmpg1 test case. 111

5.3 Power efficiency in circuits with and without separation of power conversion and regulation. CPU time for hybrid (with near-optimal efficiency), dynamic programming (DP) based, and exhaustive clustering is provided. 112

5.4 Parameter setup for circuits C_1 and C_2. 114

5.5 On-chip voltage regulation with and without separation of power conversion and regulation. 115

6.1 DC gain over a range of load currents at slow (SS, −30 °C), typical (TT, 25 °C), and fast (FF, 105 °C) corners. 127

6.2 Quiescent current with and without adaptive biasing. 136

6.3 Performance summary and comparison with previously published LDO regulators. 147

6.4 Voltage droop for different input and output voltage levels. 147

6.5 Measured quiescent current and current efficiency. 148

7.1 Timing parameters of the current controlled ring oscillator 161

7.2 Change in the duty cycle of the proposed PWM under PVT variations for the 22 nm predictive CMOS model. 169

9.1 Output load in a PNoC with four power domains. 214

List of Figures

1.1 The first transistor, (a) the inventors John Bardeen, William Shockley, and Walter Brattain, and (b) their point-contact transistor. 1

1.2 Photograph of the first integrated circuits, (a) discrete IC built by Jack Kilby of Texas Instruments in 1957, and (b) monolithic IC built by Robert Noyce of Fairchild Semiconductor in 1959. 3

1.3 Evolution of functional communication, synchronization, and power delivery based on the MtM concept of functional diversification. . . . 5

1.4 Core clock distribution of the IBM POWER7™ multi-processor [19]. 6

1.5 Evolution of functional complexity and routing methodology in microprocessors. (a) Intel 4004 processor with 2,300 transistors routed in a traditional manual way, (b) Vertical and horizontal buses in the IBM POWER4 processor with 174 million transistors, and (c) IBM POWER7 SoC with 1.2 billion transistors routed with a network-on-chip. 8

1.6 Toshiba full high definition (full HD) multiprocessor with 25 power domains [23]. 9

1.7 Current consumption in Intel microprocessors over the last four decades. 11

2.1 Speed and power consumption in recent Intel microprocessors [39]. . . 19

2.2 An SMPS, (a) boost, and (b) buck converter. 21

2.3 A buck converter composed of a spiral inductor L, MOS capacitor C, and MOSFET switches. 24

2.4 Single stage switched-capacitor converter, (a) phase 1 ($\varphi_1 : 0 \leq t \leq T_S/2$) operation, and (b) phase 2 ($\varphi_2 : T_S/2 \leq t \leq T_S$) operation. . . 26

2.5 Linear DC-DC regulator. 29

2.6 LDO regulator with a power transistor controlled in a negative feedback loop with an error amplifier. 30

3.1 Spiral model of typical iterative power delivery design process. 38

3.2 Power distribution network of the printed circuit board, package, and integrated circuit. 40
3.3 Trends for target output impedance of power distribution networks [74]. 41
3.4 Circuit level representation of a system-wide power network with an off-chip power supply. 42
3.5 Circuit level representation of a system-wide power distribution network with off-chip power supply and decoupling capacitors at the board, package, and circuit levels. 43
3.6 Impedance characteristics of a power delivery system with an off-chip power supply, and board, package, and integrated circuit level decoupling capacitors. 44
3.7 Power grid with a single power supply and two current loads, (a) circuit level representation, and (b) weighted graph representation. . . 48
3.8 Portion of power supply network with three rings of power supplies and a single current load. 50

4.1 Power delivery system with four voltage domains, (a) off-chip, (b) integrated on-chip, and (c) distributed point-of-load power supplies for voltage conversion and regulation. 56
4.2 Buck converter circuit. 59

4.3 Buck converter (a) physical area, and (b) power efficiency vs. load current for moderate, high, and ultra-high switching frequencies. . . 64

4.4 LDO circuit. 66

4.5 LDO physical area per 1 mA load. 68

4.6 LDO area for typical current loads. 69

4.7 Trends in typical (a) high performance (V_{HP}), low power (V_{LP}), and internal core primary (V_{IN}) voltage supplies, and (b) voltage conversion ratios (V_{HP}/V_{IN}) and (V_{LP}/V_{IN}). 70

4.8 LDO and buck converter (a) physical area, and (b) power efficiency for moderate, high, and ultra-high switching frequencies. 71

4.9 Power and area overhead of a linear, SMPS, PSiP, and preferred power conversion system. 73

4.10 Power delivery system with four voltage domains, utilizing (a) off-chip power supplies, (b) distributed POL power supplies, and (c) a heterogeneous system with off-chip converters and on-chip regulators. 74

4.11 Heterogeneous power delivery system with S off-chip switching converters, $L = N = \Sigma l_i$ on-chip linear regulators, and N on-chip power domains. 78

4.12 Power supply clusterings for a heterogeneous power delivery system with $S = 2$ and $L = N = 3$, (a) $K_1 = \{1, 2\}$, and (b) $K_2 = \{2, 1\}$. . . 82

5.1 Heterogeneous power delivery with multiple power domains. 87

5.2 Heterogeneous power delivery with multiple dynamically controllable power domains. 90

5.3 Heterogeneous power delivery system (a) standard deviation, and (b) average efficiency using an exhaustive power supply clustering algorithm. 98

5.4 Decrease in linear and binary power efficiency from the optimal power efficiency for (a) randomly distributed voltage levels, and (b) voltage levels grouped within three voltage ranges. 102

5.5 A single step of the recursive power supply clustering algorithm for a heterogeneous power delivery system with $S = 3$ and $L = N = 5$, (a) a single LDO ($l_3 = 1$), (b) two LDO regulators ($l_3 = 2$), and (c) three LDO regulators ($l_3 = 3$) in a high voltage SMPS cluster. 104

5.6 Power efficiency of heterogeneous power delivery system based on a recursive power supply clustering algorithm. 108

5.7 Test case ibmpg1 with (a) V_{DD} and V_{GND} voltage map, and (b) $V_{DD} - V_{GND}$ voltage drop across the circuits in metal layer $M6$. 110

6.1 Model of distributed LDO and power distribution network. 122

6.2 Load sharing in distributed power delivery system. 122

6.3 LDO topology. 124

6.4 Three current mirror OTA. 125

6.5 Small signal linear model of LDO. 125

6.6 Frequency response of near optimally compensated LDO regulator. . 129

6.7 Stability of the proposed LDO regulator as function of (a) load current I_{Load}, and (b) compensation accuracy. 131

6.8 PM with different constant compensation resistors and adaptive compensation. 132

6.9 Adaptive bias boost and compensation networks. 133

6.10 Voltage droop and quiescent current with and without adaptive biasing at (a) slow corner, (b) typical corner, and (c) fast corner. 135

6.11 Phase margin with different compensation and load capacitance for (a) $I_{Load} = 1\,\mathrm{mA}$, and (b) $I_{Load} = 10\,\mathrm{mA}$. 138

6.12 Small signal linear model of distributed power delivery system with k LDO regulators. 139

6.13 Small signal linear model of distributed power delivery system with k LDO regulators and equally shared load current. 140

6.14 Transient step response at the load for $V_{IN} = 1$ volt, $V_{OUT} = 0.7$ volts, and load current step from 52 mA to 441 mA in 10 ns, measured at (a) $T = -25\,°C$, (b) $T = 25\,°C$, and (c) $T = 125\,°C$. 144

6.15 Transient step response at the load at the typical temperature of 25 °C for $V_{OUT} = 0.7$ volts and a load current step from 52 mA to 441 mA in 10 ns, measured for (a) $V_{IN} = 0.9$ volts, (b) $V_{IN} = 1.0$ volt, and (c) $V_{IN} = 1.1$ volts. 145

6.16 Measured transient response for a load current step from 52 mA to 788 mA in 5 ns. 145

6.17 Die microphotograph of 28 nm ultra-small LDO. 148

7.1 Conventional ring oscillator. Note that an odd number of inverters is required for the system to oscillate. 151

7.2 Proposed PWM. The header circuitry has two input control signals, digital control (C_d) and analog control (C_a). C_d is used to dynamically change the individual transistors to provide a high granularity duty cycle control whereas C_a maintains a constant current from the header to the ring oscillator under PVT variations. 154

7.3 Addition based current source used as a header circuit [171]. 155

7.4 Parallel PMOS transistors replace M_3 to improve the granularity of the current control as well as behave as switch transistors to turn on different sections of the header circuitry. 156

7.5 Frequency to voltage converter proposed in [169] operating as a duty cycle to voltage converter. 158

7.6 Ring oscillator with current controlled P_{odd} transistors. 162

7.7 Ring oscillator with current controlled P_{odd} and P_{even} transistors. . . . 165

7.8 Monte Carlo simulation of (a) duty cycle, and (b) frequency distribution. 170

7.9 Duty cycle varies between 25% and 90% when the header current changes from 50 to 2 μA ($error < 4.4\%$). 171

7.10 Header current $I_{bias,A}$ changes from 40 to 11 µA. (a) Duty cycle varies between 25% and 90% ($error < 3.1\%$). (b) Period remains approximately constant ($error < 1.25\%$). 173

8.1 Power delivery system (a) with $n \geq 2$ distributed power supplies, and (b) reduced single port network. 178

8.2 Low-dropout regulator with $R_C C_C$ compensation, (a) topology, and (b) small signal linear model. 182

8.3 Distributed power delivery system with n LDO regulators driving a common power grid. 185

8.4 Summation of two horizontally shifted polynomials. 186

8.5 Model of distributed LDO and power distribution network. 188

8.6 Output impedance of individual LDO regulators loaded by different currents (between 20 mA and 100 mA) and the combined system output impedance, (a) phase $\angle Z_{OUT}$), (b) amplitude $|Z_{OUT}|$, and (c) poles and zeros. 190

8.7 Output response of a distributed power delivery system with different compensation ($C_C = 0.4pF$, $C_C = 0.5pF$, $C_C = 1pF$, and $C_C = 5pF$), showing the correlation between the (a) transient response, and (b) phase of the output impedance. 191

8.8 Conventional design flow for an analog circuit. 193

8.9 Automated PBSC-based design flow for a distributed power delivery system. 194

8.10 Output voltage of a single LDO regulator at (a) slow corner and $-30\,^{\circ}$C, (b) typical corner and $25\,^{\circ}$C, and (c) fast corner and $105\,^{\circ}$C. . 197

8.11 Transient step response at the load for $V_{IN} = 1$ volt, $V_{OUT} = 0.7$ volts, and a load current step from 52 mA to 441 mA in 10 ns, measured at (a) $T = -25\,^{\circ}$C, (b) $T = 25\,^{\circ}$C, and (c) $T = 125\,^{\circ}$C. 199

8.12 Transient step response at the load at a temperature of 25 °C for V_{OUT} = 0.7 volts and a load current step from 52 mA to 441 mA in 10 ns, measured for (a) V_{IN} = 0.9 volts, (b) V_{IN} = 1.0 volt, and (c) V_{IN} = 1.1 volts. 200

8.13 Measured transient response for a load current step from 52 mA to 788 mA in 5 ns. 200

9.1 On-chip networks based on the approach of separation of functionality, (a) network-on-chip (NoC), and (b) power network-on-chip (PNoC). 204

9.2 On-chip power network with multiple locally powered loads and three supply voltage levels (a) PNoC configuration at time t_1, and (b) PNoC configuration at time t_2. 206

9.3 On-chip power network with routers distributing the current over the power grid to the local loads. 207

9.4 Power routers for PNoC a) Simple topology with linear voltage regulator. b) Advanced topology with dynamically adaptable voltage regulator and microcontroller. 209

9.5 Preferred and supplied voltage levels in PNoC with four power domains. 212

9.6 Proposed PNoC with four power domains and four power routers connected with control switches. 213

9.7 Power router with voltage regulator, load sensor, and adaptive networks. 213

9.8 Voltage levels in PNoC with four power domains. 215

10.1 Centralized power management of an SoC with n modules. 219

10.2 Distributed PNoC scheme for power management of a single IC. . . . 220

Chapter 1

Introduction

It was in 1947 when John Bardeen and Walter Brattain in the Solid State Physics Group led by William Shockley in AT&T's Bell Labs applied two gold contacts to a crystal of germanium and observed power amplification of a signal [1] (see Figure 1.1). The first point contact transistor was born and a new era of technological



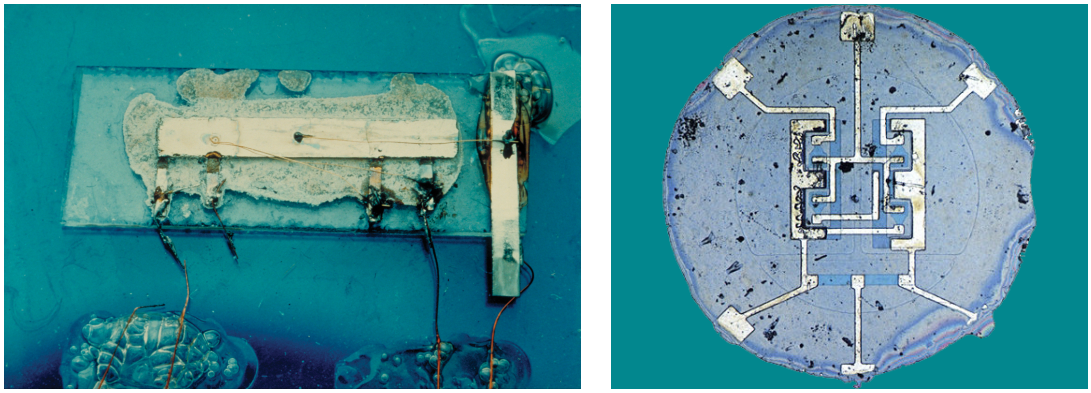
(a) (b)

Figure 1.1: The first transistor, (a) the inventors John Bardeen, William Shockley, and Walter Brattain, and (b) their point-contact transistor.

growth began [2]. The three men received the 1956 Nobel Prize in Physics for *"their researches on semiconductors and their discovery of the transistor effect"* [3]. A series of fundamental research in quantum mechanics during the early 1920's and 1930's was achieved that preceded the success of the experimental result of Bardeen and Brattain in 1947. Julius E. Lilienfield invented and patented the concept of a field effect transistor (FET) in 1926 [4]. In 1935, Oskar Heil patented a mechanism to control the resistance in a semiconducting material with an electric field [5]. These inventions were important milestones to understanding and applying solid state physics.

In the early 1950s, Bell Labs licensed the invention of the transistor and sold it for the price of $25,000, and a network of high tech companies, such as Fairchild Semiconductor, Intel, National Semiconductor, Texas Instruments, and Sony began to grow across the world, led by former employees from Bell Labs [2]. Intensive semiconductor research produced novel technologies and important milestones. In 1957, Jack Kilby of Texas Instruments built the first discrete integrated circuit (IC), shown in Figure 1.2(a), by connecting a single transistor, capacitor, and resistor with discrete wires. Kilby is acknowledged as the inventor of the microchip and won an inventor's *"Triple Crown"* for his accomplishments: the Nobel Prize in physics, the National Medal of Science, and the National Medal of Technology. Immediately after the invention of the IC, Jean Hoerni invented the planar process for manufacturing

transistors on a substrate [6], [7]. This process was used by Robert Noyce of Fairchild Semiconductor to build the first completely monolithic IC in 1959, as shown in Figure 1.2(b). Five decades later, this technology is still used today to manufacture modern ICs with several billions of transistors monolithically integrated onto the same layer of silicon [8].



(a) (b)

Figure 1.2: Photograph of the first integrated circuits, (a) discrete IC built by Jack Kilby of Texas Instruments in 1957, and (b) monolithic IC built by Robert Noyce of Fairchild Semiconductor in 1959.

Technology miniaturization according to Moore's law [9] has been driving the exponential growth of the semiconductor market over the last five decades, increasing integration levels, speed, compactness, and functionality while lowering power consumption and cost per function. Continued scaling of the physical feature size of digital circuits has increased pre - 1995 yearly sales by 15% to 17% and post - 1995 sales

by 6% [10], [11], maintaining the virtuous circle of market growth and investment in research for next generation technologies. The future of the More Moore (MM) virtuous circle has recently been challenged by a decrease in revenue growth and increase in the cost of research, development, and engineering (RD&E) of the world-wide semiconductor market [10], [11]. In addition, novel market segments such as intelligent transportation, revolutionary health care, sophisticated security systems, and smart energy applications have emerged, requiring increasingly diverse functionality such as RF, power control, passive components, sensors/actuators, biochips, optical communication, and MEMs. These non-digital heterogeneous functionalities do not scale with Moore's law (as shown in Figure 1.3), yet provide significant value to the end customer. To support future societal needs, functional diversification (More-than-Moore (MtM)) has been recognized as a primary objective of semiconductor RD&E [11]. Integration of non-digital functionalities at the board-level into system platforms such as systems-in-package (SiP) [12], systems-on-chip (SoC) [13], and TSV-based 3-D systems [14], [15] is a primary near and long term challenge of the More-than-Moore semiconductor reality. These integration trends and research opportunities are illustrated in Figure 1.3 with respect to the key bottleneck issues in high performance ICs.

Communication, synchronization, and power delivery are three essential aspects

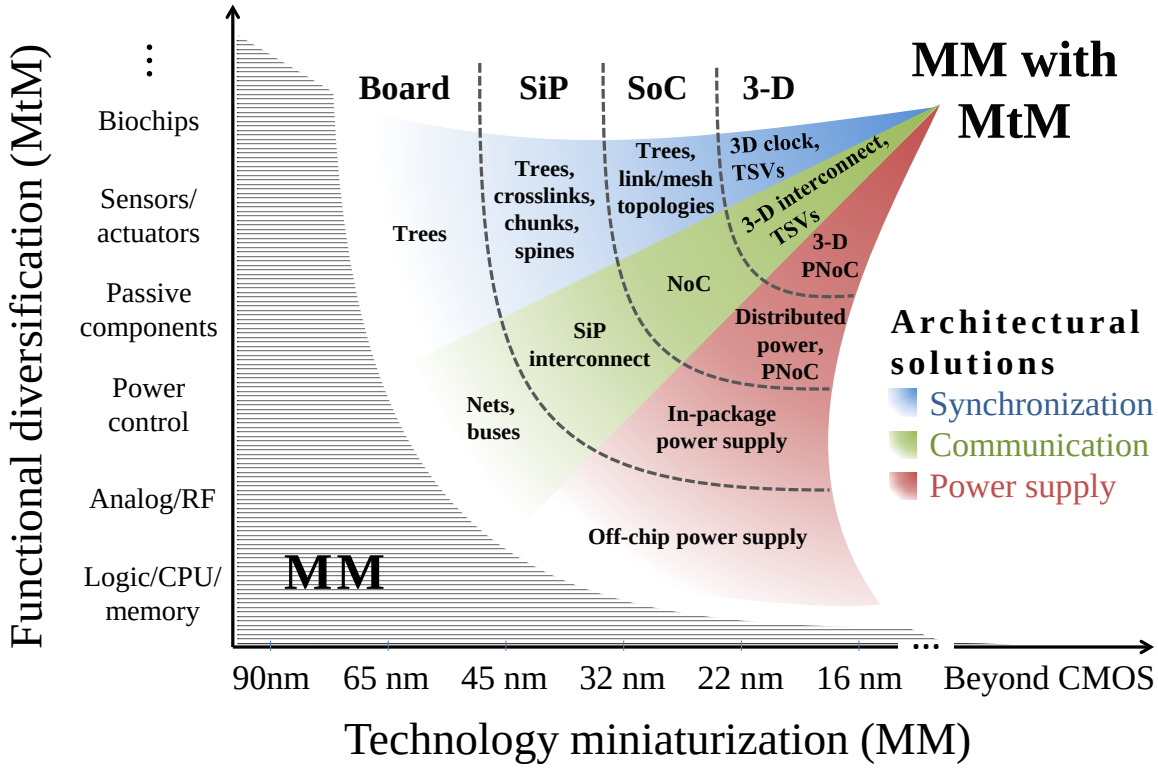


Figure 1.3: Evolution of functional communication, synchronization, and power delivery based on the MtM concept of functional diversification.

of integrated circuits. To facilitate the integration of these diverse functions, architectural, circuit, device, and material level solutions are required for each of these issues. The evolution of signaling, clock distribution, and power delivery is illustrated in Figure 1.3, exhibiting a roadmap for system solutions that scale while providing functional diversification.

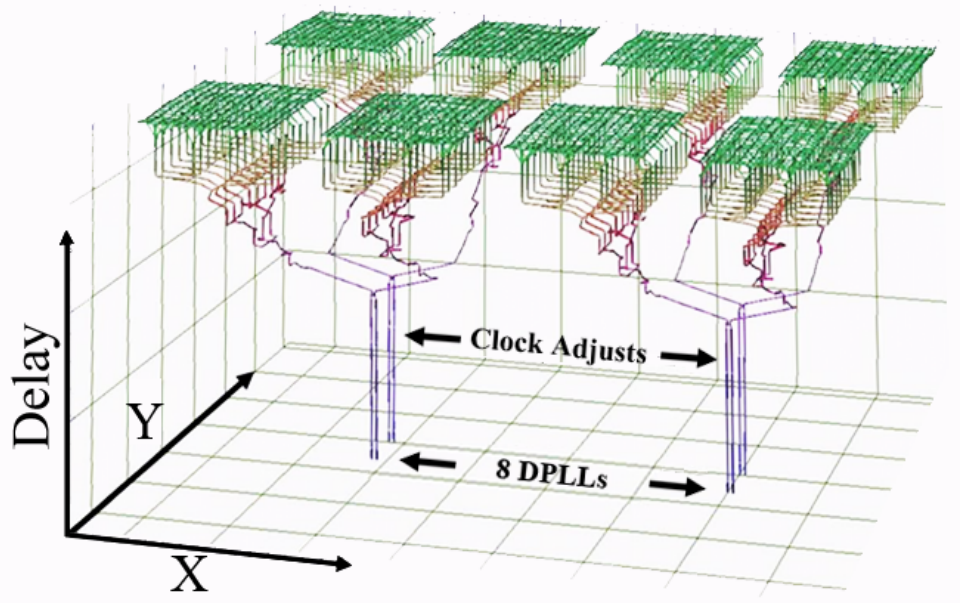
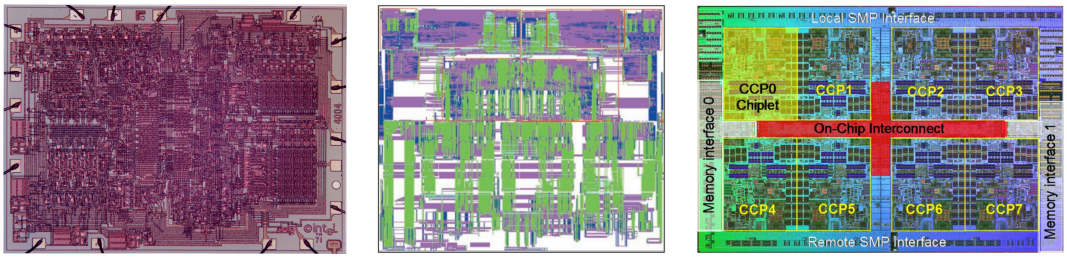


Figure 1.4: Core clock distribution of the IBM POWER7™ multi-processor [19].

To cope with synchronization challenges in SiP and SoC applications, clock distribution networks have evolved from different tree topologies to non-tree mesh topologies [16], as illustrated in Figure 1.3. Clock meshes are characterized by low clock skew and effective mitigation of process, voltage, and temperature (PVT) variations at the cost of higher power [17], [18]. As *ad hoc* clock network design approaches had become ineffective for controlling skew variations in modern high performance systems, systematic scalable methodologies and architectures have emerged. The clock distribution network within the IBM POWER7™ multi-core processor is shown in Figure 1.4 to illustrate the effectiveness of hierarchical clock distribution networks for skew mitigation within 23 clock domains. The global clock signal within the

POWER7^{TM} is distributed from eight digital phase-locked loops (DPLLs) to each of the clock domains through a binary clock distribution tree. Top level metal layers are utilized at this high level of distribution to minimize global skew. Alternatively, within each clock domain, the clock skew is locally mitigated with a non-tree mesh network at the expense of higher area and power consumption. Dynamically controllable programmable delay segments are integrated on-chip to manage mesh-to-mesh skew variations, satisfying the skew budget. Clock networks with global and local networks and inter-domain skew controllers deliver high quality clock signals at up to 14 GHz, exhibiting a scalable solution for high end microprocessors [19].

In functional communication, the interconnect network has progressed through several evolutionary stages from a "spaghetti" of nets and data busses to the emerging network-on-chip (NoC) paradigm [20], as depicted in Figure 1.3. The evolution of functional complexity and routing solutions is illustrated in Figure 1.5 with traditional interconnect wiring, bus-based routing, and network-on-chip, respectively, in the Intel 4004, IBM POWER4, and IBM POWER7 microprocessors [19], [21], [22]. The SoC paradigm shift is based on the "separation of different concerns" - a fundamental and cost effective approach for increasing performance and design productivity as the complexity of *ad hoc* solutions has become excessively high. Specifically, NoC separates computation from communication, enhancing the performance,



(a) (b) (c)

Figure 1.5: Evolution of functional complexity and routing methodology in microprocessors. (a) Intel 4004 processor with 2,300 transistors routed in a traditional manual way, (b) Vertical and horizontal buses in the IBM POWER4 processor with 174 million transistors, and (c) IBM POWER7 SoC with 1.2 billion transistors routed with a network-on-chip.

scalability, and control of quality of service (QoS), while increasing engineering productivity in the development of heterogeneous integrated systems.

For decades the semiconductor market had been driven by speed-centric concerns. To maintain integration trends, in the late 1990's and early 2000's, different design objectives have been separated and the design flow systemized in the functional and communication domains, as shown in Figure 1.4. Today, existing scalable methodologies and architectures in these domains are being integrated into 3-D technologies. Alternatively, integrated in-package and distributed on-chip power delivery is under development, as shown in Figure 1.3, although power delivery is still currently dominated by *ad hoc* approaches. The lack of a methodology, architectures, and circuits for scalable on-chip power delivery and management is at the forefront of current

MtM problems. An overview of power management in nanoscale electrical circuits is presented in Section 1.1. An outline of the proposal is provided in Section 1.2.

1.1 Power delivery in integrated circuits

The delivery of high quality power to the on-chip circuitry with minimum energy loss is a fundamental requirement of all ICs. In a modern SoC, the power supplies provide the required voltage for the ICs within the overall system (CPUs, GPUs, hard disks, storage, sensors, and others), as exemplified by the 25 power domains within the Toshiba microprocessor shown in Figure 1.6. A regulated 12 volt output voltage

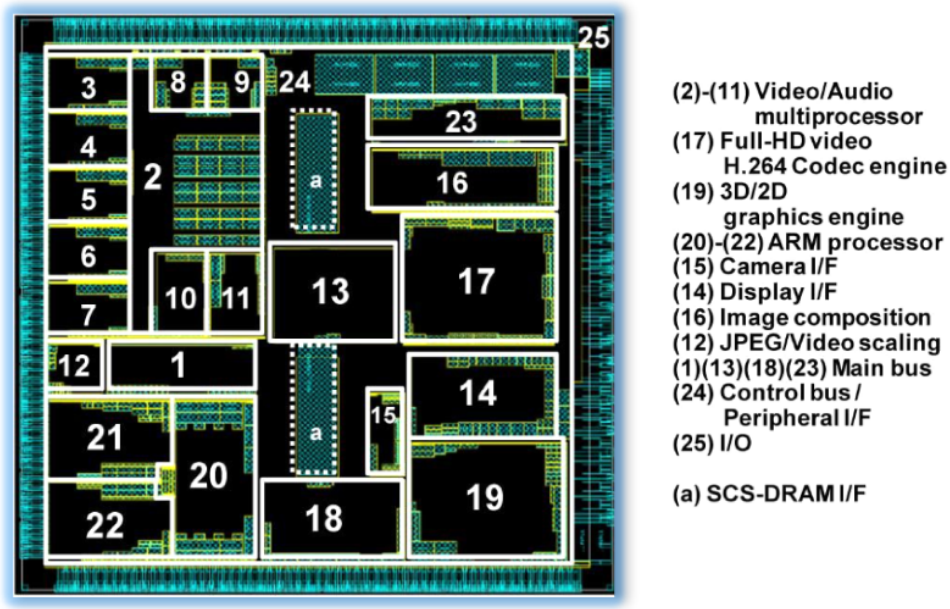


Figure 1.6: Toshiba full high definition (full HD) multiprocessor with 25 power domains [23].

is typically derived off-chip from a 48 volt battery [24]. The on-chip DC voltages are significantly lower and range from a fraction of a volt in the low power digital blocks to several volts in the input/output buffers, high precision analog blocks, and storage ICs [25]. To supply sufficient power, a higher unregulated DC voltage is typically stepped down and regulated within the power delivery system [24]. As the physical feature size of the digital circuits has scaled, on-chip integration has scaled from small scale integration (SSI) with tens of transistors to very large scale integration (VLSI) and beyond with several billions of transistors [10]. The total current load and power delivery losses have also increased with scaling of the on-chip devices, increasing power consumption and degradation in the quality of the supplied power. The quality of power and efficiency of the overall power supply system are primary concerns in high performance ICs.

Switching and linear DC-DC converters are common topologies for DC-DC conversion and regulation [24]. The operating principle of switching converters is based on storing and retrieving energy during, respectively, the on and off phases of the switching operation. Alternatively, linear power converters dissipate excessive power as heat. Historically, power was converted off-chip with large and power efficient switching mode power supplies (SMPS) and delivered on-chip via interconnect wires

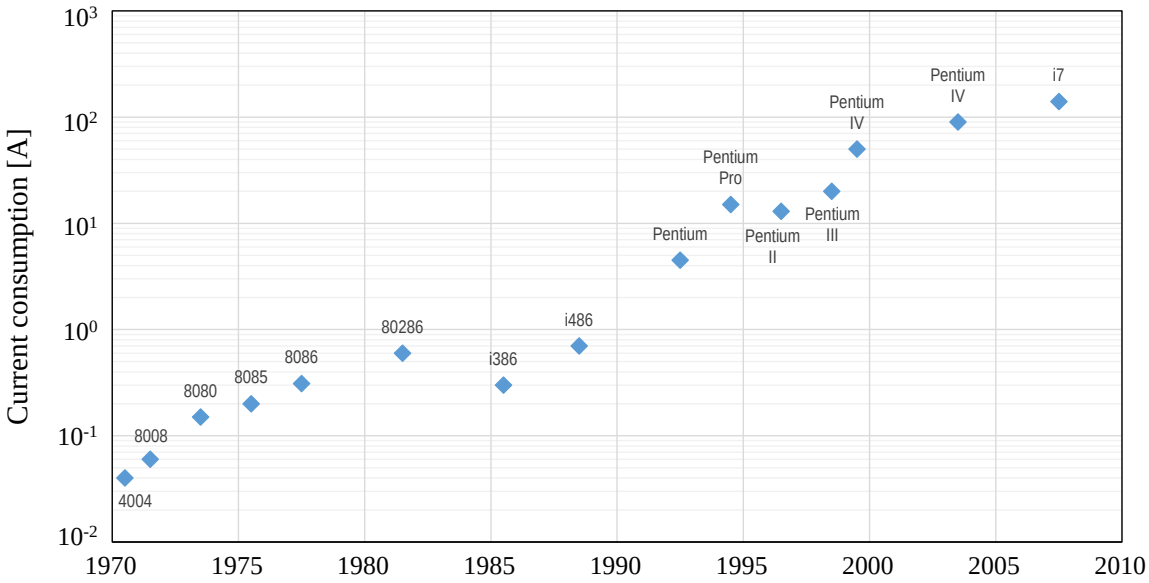


Figure 1.7: Current consumption in Intel microprocessors over the last four decades.

from the input/output (I/O) pads. Evolution of the on-chip current in Intel microprocessors is illustrated in Figure 1.7 [21], [26]–[40]. With the increasing number of on-chip devices and greater current demand, the voltage drop (proportional to the current consumption) and power dissipated by the interconnect and I/O pads (proportional to the square of the current consumption) have also increased, degrading the efficiency and quality of the on-chip power. The limited number of I/O pads is an additional concern. To provide high quality power to the loads while decreasing parasitic losses and the number of dedicated power pins, the power supply in a package (PSiP) with partially off-chip yet in package power supplies and on-chip power supply approaches have recently been developed [41].

While in-package and on-chip power integration has become a primary objective for systems integration [12], [42]–[50], research has remained focused on developing more compact and efficient power supplies, lacking a methodology to effectively integrate and manage the in-package and on-chip power delivery system. The challenge becomes greater as the diversity of modern systems increases, and dynamic voltage scaling (DVS) and dynamic voltage and frequency scaling (DVFS) become a part of the power management process. Recently, hundreds of on-chip power domains with tens of different supply voltage levels have been reported, and thousand-core ICs are being considered [23], [51]–[55]. Scalable power delivery systems, and the granularity of power management in DVS/DVFS multicore systems are limited by existing *ad hoc* approaches. To cope with this increasing design complexity and the quality and overall efficiency challenges of next generation power delivery systems, enhanced methodologies for designing scalable, hierarchical power management and delivery systems with fine granularity of dynamically controlled voltage levels are required.

1.2 Outline

Power delivery and management are fundamental to all nanoelectronic systems and have become primary concerns in high performance integrated circuits. On-chip

power supply techniques and circuits are reviewed in Chapter 2. Design objectives and the challenges of specialized power delivery circuits, microcontrollers, and ultra-small power supplies are reviewed. Switching and linear power supply topologies are also discussed.

Computer-aided design (CAD) for power delivery is reviewed in Chapter 3. Existing automated design techniques for the co-design of distributed power supplies are discussed. The architectural aspects of power delivery and management in portable mobile electronic systems are also reviewed.

To provide a high quality power delivery system, the power needs to be regulated on-chip with ultra-small locally distributed power efficient converters. Different types of power supplies are reviewed in Chapter 4. To exploit the advantages of existing power supplies, a heterogeneous power delivery system is proposed. The power efficiency of the system is shown to be a strong function of the clustering of the power supplies - the specific configuration in which the power converters and regulators are co-designed.

The co-design of power supplies to maximize overall power efficiency is computationally inefficient and impractical with exhaustive clustering approaches in real-time systems. A recursive clustering algorithm with polynomial computational complexity is proposed in Chapter 5 for an optimal real-time power distribution system with

minimum power losses. The proposed algorithm is evaluated on IBM power grid benchmark circuits and two multi-power domain circuits, demonstrating up to a 21% increase in power efficiency and orders of magnitude speedup in run time with the proposed recursive clustering algorithm.

The stability of a distributed power delivery system with multiple power supplies driving a single power network is an additional concern. A fully integrated power delivery system with distributed on-chip low-dropout (LDO) regulators developed for voltage regulation in portable mobile devices is described in Chapter 6. The circuit is fabricated in a 28 nm CMOS process. Each LDO employs adaptive bias for fast and power efficient voltage regulation, exhibiting 0.4 ns response time of the regulation loop and 98.45% current efficiency. An adaptive compensation network is also employed within the distributed power delivery system to maintain a stable system response. The proposed system is the first successful silicon demonstration of a stable fully integrated system of parallel analog LDO regulators.

With hundreds of power domains and thousands of cores, DVS and DVFS within each core are primary design objectives for efficiently managing a power budget. In addition, line regulation will become more important when hundreds of power supplies operate together on a single monolithic substrate. Efficient design techniques and circuits for adaptive control of power lines are therefore important. A

digitally controlled current starved pulse width modulator for adaptively changing the voltage of a power converter is described in Chapter 7. Analytic closed-form expressions for the operation of a pulse width modulator are provided. The accuracy and performance of the proposed pulse width modulator is evaluated with 22 nm CMOS predictive technology models under PVT variations. The proposed pulse width modulator is appropriate for dynamic voltage scaling systems due to the small on-chip area and high accuracy under PVT variations.

A multi-feedback system with parallel connected power supplies delivering current to a single grid exhibits significant design complexity and degraded stability due to complex interactions among the power supplies, power distribution network, and current loads. While the design complexity of these power delivery systems can be addressed with analog integrated circuit techniques, no straightforward method exists to evaluate the stability of these multi-feedback systems. A criterion for evaluating the stability of a distributed power delivery system is therefore needed. A passivity-based criterion is proposed in Chapter 8 for maintaining a stable power delivery system composed of multiple regulators. A distributed power delivery system with parallel connected LDO regulators is evaluated, exhibiting a stable multi-feedback response if and only if the proposed passivity-based criterion is satisfied. Application of the proposed stability criterion to automating the power delivery design flow is

also discussed.

Centralized power delivery systems have recently been used to dynamically manage power in heterogeneous high performance multicore systems, requiring a long feedback path from the individual power domains to a single power management controller. Additional power is dissipated in centralized power delivery systems due to unnecessary data transport, the slow response from the long feedback path limits real-time control over the energy budget, and all of the power management functions in one or a few locations may not scale. Alternative power management approaches should therefore be considered to efficiently manage the power budget in real-time in these distributed power delivery systems. The concept of a power network on-chip (PNoC) is introduced in Chapter 9 as a systematic methodology for on-chip power delivery and management that provides enhanced power control and real-time management of resource sharing. The proposed PNoC utilizes a modular architecture with intelligent distributed on-chip power routers to address the issues of design complexity and scalability.

Finally, directions for future work are proposed in Chapter 10. Specifically, a heterogeneous power management architecture is described based on the PNoC system, where the energy budget is distributed and locally managed with intelligent power routers. In addition, prescheduled DVS/DVFS operations are controlled by

a centralized power management circuit. Algorithms and specialized controllers for distributed power management should therefore be investigated. Integration of preventive techniques for power attacks within the PNoC platform is also proposed to secure the operation of the information and performance in sensitive ICs. The research presented in this dissertation is concluded in Chapter 11.

Chapter 2

Power Converter/Regulator Circuits

Compromises in speed and power are the new semiconductor reality of the 21st century, as illustrated in Figure 2.1 describing the lead Intel processors, Bloomfield, Lynnfield, Sandy Bridge, Ivy Bridge, and Haswell (the codenames for Intel processor microarchitectures launched between 2007 and 2013) [39]. To support the demand for decreasing cost per function, novel design concepts of functional diversification, on-chip integration, parallelism, and dynamic control have recently been adopted, posing significant circuit and system level challenges on power delivery and management in modern ICs. To cope with these challenges, traditional power delivery needs to be revised at the architectural, circuit, device, and material levels. Existing circuit level techniques to convert and regulate power are reviewed in this chapter in terms of the advantages, drawbacks, and compatibility, with a focus on on-chip integration of

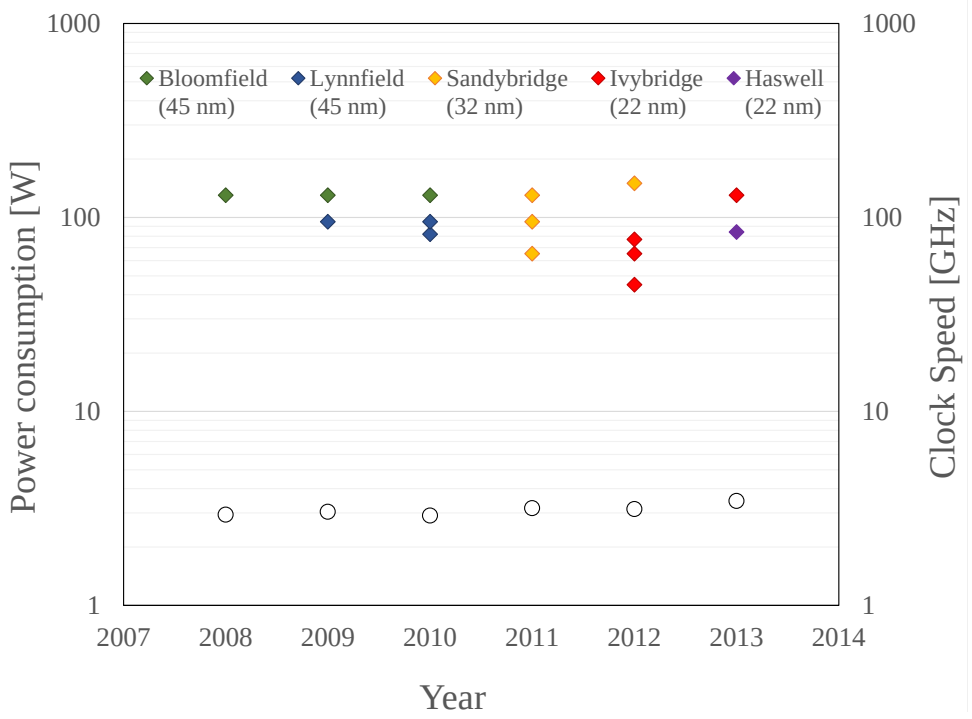


Figure 2.1: Speed and power consumption in recent Intel microprocessors [39].

heterogeneous systems.

One of the basic building blocks of converting and regulating power within a power delivery system is a DC-DC power converter. Power supplies step voltage signals either up (boost) or down (buck) and supply the required current to the load, while regulating the output voltage under PVT, line, and load variations. Historically, the power efficiency of the conversion process has been the defining characteristic of a power supply. With increasing on-chip noise and demand on the quality of the dynamically controlled power, load regulation of the power supply has also become

critical. With on-chip integration, the physical area of these power supplies is an additional concern.

Two typical strategies to convert power are to utilize passive storage elements for energy conversion (switching topologies) or to dissipate the excess energy within a resistive element (linear topologies). Two switching topologies, switching mode power supplies (SMPS) and switched-capacitors (SC), are discussed, respectively, in Sections 2.1 and 2.2. A linear topology is reviewed in Section 2.3. A comparison of these types of power supplies is provided in Section 2.4, followed by a brief summary of the chapter in Section 2.5.

2.1 Switching mode power supplies

To convert and regulate voltage under different load conditions, a typical SMPS transfers power from a voltage source to multiple current loads in two steps. The energy from the input is stored in an inductor L and capacitor C, and delivered from these elements to the load in alternating phases. A switching network reconfigures the LC connections into different phases, as illustrated in Figure 2.2, for a step up (boost) and step down (buck) converter.

The principle of energy conservation is key to understanding the operation of an SMPS converter. Assuming lossless SMPS components, the energy supplied from the

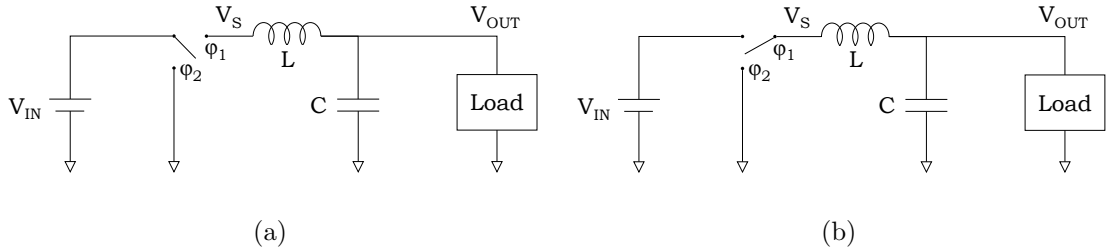


Figure 2.2: An SMPS, (a) boost, and (b) buck converter.

input is equal to the energy delivered to the load. The voltage at the output, noted as V_{OUT}, is, therefore, the average of the input voltage over the switching period T_S, yielding

$$V_{OUT} = \frac{1}{T_S} \int_0^{T_S} V_S(t)dt = \begin{cases} \dfrac{V_{IN}}{1-D} & \text{in boost converter} \quad (2.1) \\ D \cdot V_{IN} & \text{in buck converter,} \quad (2.2) \end{cases}$$

where $0 < D < 1$ is the duty cycle of the switching signal. The DC voltage at the output of a boost (buck) converter V_{OUT} is, therefore, higher (lower) than the input voltage V_{IN}, and increases with the duty cycle D in both of the boost and buck topologies. Due to the nonideal characteristics of the filter, some of the high frequency harmonics also propagate to the output, generating a voltage ripple $\Delta V_{OUT}(t)$ at the output and a current ripple $\Delta I_L(t)$ in the inductor current. In modern ICs, a *small ripple* or *linear ripple approximation* ($\Delta V_{OUT}(t) < 1\%$) is typically maintained [56]. Assuming ideal switch network characteristics and the linear ripple approximation, the current and voltage ripple in both the boost converter and buck converter

increases with smaller area and lower switching frequency $f_s = 1/T_S$ of the SMPS,

$$\Delta I_L \propto \frac{D}{2Lf_s}, \tag{2.3}$$

$$\Delta V_{OUT} \propto \frac{D}{16LCf_s^2}. \tag{2.4}$$

The peak currents of the switching devices increase with higher peak current of the inductor, increasing the size and cost of the switching converter. Large values of ΔI_L are, therefore, undesirable. The current ripple is typically significantly smaller than the total current, not exceeding 20% of the maximum load current [56]. The expressions in (2.3) and (2.4) with the design criteria $\Delta V_{OUT}(t) < 1\%$ and $\Delta I_L(t) < 20\%$ impose important constraints on the inductor and capacitor characteristics, limiting the minimum size of the LC filter and minimum switching frequency of the converter. SMPS converters have historically been designed with large LC filters to satisfy these design criteria. The physical size of a power converter is a primary concern for on-chip integration which limits the effectiveness of large SMPS converters in distributed on-chip power delivery systems. Alternatively, compact switching converters can potentially be designed to operate at higher switching frequencies. The effectiveness of these converters is discussed in Chapter 4.

Power efficiency is another important characteristic of a power converter. The

power efficiency of an SMPS converter is primarily limited by the parasitic impedance of the inductors, capacitors, and switches, approaching the theoretical limit of 100% with ideal circuit elements. In a practical switching converter, power losses due to switching activity and resistive parasitic impedances are unavoidable, limiting typical SMPS efficiencies to 85% to 95% [57], [58]. The overall power dissipated within an SMPS is the sum of the switch related power loss P_{SW}, inductor related power loss P_{IND}, and capacitor related power loss P_{CAP}. The switching losses increase with the total parasitic capacitance of the circuit elements and with the switching frequency, and are proportional to the square of the input voltage. The resistive losses increase with the total parasitic resistance of the circuit elements and are proportional to the square of the output and ripple currents, yielding

$$P_{SMPS} = P_{SW} + P_{IND} + P_{CAP} = \sum_i \alpha_i \left(R_{PAR,i}\right) I_i^2 + \sum_i \beta_i \left(C_{PAR,i}\right) f_s V_i^2, \quad (2.5)$$

where α_i and β_i are power coefficients for a specific converter, $R_{PAR,i}$ and $C_{PAR,i}$ are, respectively, the parasitic resistance and capacitance of the inductors, capacitors, and switches, and I_i and V_i are, respectively, the current flowing through the parasitic resistors and the voltage across the parasitic capacitors within the converter. These parasitic impedances are illustrated in Figure 2.3 for a buck converter with an inductor, capacitor, and two MOSFET switches. In this example, a buck

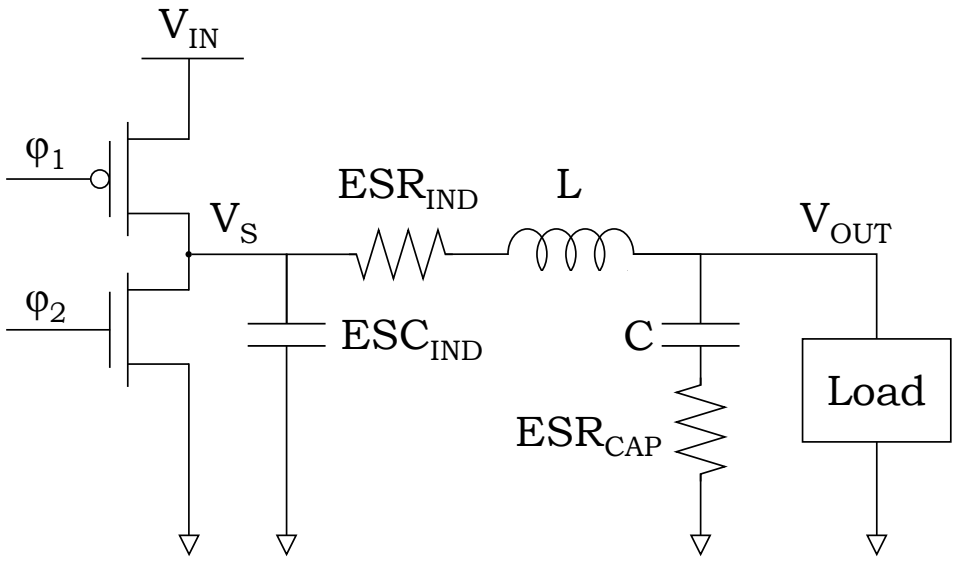


Figure 2.3: A buck converter composed of a spiral inductor L, MOS capacitor C, and MOSFET switches.

converter is comprised of two MOSFET switches with an equivalent series resistance $R_{PAR,1} = R_O$ and total capacitance $C_{PAR,1} = C_{SW}$, a spiral inductor L with parasitic series resistance $R_{PAR,2} = ESR_{IND}$ and stray capacitance $C_{PAR,2} = ESC_{IND}$, and a MOS capacitor C with a series parasitic resistance $R_{PAR,3} = ESR_{CAP}$. The overall power dissipated within a buck converter is therefore

$$\begin{aligned} P_{Buck} = &\left(\alpha_1 R_O + \alpha_2 ESR_{IND}\right)\left(I_{OUT}^2 + \frac{1}{3}\Delta I_L^2\right) + \alpha_3 ESR_{CAP}\Delta I_L^2 \\ &\left(\beta_1 C_{SW} + \beta_2 ESC_{IND}\right) f_s V_{IN}^2. \end{aligned} \tag{2.6}$$

The effect of the parasitic capacitance C_{SW} and ESC_{IND} on switching power losses increases with higher values of f_s. Alternatively, the inductor current decreases with

higher values of f_s, lowering the resistive power losses. Note that ESR_{IND} also varies with the switching frequency [59], affecting the resistive losses of an SMPS. The power efficiency of an SMPS is, therefore, a strong function of the switching frequency f_s, and depends upon the specific converter [24], [41]. An SMPS converter should switch at a frequency that maximizes the overall power efficiency of the specific converter.

2.2 Switched-capacitor converter

Another type of switching power supply is a switched-capacitor converter, also referred to as a charge pump (CP) [60]. SC converters utilize capacitors and switches to step up or step down the input voltage, based on the principles of charge conservation [24], [41]. The power is typically converted in two non-overlaping time phases, φ_1 and φ_2. In each phase, capacitors are connected in a different configuration, affecting the distribution of charge across the switched capacitors. The operation of a simple one stage CP is illustrated in Figure 2.4 [60]. During the first phase of the period T_S ($\varphi_1 : 0 \leq t \leq T_S/2$), switches S_1 and S_2 are, respectively, closed and open, and the V_{PWM} signal is low, charging capacitor C through the input source $V_{IN} = V_{DD}$ and discharging the output node by the load current I_{OUT} (see Figure 2.4(a)). At the end of φ_1, capacitor C is charged to V_{IN}, and the output node is discharged by $I_{OUT} \cdot T_S/2$. During the second phase ($\varphi_2 : T_S/2 \leq t \leq T_S$), switches

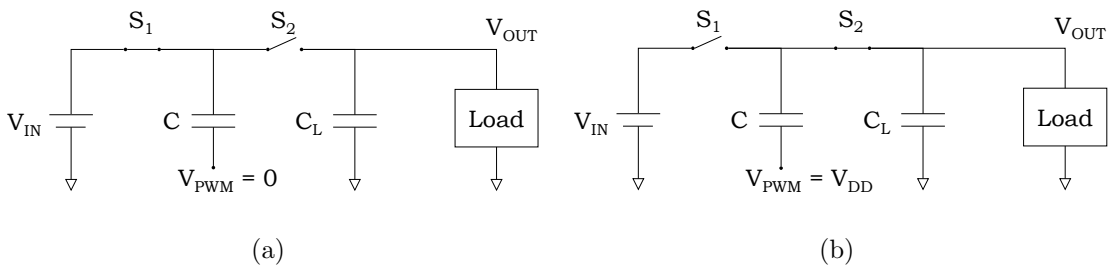


Figure 2.4: Single stage switched-capacitor converter, (a) phase 1 ($\varphi_1 : 0 \leq t \leq T_S/2$) operation, and (b) phase 2 ($\varphi_2 : T_S/2 \leq t \leq T_S$) operation.

S_1 and S_2 change state and the signal V_{PWM} switches from low to high, as shown in Figure 2.4(b)). The charge stored on C during φ_1 is redistributed between capacitor C and capacitive load C_L, and supplies the load current I_{OUT} during phase φ_2. As a result, the voltage at the output increases during each subsequent cycle up to a final asymptotic steady state value [60],

$$V_{OUT}^{(SS)} = 2V_{DD} - \Delta V = 2V_{DD} - \frac{I_{OUT}}{f_s C}, \tag{2.7}$$

where ΔV is the voltage drop due to charge sharing, and $f_s = 1/T_S$ is the switching frequency of the converter. The capacitor C behaves as a charge pump in the SC converter and, for sufficiently high switching frequencies, the output voltage in steady state is maintained close to $2V_{DD}$ [60]. The voltage drop at the output increases with higher values of I_{OUT} (see (2.7)), making load regulation difficult in SC converters with high load currents. To limit the voltage drop at the output of an SC converter,

a minimum switching frequency $f_{s,MIN}$ is determined. This voltage drop can be also reduced by increasing the size of the capacitors, trading the larger physical size of the SC for improved output voltage regulation. Large power supplies are, however, ineffective for on-chip integration. To enhance the ability of a SC converter to regulate the load, feedback circuitry is added at the expense of lower efficiency of the power conversion process. Alternatively, low current applications that require a low-to-high voltage conversion, but not necessarily excellent load regulation (such as wireless monitoring systems, non-volatile memory, and mixed-signal systems [61]–[64]) are natural applications for switched-capacitor converters.

The power efficiency of a typical switched-capacitor DC-DC converter is primarily limited by heat losses incurred when transferring charge between the switched capacitors P_{TRANS}, losses in the power switches P_{SW}, and dynamic power dissipated in the parasitic resistance P_{DYN}. While the theoretical power efficiency of other switching converters (e.g., SMPS) is 100%, the power loss due to charge transfer P_{TRAN} is unavoidable in switched-capacitors, degrading the maximum power efficiency of an SC converter by [24]

$$P_{TRANS} = \frac{f_s}{2} \cdot \frac{CC_L}{C + C_L} \cdot (V_{DD} - \Delta V)^2 . \tag{2.8}$$

The switching and dynamic components of the total power loss are strongly dependent upon the design of the SC switches and are comparable with switching and dynamic power losses in other switching converters. Alternatively, power losses due to the parasitic resistance of the wire P_{TRANS} are specific to the switched-capacitor topology and increase with switching frequency, as described by (2.8). Intuitively, to increase the power efficiency of an SC converter, the capacitors should switch at a lower frequency. This solution is however limited by the minimum frequency constraint $f_s > f_{s,MIN}$ determined from (2.7) which is related to the voltage drop and physical size of the converter. To increase the efficiency of an SC converter at over a range of frequencies dynamically reconfigurable topologies should be considered [65]. The size, load regulation, and power efficiency characteristics are all important design criteria for integrating on-chip power supplies. Due to the unfavorable tradeoff among these characteristics in an SC converter, switched-capacitor converters are not preferred for distributed on-chip integration in modern heterogeneous systems. An alternative topology that exhibits small area and excellent load regulation is described in the following section.

2.3 Linear voltage regulator

Linear regulators convert a source of DC current from one voltage level to a lower voltage level by dissipating the excess energy as waste heat in the output device. This output device acts as a variable resistor (varistor) and is controlled through a feedback loop, maintaining a steady voltage at the output. This concept is illustrated in Figure 2.5. No passive inductors and capacitors are required, as shown in Figure 2.5, making linear regulators particularly appealing for on-chip integration. Due to the low output resistance, linear regulators exhibit fast load regulation, an important characteristic for on-chip power regulation. Although the output impedance is low, high power is dissipated in the output varistor at high load currents, limiting the efficiency of the linear power conversion to the output-to-input

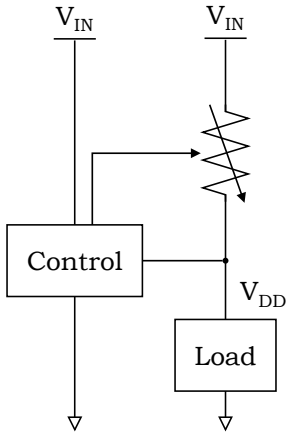


Figure 2.5: Linear DC-DC regulator.

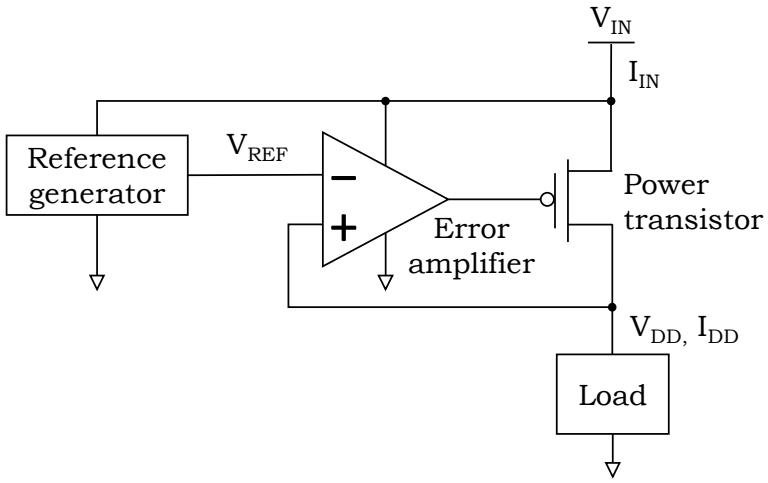


Figure 2.6: LDO regulator with a power transistor controlled in a negative feedback loop with an error amplifier.

voltage ratio V_{OUT}/V_{IN}. In addition, stepping up the input power is not possible in linear regulators.

Among these limitations, low power efficiency is the primary drawback of a linear power converter [24]. To enhance power efficiency, a linear DC-DC regulator is typically operated as a low-dropout voltage regulator (LDO) with a small difference between the input and output voltage. A block diagram of an LDO regulator is shown in Figure 2.6 with a control block and varistor with, respectively, an error amplifier and a common source connected power transistor. The total power efficiency of an LDO is a product of the voltage efficiency φ_V and current efficiency φ_I of the linear regulator,

$$\varphi = \varphi_V \cdot \varphi_I. \tag{2.9}$$

The power dissipated within an LDO regulator is composed of dynamic losses in the output power transistor and static losses in the error amplifier that limit, respectively, the voltage efficiency φ_V and current efficiency φ_I,

$$\varphi_V = \frac{V_{OUT}}{V_{IN}}, \tag{2.10}$$

$$\varphi_I = \frac{I_{OUT}}{I_{OUT} + I_Q}, \tag{2.11}$$

where I_Q is the quiescent current flowing through the error amplifier to ground. Modern LDO regulators typically achieve excellent current efficiencies, ranging between 98% and 99.99% [**?**], [66]–[68]. Thus, the voltage drop $V_{IN} - V_{OUT}$ across the power transistor is the primary cause of lower efficiency in a linear power converter. As the difference between high performance and low power voltage levels decreases [25], the typical V_{OUT}/V_{IN} ratio increases, enhancing the effectiveness of linear power supplies. While limited power efficiency and the inability to step up the voltage make linear power converters inappropriate for certain applications, the small size and excellent load regulation of LDO regulators are particularly appealing for distributed integration of on-chip power supplies.

2.4 Comparison of monolithic power supplies

As previously mentioned, the on-chip integration of multiple low voltage power supplies is a primary concern in high performance ICs. An integrated power system should deliver high quality power to multiple loads in an energy efficient manner. The load regulation and power efficiency of individual power supplies within a power delivery system are particularly important and affect the overall performance of the power delivery process. The physical size of a power supply is also critical for integrating multiple power supplies on-chip. The power efficiency, physical size, and load regulation characteristics are compared in this section for the three classical power supply topologies.

The operation of both switching mode power supplies and switched-capacitor converters is based on a two-phase principle. In both topologies, the energy is stored in passive circuit elements during one phase and restored at the output during the other phase. The capacitors and inductors both increase the physical area of the converter, making integration of an on-chip SMPS more problematic. Alternatively, SC converters are comprised of only capacitive elements, and therefore require less area, and are more suitable for on-chip integration. The high power efficiency and good load regulation characteristics of an SMPS converter are, however, significantly degraded in an SC topology, making switched-capacitor converters inefficient for

Table 2.1: Comparison of SMPS, SC, and LDO topologies for power conversion.

Power Converter	SMPS	SC	LDO
Step up conversion	Yes	Yes	No
Power efficiency	High	Medium	Limited to $\frac{V_{OUT}}{V_{IN}}$
Load regulation	Good	Poor	Good
Physical area	Large	Medium	Small
Historical applications	Microprocessors, DSPs, SRAMs, and hard disks	EEPROM, DRAM, flash, and mixed-signal	DRAM

certain applications.

Linear regulators exhibit small size due to the lack of capacitors and inductors, and excellent load regulation characteristics due to the small output impedance. The power efficiency of a linear regulator is limited by the voltage drop across the output power transistor, making this topology power inefficient when converting large differences between the input and output voltages. To maintain reasonable efficiency, low-dropout LDO regulators should be utilized for on-chip integration.

The primary electrical characteristics of SMPS, SC, and low-dropout linear topologies are summarized in Table 2.1. Switched-capacitor converters are not preferred for high performance applications due to the limited power efficiency and difficulty in regulating the output voltage under load variations. The high power efficiency and good load regulation of SMPS converters have made this topology particularly effective for power conversion in high power, high performance applications, such as microprocessors, DSPs, SRAMs, and hard disks [24]. The large physical area and

difficulty to integrate inductive elements on-chip degrade the effectiveness of SMPS converters as the need for high quality, distributed on-chip power delivery increases. Alternatively, small LDO regulators are a natural choice for on-chip integration. With continuous scaling of the physical feature size and advanced circuit solutions, thousands of ultra-small LDO regulators can potentially be distributed on-chip, close to the loads [69]. Nevertheless, novel design solutions are required to enhance the overall efficiency of LDO based power delivery systems.

To better exploit the electrical characteristics of existing power converters, different topologies have recently been combined into hybrid power supplies [70]–[72]. For example, linear power supplies can be utilized on-chip for local voltage regulation, whereas off-chip or in-package SMPS can mitigate power losses caused by large voltage drops [70]. In this scheme, the small size and excellent regulation characteristics of an LDO are combined with the high power efficiency of a switching converter, enhancing both performance and quality. Hybrid topologies not only exhibit promising qualities for heterogeneous on-chip power delivery, but also pose new research challenges. The co-design of different power converters, overall system-wide power efficiency, dynamic control, and scalability of heterogeneous power delivery systems are topics of growing importance which are discussed throughout this research proposal.

2.5 Summary

Well known switching and linear power supply topologies are reviewed in this chapter. Switching mode power supplies, switched-capacitors, and linear regulators are compared with respect to power efficiency, physical area, and load regulation characteristics. Based on the electrical and physical characteristics of these converters, the preferred applications for each topology are discussed. The effectiveness of SMPS, SC, and LDO regulators as distributed power supplies within a heterogeneous on-chip power delivery system is reviewed.

A heterogeneous power delivery system should exhibit high power efficiency and excellent load regulation characteristics. In addition, the integration of multiple on-chip power supplies poses tight limitations on the size of monolithic converters. Existing power converter topologies exhibit an undesirable tradeoff among power efficiency, load regulation, and physical size. To exploit the advantages of certain power supply topologies, a heterogeneous power delivery system with different integrated power supplies should be considered. Novel methods to co-design on-chip power regulators and converters are required.

Chapter 3

Modeling, Analysis, and Optimization Techniques for Power Delivery

As the power delivery system transforms from a lumped system with a few off-chip converters into a distributed, dynamically controlled system with many thousands of on-chip electrical components, the design, synthesis, and control objectives of the power delivery process need to be rethought at the system level. To cope with the complexity of system-wide power optimization, a scalable platform for efficient power delivery and management is required that addresses the issues of design complexity with novel circuit, firmware, methodology, and architectural level solutions. Existing techniques to model, analyze, and optimize a power delivery system are reviewed in this chapter in terms of different physical objectives (such as quality of power, physical area, and power efficiency), and design complexity with a focus on on-chip

integration of distributed heterogeneous real-time systems.

The generation and distribution of high quality power to the load circuitry are two primary issues in the power delivery process. The interconnect network distributing power within a modern microprocessor typically contains many million (to several billion) nodes. Accuracy and computationally efficiency are critical in the modeling, analysis, and optimization of these complex power delivery systems. To cope with the high complexity of modern power delivery systems, iterative analysis and optimization techniques are often utilized to enhance system accuracy. A spiral model of a power delivery design process with iterative analysis and optimization stages [73] is illustrated in Figure 3.1. Design variables, constraints, and objectives need to be determined in advance. Examples of design variables in power delivery systems are the type, number, and location of the passive and active power supplies. The objectives of the power delivery process vary from enhancing quality of power and energy efficiency to lowering on-chip area and computational run time in dynamically controlled systems. The variables and constraints of the power delivery process are related to the design objectives through the models, trading off fidelity of the power delivery process with analysis time. Once the variables, constraints, objectives, and models are determined, an iterative process of analysis and optimization of a power delivery system begins, increasing the accuracy and run time with each subsequent

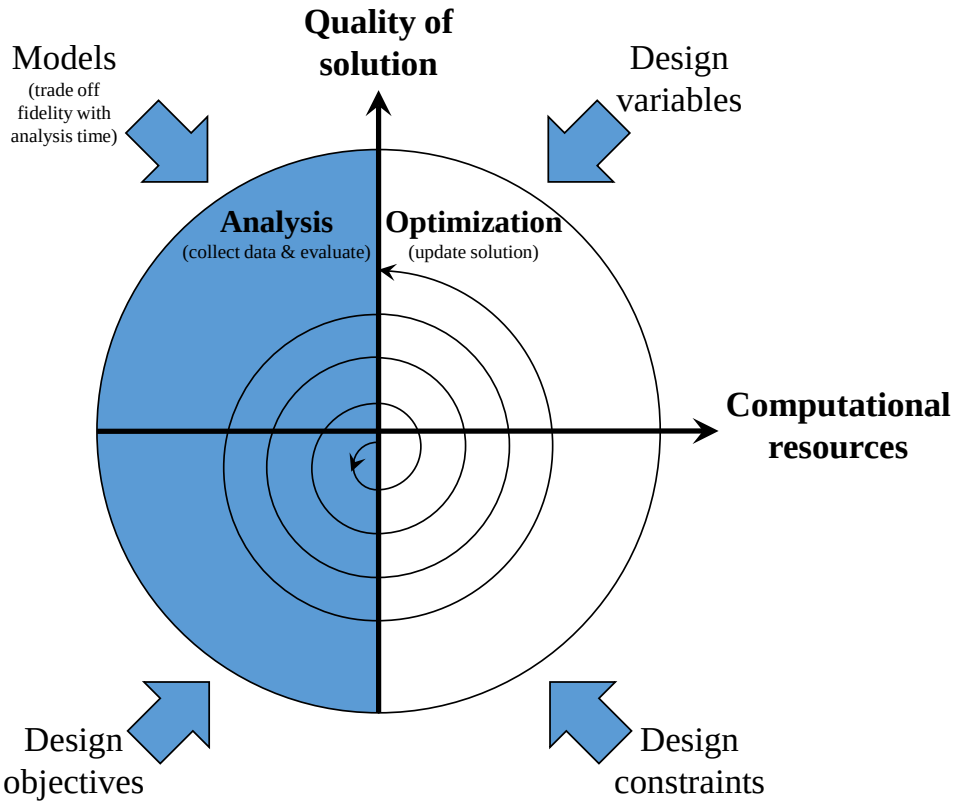


Figure 3.1: Spiral model of typical iterative power delivery design process.

iteration. Models that exhibit both high accuracy and low computationally complexity are important for designing complex power delivery systems. Existing modeling approaches for power delivery systems are reviewed in Section 3.1. Design techniques for analyzing and optimizing the power delivery process are discussed, respectively, in Sections 3.2 and 3.3. The chapter is summarized in Section 3.4.

3.1 Modeling power delivery systems

A typical power delivery system is a complex hierarchical structure that generates, regulates, and distributes power at the PCB, package, and IC levels. The current from the power supplies is delivered to the load via interconnect networks at the printed circuit board (PCB), package, and IC, referred to as a power distribution network, as illustrated in Figure 3.2. With the increase in current consumption and operational frequencies, the power and ground noise has become significant, degrading the quality of the distributed power. A typical approach for modeling a noisy power distribution network is discussed in this section. Existing solutions to enhance the quality of delivered power are reviewed. A model of the power network with the proposed solutions is also discussed.

The quality of the distributed power is characterized by the voltage drop ΔV in the power network (see Figure 3.2), due to the resistive R and inductive L network impedances, total load current $I(t)$, and rate of change in the power network current $di(t)/dt$,

$$\Delta V = I(t)R + L\frac{di(t)}{dt}. \tag{3.1}$$

While the total power and ground noise increases with both the ohmic voltage drop IR, and inductive *simultaneous switching noise* (SSN) $L\ di/dt$, the peak noise is not

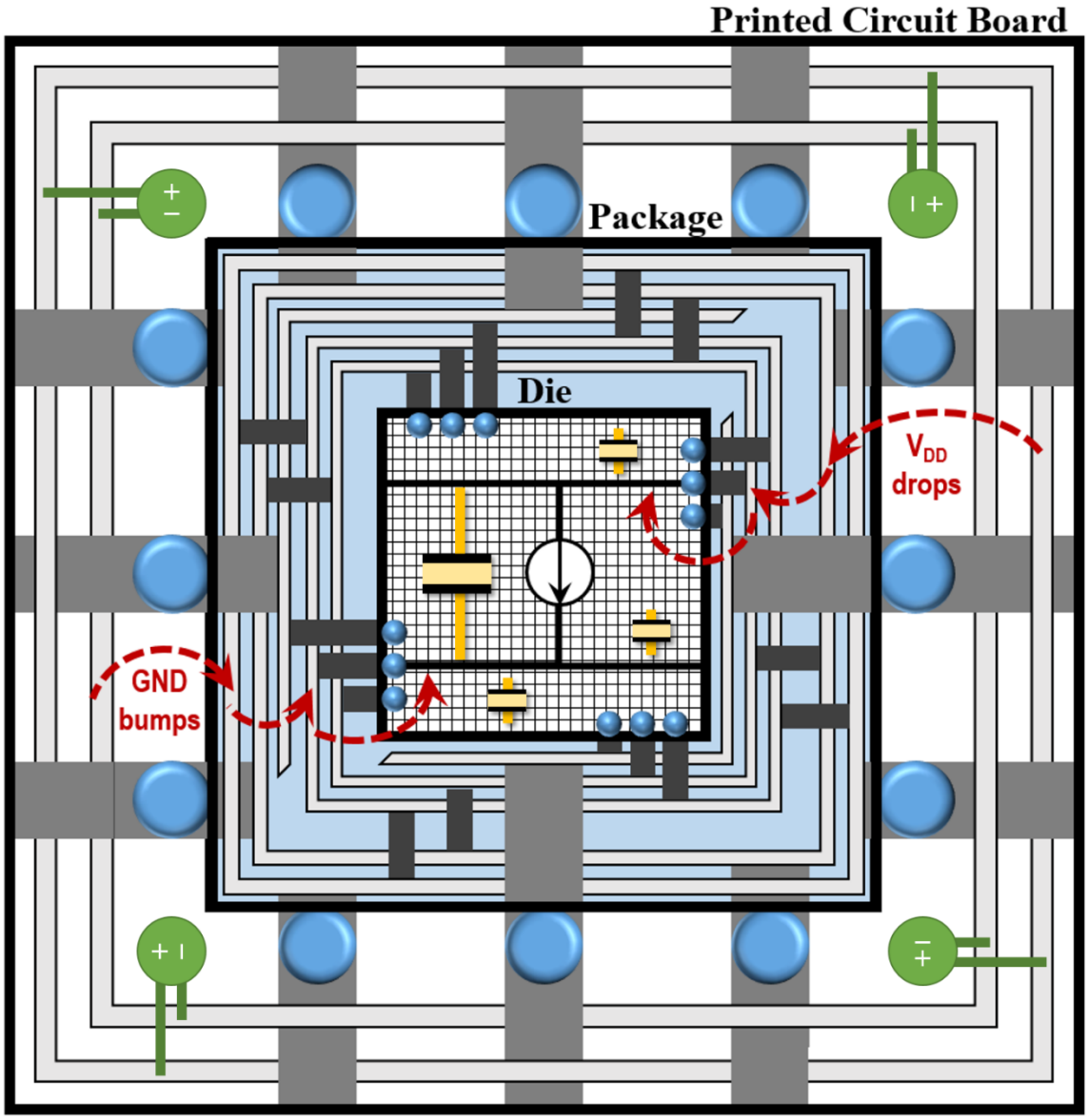


Figure 3.2: Power distribution network of the printed circuit board, package, and integrated circuit.

necessarily the sum of the greatest individual IR and $L\ di/dt$ drops. Pessimism in tolerating the maximum peak noise is avoided when simultaneously analyzing these

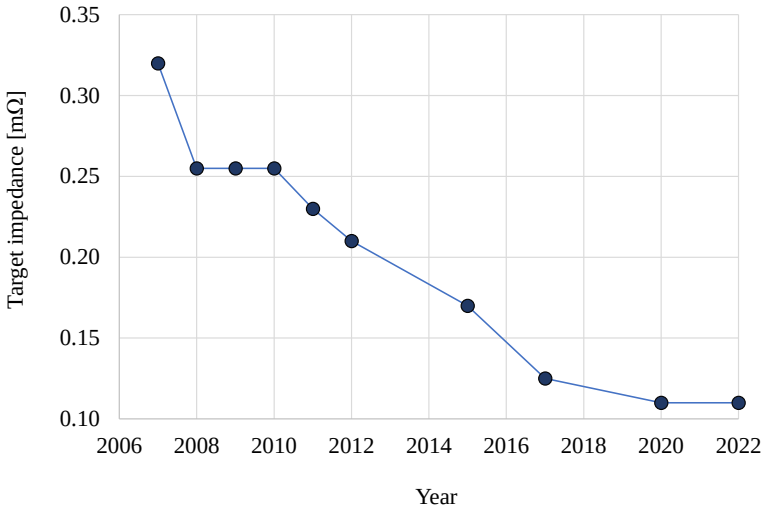


Figure 3.3: Trends for target output impedance of power distribution networks [74].

two types of noise.

Power and ground noise has increased with each technology generation with higher $I(t)$ and faster di/dt. The target power network impedance (Z_{target}) continues to drop for each technology generation [74], adding complexity to designing the power delivery system, as shown in Figure 3.3. Accurate modeling techniques are required to satisfy these stringent impedance constraints.

A power delivery system with an off-chip power supply and power distribution networks at the PCB, package, and IC levels is typically modeled as a series power and ground plane composed of RL impedances and parallel interplane capacitances, as illustrated in Figure 3.4. The output impedance of a typical power delivery system with total resistance R_{TOT} and total inductance L_{TOT} is $R_{TOT} + j\omega L_{TOT}$,

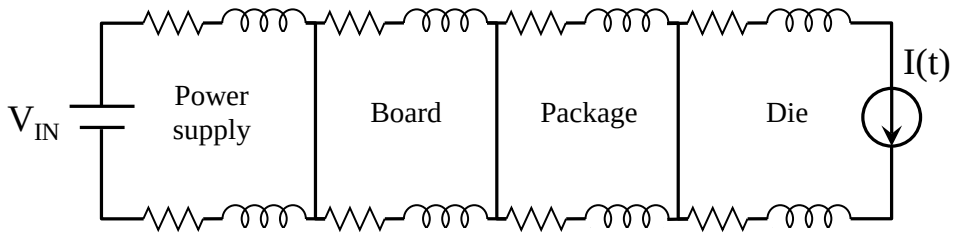


Figure 3.4: Circuit level representation of a system-wide power network with an off-chip power supply.

assuming the parasitic interplane capacitances are negligible [24]. The power delivery system exhibits resistive behavior at low frequencies $\omega < R_{TOT}/L_{TOT}$, and inductive behavior at high frequencies $\omega > R_{TOT}/L_{TOT}$, increasing at a constant rate for $\omega > R_{TOT}/L_{TOT}$ and exceeding the target power network impedance at $\omega > Z_{target}/L_{TOT}$. The operational frequency of a power delivery system comprised of off-chip power supplies and a power distributed network is, therefore, limited to $\omega < \omega_{max} = Z_{target}/L_{TOT}$.

To increase the operational frequency range, decoupling capacitors have traditionally been used at multiple levels of the power distribution network [75], [76]. To model the behavior of these capacitors over a wide range of frequencies, the decoupling capacitance C is typically connected in series with an effective series resistance (ESR) and effective series inductance (ESL) [77], as illustrated in Figure 3.5. The RLC components of the decoupling capacitors form series and parallel connections with other resistive and inductive components within the power delivery network.

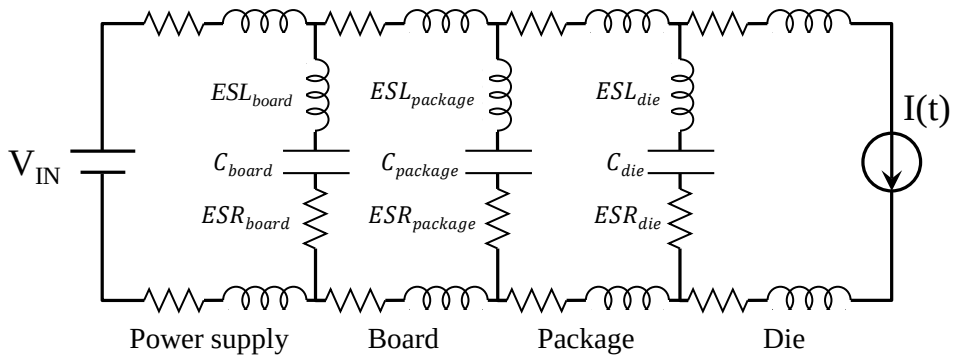


Figure 3.5: Circuit level representation of a system-wide power distribution network with off-chip power supply and decoupling capacitors at the board, package, and circuit levels.

The *resonant* and *antiresonant* behavior of, respectively, series and parallel RLC circuits is important to understand the effects of decoupling capacitors on the power delivery system [24]. The total impedance of a series RLC circuit, $R+j\omega L+1/j\omega C$, exhibits resonance with a pronounced maximum at the resonant frequency ω_{res}. Alternatively, the total admittance of a parallel RLC circuit, $1/R + 1/j\omega L + j\omega C$, contributes to the antiresonant behavior of the power network impedance with a pronounced minimum at the antiresonant frequency ω_{ares}. The magnitude of the output impedance is, therefore, constrained by the resonant and antiresonant behavior of the decoupling capacitors at the board, package, and circuit levels, as illustrated in Figure 3.6. To mitigate high frequency power and ground noise, the effective output impedance of the power network is determined by the least inductive, shortest loop at a sufficiently high frequency. The maximum power delivery frequency with the

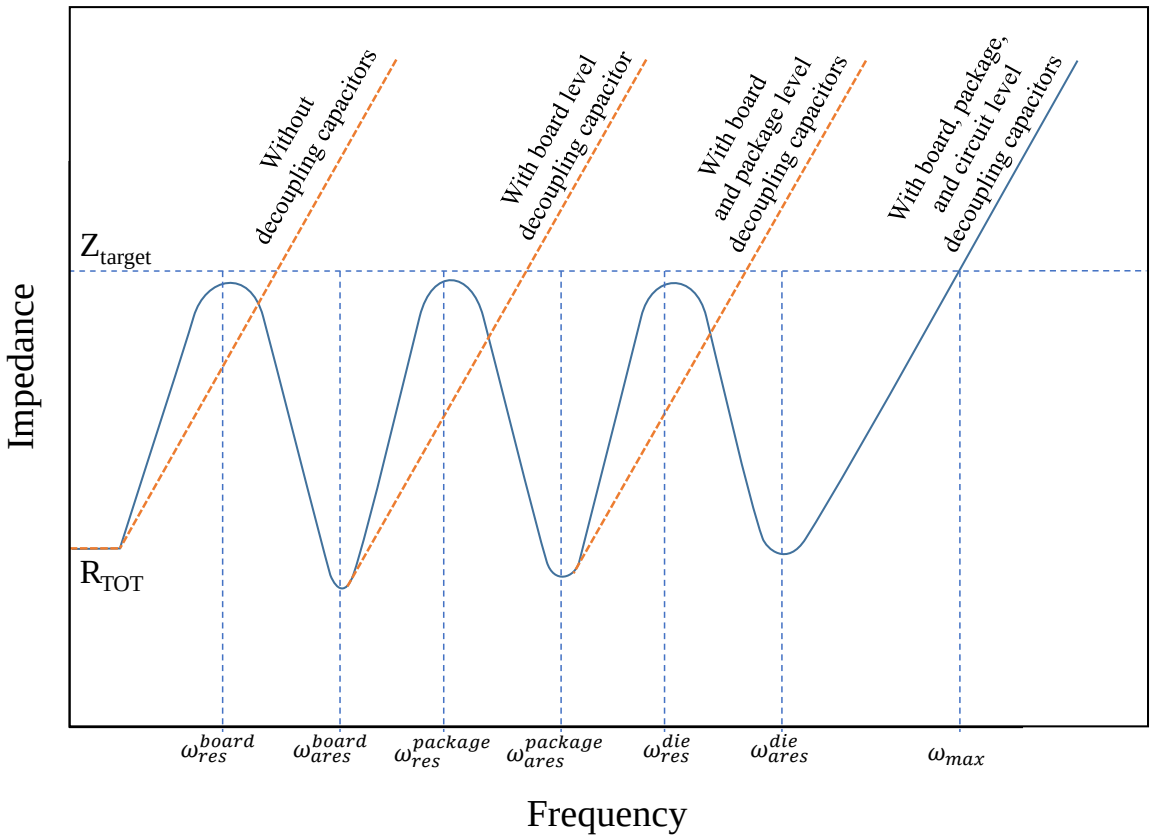


Figure 3.6: Impedance characteristics of a power delivery system with an off-chip power supply, and board, package, and integrated circuit level decoupling capacitors.

board, package, and circuit level decoupling capacitors is, therefore, limited by the impedance between the circuit level decoupling capacitor and the switching circuit, $\omega_{max} = Z_{target}/(L_{die} + ESL_{die})$, yielding a wider operational frequency range than without decoupling capacitors.

The size and parasitic impedance of the decoupling capacitors play an important role in the design of an efficient *hierarchical* power delivery system. In modern ICs,

the decoupling capacitors are sized and placed to satisfy the target impedance over a broad range of frequencies. To maintain power integrity in high performance ICs, many tens of thousands of decoupling capacitors are distributed on-chip [78]–[81].

Another approach to mitigate power and ground noise is to regulate the power closer to the load by integrating active power supplies in the package and/or on-chip. To integrate a power converter in-package or on-chip, advanced passive components, packaging technologies, and circuit topologies are essential. In-package and on-chip power converters have recently been considered with respect to cost, complexity, and performance [42], [66], [68], [71], [82]–[98]. With these power converters integrated on-chip, a power supply system with multiple on-chip power converters can be developed to improve the quality of the power delivered within an IC. Power delivery systems with tens to hundreds of power supplies integrated at different hierarchical levels will someday be integrated on-chip for delivering high quality power in high performance ICs.

3.2 Analysis of power delivery system

In a typical hierarchical system, the number of nodes increases exponentially at lower hierarchical levels, complicating the design and analysis process. Specifically, in power delivery systems in large scale ICs, power grid analysis at the circuit level is

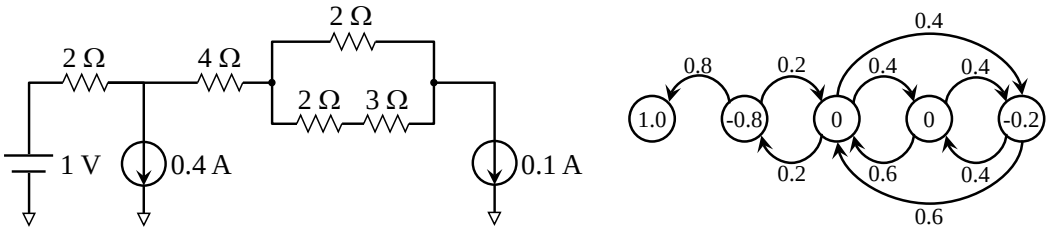
highly complicated, and straightforward nodal analysis of equations with more than many millions or billions of components is computationally infeasible for multiple power supplies and billions of highly nonlinear loads [99]. The challenge becomes greater as requirements on the granularity of the power management process in DVS/DVFS real-time multicore systems increase. Specialized techniques are, therefore, applied to provide accurate, computationally efficient power delivery analysis. While direct solvers (e.g., SPICE) are deterministic and robust, these techniques require huge storage and computational resources, and are inefficient for real-time power delivery systems. Alternatively, with numerical techniques that produce iterative methods, a partial solution is determined after each subsequent iteration, finally converging to a useful result. The special characteristics of practical power grids, such as spatial locality, smoothness of the voltage differences, and symmetry of the power grids, are exploited to ensure fast convergence with iterative analysis techniques. Approaches that exploit hierarchy and statistical methods to improve the computational complexity of the numerical analysis process are reviewed, respectively, in Sections 3.2.1 and 3.2.2. A computationally efficient method to analyze power grids based on the effective resistance between circuit components is discussed in Section 3.2.3.

3.2.1 Hierarchical power grid analysis

Hierarchical partitioning is inspired by the divide-and-conquer approach and is commonly used to speed up the power grid analysis process [100], [101]. The key idea behind hierarchical analysis is partitioning a power grid into local sub-grids of reduced size, generating macromodels for the local partitions, and connecting these macromodels to a global grid with a specialized interface [100], [101]. The size of the power grid analysis problem can be significantly decreased by applying hierarchy. In addition, multiple sub-grids are analyzed in parallel with existing iterative techniques, enhancing the overall speed of the power network analysis process. Changes within a specific partition affect only the macromodel of the modified partition, enhancing the flexibility and scalability of the hierarchical analysis process.

3.2.2 Statistical power grid analysis

Random walk is another approach for computationally efficient power grid analysis [102]–[105]. In this approach, the power grid is modeled as a weighted graph, where the impedance between adjacent nodes determines the transition probability between these nodes. The power supplies and current loads are represented as nodes within a graph with, respectively, positive (reward) and negative (cost) values. The magnitude of these positive and negative node values depends upon, respectively,



(a) (b)

Figure 3.7: Power grid with a single power supply and two current loads, (a) circuit level representation, and (b) weighted graph representation.

the level of the voltage and the current sunk at a node. A power grid with a single power supply and two current loads is illustrated in Figure 3.7 with the corresponding weighted graph. Based on the weighted graph, a random walk solver is applied to determine a path from any arbitrary node to a reward node. The value of the original random node and the voltage at the corresponding circuit node are determined based on the path to the reward node and the reward value [102], [103]. The average node values over a sufficiently large number of iterations represent the solution of the power grid, according to the central limit theorem [106]. The computational efficiency of random walk-based analysis increases with the number of reward nodes, exhibiting fast power grid analysis with a sufficiently large number of power supplies. Alternatively, for large grids with few power supplies and real-time power management, this technique is less preferable due to the long computational run time.

3.2.3 Effective resistance-based power analysis

Practical power grids in high performance ICs can be treated as locally uniform, globally non-uniform resistive meshes. IR voltage drop analysis for practical resistive power grids with multiple power supplies and current loads have recently been proposed in [107], [108], exploiting the principle of spatial locality [109]–[113]. Intuitively, most of the current is provided to a node by those power supplies located in close proximity with that node [112]. The current contribution decreases significantly with increasing distance between the load and power supply. To illustrate an effective power supply region, consider a current load connected to a power grid, as exemplified in Figure 3.8. The current from the four power supplies within the first ring and the twelve power supplies within the second ring exhibits, respectively, 60% and 30% of the total current supplied to the load, as in Figure 3.8 [107]. Alternatively, each power supply within the third ring contributes less than 1% of the total current [107]. The size of the power grid partition that effectively supplies power to an individual node can therefore be significantly smaller for each node, reducing the overall computational complexity of the power grid analysis process. Distant partitions of a power grid can be analyzed in parallel, further reducing the computational run time of the analysis process. An effective resistance-based analysis exhibits high accuracy as compared with SPICE, and significant computational speedup when compared with

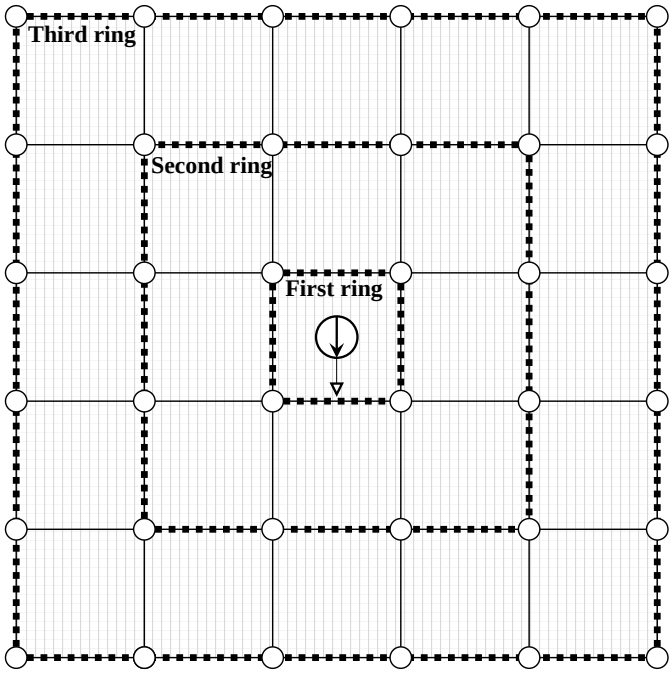


Figure 3.8: Portion of power supply network with three rings of power supplies and a single current load.

other iterative and random walk-based techniques [107]. The proposed technique is a promising approach for accurate, efficient analysis of power delivery systems.

3.3 Optimization of power delivery systems

To mitigate power and ground noise in high performance ICs, the concept of distributed power delivery and management should be adopted in next generation power delivery systems, providing local real-time conversion, regulation, and dynamic power control. Distributed power delivery requires the co-design of hundreds

of power converters at the board, package, and circuit levels with many thousands of decoupling capacitors, and billions of nonlinear current loads within multiple power domains, significantly increasing the design complexity of existing power delivery systems. Several techniques have recently been proposed for efficient power delivery, focusing on optimizing the physical area of the power network [114], [115], and placing on-chip decoupling capacitors [81], [116], [117], and on-chip power supplies [118]. The effect of the individual components (e.g., decoupling capacitors, on-chip voltage regulators, and power networks) on noise levels and physical area have been investigated in these works [81], [114]–[118], and the computationally efficient co-design of power networks with on-chip decoupling capacitors and power supplies has been proposed. A recent approach that considers the combined effects of on-chip regulators and decoupling capacitors, optimizing the power delivery process for a variety of design objectives, is reviewed below.

A facility location optimization algorithm [119]–[121] is exploited in [69] to determine the optimum location of the on-chip voltage regulators and decoupling capacitors, subject to different optimization goals (e.g., minimum voltage drop, physical area, response time, or power consumption). This algorithm provides a means to optimize a power delivery system at the circuit level. The optimization function

$F(n, m, k)$ for a power delivery system with n power supplies, k decoupling capacitors, and m load circuits is proposed in [69] that considers the parasitic impedance of the power network, output impedance of the power supplies, effective series resistance of the decoupling capacitors, load current characteristics, and response time requirements. A closed-form expression of the effective resistance R_{EFF} between two power grid nodes $node_1$ and $node_2$ located, respectively, at (x_1, y_1) and (x_2, y_2) is proposed in [107] and utilized in [69] to analyze individual power grid partitions under the assumption of spatial locality,

$$R_{EFF}(node_1, node_2) = \frac{r}{2\pi}\left[\ln\left((x_1 - x_2)^2 + (y_1 - y_2)^2\right) + 3.44388\right] - 0.033425, \quad (3.2)$$

where r is the unit resistance within the power grid. The optimum location of the power supplies and decoupling capacitors is determined from $F(n, m, k)$ based on specific design objectives, such as the minimum voltage drop, response time, on-chip area, and/or power consumption, exhibiting a flexible application-specific power grid analysis process. The proposed optimization of power delivery systems is demonstrated in [69] with ISPD'11 benchmark circuits [122], where the local voltage fluctuations within a system is minimized in feasible run time (e.g., less than three hours).

The efficiency of the overall power delivery system is a significant concern in high

performance integrated systems, and is primarily determined by the interactions among different types of power supplies. To address system-wide power efficiency, the hierarchical structure of the power delivery system and interactions among the power supplies at different hierarchical levels should be considered. A power efficient methodology to co-design different types of power supplies is discussed in this proposal in Chapters 4 and 5.

Distributed dynamic power management is an additional concern in DVS/DVFS real-time multicore systems, and is impractical with existing *ad hoc* approaches. A distributed architecture and systematic methodology for scalable heterogeneous on-chip power delivery are required to provide real-time, efficient control of the quality of power. A fine grained power management framework comprising a variety of circuits, algorithms, and architectural level techniques is described in this proposal to dynamically control the power delivery process, while optimally allocating power among different power domains at run time. The development of these novel capabilities will fundamentally change the manner in which power is delivered and managed in complex, high performance heterogeneous systems.

3.4 Summary

The generation and distribution of power by different types of power supplies and power networks are two primary issues in the power delivery process. To mitigate power and ground noise within power distribution networks, multiple active (such as power converters) and passive (such as decoupling capacitors) sources of charge are typically integrated at different levels of hierarchy within a power delivery system, increasing overall design complexity. The number of nodes in a typical power delivery system may exceed many millions (or billions) of nodes. To cope with this design complexity while maintaining high quality power in these complex power delivery systems, accurate and computationally efficient models, and effective analysis and optimization techniques are required. Existing modeling, analysis, and optimization techniques for power delivery systems with multiple power supplies and decoupling capacitors integrated at the board, package, and circuit levels are the subject of this chapter.

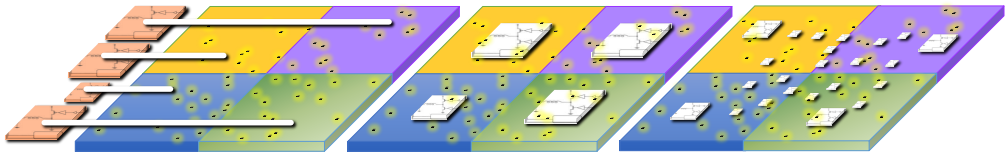
Chapter 4

Heterogeneous Power Delivery

To supply high quality power with minimum energy losses within multiple on-chip voltage domains, power conversion and regulation resources should be efficiently managed [41]. The design complexity of a power delivery system increases with greater requirements on the quality of the power supply, limitations of the passive elements, board and package parasitic impedances, and limited number of I/O pins. Furthermore, to effectively exploit the power delay tradeoff, additional power management techniques such as dynamic voltage scaling and dynamic voltage and frequency scaling are employed, further increasing the design complexity of the power delivery system. Thus, to efficiently manage the power delivered to a modern system-on-chip, a methodology to distribute and manage the power supplies is required.

Traditionally, power is managed off-chip with energy efficient power converters (see Figure 4.1(a)), delivering high quality DC voltage and current to the electrical

grid that reliably distributes the on-chip power. The supply voltage, current density,



(a) (b) (c)

Figure 4.1: Power delivery system with four voltage domains, (a) off-chip, (b) integrated on-chip, and (c) distributed point-of-load power supplies for voltage conversion and regulation.

and parasitic impedance, however, scale aggressively with each technology generation, degrading the quality of the power delivered from the off-chip power supplies to the on-chip load circuitry. The power supply in a package (PSiP) approach with partially off-chip yet in package power supplies has recently been considered as an intermediate power supply technology with respect to cost, complexity, and performance [42]. The power is regulated on-chip to lower the parasitic impedance of both the board and package (see Figure 4.1(b)). To fully integrate a power converter on-chip, advanced passive components, packaging technologies, and circuit topologies are essential. Recently, several power converters suitable for on-chip integration have been fabricated [66], [68], [71], [82]–[98]. Based on these power converters, a power supply system with several on-chip power converters can be developed to improve the quality of the power delivered within an IC.

On-chip power supply integration is an important cornerstone to the power supply

design process. A single on-chip power converter is however not capable of supplying sufficient, high quality regulated current to the billions of current loads within the tens of on-chip voltage domains. To maintain a high quality power supply despite increasing on-chip parasitic impedances, hundreds of ultra-small power converters should ultimately be integrated on-chip, close to the loads within the individual multiple voltage domains [71], [82]–[84]. A distributed point-of-load (POL) power supply system is illustrated in Figure 4.1(c).

While the quality of the power supply can be efficiently addressed with a distributed multi-voltage domain system, the limited power efficiency of the on-chip converters is a primary concern for the POL approach. The high power efficiency of the off-chip power converters is traded off for small area and locally regulated current and voltage. To address the concerns of a POL power supply system, existing power converter topologies are described and compared in Section 4.1. Heterogeneous power delivery is introduced in Section 4.2 to both decrease the noise and increase the efficiency of the overall supplied power. The chapter is summarized in Section 4.3.

4.1 Power converter topologies

Switching and linear DC-DC converters are commonly used topologies for DC-DC conversion and regulation. Historically, a large switching mode power supply (SMPS) is preferred over a compact linear power supply due to the high, ideally 100 %, power efficiency of an SMPS. With on-chip power converters, strict area constraints are imposed on the DC-DC converters, affecting the choice of power supply topology. Compact switching power converters can potentially be designed at higher switching frequencies. The parasitic impedance in these converters however increases, degrading the power efficiency of the power delivery system. The physical size and power efficiency of switching and linear topologies are discussed, respectively, in Sections 4.1.1 and 4.1.2. Some conclusions reviewing the preferable choice of on-chip power supply topology are provided in Section 4.1.3.

4.1.1 Switching converters

A typical switching mode power supply converts an input voltage V_{IN} to an output voltage V_{DD}, supplying the required current I_{DD} to the load circuitry. These converters are operated by a switching signal fed into passive energy storage components through a power MOSFET controlled by a pulse width modulator (PWM). A common step down SMPS converter operating as a buck converter is shown in Figure

4.2. The stored input energy is restored at the output at the required voltage level,

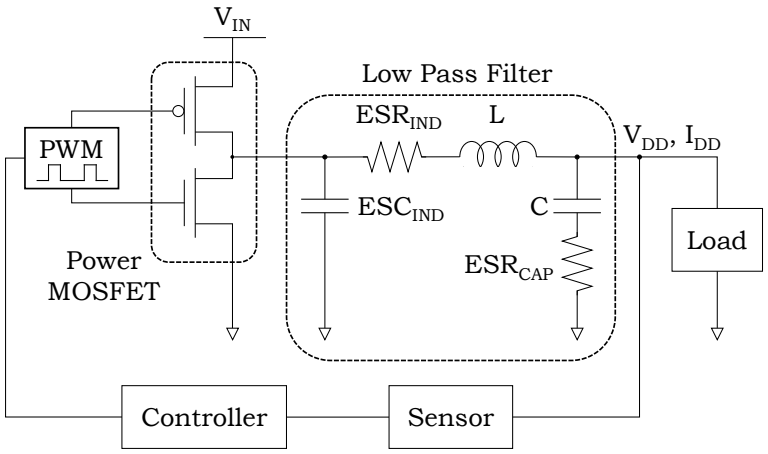


Figure 4.2: Buck converter circuit.

maintaining high power efficiency up to a frequency f_s of a few megahertz [56]. The operational mode of a buck converter, output voltage, output current, and transient performance are affected by the output LC filter and controller in the feedback loop, as illustrated in Figure 4.2. The on-chip integration of SMPS converters is greatly complicated due to I/O limitations, and constraints related to the physical size of the passive elements [48]. The area required by the passive components to achieve a specific impedance is inversely proportional to the frequency, and can be reduced in on-chip converters by operating at ultra-high switching frequencies. Conversely, an SMPS operating at a high frequency is more affected by the parasitic impedances, degrading the power efficiency of the converter. The area of a buck converter is

dominated by the size of the passive elements and is

$$A_{Buck} \approx \frac{L}{L_{\square}} + \frac{C}{C_{\square}}, \tag{4.1}$$

where $L_{\square}$ and $C_{\square}$ are, respectively, the inductance and capacitance per square micrometer of the LC filter. Voltage regulation is a primary concern for POL power delivery. In Discontinuous Conduction Mode (DCM) [123], the current ripple $\gamma_i I_{DD}$ within the inductor L exceeds the output current I_{DD}, and the voltage V_{DD} at the output of a converter becomes load dependent, degrading the quality of the delivered power. To support high load regulation, the buck converter is assumed to be loaded with an output current I_{DD} that exceeds the current ripple ($\gamma_i I_{DD} \leq I_{DD}$), yielding expressions for the inductor and capacitor operating in the Continuous Conduction Mode (CCM) [85], [124],

$$L = \frac{V_{IN} - V_{DD}}{2 f_s \gamma_i I_{DD}} \cdot \frac{V_{DD}}{V_{IN}}, \tag{4.2}$$

$$C = \frac{\gamma_i I_{DD}}{8 f_s \gamma_v V_{DD}}, \tag{4.3}$$

where $\gamma_v V_{DD}$ is the voltage ripple at the converter output, and V_{DD} is the voltage at the load. To satisfy the tight load regulation specifications, the output voltage ripple is assumed to range up to 10 % of V_{DD} ($\gamma_v = 0.1$). Substituting (4.2) and (4.3) into

(4.4), the area of a buck converter is

$$A_{Buck} \approx \left(\frac{(V_{IN}-V_{DD})V_{DD}}{2L_{\square}V_{IN}f_s}\right)\frac{1}{\gamma_i I_{DD}} + \left(\frac{1}{8C_{\square}\gamma_v V_{DD}f_s}\right)\gamma_i I_{DD}. \tag{4.4}$$

At low values of current ripple, the area of a buck converter is dominated by the inductor and increases with smaller values of $\gamma_i I_{DD}$. Alternatively, at larger values of $\gamma_i I_{DD}$, the area of a buck converter is dominated by the size of the capacitor and is proportional to the current ripple. An optimum ripple current $\gamma_{i,OPT} I_{DD}$ therefore exists that minimizes the area of a buck converter for a target output voltage ripple $\gamma_v V_{DD}$, and input and output voltage levels,

$$\gamma_{i,OPT} I_{DD} = \begin{cases} 2\sqrt{\gamma_v\left(1-\frac{V_{DD}}{V_{IN}}\right)\frac{C_{\square}}{L_{\square}}}\cdot V_{DD} \triangleq \gamma_G G \cdot V_{DD}, & \gamma_G G \cdot V_{DD} \le I_{DD} \quad (4.5) \\ I_{DD}, & \gamma_G G \cdot V_{DD} > I_{DD}, \quad (4.6) \end{cases}$$

where $\gamma_G G$ is the output conductance ripple and depends upon the technology parameters, converted voltages, and regulation specification. The minimum area of the buck converter is therefore

$$A_{Buck,MIN} \approx \frac{1}{2f_s}\begin{cases} \sqrt{\frac{1}{\gamma_v}\left(1-\frac{V_{DD}}{V_{IN}}\right)\frac{1}{L_{\square}C_{\square}}}, & \gamma_G G \cdot V_{DD} \le I_{DD} \quad (4.7) \\ \left[\left(1-\frac{V_{DD}}{V_{IN}}\right)\frac{1}{L_{\square}}\cdot\frac{V_{DD}}{I_{DD}} + \left(\frac{1}{4\gamma_v}\frac{1}{C_{\square}}\cdot\frac{V_{DD}}{I_{DD}}\right)\right] \approx \\ \left(1-\frac{V_{DD}}{V_{IN}}\right)\frac{V_{DD}}{L_{\square}I_{DD}}, & \gamma_G G \cdot V_{DD} > I_{DD}. \quad (4.8) \end{cases}$$

Thus, in CCM at low current loads ($I_{DD} < \gamma_G G \cdot V_{DD}$), the minimum area of a buck converter is dominated by the inductance characteristics and increases with smaller values of I_{DD}. However, for values of I_{DD} larger than $\gamma_G G \cdot V_{DD}$, the minimum size of a buck converter does not strongly depend on I_{DD}. Alternatively, both the power MOSFET losses and power dissipated in the LC filter are dominant at different frequencies, conversion voltages, and current levels in CCM. The power dissipated in the power MOSFET comprises the MOSFET switching power ($\propto f_s V_{IN}^2$), and the resistive power ($\propto R_{ON} I_{DD}^2$) dissipated by the effective resistor R_{ON} of the MOSFET, yielding

$$P_{Buck,MOS} = \frac{l_{min}^2}{\mu R_{ON}(V_{IN} - V_T)} \cdot f_s V_{IN}^2 + \frac{4}{3} R_{ON} \frac{V_{DD}}{V_{IN}} I_{DD}^2, \tag{4.9}$$

where l_{min} is the minimum channel length, μ is the MOSFET carrier mobility, and V_T is the threshold voltage [125]. From (4.9), increasing the effective resistance of the MOSFET reduces the switching power dissipation, while increasing the resistive loss. Thus, an optimum MOSFET resistance R_{ON}^{OPT} exists that minimizes the power dissipated in an MOSFET, yielding

$$R_{ON}^{OPT} = \sqrt{\frac{3}{4} \frac{l_{min}^2}{\mu(V_{IN} - V_T)} \cdot f_s \frac{V_{IN}}{V_{DD}}} \cdot \frac{V_{IN}}{I_{DD}}, \tag{4.10}$$

and

$$P_{Buck,MOS}^{MIN} = 2I_{DD}\sqrt{\frac{4}{3}\frac{l_{min}^2}{\mu(V_{IN}-V_T)}\cdot f_s V_{IN}V_{DD}}. \tag{4.11}$$

The power dissipated in an LC filter [125] comprises the power losses due to the resistive (ESR_{IND}) and capacitive (ESC_{IND}) parasitic impedances of the inductor,

$$P_{Buck,IND} = \frac{4}{3}ESR_{IND}\cdot I_{DD}^2 + ESC_{IND}f_s\cdot V_{IN}^2, \tag{4.12}$$

and the power losses due to the parasitic resistance of the capacitor (ESR_{CAP}),

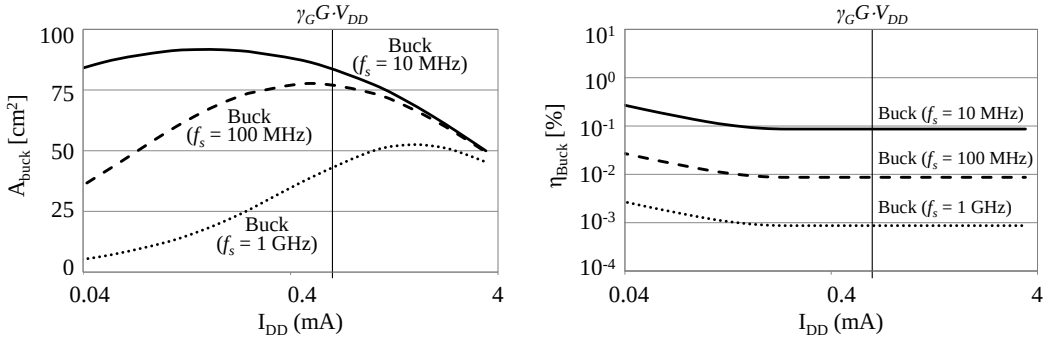
$$P_{Buck,CAP} = ESR_{CAP}(\gamma_i I_{DD})^2. \tag{4.13}$$

The total power dissipation and power efficiency $\frac{P_{Load}}{P_{Load}+P_{Buck}}$ of the buck converter are, respectively,

$$\begin{aligned} P_{Buck} = &\left(\frac{4}{3}ESR_{IND} + ESR_{CAP}\right)\cdot I_{DD}^2 + ESC_{IND}\cdot f_s\cdot V_{IN}^2 \\ &+ 2\sqrt{\frac{4}{3}\frac{l_{min}^2}{\mu(V_{IN}-V_T)}\cdot f_s V_{IN}V_{DD}}\cdot I_{DD}, \end{aligned} \tag{4.14}$$

$$\varphi_{Buck} = \frac{P_{Load}}{P_{Load}+P_{Buck}} = \frac{I_{DD}V_{DD}}{I_{DD}V_{DD}+P_{Buck}} \tag{4.15}$$

Typical passive component parameters, represented by [126]–[128] and technology parameters [25], are assumed to demonstrate power and area tradeoffs and trends in buck converters. Current load levels from a few milliamperes to several amperes, and input and output voltages of, respectively, 1 volt and 0.7 volts, are considered. The physical area [see (4.7)] and power efficiency [see (4.15)] trends are depicted in Figure 4.3 for moderate (10 MHz), high (100 MHz), and ultra-high (1 GHz) switching frequencies.



(a) (b)

Figure 4.3: Buck converter (a) physical area, and (b) power efficiency vs. load current for moderate, high, and ultra-high switching frequencies.

At low current loads, the power losses of a buck converter in CCM are dominated by the parasitic capacitance of the inductor (ESC_{IND}), decreasing the power efficiency at lower I_{DD} and larger converter size ($A_{Buck} \propto 1/I_{DD}$ for $I_{DD} < \gamma_G G \cdot V_{DD}$). Alternatively, at high current loads, the power efficiency is dominated by the parasitic

resistance of the inductor (ESR_{IND}) and capacitor (ESR_{CAP}), increasing the power losses of a buck converter at higher values of I_{DD}. Thus, a buck converter exhibits a parabolic shaped power efficiency with current in CCM, while the physical size of the converter is reduced at higher currents. Therefore, by targeting high switching frequencies, the preferred current load can be determined to convert a voltage with minimum power losses and area for a specific value of switching frequency f_s. For example, as shown in Figure 4.3, a preferable current exists for $f_s = 100\,\text{MHz}$ and $f_s = 1\,\text{GHz}$ since the maximum power efficiency is reached at $I_{DD} > \gamma_G G \cdot V_{DD}$, but not at $f_s = 10\,\text{MHz}$. The minimum power loss in (4.15) is proportional to $\sqrt{f_s}$, significantly degrading the power efficiency at high frequencies. Alternatively, the size of the power supply converter is proportional to $1/f_s$, and decreases at higher frequencies, exhibiting an undesirable tradeoff between the power efficiency and physical size of a buck converter. The high power efficiency of traditional large power converters operating at low frequencies is therefore traded off for smaller physical size at ultra-high switching frequencies.

4.1.2 Linear converters

To supply a specific voltage V_{DD} and current I_{DD} to the load circuitry, a linear power supply converts an input DC voltage V_{IN} using a resistive voltage divider controlled by feedback from the output. The primary drawback of a linear topology is the resistive power losses that increase with a larger $V_{IN} - V_{DD}$ voltage drop, which limit the power efficiency to V_{DD}/V_{IN}. Alternatively, linear converters exhibit a relatively small area, an important characteristic for on-chip integration. A low dropout (LDO) DC-DC regulator, depicted in Figure 4.4, is a standard linear converter that operates with a low $V_{IN} - V_{DD}$ voltage drop. The total current that flows through the linear converter is I_{IN}.

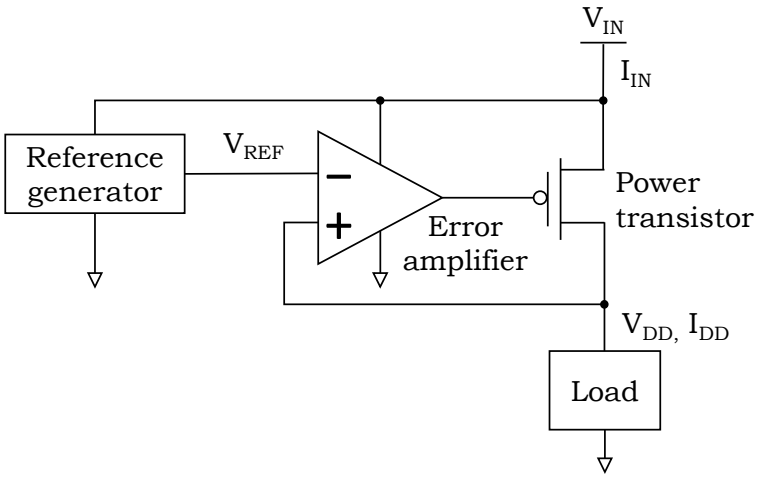


Figure 4.4: LDO circuit.

The total current supplied by a linear converter comprises the useful LDO current

I_{DD} that flows to the load, and the short-circuit $I_{IN} - I_{DD}$ current dissipated in the bandgap voltage reference and error amplifier. Power and area efficient voltage references have recently been reported [66], [68], [87], [88]. The total LDO current is, therefore, dominated by the error amplifier and power transistor currents. To mitigate transient voltage peaks while supporting fast changes in the load current, larger currents should be utilized within the error amplifier, increasing the short-circuit current. Alternatively, to satisfy the current load requirements in modern high performance circuits, high currents of up to several amperes are required by the load circuitry. The current flow within an LDO is therefore dominated by the load current I_{DD}. In this case, both the area and power dissipation of a linear converter are primarily dictated by the size of the output power transistor and the dissipated power. Thus, the area of an LDO is proportional to the width W of the output transistor, yielding

$$A_{Linear} \propto W l_{min} \propto \frac{I_{DD} \cdot l_{min}^2}{\mu C_{OX}(V_{IN} - V_{AMP} - V_T)^2}, \tag{4.16}$$

where l_{min} is the minimum channel length, μ is the MOSFET carrier mobility, C_{OX} is the gate oxide capacitance, and V_{AMP} is the output from the error amplifier. To accommodate the effect of the line and load specifications that may significantly affect the physical size of an LDO, a typical area per 1 mA load [66], [68], [71], [82]–[88]

(see Figure 4.5) is considered for those LDOs with a high current load, exhibiting a parabolic trend of area with minimum technology length ($A_{Linear}/I_{DD} \propto l_{min}^2$). The

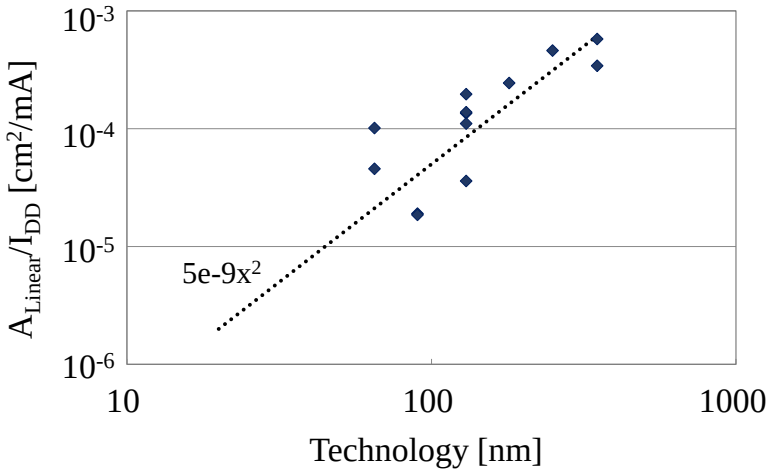


Figure 4.5: LDO physical area per 1 mA load.

ratio $A_{Linear}/I_{DD} = 5 \times 10^{-6}\,\text{mm}^2/\text{mA}$ corresponds to the 28 nm technology node considered in Figure 4.6. Typical 28 nm CMOS technology parameters [25], and input and load voltages are assumed in this analysis to demonstrate the need for a large power transistor to supply high current to the load (see Figure 4.6). The size of the linear converter ranges from $60 \times 60\,\mu\text{m}^2$ for $I_{DD} = 0.5$ amperes to $150 \times 150\,\mu\text{m}^2$ for $I_{DD} = 3.5$ amperes (see Figure 4.6), which can be further reduced with technology scaling ($A_{Linear} \propto l_{min}^2$) and advanced design solutions [66], [68], [71], [82]–[98]. The current can therefore be supplied to the load with an LDO that is orders of magnitude

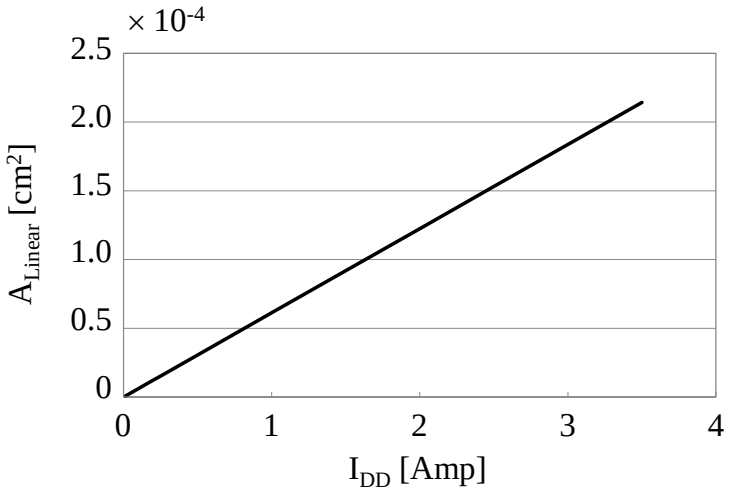
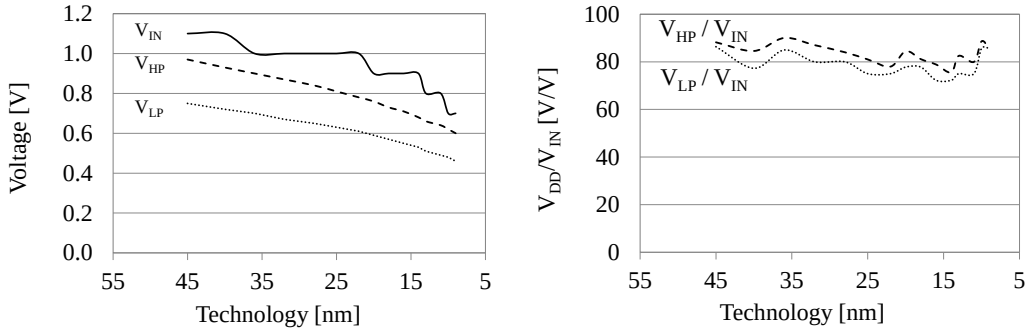


Figure 4.6: LDO area for typical current loads.

smaller than a corresponding buck converter. The power dissipation of an LDO is

$$P_{Linear} \approx (V_{IN} - V_{DD})\, I_{DD}. \tag{4.17}$$

Thus, the power loss in a linear converter increases with a higher $V_{IN} - V_{DD}$ drop, degrading the power efficiency of the converter. Recent supply voltage trends are illustrated in Figure 4.7 for the internal core primary voltage V_{IN}, and typical high and low V_{DD} levels [25], yielding efficiency bounds within the 70 % to 90 % range of the V_{DD}/V_{IN} ratio shown in Figure 4.7. Thus, a moderate LDO power efficiency $\varphi_{Linear} = V_{DD}/V_{IN}$ of at least 70 % can be predicted. Sub/near threshold computing is a promising technique to reduce the power consumed by an IC [129], [130]. To



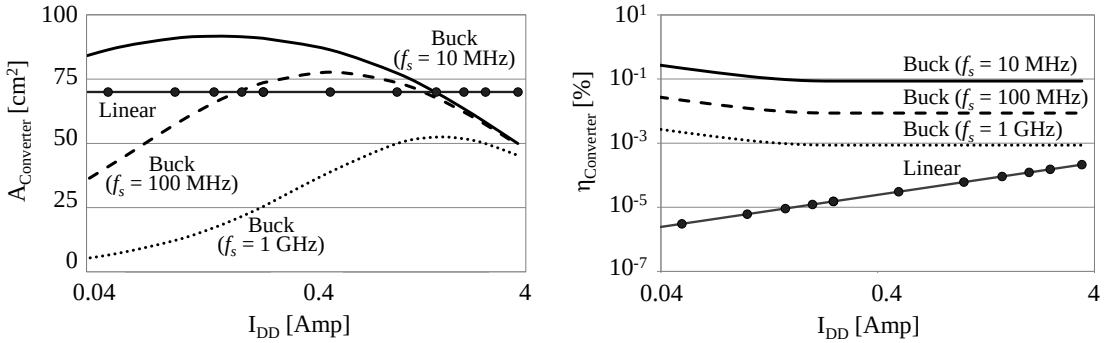
(a) (b)

Figure 4.7: Trends in typical (a) high performance (V_{HP}), low power (V_{LP}), and internal core primary (V_{IN}) voltage supplies, and (b) voltage conversion ratios (V_{HP}/V_{IN}) and (V_{LP}/V_{IN}).

provide a stable supply voltage at sub/near threshold levels, tunable low noise voltage regulation below 0.5 volts is required. To provide these output voltages with sufficient power efficiency, analog LDO regulator should operate with low input voltages (e.g., V_{IN} ¡ 0.7 volts to provide V_{DD} of 0.5 volts with at least 70 % power efficiency). A conventional analog LDO, however, fails to operate at these low input voltages due to the insufficient voltage drop across the individual transistors within the current mirror of the LDO. A digital LDO can be used to suppress the analog nature of a conventional LDO [131], [132].

4.1.3 Comparison of power supply topologies

The physical area and power efficiency of an LDO and buck converter is shown in Figure 4.8. Buck converters that are more power efficient than an alternative



(a) (b)

Figure 4.8: LDO and buck converter (a) physical area, and (b) power efficiency for moderate, high, and ultra-high switching frequencies.

LDO can operate at lower switching frequencies. These buck converters are, however, inappropriate for on-chip power conversion due to the large physical size and technology constraints of the passive elements that make on-chip integration even more difficult. Alternatively, compact buck converters can operate at high switching frequencies. These buck converters, however, exhibit a lower power efficiency and are therefore less effective for on-chip integration. Thus, to deliver high quality power to the load circuitry under typical area constraints, on-chip linear regulators should

be considered. The moderate power efficiency of an LDO becomes a significant constraint when the power consumption at the load increases. For example, converting 2 volts into 1 volt while delivering 1 µAmp to the current load results in a 50 % power efficiency and $1\,\mathrm{V} \cdot 1\,\mu\mathrm{A} = 1\,\mu$Watt power loss that can possibly be absorbed by the power delivery system. Alternatively, converting 1.25 volts into 1 volt while delivering 1 mA to the current load results in 80 % power efficiency and a significant $0.25\,\mathrm{V} \cdot 1\,\mathrm{mA} = 250\,\mu$Watt power loss that is difficult to mitigate. Thus, linear regulators are preferable to switching power supplies, mainly for small input-output voltage differences. A heterogeneous power delivery system that efficiently exploits the power and area characteristics of linear and switching converters is desirable to enhance the power supply quality and efficiency while satisfying on-chip area constraints.

4.2 Heterogeneous power delivery system

Both linear and switching power regulators are characterized by an undesirable power-area tradeoff, exhibiting either high power in compact linear regulators or large area in power efficient SMPS, as depicted in Figure 4.9. Thus, the overhead of a power delivery system composed of only switching or linear regulators is significant. Several power delivery solutions exist that exhibit intermediate power losses and area as compared to either linear or traditional SMPS systems. For example, in

a PSiP system, lower power losses as compared to a linear system, and smaller area as compared to a traditional off-chip SMPS system, are traded off for greater design complexity. A desirable power delivery system minimizes power losses while satisfying on-chip area constraints, yielding both high power efficiency and small area, as depicted in Figure 4.9.

To exploit the advantages of switching and linear converters, a heterogeneous power delivery system is considered that converts the power in off-chip switching power supplies and regulates the on-chip power with compact linear power supplies, minimizing LDO voltage drops and on-chip power losses. In a heterogeneous power delivery system, the area overhead is primarily constrained by the compact LDOs that regulate the on-chip power, while the power overhead is dictated by the power efficient switching converters. Power conversion is therefore decoupled from power

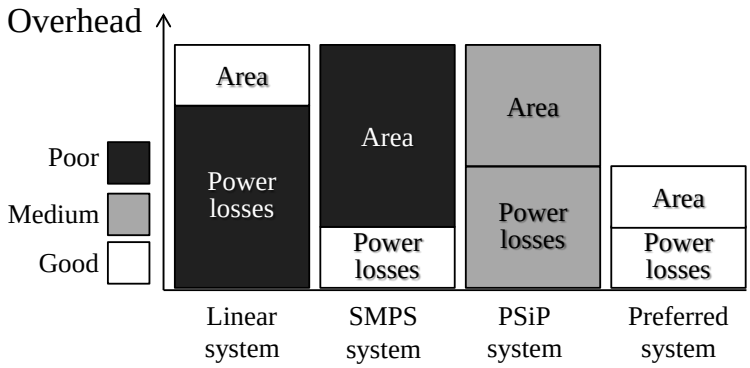


Figure 4.9: Power and area overhead of a linear, SMPS, PSiP, and preferred power conversion system.

regulation, lowering the power and area overhead of the overall power delivery system. A heterogeneous power delivery system moderates the drawbacks and exploits the advantages of the historically power efficient power supplies that both convert and regulate the power off-chip with more recent trends for area efficient distributed power supplies that both convert and regulate the power on-chip. Off-chip, on-chip distributed, and heterogeneous power delivery topologies are illustrated in Figure 4.10.

Consider a heterogeneous power delivery system with L on-chip LDOs and S off-chip SMPSs that deliver power to N voltage domains $\{(V_{DD}^{(i)}, I_{DD}^{(i)})\}_{i=1}^{N}$ with an operating voltage $V_{DD}^{(i)}$ and current $I_{DD}^{(i)}$. To supply the required voltages, $V_{DD}^{(i)} \neq V_{DD}^{(j)}$ $\forall i \neq j$, the number of on-chip power supplies L should be equal to or greater than the number of voltage domains $N \leq L$. Alternatively, each SMPS drives one or more LDOs, yielding the relation, $S \leq L$. The effect of the number of on-chip power

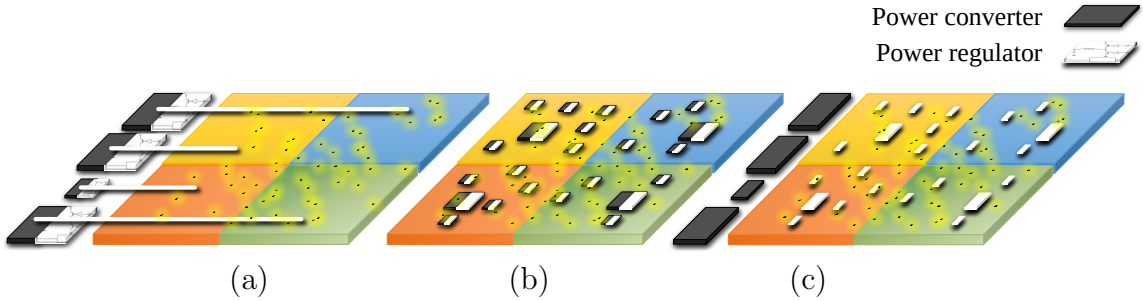


Figure 4.10: Power delivery system with four voltage domains, utilizing (a) off-chip power supplies, (b) distributed POL power supplies, and (c) a heterogeneous system with off-chip converters and on-chip regulators.

regulators, off-chip power converters, and distribution of the on-chip power supplies in a heterogeneous power delivery system is described, respectively, in Sections 4.2.1, 4.2.2, and 4.2.3.

4.2.1 Number of on-chip power regulators

The area of an LDO is proportional to the current load [see (4.16)], and the power efficiency is primarily dictated by the current load and voltage drop V_{Drop} across the power transistor within the LDO [see (4.17)]. Thus, a single LDO that provides a specific current and voltage to a load consumes approximately the same area and dissipates similar power as numerous LDOs providing the same total current and voltage to a load. Consider l_k on-chip distributed LDOs to maintain a regulated voltage V_{DD} and load current I_{DD} within a specific voltage domain (V_{DD}, I_{DD}). Let I_i $(i = 1, \ldots, l_k)$ be a local current load supplied by a single LDO within the domain, such that $\sum I_i = I_{DD}$. The LDO area A_i is assumed here to be linearly proportional to the supply current I_i within a specific current range [see (4.16)], $A_i = \alpha \cdot I_i$. The l_k LDOs form a distributed on-chip power regulation system with a total size, $A = \sum A_i = \alpha \cdot \sum I_i = \alpha \cdot I_{DD}$. Thus, the total area of the distributed regulation system does not strongly depend on l_k, the number of LDOs. To maximize the power efficiency of a system, all of the LDOs operate at the minimum voltage drop V_{Drop},

exhibiting a total power loss $V_{Drop} \cdot \sum I_i = V_{Drop} \cdot I_{DD}$ which is independent of l_k. Alternatively, the distance between an LDO and a current load is reduced at higher values of l_k, decreasing the on-chip voltage drops and increasing the quality of the supplied power.

4.2.2 Number of off-chip power converters

Intuitively, the number of off-chip voltage levels increases with the larger number of off-chip converters, increasing the granularity of the voltage levels supplied to the on-chip regulators and lowering the voltage drop across the hundreds of ultra-small regulators distributed on-chip. To minimize the voltage drop across an on-chip linear regulator, each off-chip SMPS converter should drive a single on-chip LDO. In practice, however, the number of power converters that can be placed off-chip is limited. Thus, each off-chip SMPS supplies power to several on-chip LDOs within an SMPS cluster. As a result, the voltage drop across the on-chip regulators is greater, degrading the overall power efficiency of the system. The upper and lower bounds of the power efficiency of a heterogeneous system for a specific number of SMPS are described in this section.

Given N power domains $\{(V_{DD}^{(i)}, I_{DD}^{(i)})\}_{i=1}^{N}$ sorted by the supply voltages $V_{DD}^{(i)} < V_{DD}^{(j)}$ $\forall i < j$, and l_k power supplies in the k^{th} power domain, $L = \sum_{k=1}^{k=N} l_k$ linear power

supplies should be distributed on-chip to deliver high quality power to the load circuitry. To explore the area-power efficiency tradeoff in a heterogeneous power delivery system, a single linear regulator is assumed capable of providing sufficient high quality current within a power domain, yielding $l_k = 1\ \forall k$ and $L = N$. The voltage supplied by an LDO to a power domain cannot be stepped up by an LDO. The output voltage of each SMPS is therefore higher than the voltage within the individual power domains, increasing the voltage drop across the LDOs within an SMPS cluster, degrading power efficiency.

An expression for determining the optimal LDO clustering within the SMPS clusters is presented below. Consider a system with S off-chip or in-package SMPS converters and L on-chip LDOs, delivering power to L power domains with N different supply voltages $\{(V_{DD}^{(i)}, I_{DD}^{(i)})\}_{i=1}^{N}$. Intuitively, LDOs that regulate power domains with similar supply voltages should be assigned to the same voltage cluster. Thus, to explore the power efficiency of a heterogeneous power delivery system, $L = N \geq S$ is assumed. The i^{th} SMPS supplies power to l_i LDOs ($\Sigma l_i = L = N$), forming the i^{th} voltage cluster. The proposed heterogeneous system is illustrated in Figure 4.11 with off-chip converters. Note that the SMPS, LDO, and supply voltages are assumed to

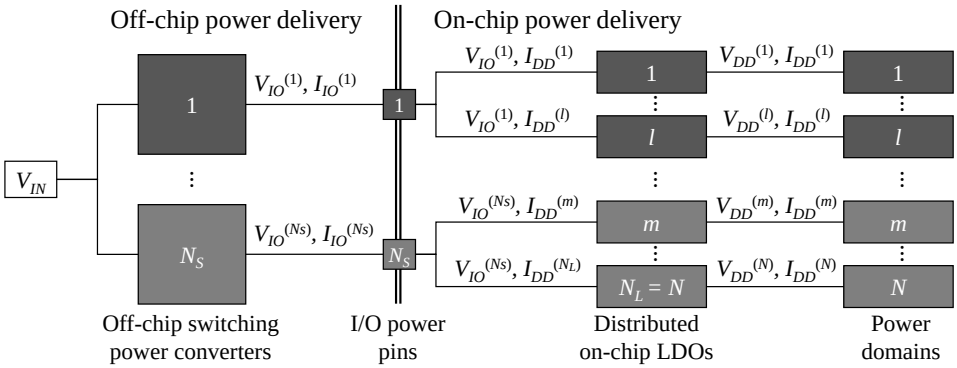


Figure 4.11: Heterogeneous power delivery system with S off-chip switching converters, $L = N = \Sigma l_i$ on-chip linear regulators, and N on-chip power domains.

be ordered such that

$$V_{SMPS}^{(i)} < V_{SMPS}^{(j)} \text{ if } \{i < j\}, \tag{4.18}$$

$$V_{LDO}^{(i,m)} < V_{LDO}^{(j,n)} \quad \text{if } \{i < j\} \text{ or } \{i = j, m < n\}, \tag{4.19}$$

$$V_{DD}^{(i)} < V_{DD}^{(j)} \qquad \text{if } \{i < j\}, \tag{4.20}$$

where $V_{SMPS}^{(i)}$ is the output voltage of the SMPS in the i^{th} cluster, $V_{LDO}^{(i,m)}$ is the output voltage of the m^{th} LDO in the i^{th} cluster, and $V_{DD}^{(j)}$ is the voltage supplied to the j^{th} power domain.

To increase the power efficiency of a heterogeneous power delivery system, the voltage drops across the distributed on-chip LDOs should be reduced. The granularity of the converted voltage levels supplied on-chip increases with additional

off-chip SMPS converters, reducing the power losses within the on-chip LDOs. At the limit, $S = L = N$ switching power converters are placed off-chip, providing voltages $\{V_{SMPS}^{(i)}\}_{i=1}^{N}$ at the I/O power pins. In the configuration with $S = L = N$, the on-chip LDOs operate with a minimum output voltage drop V_{Drop}, yielding

$$V_{SMPS}^{(i)} = V_{DD}^{(i)} + V_{Drop}, \qquad i = 1, ..., N, \tag{4.21}$$

where V_{Drop} is the voltage dropout of the output transistor within the LDO. Assuming ideal power efficiency of the off-chip SMPS, the power efficiency of a system with the maximum number of SMPS converters ($S = L$) is

$$\varphi_{S=L=N} = \frac{P_{Load}}{P_{IN}} = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{S} V_{SMPS}^{(i)} I_{SMPS}^{(i)}} = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{N} \left(V_{DD}^{(i)} + V_{Drop} \right) I_{DD}^{(i)}}. \tag{4.22}$$

In this case, the power efficiency is only limited by the dropout voltage of the transistor, and exhibits a high power efficiency for low V_{Drop} devices.

Area and I/O power pin constraints exist, however, that limit the number of off-chip power supplies, degrading the overall power efficiency. Let S be the maximum number of off-chip switching power converters in a heterogeneous power delivery system. To minimize the voltage drop across the on-chip LDOs for the worst case

power efficiency scenario where $S = 1$, the off-chip SMPS produces a voltage $V_{SMPS}^{(1)}$ that is higher than the maximum domain voltage by one dropout voltage V_{Drop},

$$V_{SMPS}^{(1)} = \max_{1 \leq i \leq N} \left\{ V_{DD}^{(i)} \right\} + V_{Drop}, \tag{4.23}$$

exhibiting a power efficiency,

$$\varphi_{S=1,L=N} = \frac{P_{Load}}{P_{IN}} = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{V_{SMPS}^{(1)} I_{SMPS}^{(1)}} = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\max_{1 \leq i \leq N} \left\{ V_{DD}^{(i)} + V_{Drop} \right\} \sum_{i=1}^{N} I_{DD}^{(i)}}. \tag{4.24}$$

In a system with a single off-chip SMPS, the power loss within each domain, in addition to the V_{Drop} drop, is determined by the difference between the domain voltage and maximum voltage in the system. Those power domains with lower voltages exhibit greater power losses, significantly degrading the power efficiency of a heterogeneous system. The upper and lower bounds of the power efficiency of a heterogeneous system for a specific number of switching power converters are given, respectively, by (4.24) and (4.22), yielding

$$\frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\max_{1 \leq i \leq N} \left\{ V_{DD}^{(i)} + V_{Drop} \right\} \sum_{i=1}^{N} I_{DD}^{(i)}} \leq \varphi \leq \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{N} \left(V_{DD}^{(i)} + V_{Drop} \right) I_{DD}^{(i)}}. \tag{4.25}$$

Thus, the power efficiency of a heterogeneous system is a strong function of the number of off-chip power converters.

4.2.3 Power supply clusters

In a practical heterogeneous power delivery system, the number of SMPS converters is smaller than the number of on-chip LDO regulators ($S < L$). Several options therefore exist to include the on-chip LDOs within SMPS clusters, affecting the power efficiency of the overall power delivery system. To illustrate the effect of the clustering topology on the power efficiency of a power delivery system, a heterogeneous system is considered with two switching converters and three linear regulators, supplying equal current I_{DD} to three power domains $\{V_{DD}^{(i)}\} = \{1.8\,\text{volts}, 1.1\,\text{volts}, 1.0\,\text{volt}\}$. Assume $V_{Drop} = 0.1\,\text{volts}$. The power supply clusterings $K_1 = \{1, 2\}$ and $K_2 = \{2, 1\}$ for the heterogeneous system with $S = 2$ and $L = N = 3$ are shown in Figure 4.12. The voltage at the output of the switching converters is determined from (4.27), yielding a power efficiency, $\varphi(K_1) = P_{Load}/P_{IN} = 91\,\%$ and $\varphi(K_2) = P_{Load}/P_{IN} = 80\,\%$. Determining the optimal clustering of the on-chip power supplies is a primary challenge in a heterogeneous power efficient system.

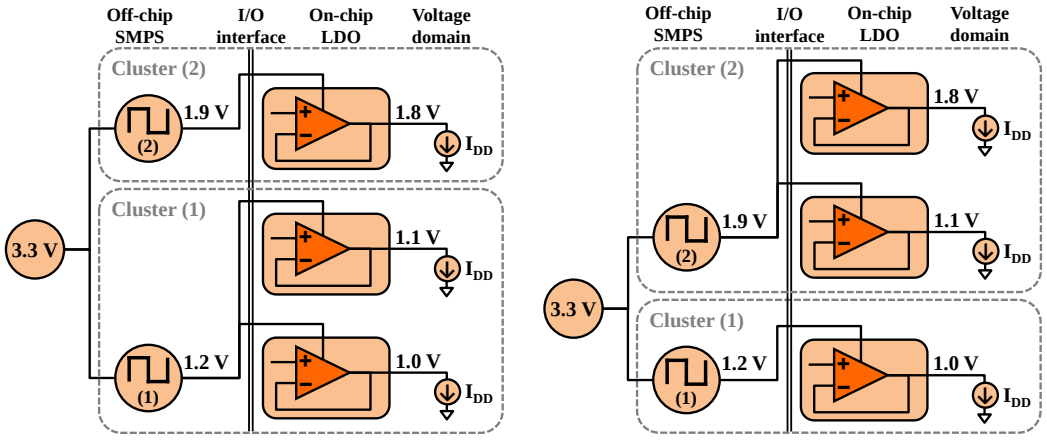


Figure 4.12: Power supply clusterings for a heterogeneous power delivery system with $S = 2$ and $L = N = 3$, (a) $K_1 = \{1, 2\}$, and (b) $K_2 = \{2, 1\}$.

The power efficiency of a general heterogeneous power delivery system, as illustrated in Figure 4.11, is

$$\varphi = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{S} \left(V_{SMPS}^{(i)} \cdot \sum_{m=1}^{l_i} I_{LDO}^{(i,m)} \right)}. \tag{4.26}$$

The voltage supplied by an LDO to a power domain cannot be stepped up. The output voltage of each SMPS is therefore higher than the output voltage of the LDOs within an SMPS cluster, yielding $V_{SMPS}^{(i)} \geq V_{LDO}^{(i,m)},\ \forall 1 \leq m \leq l_i$. Alternatively, the power efficiency of a power delivery system increases with smaller LDO dropout voltages ($V_{SMPS}^{(i)} - V_{LDO}^{(i,m)},\ \forall 1 \leq m \leq l_i$). Thus, to minimize the power loss within

the i^{th} SMPS cluster, the output voltage of an SMPS is

$$V_{SMPS}^{(i)} = \max_{1 \leq m \leq l_i} V_{LDO}^{(i,m)} + V_{Drop}, \tag{4.27}$$

The preferred SMPS output voltage in (4.27) with power efficiency φ described by (4.26) yields the optimum power efficiency for a specific choice of clusters,

$$\varphi = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{S} \left(\max_{1 \leq m \leq l_i} V_{LDO}^{(i,m)} + V_{Drop} \right) \sum_{m=1}^{l_i} I_{LDO}^{(i,m)}}. \tag{4.28}$$

For each SMPS converter, the voltage drop across the driven LDOs increases with a wider range of voltages included within that SMPS cluster, increasing the overall power dissipation. Intuitively, for any power supply clustering, adding a power domain with a specific voltage within a SMPS cluster that includes a similar voltage range results in a lower voltage drop and power loss than including the same power domain in a SMPS cluster with a significantly different range of voltages. Thus, the choice of power clustering directly affects the efficiency of the power delivery system. To minimize power losses in a heterogeneous power delivery system, a power distribution network with a higher φ is preferred. The optimal solution with minimum power losses can be obtained by comparing the power efficiency φ [see (5.2)] for all

possible clusters $\{K_i\}$, and choosing the configuration with the maximum efficiency φ^{OPT},

$$\varphi^{OPT} = \max_{\{K_i\}} \{\varphi\} = \max_{\{K_i\}} \left\{ \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{S} \left(\max_{1 \le m \le l_i} V_{LDO}^{(i,m)} + V_{Drop} \right) \sum_{m=1}^{l_i} I_{LDO}^{(i,m)}} \right\}. \tag{4.29}$$

The power efficiency of a heterogeneous system is also a strong function of the current distribution, which is not necessarily equally distributed to the individual power domains. Optimizing the power efficiency of a heterogeneous system based on the current distribution within the power domains requires additional assumptions regarding the behavior and specifications of the currents. The purpose here is to provide a framework for an efficient system-wide power delivery methodology and specific rules for delivering power.

4.3 Summary

On-chip power integration is necessary for delivering high quality power to modern high performance circuits. With on-chip power supplies, new design challenges have arisen that require advanced circuit design solutions. A power delivery topology is therefore required that minimizes power conversion and regulation losses while

satisfying specific design constraints. The tradeoff between power efficiency and area for switching and linear power supplies is discussed in this chapter. To convert power with minimum power losses while avoiding area consuming on-chip passive components, power efficient SMPS should be placed off-chip. In addition, area efficient LDOs should be employed on-chip to regulate and deliver the converted power to the load circuitry, reducing power losses from the low voltages dropped across the LDOs. Thus, to maintain high quality on-chip power delivery, the power conversion and regulation operations should be separated. Based on this separation principle of power conversion and regulation, a system-wide efficient heterogeneous power delivery system is proposed.

Chapter 5

Energy Efficient Adaptive Clustering of Power Supplies in Heterogeneous Dynamically Controlled Systems

The delivery of high quality power to the on-chip circuitry with minimum energy loss is a fundamental requirement of all integrated circuits. A principle of separation of power conversion and regulation has recently been introduced [70] that addresses the issue of power efficiency in distributed power supply systems. To optimize the power efficiency of the overall system, power should be primarily converted with a few power efficient switching supplies, delivered to on-chip voltage clusters, and regulated with linear low dropout regulators within the individual power domains. The separation principle with multiple voltage clusters is illustrated in Figure 5.1

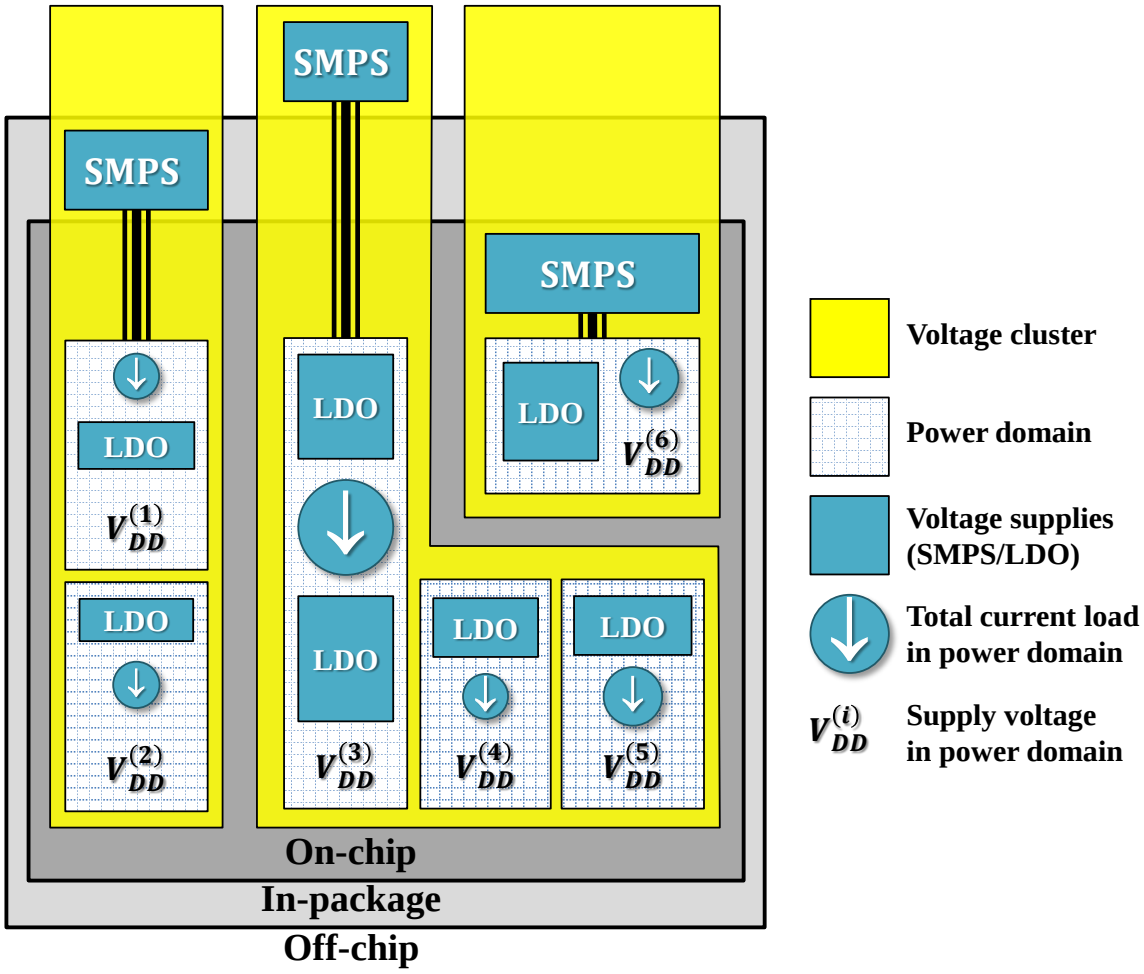


Figure 5.1: Heterogeneous power delivery with multiple power domains.

by a heterogeneous power delivery system with multiple power domains, off-chip/in-package/on-chip SMPS power converters, and on-chip LDO power regulators.

Several schemes for heterogeneous power delivery [69], [72], [82], [118], [133] that consider tens to hundreds of on-chip power regulators have recently been proposed.

Optimizing the power delivery process in terms of the co-design of the on-chip voltage regulators, decoupling capacitors, and current loads have been proposed in [69], [72], [82], [118], [133]. The co-design of hundreds to thousands of on-chip regulators with multiple switching converters is a new design objective. In energy efficient systems, the voltage and current are dynamically scaled within the individual power domains, affecting the voltage drop across the LDO regulators and the overall efficiency of the power delivery system. Optimal real-time clustering of the power supplies decreases the voltage drop within the LDO regulators, increasing overall power efficiency. Exhaustive approaches for clustering power supplies are computationally impractical in DVS systems with hundreds to thousands of power domains. Other existing approaches for on-chip power delivery are *ad hoc* in nature and not optimal. A computationally efficient methodology to co-design in run time switching converters and on-chip LDO regulators within a heterogeneous system is proposed in this chapter, achieving high quality power and efficiency within limited on-chip area. The power savings with the proposed approach are evaluated with IBM power grid benchmarks, demonstrating up to 24% increase in power efficiency with the proposed voltage clusters [134], [135]. Significant speedup is exhibited with the proposed recursive clustering algorithm, exhibiting polynomial computational complexity.

The rest of the chapter is organized as follows. The effects of the number of power

supplies, on-chip power supply voltages, and clustering topology on the power efficiency of a dynamically controlled heterogeneous power delivery systems is described in Section 5.1. Computationally efficient algorithms for optimal and near-optimal clusters of power supplies are demonstrated in Section 5.2. Separation of power conversion and regulations in benchmark circuits is evaluated in Section 9.2. The chapter is concluded in Section 5.4.

5.1 Dynamic control in heterogeneous power delivery systems

DVS is a primary objective for efficiently managing the power budget in hundred- and thousand-core ICs, further increasing the design complexity of the power delivery system. As the voltage supplied by an LDO to a power domain changes, the voltage dropout within the LDO varies. As a result, the power saved during low power operation is dissipated within the regulators. Varying the load currents affects the efficiency of the power supplies in a similar way. Thus, in a system with fixed power supply clusters, the energy efficiency of the power delivery system is not optimal. To avoid excessive dissipation of power, SMPS clusters should be dynamically reconfigured in every time slot Δt based on the temporarily required voltage and current

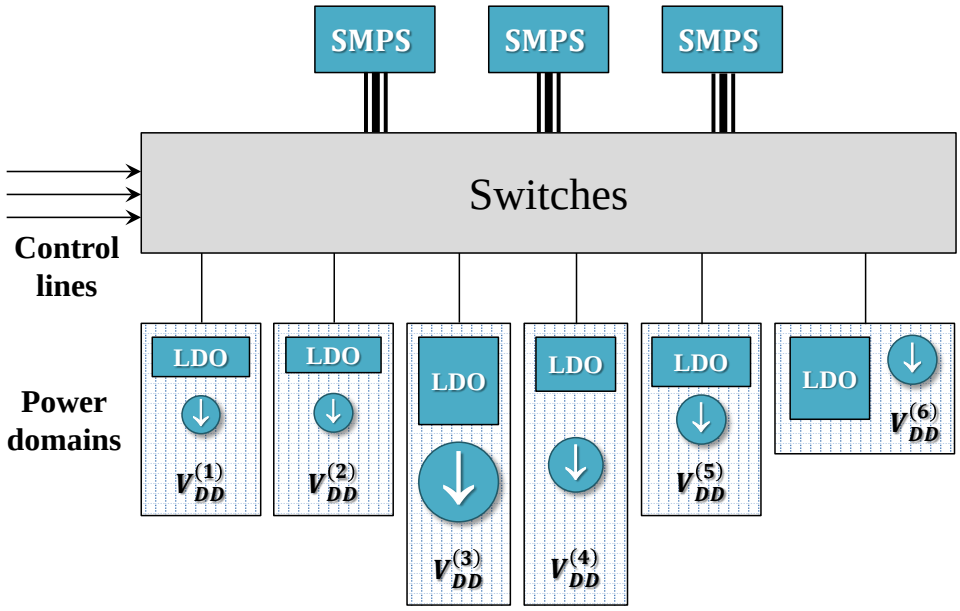


Figure 5.2: Heterogeneous power delivery with multiple dynamically controllable power domains.

levels within the individual power domains. A heterogeneous system for real-time power management in modern high performance integrated circuits is illustrated in Figure 5.2.

The optimal power supply clusters need to be determined during each control time slot, decreasing the voltage dropout within the regulators and increasing the overall energy efficiency within shorter time slots. Alternatively, the duration of the control time slot Δt is inversely proportional to the power dissipated by a MOSFET

switch in the i^{th} power domain,

$$P_{SW}^{(i)} = f_{sw} \cdot t_{ave} \cdot V_{Off} \cdot I_{DD}^{(i)}, \tag{5.1}$$

where $f_{sw} = 1/\Delta t$ is the system switching frequency, t_{ave} is the MOSFET average on/off time, V_{Off} is the switch off voltage, and $I_{LDO}^{(i,m)} + I_{LDO,Q}^{(i,m)} \approx I_{LDO}^{(i,m)}$. Power losses of the proposed system, therefore, increase with higher switching frequencies, and are considered here to determine the preferred duration of the control time slot $\Delta t = 1/f_{sw}$. In a power efficient system, power switching losses should not exceed the power savings due to dynamic control over the power delivery process. The optimal power efficiency in (4.26) with switching power losses is

$$\varphi = \frac{\sum_{i=1}^{N} V_{DD}^{(i)} I_{DD}^{(i)}}{\sum_{i=1}^{S} \left(V_{SMPS}^{(i)} + f_{sw} \cdot t_{ave} \cdot V_{Off} \right) \sum_{m=1}^{l_i} I_{LDO}^{(i,m)}}. \tag{5.2}$$

Both analog and digitally controlled LDO regulators with voltage drops as low as 0.15 volts down to 0.05 volts have recently been reported [67], [136]–[138], yielding $V_{Drop} \approx V_{Off}$. Thus, for a sufficiently long control time slot $\Delta t = 1/f_{sw} >> t_{ave}$, the power dissipated within the switches is significantly lower than the power dissipated within the LDO and can therefore be neglected. Modern MOSFET switches are

capable of switching within tens of nanoseconds [139], exhibiting a practical target for the $\Delta t >> t_{ave}$ requirement in dynamically controlled heterogeneous systems. Alternatively, a real-time power delivery management system poses a significant computational challenge. Thus, a computationally efficient method to co-design the on-chip power supplies in modern high performance circuits is required.

5.2 Computationally efficient power supply clustering

The optimal clustering topology with minimum power losses can be obtained by exhaustively comparing the power efficiency φ (see (5.2)) for all possible clusterings, and choosing the configuration with the maximum efficiency. The number of possible clusterings, however, grows exponentially with S, producing a computationally infeasible solution. To efficiently determine the preferable power supply clusters, an alternative computationally efficient solution is required. Binary and linear near-optimal power supply clusterings with, respectively, $\mathcal{O}(S)$ and $\mathcal{O}(N)$ are described in Section 5.2.1. An optimal power supply clustering with dynamic programming with $\mathcal{O}(N^2 \cdot S)$ is described in Section 5.2.2.

5.2.1 Near-optimal power supply clustering

Intuitively, to reduce the voltage drop across the on-chip LDOs, LDOs that regulate the voltage domains with a small difference in voltage levels should be assembled into a voltage cluster driven by the same SMPS, minimizing the voltage range within each cluster. A binary power supply clustering, based on a greedy algorithm, identifies in each step the voltage cluster with the widest voltage range and distributes the LDOs into two separate clusters. Pseudo-code of the algorithm is provided in Algorithm 5.1. The algorithm produces a set of S SMPS voltage clusters $List\_of\_Clusters$ with a binary clustering of power supplies. The third step is executed S times, yielding an algorithm that exhibits linear complexity $\mathcal{O}(S)$ with the number of switching converters. The primary weakness of the binary power supply clustering is the greedy nature of the algorithm. The number of voltage clusters S is only considered when the algorithm is terminated, reducing the power efficiency of the overall power delivery system. Consider a heterogeneous power delivery system with three switching converters and four LDO regulators that supply power to four voltage domains. The voltage and current levels within the voltage domains are (1 volt, 1 ampere), (1.49 volts, 1 ampere), (1.51 volts, 1 ampere), and (2 volts, 1 ampere). The optimal and binary power supply clusterings, SMPS output voltages, and power efficiency are listed in Table 5.1, exhibiting, respectively, 93% and 87% power efficiency for

Algorithm 5.1 Algorithm for binary power supply clustering.

1: **procedure** BINARY_CLUSTERING($\left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}$)
2: $List\_of\_Clusters \leftarrow \{\}$;
3: $Next\_Cluster \leftarrow \left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}$;
4: $(Low\_Cluster, High\_Cluster) \leftarrow$ DISTRIBUTE_A_CLUSTER($Next\_Cluster$);
5: $List\_of\_Clusters \leftarrow \{List\_of\_Clusters, Low\_Cluster, High\_Cluster\}$;
6: **if** (number of clusters in $List\_of\_Clusters < S$) **then**
7: **for all** ($Cluster$'s in $List\_of\_Clusters$) **do**
8: Find $Cluster$ such that $(\max\{Cluster\} - \min\{Cluster\})$ is maximal;
9: **end for**
10: $Next\_Cluster \leftarrow Cluster$;
11: Return to line 4;
12: **end if**
13: **return** $List\_of\_Clusters$;
14: **end procedure**

15: **procedure** DISTRIBUTE_A_CLUSTER($Next\_Cluster$)
16: $V_{Mean} \leftarrow {}^1\!/_2\,(\min\{Next\_Cluster\} + \max\{Next\_Cluster\})$;
17: $Low\_Cluster \leftarrow \left\{V_{DD}^{(i)} \in Next\_Cluster \mid V_{DD}^{(i)} \leq V_{Mean}\right\}$;
18: $High\_Cluster \leftarrow \left\{V_{DD}^{(i)} \in Next\_Cluster \mid V_{DD}^{(i)} > V_{Mean}\right\}$;
19: **return** $(Low\_Cluster, High\_Cluster)$;
20: **end procedure**

Table 5.1: Power supply clustering for a heterogeneous power delivery system with $S = 3$, $L = N = 4$, $V_{Drop} = 0.1$ volts and voltage domains (1 volt, 1 ampere), (1.49 volts, 1 ampere), (1.51 volts, 1 ampere), and (2 volts, 1 ampere).

Clustering Type	Power supply clustering $K = \{l_1, l_2, l_3\}$	SMPS voltages [V] $\left\{V_{SMPS}^{(i)} \mid i = 1, 2, 3\right\}$	Power efficiency [%]
Binary	$\{2, 1, 1\}$	$\{1.59, 1.61, 2.10\}$	87
Optimal	$\{1, 2, 1\}$	$\{1.10, 1.61, 2.10\}$	93

$V_{Drop} = 0.1$ volts.

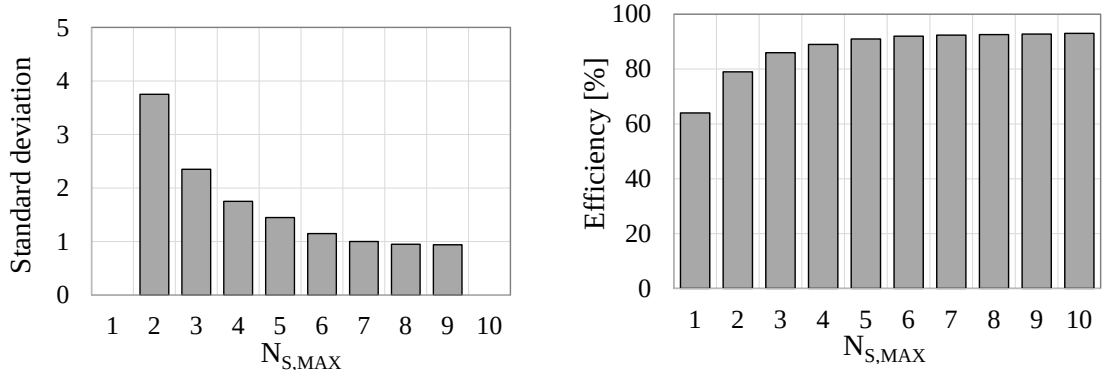
Alternatively, a linear power supply clustering produces a topology by linearly distributing the LDOs within S voltage clusters, as described by the algorithm represented by the pseudo-code provided in Algorithm 5.2. If less than S SMPS voltage clusters are produced within steps 1 through 3 in the linear power supply clustering algorithm, the linearly generated clusters are distributed into additional clusters using a binary algorithm. This algorithm produces a set of S SMPS voltage clusters $List\_of\_Clusters$ with a linear power supply clustering. In the worst case, the third and fourth steps are executed, respectively, N and S times, yielding an algorithm complexity that is linear with the number of voltage domains, $\mathcal{O}(N)$.

To compare the power efficiency of the near-optimal and optimal power delivery networks, a heterogeneous power delivery system with a small number of voltage

Algorithm 5.2 Algorithm for linear power supply clustering.

1: **procedure** LINEAR_CLUSTERING(sorted supply voltages $\left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}$)
2: $List\_of\_Clusters \leftarrow \{\}$;
3: $Cluster\_Range \leftarrow \left(\max\left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\} - \min\left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}\right)/S$;
4: **for each** $\left(V_{DD} \in \left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}\right)$ **do**
5: $k \leftarrow \left\lfloor \left(V_{DD} - \min\left\{V_{DD}^{(i)} \mid i = 1 \ldots N\right\}\right)/Cluster\_Range \right\rfloor + 1$;
6: Add V_{DD} to the k^{th} cluster in $List\_of\_Clusters$;
7: **end for**
8: **if** (number of clusters in $List\_of\_Clusters < S$) **then**
9: **for all** ($Cluster$'s in $List\_of\_Clusters$) **do**
10: Find $Cluster$ such that $(\max\{Cluster\} - \min\{Cluster\})$ is maximal;
11: **end for**
12: $Next\_Cluster \leftarrow Cluster$;
13: $(Low\_Cluster, High\_Cluster) \leftarrow$ DISTRIBUTE_A_CLUSTER($Next\_Cluster$);
14: $List\_of\_Clusters \leftarrow \{List\_of\_Clusters, Low\_Cluster, High\_Cluster\}$;
15: Return to line 8;
16: **end if**
17: **return** $List\_of\_Clusters$;
18: **end procedure**

domains is considered initially due to the computational complexity of the exhaustive optimal algorithm. The exhaustive algorithm determines the most power efficient clustering by comparing the power efficiency of all possible clusterings. The efficiency of the optimal power network produced by the exhaustive algorithm is compared here to the power efficiency of the near-optimal clustering algorithms. To estimate the power efficiency of the optimal power supply clusters, a heterogeneous power delivery system S_H with ten voltage domains ($N = 10$) and ten on-chip linear regulators ($L = 10$) is considered. The maximum number of off-chip switching converters is evaluated for one to ten converters ($1 \leq S \leq 10$). A voltage threshold of $V_{Drop} = 0.1$ volts, and domain voltages and currents of, respectively, 0.5 volts to 2 volts and 0.5 amperes to 3.5 amperes, are considered. Simulation results are sampled for 100 iterations. The power efficiency of a heterogeneous power delivery system with the power supply clusters, determined by an exhaustive analysis, is presented in Figure 5.3(a). A power efficiency above 80% is demonstrated for $S \geq 2$, and a maximum 93% power efficiency is achieved for $S = N$. Thus, the power efficiency of a heterogeneous power delivery system with an optimal power clustering exhibits a reasonable power efficiency of 80%, using only two off-chip switching converters. The efficiency increases rapidly with additional off-chip converters. Based on the Monte Carlo integration technique [140], the average error in the efficiency is bounded by



(a) (b)

Figure 5.3: Heterogeneous power delivery system (a) standard deviation, and (b) average efficiency using an exhaustive power supply clustering algorithm.

$\sigma_M/\sqrt{M}$, where σ_M is the standard deviation of a power efficiency sample and M is the number of samples. The standard deviation of the power efficiency is shown in Figure 5.3(b) for $2 \leq S \leq 9$. Values of σ_M range from 3.7 for $S = 2$ to 0.9 for $S = 9$, bounding the power efficiency error for $M = 100$ by, respectively, 0.37% to 0.09%. Power supply clustering for $S = 1$ and $S = N$ is explicit, yielding no error in the power efficiency. To evaluate the power efficiency of the near-optimal power supply clustering topologies described in Section IV, algorithms for binary and linear power supply clusterings have also been evaluated in Matlab. The same heterogeneous system S_H is considered for both linear and binary distributed power supplies. For a heterogeneous system with a single off-chip SMPS converter ($S = 1$) or maximum number of off-chip SMPS converters ($S = L$), the linear, binary, and optimal

clustering of the on-chip LDO regulators is identical. For $S = 1$, all of the LDOs are driven by a single SMPS converter, while for $S = L$, each LDO is driven by a different SMPS converter. Thus, the power efficiency of a heterogeneous system with $S = 1$ or $S = L$ is optimal with either a linear or binary power supply clustering. Alternatively, for $S < L$, a linear and binary clustering of the power supplies may differ from the exhaustive optimal solution, exhibiting a lower than optimal power efficiency. Due to the uniform nature of the linear approach, a linear clustering of the on-chip LDO regulators within the off-chip SMPS converters exhibits near optimal efficiency for power delivery systems with near uniformly distributed domain voltages. Alternatively, for a power delivery system with domain voltages that exhibit significant deviation from a uniform distribution, the power efficiency with the binary power supply clustering may be higher than with the linear clustering. This behavior is due to the greedy nature of the binary approach that iteratively identifies the on-chip power supply cluster with the lowest power efficiency and splits the cluster, increasing the overall efficiency of the system. To demonstrate the power efficiency of the binary and linear clusterings, the reduction in efficiency with both the binary and linear power supply clustering is simulated for two different power profiles, exhibiting a maximum 4% drop in power efficiency. The optimal solution with zero reduction in power efficiency is demonstrated for both power profiles in

Figure 5.4 for $S = 1$ and $S = L$. In the first power profile, the voltage levels are assumed to be randomly distributed between 0.5 volts and 2 volts, yielding an average power efficiency generated from over 100 iterations, as depicted in Figure 5.4(a). In this case, for $1 < S < L$, the exhaustive optimal solution produces a power supply that is uniformly distributed, while the linear power supply clustering yields a higher power efficiency. In the second power profile, the voltage levels are assumed to be normally distributed within each of the [0.5, 1.5), [1.5, 1.8), and [1.8, 2] ranges, prioritizing the mean value of the groups. Due to the non-uniform clustered nature of the voltage domain profile, for a heterogeneous system with three off-chip SMPS converters, intuitively, the on-chip LDO regulators should be non-uniformly distributed into three clusters covering the ranges [0.5, 1.5), [1.5, 1.8), and [1.8, 2]. In this case, a system with uniformly distributed clusters with voltage ranges [0.5, 1), [1, 1.5), [1.5, 2) is less power efficient. This heterogeneous system is therefore more suitable for a binary power supply clustering rather than a linear power supply clustering. The average power efficiency for the second power profile, generated from over 100 iterations, is depicted in Figure 5.4(b). In this case, specifically for $S = 3$, the optimal solution produces three non-uniform SMPS clusters, covering the three ranges, [0.5, 1.5), [1.5, 1.8), and [1.8, 2]. The binary power supply clustering with $S = 3$ also produces three SMPS clusters with voltage ranges, [0.5, 1.25),

[1.25, 1.625), and [1.625, 2], exhibiting a higher power efficiency than the efficiency produced by a linear power supply clustering. Based on a Monte Carlo integration technique, the error in estimating the drop in power efficiency, illustrated in Figure 5.4, is smaller than 0.63% for all values of S. Due to the greedy nature of the binary power supply clustering, the binary algorithm is better for those voltage domain levels grouped near specific voltage levels. Alternatively, the number of SMPS clusters S is only considered at the termination of the binary algorithm, potentially reducing the effectiveness of the binary clustering algorithm in those systems with uniformly distributed voltage domains. As expected, for most values of S and power supply specifications, the drop in power efficiency for the linear power supply clustering algorithm is lower than with the binary approach. However, the second power profile that forms three non-uniform voltage groups is better addressed by the binary power supply clustering algorithm, producing a more efficient heterogeneous power delivery system for $S = 3$. Thus, a heterogeneous power delivery system with a higher power efficiency is usually produced with a linear power supply clustering. However, for certain power profiles, a binary power supply clustering is preferable. To increase the power efficiency of a heterogeneous power delivery system, a combined hybrid approach should be employed. The power efficiency should be evaluated with both

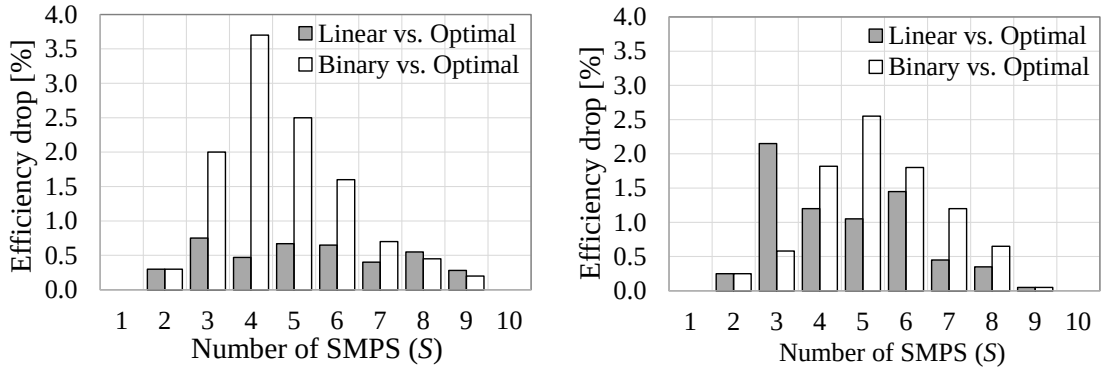


Figure 5.4: Decrease in linear and binary power efficiency from the optimal power efficiency for (a) randomly distributed voltage levels, and (b) voltage levels grouped within three voltage ranges.

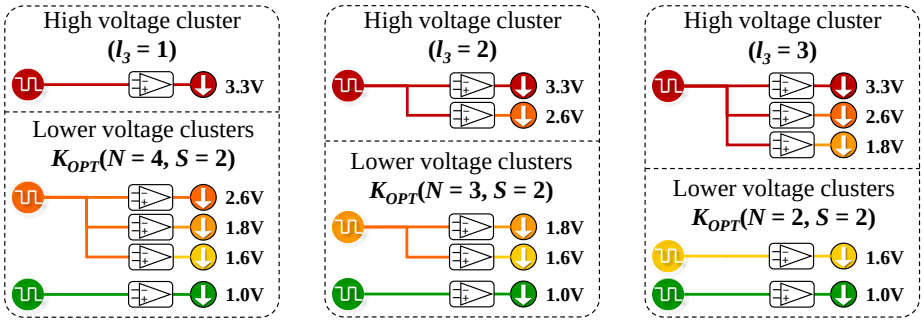
the binary and linear algorithms, and the configuration with the higher power efficiency should be employed. Analyzing the results depicted in Figure 5.4 based on this combined hybrid approach, the drop in power efficiency from the optimal solution is reduced to 1.5%, yielding a computationally efficient, $\mathcal{O}(S + N)$ complexity, near-optimal, and high fidelity power supply clustering.

5.2.2 Power supply clustering with dynamic programming

Determining the optimal clusters of the on-chip power supplies is an important challenge in a heterogeneous power efficient system. An optimal power supply clustering algorithm with $\mathcal{O}(N^2 \cdot S)$ is described in this section. A recursive analytic expression is provided for power supply clustering with L LDO regulators and S SMPS in smaller power delivery systems ($l < L$ LDO regulators and $s = S - 1$ SMPS). Power supply clusters are determined recursively with dynamic programming. Recursive power supply clustering, similar to exhaustive power supply clustering, yields the optimal set of power supply clusters.

The key idea behind the proposed algorithm is determining the number of voltage regulators within a high voltage SMPS cluster in $\mathcal{O}(N)$. Once the number of LDO regulators in a high voltage SMPS cluster is determined, the problem of power supply clustering is reformulated for the remaining LDO regulators and a smaller number of SMPS clusters. To exemplify the proposed solution, consider a heterogeneous system with three switching converters ($S = 3$) and five linear regulators ($L = 5$), supplying equal current I_{DD} to five power domains ($N = 5$) $\{V_{DD}^{(i)}\} = \{3.3$ volts, 2.6 volts, 1.8 volts, 1.6 volts, 1.0 volt$\}$. Dynamic programming is used to determine the optimum set of power supply clusters. The optimum clustering $K_{OPT}(5, 3) = \{l_1, l_2, l_3\}$, $\Sigma l_i = 5$ is determined recursively based on the number of

LDO regulators in the high voltage cluster l_3, and lower order optimal supply clustering $K_{OPT}(4,2)$, $K_{OPT}(3,2)$, and $K_{OPT}(2,2)$. A single recursive step is illustrated in Figure 5.5, demonstrating three possible alternatives for clustering with one ($l_3 = 1$), two ($l_3 = 2$), and three ($l_3 = 3$) LDO regulators within the high voltage SMPS cluster. Given the lower order clustering $K_{OPT}(4,2)$, $K_{OPT}(3,2)$, and $K_{OPT}(2,2)$,



(a) (b) (c)

Figure 5.5: A single step of the recursive power supply clustering algorithm for a heterogeneous power delivery system with $S = 3$ and $L = N = 5$, (a) a single LDO ($l_3 = 1$), (b) two LDO regulators ($l_3 = 2$), and (c) three LDO regulators ($l_3 = 3$) in a high voltage SMPS cluster.

the optimum clustering $K_{OPT}(5,3)$ is determined with linear computational complexity by comparing the power efficiencies $\varphi\{K_{OPT}(4,2),1\}$, $\varphi\{K_{OPT}(3,2),2\}$, and $\varphi\{K_{OPT}(2,2),3\}$, and choosing the clustering topology that minimizes the power losses.

For a general clustering algorithm, consider clustering L on-chip LDO regulators within S SMPS clusters to deliver power to L power domains with $N = L$ different

supply voltages. The optimal clustering topology of a system with N different supply voltages and S SMPS $K_{OPT}(N,S) = \{l_i\}_{i=1}^{S}$, $\sum l_i = N$ is determined recursively by

$$K_{OPT}(N,S) = \{K_{OPT}(N - n_0, S-1), l_S\}, \tag{5.3}$$

with the initial conditions,

$$K_{OPT}(N,2) = \{N - l_S, l_S\}, \tag{5.4}$$

$$K_{OPT}(N, S=N) = \{1, 2, ..., N\}, \tag{5.5}$$

where $1 \leq l_S \leq (N-S)$ is the number of LDO regulators in the high voltage SMPS cluster. To maximize the overall power efficiency of the system, the number of LDO regulators in the last SMPS cluster is

$$\varphi(K_{OPT}(N,S)) = \max_{l_S} \varphi\left(\{K_{OPT}(N - l_S, S-1), l_S\}\right). \tag{5.6}$$

Once the power supply clusters are recursively determined with dynamic programming, the maximum voltage level within each SMPS cluster determines the SMPS output and LDO input voltage based on (4.27). Pseudo-code of the algorithm for determining the LDO input voltages based on the proposed clustering is shown in

Algorithm 5.3.

The LDO input voltages in a system with a single ($S = 1$) switching converter and the maximum number of SMPS ($S = N$) are determined, respectively, at lines 2 to 4 and 5 to 7. To determine the optimal clustering of a general system with ($1 < S < N$) switching converters, lines 8 through 42 are executed. The LDO input voltages for all of the systems with $s \leq S$ SMPS and $l \leq L$ LDO regulators are determined progressively and stored in matrix all_V_{LDO}. The matrix is allocated and initiated based on (5.4) and (5.5) at lines 9 to 17. The voltage levels at the LDO input voltages are determined in a loop (see lines 19 to 20) for systems with a progressively increasing number of power supplies. All of the high voltage cluster configurations with a different numbers of LDO regulators are determined at lines 26 to 28. The power efficiency of different configurations is compared at lines 29 to 33, determining the most power efficient system. The number of efficiency comparisons required to determine the optimal set of clusters $K_{OPT}(N, S)$ given all of the optimal clusterings of lower order $K_{OPT}(n < N, s < S)$ is $N - S$. The computational complexity to determine the most power efficient clusters with $N = L$ LDO regulators and S SMPS converters is therefore

$$\sum_{s=1}^{S} \left(\sum_{n=s}^{N} \mathcal{O}(n-s) \right) = \mathcal{O}(N^2 \cdot S), \quad N \geq S. \tag{5.7}$$

Algorithm 5.3 Dynamic programming algorithm to determine LDO input voltages for power efficient clustering.

procedure DP_CLUSTERING($\left\{V_{DD}^{(i)}, I_{DD}^{(i)} \mid i = 1 \ldots N\right\}$,$S$,$V_{Drop}$)
  if $S == 1$ **then**
    % There is only one cluster
    $V_{LDO}(:) \leftarrow \max V_{DD}(:) + V_{Drop}$;
  else if $S == N$ **then**
    % The number of clusters equals the number of voltage levels
    $V_{LDO}(:) \leftarrow V_{DD}(:) + V_{Drop}$;
  else
    % Initiate a matrix to store all the clusters for lower order systems
    $all\_V_{LDO}(:) \leftarrow zeros(S, L = N, N)$;
    % Update initial conditions based on (5.4) and (5.5)
    for $l = 1$ **to** N **do**
      $all\_V_{LDO}(1, l, 1 : l) \leftarrow ones(1, l) \cdot (V_{DD}(l) + V_{Drop})$;
    end for
    for $s = 1$ **to** S **do**
      $all\_V_{LDO}(s, s, 1 : s) \leftarrow V_{DD}(1 : s) + V_{Drop}$;
    end for
    % Find all the lower order ($s \leq S$ and $l \leq L$) clusters
    for $(s = 2$ **to** $S)$ **do**
      for $(l = s + 1$ **to** $N)$ **do**
        $all\_V_{LDO,OPT}(:) \leftarrow zeros(1, N)$;
        $\varphi_{OPT} \leftarrow 0$;
        for $(l_s = 1$ **to** $l - s + 1)$ **do**
          % l_s and $(l - l_s)$ are the number of LDO regulators in, respectively,
          % the highest voltage cluster and the rest $(s - 1)$ clusters
          $V_{LDO,TMP}(1 : l - l_s) \leftarrow all\_V_{LDO}(s - 1, l - l_s, :)$;
          $V_{LDO,TMP}(1 - l_s + 1 : l) \leftarrow V_{DD}(l) + V_{Drop}$;
          $\varphi_{TMP} \leftarrow 100 \cdot [\sum V_{DD}(1 : l) \cdot I_{DD}(1 : l)] \,/\, [\sum V_{LDO,TMP}(1 : l) \cdot I_{DD}(1 : l)]$;
          if $(\varphi_{TMP} > \varphi_{OPT})$ **then**
            % Store the clustering with the highest efficiency
            $V_{LDO,OPT} \leftarrow V_{LDO,TMP}$;
            $\varphi_{OPT} \leftarrow \varphi_{TMP}$;
          end if
        end for
        % Update all_V_{LDO} with the optimal clusters
        $all\_V_{LDO}(s, l, :) \leftarrow V_{LDO,OPT}$;
      end for
    end for
    $V_{LDO}(:) \leftarrow all\_V_{LDO}(S, L = N, :)$;
  end if
  return V_{LDO};
end procedure

To estimate the power efficiency of the recursive power supply clustering algorithm, a heterogeneous power delivery system with 25 supply voltage levels ($N = 25$) and 25 on-chip linear regulators ($L = 25$) is considered. The number of off-chip switching converters is evaluated for one to 25 converters ($1 \leq S \leq 25$). A voltage drop of $V_{Drop} = 0.1$ volts and 100 random profiles of domain voltages and currents of, respectively, 0.5 volts to 2 volts and 0.5 amperes to 3.5 amperes are considered. The maximum power efficiency with an average domain voltage and current of, respectively, 1.25 volts and 2 amperes is evaluated based on (4.22), yielding 93% power efficiency for $S = 25$. The power efficiency of a heterogeneous power delivery system with the power supply clusters determined by a recursive analysis is illustrated in Figure 5.6. As expected, a maximum 93% power efficiency is achieved for $S = N$. An

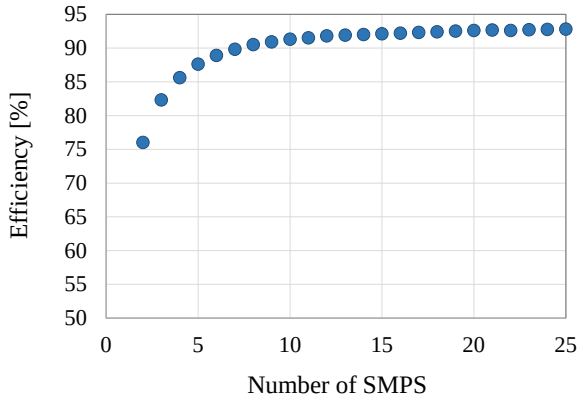


Figure 5.6: Power efficiency of heterogeneous power delivery system based on a recursive power supply clustering algorithm.

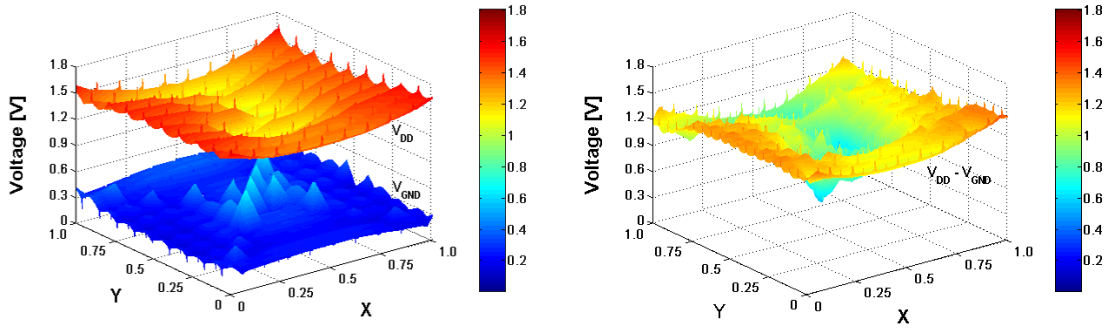
average power efficiency above 82% is demonstrated for $S \geq 3$. Thus, the power efficiency of a heterogeneous power delivery system with an optimal voltage clustering exhibits a reasonable power efficiency, despite only three switching converters. The efficiency increases rapidly and saturates with additional SMPS converters. Thus, an excessive number of off-chip or in-package converters is avoided with the proposed power supplies clustering, producing a power and area efficient co-design of the overall system of power converters and regulators.

5.3 Co-design of power delivery system in circuit benchmarks

The optimal exhaustive power delivery network, the hybrid near-optimal, and the recursive (based on dynamic programming) optimal solutions described in Section 5.2 are evaluated in Matlab. To determine the power efficiency with and without power separation, several test cases are evaluated based on IBM power grid benchmarks [141]. The test cases and simulation results with the proposed linear, binary, and recursive clustering algorithms are described in Section 5.3.1. Power delivery with on-chip regulation in circuits with multiple power domains [118] is considered in Section 5.3.2.

5.3.1 Power supply clustering of IBM power grid benchmarks

Five test cases have been considered based on IBM power grid benchmarks [141] to evaluate the efficiency of the power separation principle in circuits with hundreds of power domains and tens of different supply voltages. The voltage map of a V_{DD} and V_{GND} $M6$ metal layer at normalized (x, y) locations is presented in Figure 5.7(a). The actual voltage drop in the metal layer $M6$ is $V_{DD} - V_{GND}$, as shown in Figure 5.7(b) for the ibmpg1 benchmark. Different circuits in ibmpg1 operate with different



(a) (b)

Figure 5.7: Test case ibmpg1 with (a) V_{DD} and V_{GND} voltage map, and (b) $V_{DD} - V_{GND}$ voltage drop across the circuits in metal layer $M6$.

supply voltages, varying from 0.5 volts to 1.5 volts. A total number of 66 voltage domains with voltage levels ranging from 0.5 volts to 1.8 volts with a 0.02 volt shift $\{V_{DD}^{(i)} = 0.5V + i \cdot 0.02V\}_{i=0}^{i=65}$ is considered in this analysis. Each of the benchmarks

Table 5.2: Power grid specifications for 0.8 volt, 1.0 volt, 1.2 volt, 1.4 volt, and 1.6 volt domains based on ibmpg1 test case.

Voltage domain id	15	25	35	45	55
Supply voltage [V]	0.8	1.0	1.2	1.4	1.6
Number of nodes	636	820	1297	19	0
Area [%]	2.12	2.73	4.320	0.062	0
Supply current [mA]	212	273	432	6.2	0

is partitioned into voltage domains based on the voltage maps, and the area of each domain is determined. The current within a benchmark circuit is assumed to be uniformly distributed. The current load of a domain is therefore assumed to be proportional to the area of the domain. The voltage within each domain is regulated by an LDO, ensuring that the total number of on-chip LDO regulators is the same as the number of voltage domains. To illustrate the process in which a test case is generated, specifications for several voltage domains in ibmpg1 are listed in Table 5.2. The total number of nodes processed in ibmpg1 is 30,027. A total current of 10 amperes is assumed to be consumed by the circuits represented by ibmpg1.

Test cases for the IBM benchmarks, ibmpg2, ibmpg3, ibmpgnew1, and ibmpgnew2, are generated in a similar way. The proposed power supply clustering algorithm is demonstrated in Matlab and applied to all of the test cases on a multi-core system with four Intel(R) Core(TM) i3-2120 CPU @ 3.30 GHz processors and 2.498 MB memory. A voltage drop of 0.1 volts within an LDO is assumed. The power grid specifications and simulation results with and without power supply clustering are

Table 5.3: Power efficiency in circuits with and without separation of power conversion and regulation. CPU time for hybrid (with near-optimal efficiency), dynamic programming (DP) based, and exhaustive clustering is provided.

Benchmark	Voltage domains/ LDO regulators (N)		Power efficiency with S voltage clusters [%]					CPU time [s] ($S = 1/2N$)		
	Number	Voltage range [V]	Without power separation	With power separation				Hybrid clusters	DP clusters	Exhaustive clusters
				$S = 2$	$S = 3$	$S = 5$	$S = 10$			
ibmpg1	49	(0.50,1.46)	68.1	78.3	82.4	86.5	89.2	0.023	0.299	> 2,000
ibmpg2	21	(1.12,1.52)	77.2	87.0	89.2	91.0	92.1	0.018	0.038	> 2,000
ibmpg3	15	(1.44,1.70)	88.6	91.9	92.8	93.6	94.1	0.022	0.012	> 2,000
ibmpgnew1	11	(1.52,1.72)	90.3	92.9	93.5	94.1	94.3	0.022	0.011	5.537
ibmpgnew2	11	(1.52,1.72)	90.4	92.9	93.5	94.1	94.3	0.020	0.011	5.477

listed in Table 5.3.

Power clustering with the proposed algorithms exhibits orders of magnitude shorter CPU time than the exhaustive approach. Those power grids with a range of LDO output voltages up to 0.2 volts (ibmpgnew1 and ibmpgnew2) exhibit a high power efficiency of 93% despite only two SMPS clusters. The power efficiency of these grids increases by 2.5% as compared to a power delivery system without power separation. Increasing the power efficiency in those power grids with a large number of SMPS clusters (94% with ten switching converters) requires excessive area. Alternatively, the ibmpg1 benchmark exhibits a wider range of LDO output voltages, 0.5 volts to 1.5 volts, and therefore, a low power efficiency of 68% without power supply clustering. The effectiveness of power separation is shown to be significant in ibmpg1 with a 10.2% and 21.1% increase in power efficiency with, respectively, $S = 2$ and $S = 10$ SMPS clusters as compared to $S = 1$. Separation of power

conversion and regulation is therefore particularly important in those systems with a wide range of on-chip supply voltages and voltage drops. To provide high quality power in dynamically scaled multi-voltage circuits, the efficiency of the power supply clustering is dynamically evaluated. The proposed power supply clustering DP algorithm exhibits an order of magnitude smaller CPU run time as compared with the exhaustive method, while providing identical clusters. Additional clustering speedup is achieved with the proposed hybrid approach with near-optimal power efficiency. With the proposed approach, the switching converters and linear regulators can be co-designed in run time for power and area efficient management of the energy budget.

5.3.2 Power supply clustering and existing power delivery solutions

The separation principle is illustrated in circuits C_1 and C_2 with multiple power domains $V_{DD}^{(A)} = 1.4\,\text{volts}$, $V_{DD}^{(B)} = 1.2\,\text{volts}$, and $V_{DD}^{(C)} = 1.0\,\text{volt}$, and the on-chip voltage conversion and regulation scheme used in [118]. A maximum voltage drop (V_{Drop}) of 5% of the input voltage at the I/O interface (V_{SMPS}) is allowed at POLs within all of the power domains ($V_{Drop} \leq 0.05 V_{SMPS}$). The total area of the circuits, C_1 and C_2, is similar. The area of C_1 is dominated by the low power domain C, while

the area of C_2 is dominated by the high performance power domain A. The current density J and normalized area per power domain are listed in Table 5.4 for both circuits. To support on-chip voltage conversion and regulation within the power do-

Table 5.4: Parameter setup for circuits C_1 and C_2.

Power domain	J [A/unit area]	Normalized area	
		Circuit C_1	Circuit C_2
A	2/3 V	1	6
B	1/2 V	2	2
C	1/3 V	6	1

mains A, B, and C in C_1 (C_2), a single input voltage $V_{SMPS} = 1.45$ volts ($V_{SMPS} = 1.50$ volts) is used in the original configuration without power separation [118]. Note that the input voltage $V_{SMPS} = 1.45$ volts ($V_{SMPS} = 1.50$ volts) is higher than the highest supply voltage $V_{DD}^{(A)} = 1.4$ volts to maintain the required margin for on-chip voltage regulation with a reasonable number of 50 (60) on-chip LDO regulators in C_1 (C_2). Alternatively, the power can be converted separately off-chip or in-package for each power domain and regulated on-chip within each power domain. The input voltages and power efficiency for on-chip voltage regulation with and without separation of power conversion and regulation is listed in Table 5.5 for both circuits. In the proposed configuration, three off-chip or in-package SMPS supply three different voltages, $V_{SMPS}^{(A)} = 1.47$ volts, $V_{SMPS}^{(B)} = 1.26$ volts, and $V_{SMPS}^{(C)} = 1.05$ volts to, respectively, power domains A, B, and C, yielding a power efficiency as high as 95.2%.

Table 5.5: On-chip voltage regulation with and without separation of power conversion and regulation.

Circuit	Without power separation ([118])		With power separation (current work)	
	V_{SMPS}	φ	$V_{SMPS}^{(A)}$, $V_{SMPS}^{(B)}$, $V_{SMPS}^{(C)}$	φ
C_1	1.45 V	77.3%	1.47 V, 1.26 V, 1.05 V	95.2%
C_2	1.50 V	90.7%	1.47 V, 1.26 V, 1.05 V	95.2%

To regulate the on-chip voltage with the required number of LDO regulators, the voltage at the output of each SMPS is designed with a margin of 5% of the required on-chip supply voltage ($V_{DD}^{(i)} = 1.05V_{DD}^{(i)}$). Without separating power conversion and regulation, the choice of off-chip supply voltage is determined by the supply voltage in the high performance power domain, exhibiting a higher voltage drop (V_{SMPS} - $V_{DD}^{(C)} = 0.45$ volts in C_1) and a lower power efficiency (77.3% in C_1) in low power circuits. Alternatively, an enhanced choice of $V_{SMPS}^{(i)}$ voltages is possible by separating power conversion and regulation, exhibiting a lower voltage drop ($V_{SMPS}^{(i)} - V_{DD}^{(i)} \leq 0.07$ V in C_1) and a higher power efficiency (95.2% in C_1).

5.4 Summary

On-chip power regulation and delivery are necessary for delivering high quality power to modern high performance integrated circuits. Dynamic power management

is employed in modern systems to efficiently manage the energy budget. For efficient adaptive power management, power should be converted with efficient power converters and dynamically regulated with ultra-small on-chip regulators. To avoid excessive usage of switching converters, the power efficiency as a function of the number of SMPS clusters is determined. The preferred number of power converters should be determined with the proposed function based on the application-specific efficiency objectives.

Dynamically co-designing tens of power converters with hundreds to thousands of on-chip regulators is a primary concern in an energy efficient power delivery system. An exhaustive solution that determines the on-chip power supply clusters with the highest power efficiency is, however, computationally inefficient and impractical in real-time systems. Thus, algorithms to cluster a heterogeneous power supply system with polynomial computational complexity are presented. An order of magnitude speedup is exhibited with the proposed algorithms as compared with exhaustive clustering. Up to a 21% increase in power efficiency is demonstrated based on the IBM benchmarks with more than two switching converters. A dynamically controlled heterogeneous integrated power delivery system is shown to be a computationally and power efficient alternative to existing *ad hoc* methodologies that employ either switching or linear on-chip power supplies.

To achieve a power efficient system, power should be primarily converted off-chip,

in-package, and/or on-chip with power efficient switching supplies, and regulated with ultra-small linear low dropout regulators at the point-of-load. To dynamically co-design tens of power converters with hundred to thousands of on-chip regulators, the optimal clustering of the on-chip LDO regulators within the SMPS voltage clusters is determined that maximizes in run time the power efficiency of the overall power delivery system. Additional speedup is achieved with the proposed heuristics, yielding near-optimal power efficiency.

Chapter 6

Distributed Power Delivery with 28 nm Ultra-Small LDO Regulator

The quality of the power supply in portable electronic systems can be efficiently addressed with point-of-load (POL) distributed power delivery [69], [82], which requires the on-chip integration of multiple power supplies. Several low-dropout (LDO) regulators suitable for on-chip integration have recently been fabricated [66]–[68], [88], [138], [142]–[151], exhibiting fast load regulation and high current efficiency (i.e., the ratio of the load current and input current). Due to these characteristics, the LDO is a key component in on-chip power management.

To achieve a fast transient response for a load current step of hundreds of microamperes, the quiescent current of an LDO is typically increased [88], [152], lowering the current efficiency. Dynamically biased shunt feedback, proposed in [68], has been applied to achieve high current efficiency and system stability over a wide

range of load currents. While the impedance-attenuated-buffer in [68] is dynamically biased, the error amplifier in [68] is statically biased, making simultaneous optimization of the LDO speed and current consumption difficult. Alternatively, adaptive biasing techniques have been proposed that boost the bias current during fast output transitions [67], [149], [152], yielding a promising technique for fast and power efficient load regulation. The LDO in [67], however, utilizes a 1 µF off-chip capacitor to stabilize the voltage regulation, significantly increasing the response time of the regulation loop. Alternatively, a flipped voltage follower (FVF) LDO compensated by a single Miller capacitor is proposed in [66] that achieves excellent current efficiency of 99.99%, good load regulation (0.1 mV/mA), and moderate regulation speed without an off-chip capacitor [66].

With the increasing number of power domains and high granularity of the on-chip supply voltages [54], multiple ultra-small voltage regulators will ultimately be integrated on-chip [69], [70]. The physical size of the LDO therefore becomes a primary issue in power management ICs. A nanoscale voltage regulator is expected to exhibit a smaller physical area and improved large- and small-signal characteristics. Alternatively, significant process, voltage, and temperature (PVT) variations pose new stability challenges on the co-design of these ultra-small on-chip voltage regulators. Parallel voltage regulation where multiple regulators are connected to the same power

grid has recently attracted significant attention, both from academia [153]–[157] and industry [158]–[160]. Satisfying small area, high power efficiency, and stability is however more challenging with parallel voltage regulation. Existing on-chip voltage regulator topologies do not simultaneously overcome these three challenges.

A power delivery and regulation system with six ultra-small 28 nm LDO regulators distributed on-chip is described in this chapter. The proposed distributed power delivery system features an adaptive current boost bias and an adaptive RC compensation network controlled individually within each LDO regulator, increasing the power efficiency and stability of the overall system over a wide range of load currents and PVT variations. As compared to other state-of-the-art LDO regulators providing fast voltage regulation [66]–[68], [88], a single LDO within the proposed power delivery system (including all capacitors and a bias generator) is 2.24 times smaller. The proposed power delivery system delivers 3.9 to 15.8 times more load current, while exhibiting a similar current efficiency. The proposed distributed power delivery system has been tested under a wide range of PVT variations, yielding a stable and fast loop response. Although parallel voltage regulation has previously been demonstrated using eight digital LDO regulators with 77.5% current efficiency [159], to the best of our knowledge, the proposed system is the first successful silicon demonstration of stable parallel analog LDO regulators without off-chip compensation, and

exhibits 99.49% current efficiency.

The rest of the chapter is organized as follows. The proposed power delivery system with six fully integrated LDO regulators with adaptive current boost bias and RC compensation networks is described in Section 6.1. Measured performance results are reviewed in Section 6.2. The chapter is concluded in Section 6.3.

6.1 Power delivery system

A power delivery system with six fully integrated LDO regulators is described in this section. The proposed system converts any input voltage ranging from 0.9 volts to 1.1 volts into a target output voltage ranging from 0.6 volts to 0.8 volts, supplying up to 788 mA to the load. A model of the power delivery system with six LDO regulators and a distributed power delivery network is shown in Figure 6.1. Current sharing is a primary concern in a distributed power delivery system. Each LDO contributes differently to the voltage regulation of a power network based on the position of the active current loads and the level of consumed current. Load sharing among the LDO regulators is illustrated in Figure 6.2 with a single current load (at the upper right corner of the power network) switching between 18 mA and 450 mA. The LDO at the upper right corner (in Figure 6.1) is located in close proximity with the current load and supplies the largest portion (up to 160 mA) of

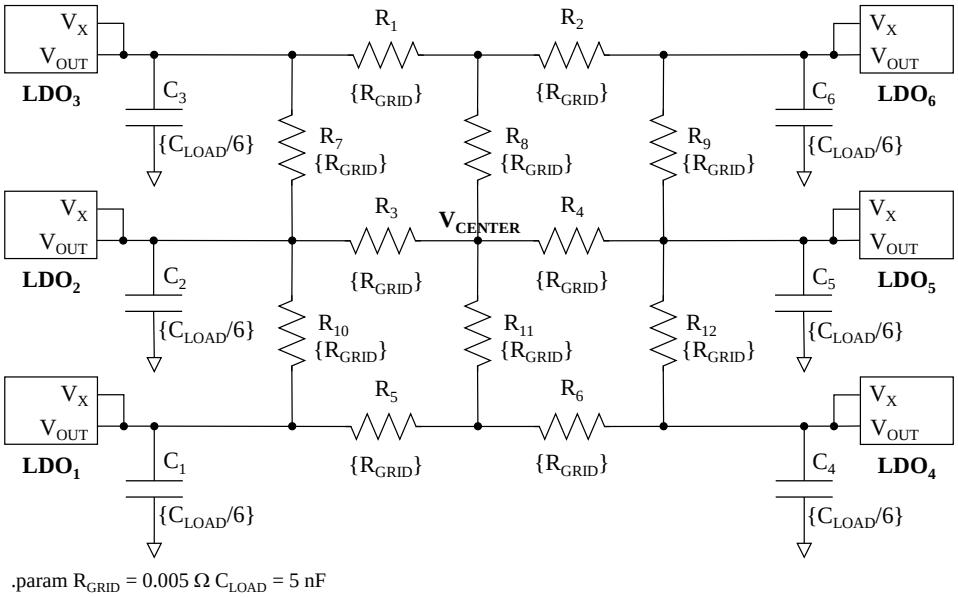


Figure 6.1: Model of distributed LDO and power distribution network.

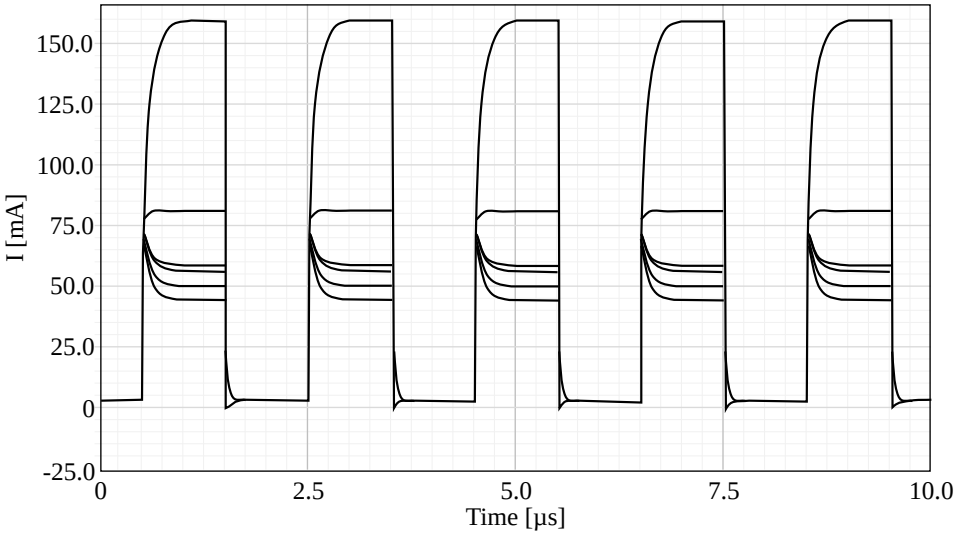


Figure 6.2: Load sharing in distributed power delivery system.

the total current requirements, which is higher by a factor of two than the average current load supplied by a single LDO. Alternatively, the remote LDO at the bottom left corner supplies significantly less current (up to 40 mA), only half of the average LDO load current. In modern high performance circuits, the load map may change significantly over time [141] and under PVT variations. Mechanisms are required to co-design the distributed on-chip regulators to dynamically stabilize the power delivery system over time. Adaptive mechanisms are described in this chapter that respond to load variations at the output of each of the LDO regulators, increasing the power efficiency of the system and enhancing the performance and stability.

An adaptive current boost bias and an adaptive RC compensation network are included within each LDO to, respectively, enhance the slew rate with low power overhead, and stabilize the power regulation over a wide range of load currents and PVT variations. The operation of the proposed dynamic mechanisms is controlled within the individual LDO regulators, providing fine grain regulation of the power voltage. Alternatively, both mechanisms within each LDO are adaptively triggered by the same sensing circuit, exhibiting a more compact power delivery system. The circuit topology of the proposed LDO is shown in Figure 6.3. The components of the proposed power delivery system are described in the following subsections.

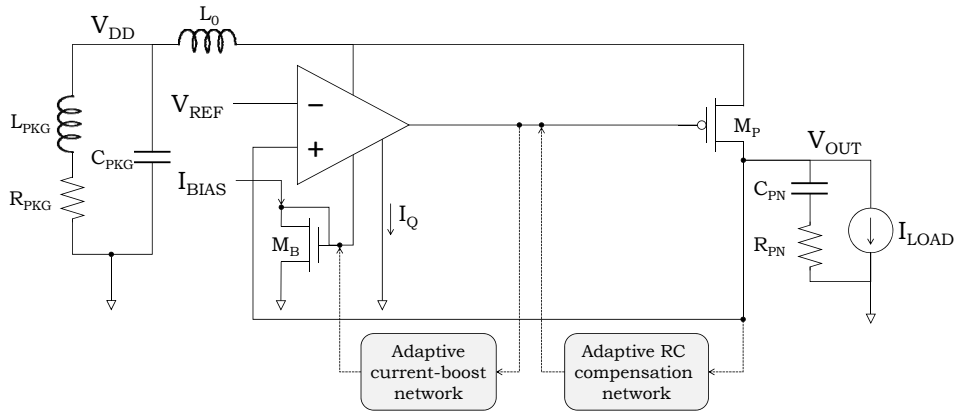


Figure 6.3: LDO topology.

6.1.1 Op amp based LDO

The open loop output resistance, load capacitance, and control loop gain and bandwidth are important criteria when developing a fast LDO. To address these challenging transient requirements, a three current mirror operational transconductance amplier (OTA) topology [147] is used within each LDO, as shown in Figure 6.4. A linear model that considers the effects of the open loop output resistance, load capacitance, and control loop gain and bandwidth is used to model the behavior of the three current mirror OTA. Miller compensation is used to achieve a dominant pole. The proposed model is shown in Figure 6.5. The open loop gain of the LDO regulator is

$$A(s) = \frac{V_{OUT}}{V_{IN}} = \frac{-(g_{m1}R_1)\cdot(g_{m2}R_2)(1+N\cdot s)}{1+D_1\cdot s+D_2\cdot s^2+D_3\cdot s^3}, \tag{6.1}$$

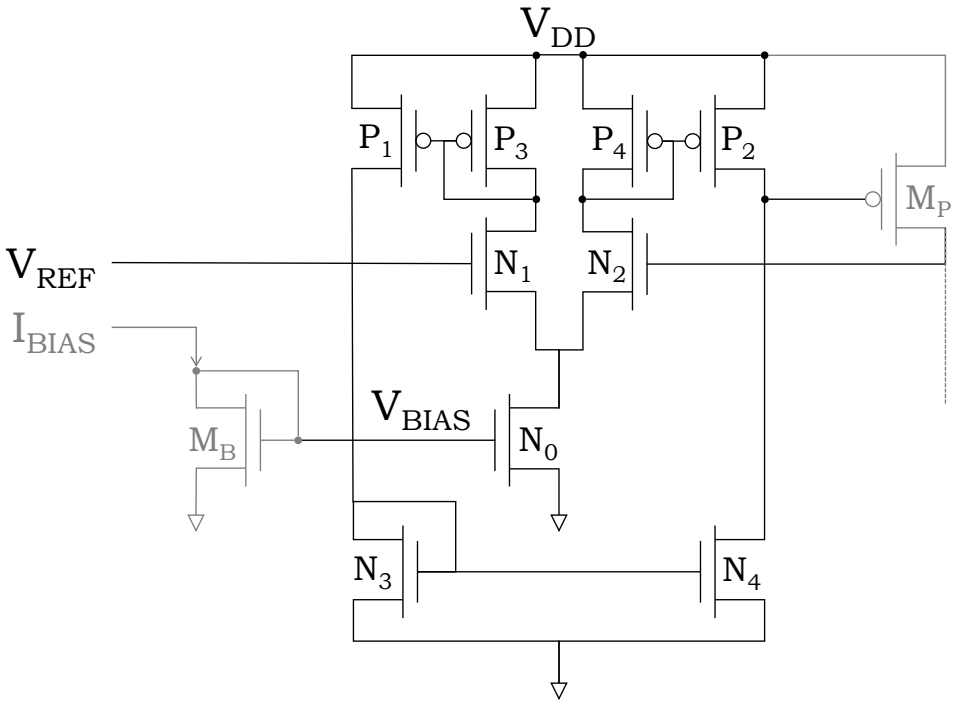


Figure 6.4: Three current mirror OTA.

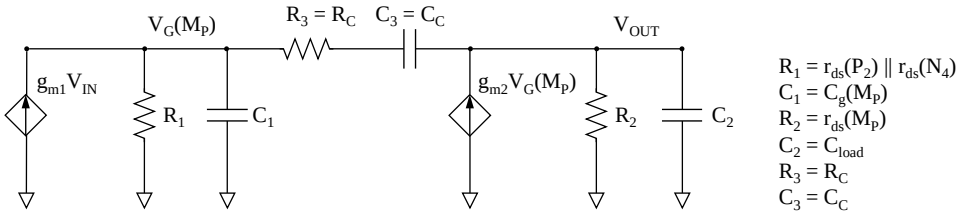


Figure 6.5: Small signal linear model of LDO.

where

$$N = (R_3 - \frac{1}{g_{m2}})C_3, \tag{6.2}$$

$$D_1 = R_1C_1 + R_2C_2 + R_3C_3 + R_1C_3 + R_2C_3(1 + g_{m2}R_1), \tag{6.3}$$

$$D_2 = R_1R_2C_1C_2 + R_1C_1R_3C_3 + R_2C_2R_3C_3 + R_1R_2(C_1 + C_2)C_3, \tag{6.4}$$

$$D_3 = R_1C_1R_2C_2R_3C_3. \tag{6.5}$$

The zero of the LDO regulator is formed by the compensation network at the frequency f(z),

$$f(z) = \frac{1}{2\pi(R_3 - \frac{1}{g_{m2}})C_3} \approx \frac{1}{2\pi(R_CC_C)}. \tag{6.6}$$

The dominant pole frequency $f(p_1)$ is assumed to be significantly lower than the frequency of the other poles, $f(p_2)$ and $f(p_3)$ ($f(p_1) << f(p_2), f(p_3)$). All of the poles are assumed to be real and approximated over a feasible range of g_{mi}, R_i, and C_i components, yielding,

$$\begin{aligned} f(p_1) &\approx \frac{1}{2\pi(g_{m2}R_2)(R_1C_3)} \\ &= \frac{1}{2\pi[g_{m2}r_{ds}(M_P)][(r_{ds}(P_2)\|r_{ds}(N_4))C_C]}, \end{aligned} \tag{6.7}$$

$$f(p_2) \approx \frac{g_{m2}}{2\pi C_2} = \frac{g_{m2}}{2\pi C_{Load}}, \tag{6.8}$$

$$f(p_3) \approx \frac{1}{2\pi R_3C_1} = \frac{1}{2\pi R_CC_g(M_P)}. \tag{6.9}$$

The DC gain of the LDO regulator $A_0 = (g_{m1}R_1g_{m2}R_2)$ is listed in Table 6.1, exhibiting an average gain of 57 dB and less than 1% variations over a wide range of

process, temperature, and load variations.

Table 6.1: DC gain over a range of load currents at slow (SS, −30 °C), typical (TT, 25 °C), and fast (FF, 105 °C) corners.

Process	Temperature	I_{Load} [mA]	DC gain [dB]
SS	−30 °C	70	58.73
		20	60.38
		1	61.20
TT	25 °C	100	56.73
		25	58.38
		3	57.20
FF	105 °C	150	51.23
		100	53.40
		70	54.35

To analyze the stability and the proposed compensation of a single LDO regulator, the small signal transconductance and drain source resistance of the output device are assumed to be, respectively, $g_{m2} \propto \sqrt{I_{Load}}$ and $R_2 \propto 1/I_{Load}$. Other parameters are assumed to be approximately independent of the load current in the region of interest. Under these assumptions, the frequency of the first and second poles increases with $\sqrt{I_{Load}}$, while the zero frequency $f(z)$ and third pole frequency $f(p_3)$ are approximately constant under load current variations. The value of R_C is chosen to ensure that the frequency of the third pole $f(p_3)$ is larger than the unity gain frequency in the region of interest, yielding a second order system to enhance stability.

To increase stability over a wide range of load capacitance, the dominant pole is

determined by the compensation capacitor, yielding $f(p_1) < f(p_2)$ and, therefore,

$$C_C > \frac{C_2}{g_{m2}^2 R_2 R_1} > 3\,\text{pF}. \tag{6.10}$$

The maximum phase margin is achieved when the second pole is canceled by the zero, yielding a first order system under the following constraint on the compensation network,

$$R_C = \frac{C_2}{g_{m2} C_C} < 3\,\text{k}\Omega. \tag{6.11}$$

Finally, the first order system exhibits a unity gain at $f_t \approx A_0 \cdot f(p_1) = g_{m1}/2\pi C_C < 130MHz$, fulfilling the requirement $f_t < f(p_3)$ under the constraints, (6.10) and (6.11).

$$R_C < \frac{C_C}{g_{m1} C_1} < 2\,\text{k}\Omega. \tag{6.12}$$

Under the constraints, (6.10), (6.11), and (6.12), the proposed LDO regulator is a first order system with a phase margin between 45° and 90° (the PM is reduced over three decades by 90° due to p_1 and a portion of 45° due to p_3) and a bandwidth $f(p_1)$, as shown in Figure 6.6.

The transconductance of the output device g_{m2} increases, however, with $\sqrt{I_{Load}}$. Thus, under current load variations the second pole is shifted away from the zero frequency, violating the first order assumption, and degrading the stability of the

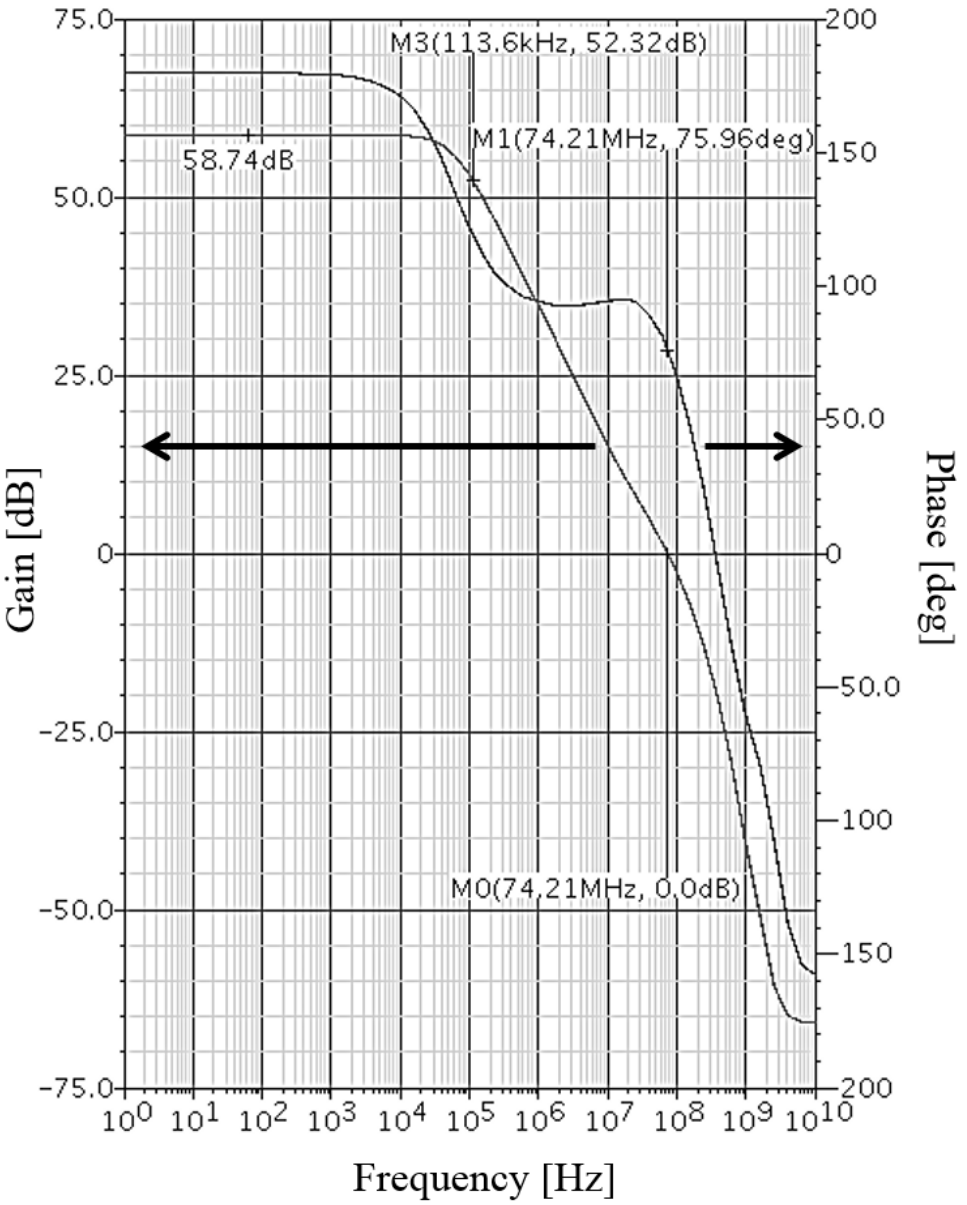


Figure 6.6: Frequency response of near optimally compensated LDO regulator.

LDO regulator. The behavior of the PM is, therefore, primarily determined by the variations of the frequency of the second pole, as shown in Figure 6.7, linearly

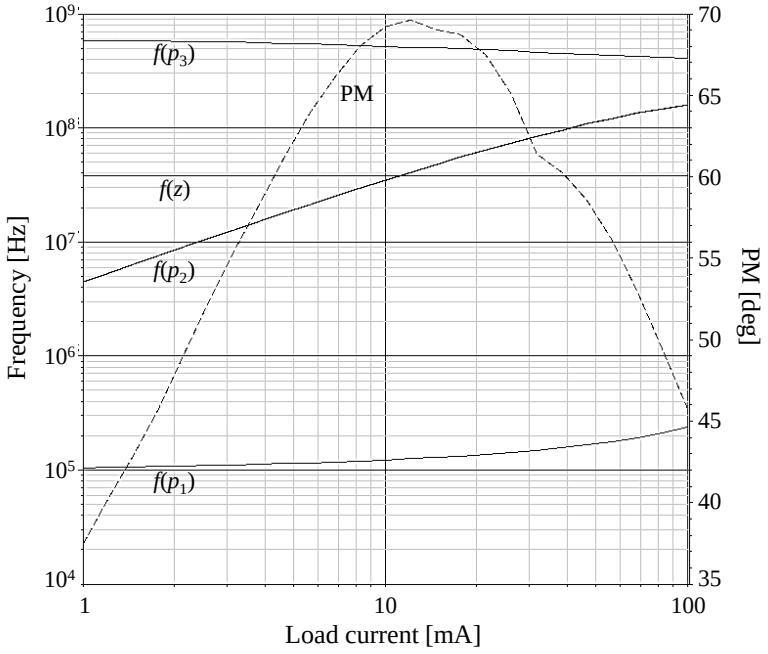
decreasing with a larger $|\log(f(p_2)/f(z))|$ ratio,

$$\begin{aligned} PM(f(z)) - PM(f(p_2)) &\propto \left|\log\left(\frac{f(p_2)}{f(z)}\right)\right| \\ &= \left|\log\left(\frac{g_{m2}R_C C_C}{C_{Load}}\right)\right|. \end{aligned} \tag{6.13}$$

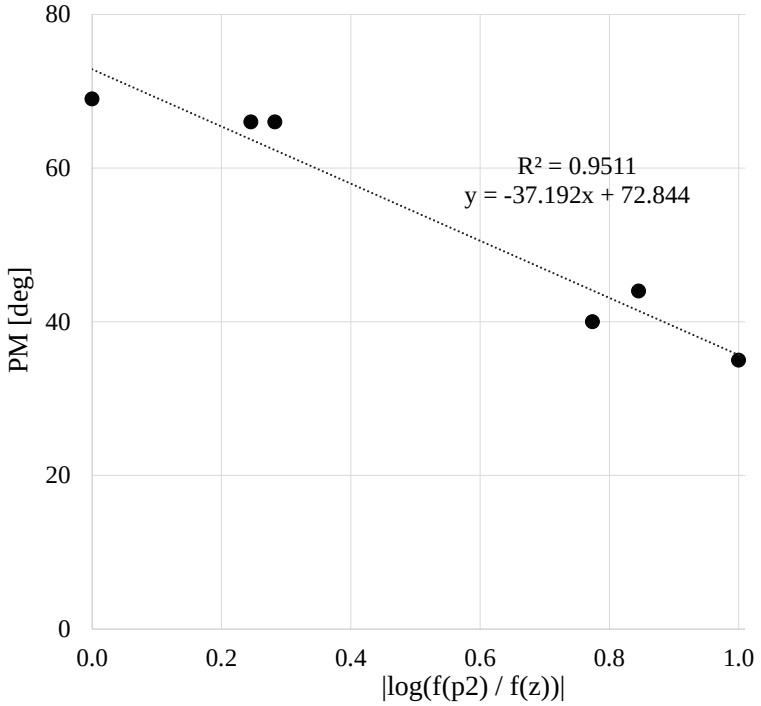
Note the high accuracy of this linear approximation ($R^2 = 0.9511$).

To maximize the stability of an LDO regulator over a range of load currents, the compensation should be modified with changing transconductance g_{m2}, maintaining $g_{m2}R_C C_C/C_{Load}$ **to** 1. The phase margin is shown in Figure 6.8 with two different compensation resistors, $R_C = 0.7\,\text{k}\Omega$ and $R_C = 1.7\,\text{k}\Omega$, and a compensation capacitor of $C_C = 8.5\,\text{pF}$ for a range of low load currents. At low currents of $I_{Load} < 3\,\text{mA}$, compensation with a larger resistor ($R_C = 1.7\,\text{k}\Omega$) results in a higher phase margin. At higher currents of $I_{Load} > 3\,\text{mA}$, a lower compensation resistance ($R_C = 0.7\,\text{k}\Omega$) is preferred. Ultimately, the compensation network is adaptively modified with the load current, increasing the phase margin over a wide range of load and PVT variations, as described in Section 6.1.3.

The speed of a three current mirror OTA topology, shown in Figure 6.4, is limited by the bias current that flows into the input differential pair. To produce fast transitions at the load, a higher bias current is preferred. Alternatively, to lower power losses, the OTA should operate under low bias currents. To enhance the loop



(a)



(b)

Figure 6.7: Stability of the proposed LDO regulator as function of (a) load current I_{Load}, and (b) compensation accuracy.

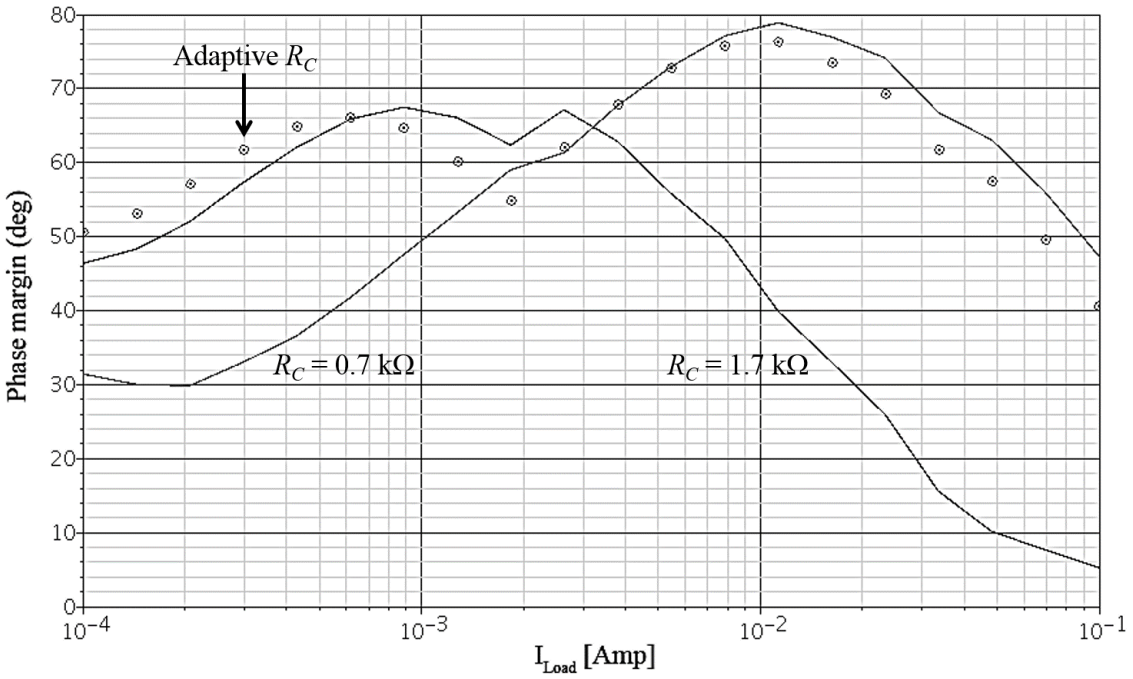


Figure 6.8: PM with different constant compensation resistors and adaptive compensation.

response while mitigating power dissipation, an adaptive bias is employed in the proposed power delivery system, as described in Section 6.1.2.

Distributed power delivery is exploited in the proposed power delivery system to regulate power close to the load, mitigating variations within the power distribution network. Scalability of the proposed power delivery system with the number of distributed LDO regulators is discussed in Section 6.1.4.

6.1.2 Adaptive bias

A self-adaptive bias current mechanism is described in this section that temporarily boosts the bias current to mitigate fast fluctuations while lowering power losses. The proposed current boost circuit is composed of a sensor block that follows the output voltage at the drain of transistor M_P, and a current boost block that controls the current through the differential pair, as shown in Figure 6.9. The current boost

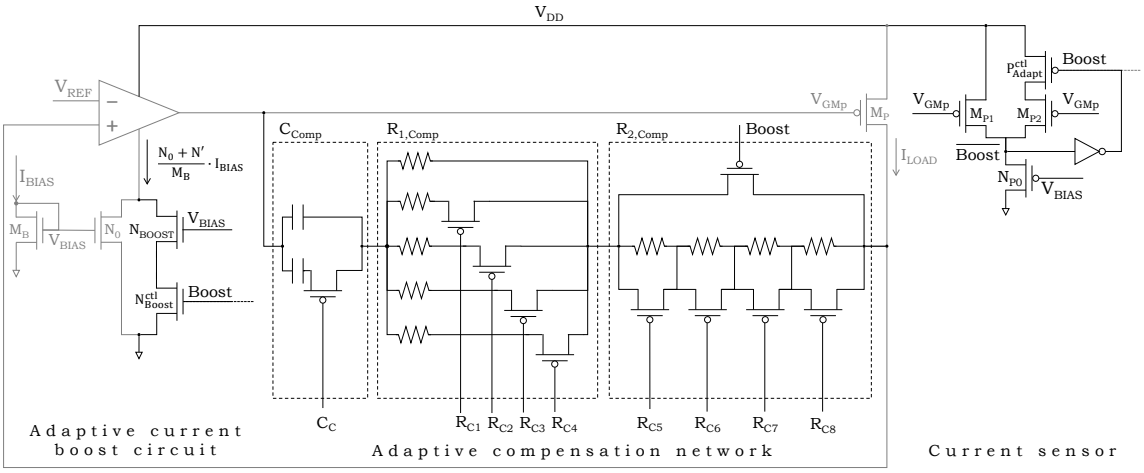
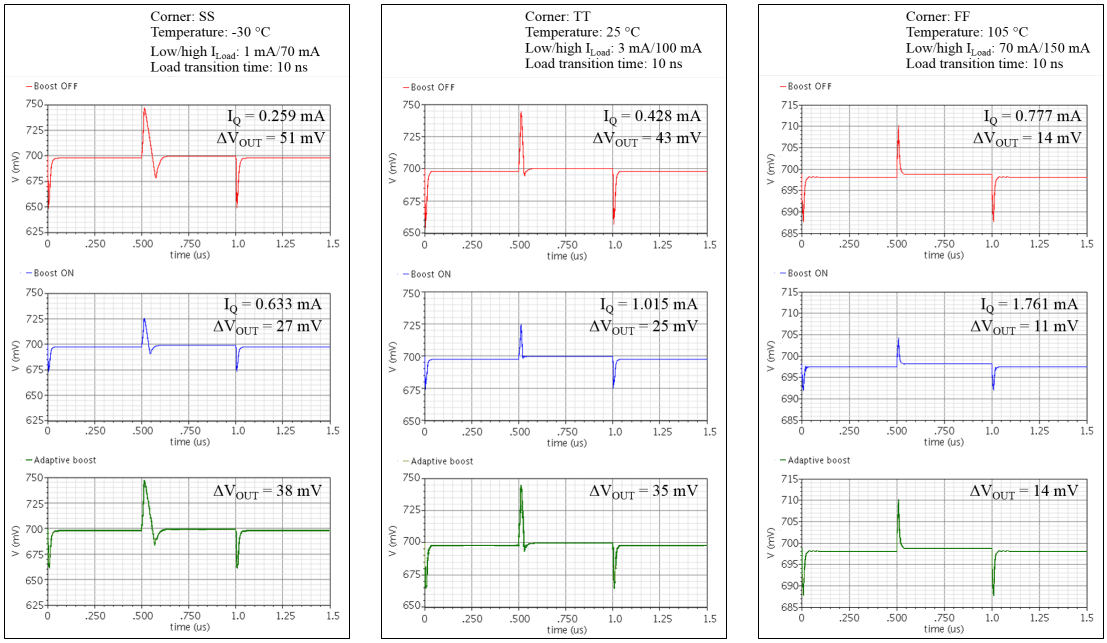


Figure 6.9: Adaptive bias boost and compensation networks.

transistor N_{Boost} is connected in parallel with the bias transistor N_0, and controlled by the Boost line. During the boost mode of operation (the Boost voltage is high), the current into the differential pair is raised, reducing the time response of the LDO. Alternatively, during regular mode (the Boost voltage is low), transistor N_{Boost} is off

and no additional current flows into the differential pair, enhancing the power efficiency of the LDO. The Boost line is controlled by the sensor block. The voltage at the $\overline{\text{Boost}}$ node follows the output voltage. When the output voltage drops, the voltage on the Boost line in the sensor block increases. The boost mode is therefore activated during the low-to-high load current transition, exhibiting faster regulation and a lower voltage droop at the output of the LDO. At other times, the current boost circuit is deactivated, increasing the power efficiency of the LDO.

To evaluate the performance of the proposed adaptive biasing technique, the load current is switched in 10 ns from 1 mA to 70 mA, from 3 mA to 100 mA, and from 70 mA to 150 mA at, respectively, the slow, typical, and fast corners. The voltage droop ΔV_{OUT} and quiescent current I_Q are recorded for three different adaptive modes. In the first mode, the current boost mechanism is disabled. In the second mode, the bias current is increased at a constant rate. In the third mode, the bias current is adaptively boosted up during the load transitions, and lowered during the steady state operation. Simulation results for each of the modes are shown in Figure 6.10 for all three corners. Due to enhanced biasing, the voltage droop is decreased by 35% (from 43 mV to 25 mV) at the expense of a significant increase of 137% in current consumption. In the proposed network, the bias current is adaptively enhanced under light loads, exhibiting an 18.6% decrease in voltage droop while



(a) (b) (c)

Figure 6.10: Voltage droop and quiescent current with and without adaptive biasing at (a) slow corner, (b) typical corner, and (c) fast corner.

avoiding excessive power loss over time.

The proposed power delivery system is designed for modern high performance circuits that draw significant leakage current from the power regulators. Minimum load currents of 1 mA, 3 mA, and 70 mA are assumed for a single LDO regulator for, respectively, the slow, typical, and fast corners. Quiescent current simulations for different load currents are listed in Table 6.2 with and without adaptive biasing. Without adaptive biasing, an average quiescent current of 423 µA with less than 2%

Table 6.2: Quiescent current with and without adaptive biasing.

I_{Load} [mA]	−30 °C		25 °C		125 °C	
	Adapt. I_Q [mA]	Const. I_Q [mA]	Adapt. I_Q [mA]	Const. I_Q [mA]	Adapt. I_Q [mA]	Const. I_Q [mA]
0.5	0.841	0.347	1.001	0.416	1.223	0.524
1.0	0.844	0.35	1.005	0.419	1.244	0.529
1.5	0.359	0.359	1.008	0.421	1.253	0.532
2.0	0.353	0.353	0.433	0.433	1.258	0.538
2.5	0.352	0.352	0.424	0.424	0.554	0.554
5.0	0.353	0.353	0.424	0.424	0.543	0.543
100	0.353	0.353	0.424	0.424	0.543	0.543

variations is demonstrated at 25 °C for all load currents. This current is increased to 1 mA by the adaptive biasing at light loads of less than 1 mA, 1.5 mA, and 2 mA at, respectively, the slow, typical, and fast corners.

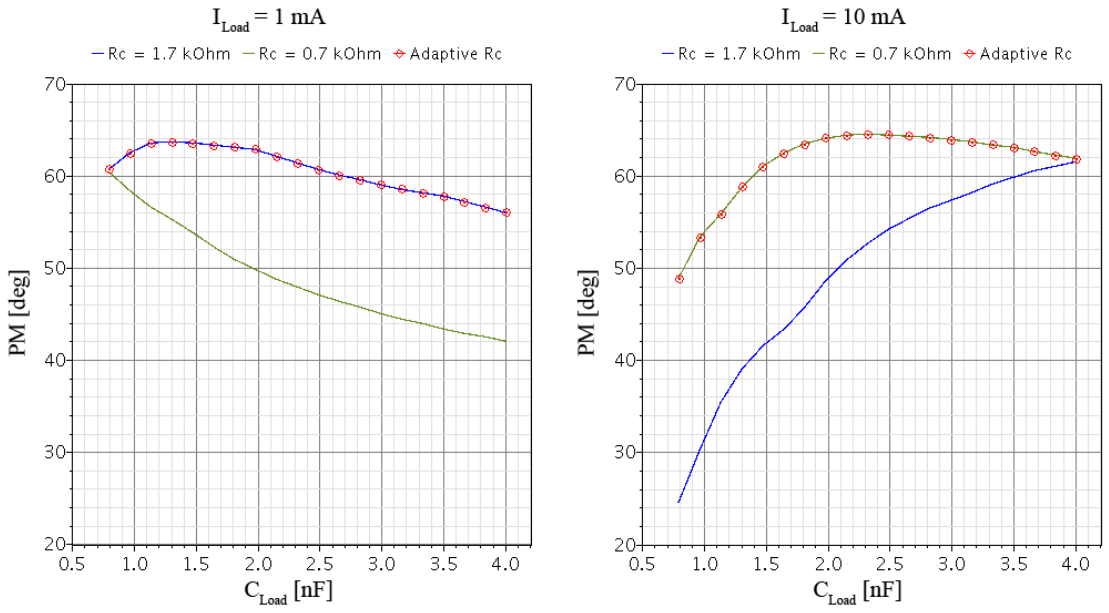
6.1.3 Adaptive compensation network

The large gate capacitance of the pass transistor together with the wide range of possible values of C_{Load} produce a complicated transfer system of poles and zeros. The low frequency non-dominant poles within the unity gain frequency of the feedback loop create a negative phase shift, degrading the stability of the overall system. To compensate for the negative phase shift, the Miller compensation technique [66] is used. However, the wide range in load capacitance and currents, and signicant PVT variations make compensation with fixed RC values impractical in

nanoscale technologies. A digitally configurable compensation network is therefore used that adaptively compensate a low frequency non-dominant pole, maintaining system stability for all values of C_{Load} and I_{Load}. This compensation network is particularly important to maintain stability when multiple LDO regulators are connected in parallel to the same power grid. Conventional LDO regulators without the proposed compensation network can easily become unstable from device mismatch, offset voltage, and varying load current when connected in parallel [155], [157], [159]. The compensation network is comprised of a capacitive block connected in series with two resistive blocks, as shown in Figure 6.9. The capacitive (C_{Comp}) and resistive blocks ($R_{1,Comp}$ and $R_{2,Comp}$) are digitally controlled by, respectively, the control signals C_C and RC_i, i = 1, ..., 8. These RC impedances are digitally configured after fabrication to mitigate any process variations. The second resistive block is also controlled by the Boost signal, which is adaptively activated (bypassed) when the Boost signal is high (low). During the high-to-low current load transition, the output impedance increases. Thus, the pole introduced by the load is pushed to a lower frequency and is no longer optimally compensated by the system zero, degrading the stability of the LDO. Alternatively, the Boost signal is activated during this transition, increasing the compensation impedance, $(R_{1,Comp} + R_{2,Comp}) \cdot C_{Comp}$. As a result, the zero introduced by the compensation network is also pushed

to a lower frequency, maintaining a stable response. At other times, the Boost signal is deactivated, and the LDO is stabilized with $\mathrm{R_{1,Comp}} \cdot \mathrm{C_{Comp}}$.

To illustrate the effect of the compensation on the LDO performance, the phase margin of a single LDO is presented in Figure 6.11 over a range of C_{Load} values for two load currents, $I_{Load} = 1\,\mathrm{mA}$ and $I_{Load} = 10\,\mathrm{mA}$. For a light load current of



(a) (b)

Figure 6.11: Phase margin with different compensation and load capacitance for (a) $I_{Load} = 1\,\mathrm{mA}$, and (b) $I_{Load} = 10\,\mathrm{mA}$.

1 mA, the phase margin increases with higher compensation resistance $\mathrm{PM}(R_C = 1.7\,\mathrm{k\Omega}) > \mathrm{PM}(R_C = 0.7\,\mathrm{k\Omega})$. Alternatively, for a higher load current of 10 mA, a smaller compensation resistor is preferable. The proposed adaptive compensation

illustrated in Figure 6.11 exhibits a higher phase margin as compared with non-adjustable compensation. The same behavior can be observed in Figure 6.8, where the compensation network is adaptively reconfigured as a function of the load current, yielding the largest PM as compared to non-adjustable compensation networks.

6.1.4 Distributed power delivery

A model of the distributed power delivery system with k LDO regulators is shown in Figure 6.12. The LDO output devices are connected in parallel at the output node

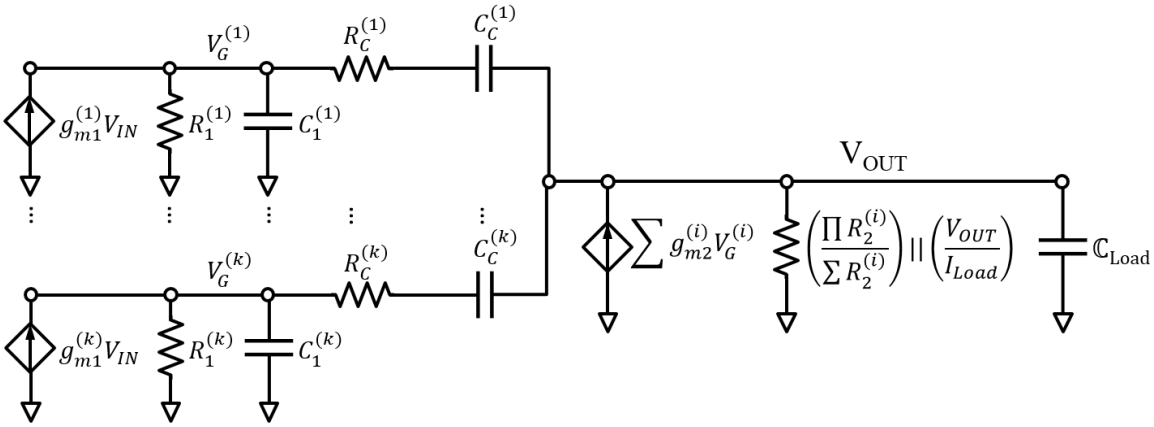


Figure 6.12: Small signal linear model of distributed power delivery system with k LDO regulators.

loaded by $k \cdot C_{Load}$, and are driven by the total current from the individual error amplifiers. With equally shared load current $k \cdot I_{Load}$, all of the distributed LDO regulators exhibit similar behaviour, yielding the simplified model shown in Figure 6.13 with $g_{m1,2}^{(i)} = g_{m1,2}^{0}$, $R_{1,2,3}^{(i)} = R_{1,2,3}^{0}$, and $C_{1,2,3}^{(i)} = C_{1,2,3}^{0}$, $\forall i = 1, ..., k$. The phase

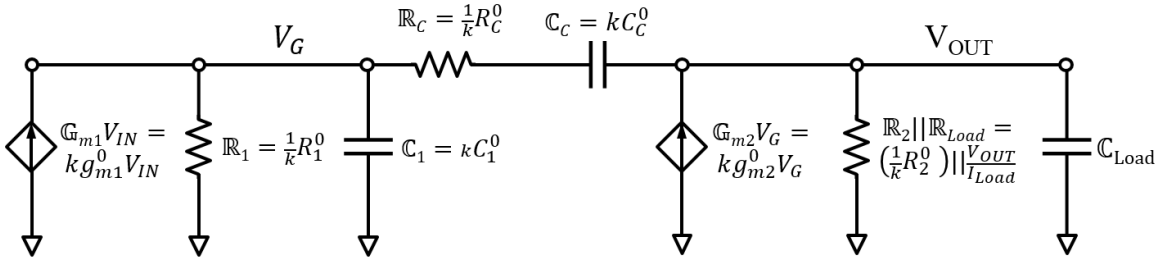


Figure 6.13: Small signal linear model of distributed power delivery system with k LDO regulators and equally shared load current.

margin of a distributed power delivery system with k LDO regulators and equally shared load is determined from (6.13) by

$$\begin{aligned} PM(f(z)) - PM(f(p_2)) &\propto \left|\log\left(\frac{\mathbb{G}_{m2}\mathbb{R}_C\mathbb{C}_C}{\mathbb{C}_{Load}}\right)\right| \\ &= \left|\log\left(\frac{g_{m2}^0 R_C^0 C_C^0}{C_{Load}^0}\right)\right|, \end{aligned} \tag{6.14}$$

exhibiting similar behavior to a power delivery system with a single LDO regulator. Note that increasing a high load current by a factor of k in a single LDO system with load capacitance C_{Load} lowers the phase margin of the system by $\log\left(\sqrt{k}\right) = (1/2)\log(k)$. Alternatively, the same increase in load current in a distributed power delivery system with k LDO regulators and similar load capacitance C_{Load} lowers the phase margin by $\log(k)$. Stability over a wide range of load currents is, therefore, more challenging with parallel load regulation. In addition, the stability of a distributed power delivery system is limited by the lowest PM among all of the LDO

regulators, exhibiting a strong function of the load current sharing. Under load current variations, the load at a single LDO regulator can be n times lower/higher than the average load current, decreasing/increasing the output transconductance g_{m2} by a factor of $\sqrt{n}$. The worst case stability of a distributed power delivery system under load current variations is, therefore,

$$PM(f(z)) - PM(f(p_2)) \propto \left|\log\left(\frac{g_{m2}^0 R_C^0 C_C^0}{C_{Load}^0}\right)\right| + \frac{1}{2}\log(n). \tag{6.15}$$

Based on typical load current variations shown in Figure 6.2 and the linear curve fitting in Figure 6.7(b), the proposed distributed power delivery system can exhibit current sharing variations of up to $n = 2$, yielding $37.2\log\sqrt{2} = 5.6°$ degradation in phase margin. To address the worst case current sharing variations and a wide range of PVT variations, each LDO regulator is optimally compensated around $I_{Load} = 10\,\text{mA}$ with $R_C C_C = 700\,\Omega \cdot 6\,\text{pF}$ to provide a stable response with $40° < PM < 70°$ for high load currents of $10\,\text{mA} < I_{Load} < 150\,\text{mA}$. Alternatively, at low load currents, sensed by the current sensor (see Figure 6.9), the compensation is adaptively increased, enhancing the stability of the system. Due to the distributive nature of the proposed power delivery system, adaptive compensation and bias are activated individually within each LDO regulator based on the specific locally sensed load currents, providing fine grain control over the local adaptive mechanisms. The

same load sensing circuit within each LDO regulator is used to trigger both the adaptive compensation and bias mechanisms, exhibiting a more compact power delivery system.

6.2 Test Results

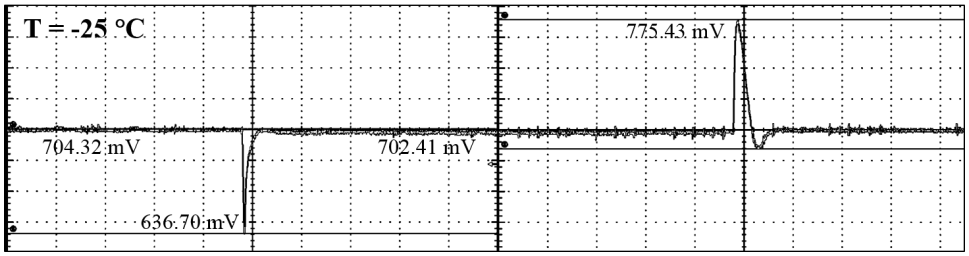
The power delivery system with six LDO regulators has been fabricated in an advanced 28 nm CMOS technology. The on-chip regulators simultaneously drive a power network, delivering power to the on-chip integrated circuits within a commercial mobile device. All of the measurements are performed on LDO regulators within the distributed power delivery system.

Modern integrated circuits exhibit aggressive transient characteristics and are expected to mitigate PVT variations. To illustrate the mitigation of voltage and temperature variations in the proposed power delivery system, the load current of the distributed power delivery system with six LDO regulators is stepped from 52 mA to 441 mA in 10 ns, drawing an average high (low) current of 73.5 mA (8.67 mA) from each LDO regulator. Due to load sharing variations (see Figure 6.2), the load current of a single LDO regulator can be increased or decreased by a factor of two (and more under PVT variations) as compared to the nominal value, exhibiting currents of up to 147 mA and down to 4.3 mA. The magnitude of these extremely fast load

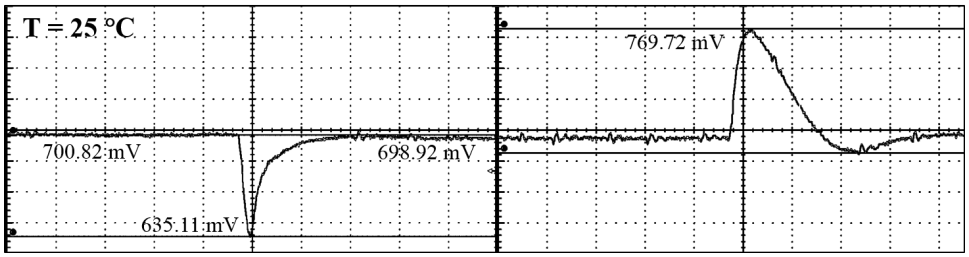
changes is therefore limited under voltage and temperature variations as compared to full range operation. The measured transient response for nominal input and output voltages of, respectively, 1.0 volt and 0.7 volts is illustrated in Figure 6.14 for $-25\,^{\circ}\mathrm{C}$, $25\,^{\circ}\mathrm{C}$, and $125\,^{\circ}\mathrm{C}$. To evaluate the proposed system under line variations, the output is tested at $25\,^{\circ}\mathrm{C}$ under $\pm 10\%$ input voltage variations. The measured transient response is illustrated in Figure 6.15. For both types of variations, the proposed distributed power delivery system exhibits a stable response over a wide range of temperatures with less than 10% voltage droop.

To demonstrate the stability of the proposed distributed power delivery system under maximum load currents and extremely fast load transitions, the load current of the system is stepped from 52 mA to 788 mA in 5 ns at $25\,^{\circ}\mathrm{C}$. The measured transient response for nominal input and output voltages of, respectively, 1.0 volt and 0.7 volts is illustrated in Figure 6.16, exhibiting a stable response and voltage droop of 0.1 volts. The system response time based on the equivalent parasitic capacitance of the load circuit ($C_{Load} = 472\,\mathrm{pF}$), maximum load current ($I_{Load,MAX} = 788\,\mathrm{mA}$), and voltage droop ($\Delta V_{OUT} = 100\,\mathrm{mV}$) is typically evaluated as $T_R = C_{TOT} \cdot \Delta V_{OUT} / I_{Load,MAX} = 0.064\,\mathrm{ns}$.

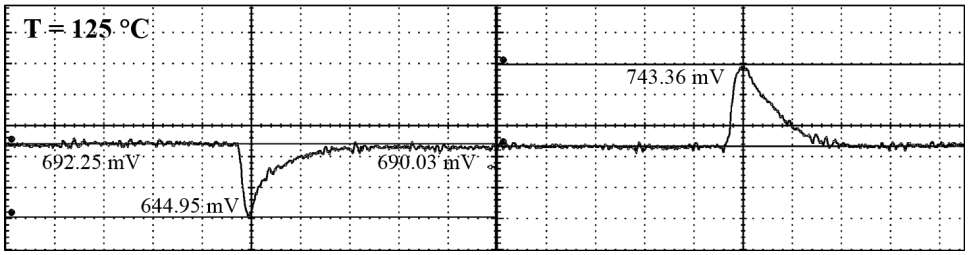
The proposed system of parallel LDO regulators yields the shortest transient response time as compared with existing LDO regulators [66]–[68], [88]. The voltage



(a)



(b)

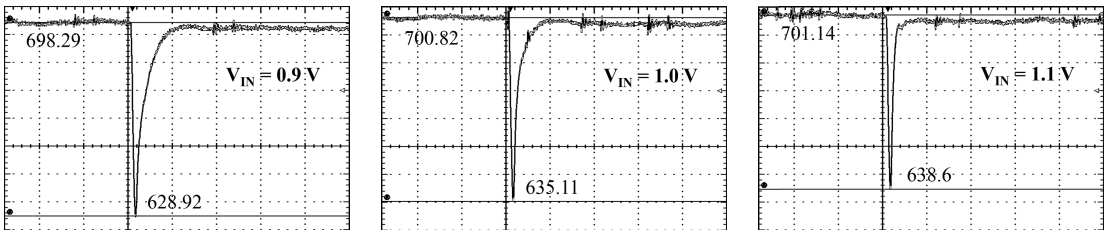


(c)

Figure 6.14: Transient step response at the load for $V_{IN} = 1$ volt, $V_{OUT} = 0.7$ volts, and load current step from 52 mA to 441 mA in 10 ns, measured at (a) $T = -25\,^{\circ}\mathrm{C}$, (b) $T = 25\,^{\circ}\mathrm{C}$, and (c) $T = 125\,^{\circ}\mathrm{C}$.

droop is a strong function of the magnitude and transition time of the load step.

Only the magnitude of the load current is, however, typically considered in T_R. For



(a) (b) (c)

Figure 6.15: Transient step response at the load at the typical temperature of 25 °C for V_{OUT} = 0.7 volts and a load current step from 52 mA to 441 mA in 10 ns, measured for (a) V_{IN} = 0.9 volts, (b) V_{IN} = 1.0 volt, and (c) V_{IN} = 1.1 volts.

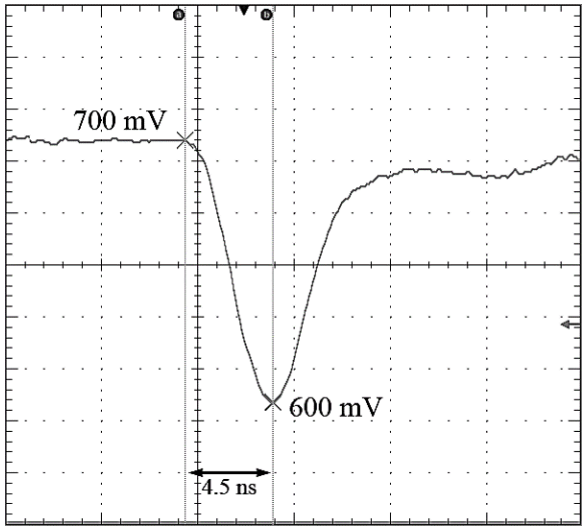


Figure 6.16: Measured transient response for a load current step from 52 mA to 788 mA in 5 ns.

a fair comparison, T_R is normalized to K, the ratio between the load transition time of the LDO regulator Δt and a 1 ns transition time (K = Δt / 1 ns). In addition, the response time of the LDO regulator is normalized to an estimated fan-out of four FO4 delay (TG), canceling the advantages of technology scaling. The LDO regulators

[66]–[68], [88] and the proposed power delivery system are compared based on the normalized, technology independent loop response $(T_R)_{Norm} = T_R \times ({}^{\Delta t}/_{1ns})/T_G$. The proposed system exhibits a smaller $(T_R)_{Norm}$ than the LDO regulators described in [66]–[68], yielding a better response time to a load transition, while exhibiting a similar current efficiency (99.49% vs. 99.99% in [66]). The loop response time in these regulators is increased by the size of the off-chip capacitor (1 µF in [67], [68]) or by a small bias current [66]. Alternatively, the speedup in the loop response achieved in [88] requires a significant increase in bias current, degrading the power efficiency of the LDO regulator (94%). The response time, power efficiency, and other primary parameters of the proposed power delivery system are listed in Table 6.3 and compared to fully integrated on-chip regulators [66], [88].

The proposed power delivery system converts an input voltage between 0.9 volts and 1.1 volts into any required output voltage between 0.6 volts to 0.8 volts, while exhibiting a stable response and less than 69.4 mV voltage droop at 25 °C for all of the input and output voltages within the range and a load step from 52 mA to 441 mA in 10 ns. The voltage droop for input and output voltages of, respectively, 0.9 volts to 1.1 volts and 0.6 volts to 0.8 volts is listed in Table 6.4. Note the measured transient step response at the output of a single LDO regulator for a distributed power delivery system with the input and output voltage of, respectively, 0.9 volts

Table 6.3: Performance summary and comparison with previously published LDO regulators.

Parameters	Unit	[88] Hazucha	[66] Guo	This work
Technology	μm	0.09	0.09	0.028
Active area	mm^2	0.008	0.019	0.00357
Input voltage	volt	1.2	0.75 to 1.2	0.9 to 1.1
Output voltage	volt	0.9	0.5 to 1	0.6 to 0.8
Minimum dropout voltage	volt	0.3	0.2	0.1
Maximum load $I_{Load,MAX}$	mA	100	100	788
Load regulation	mV/mA	1	0.1	0.023 to 0.027
On-chip capacitance	pF	600	7	5.91 to 8.37
Load circuit capacitance	pF	0	50*	472*
Quiescent current I_Q	μA	6000	8	4000
Voltage droop ΔV_{OUT}	mV	90	200	100
Output transition time T_R	ns	0.54	0.114	0.064
Normalized load transition $^{\Delta t}/_{1ns}$	ns/ns	0.1	100	5
FO4 delay T_G	ns	45	45	14
Normalized response time $(T_R)_{Norm}$	ns/ns	1.2	253	23
Current efficiency	%	94.3	99.99	99.49

*Estimated based on equivalent parasitic capacitance of the load circuitry.

Table 6.4: Voltage droop for different input and output voltage levels.

V_{IN} \ V_{OUT}	0.6 V	0.7 V	0.8 V
0.9 V	64.6 mV	69.4 mV	68.6 mV
1.0 V	61.9 mV	65.9 mV	67.9 mV
1.1 V	60.4 mV	62.5 mV	67.9 mV

and 0.8 volts, exhibiting a voltage dropout of 0.1 volts and a 68.57 voltage droop at the output.

The quiescent current of the proposed power delivery system of six distributed LDO regulators is listed in Table 6.5, yielding up to 99.49% current efficiency at 25 °C. A die microphotograph of the LDO is shown in Figure 6.17. The area occupied by the LDO with all capacitors is 85 µm × 42 µm, significantly smaller than the LDO

Table 6.5: Measured quiescent current and current efficiency.

Parameter	Comment	Temperature		
		−30 °C	25 °C	105 °C
I_Q [mA]	Distributed system	3.0	4.0	7.0
	Single LDO (average)	0.5	0.8	1.17
Efficiency [%]	$I_{Load,MAX}$ = 788 mA	99.62	99.49	99.11

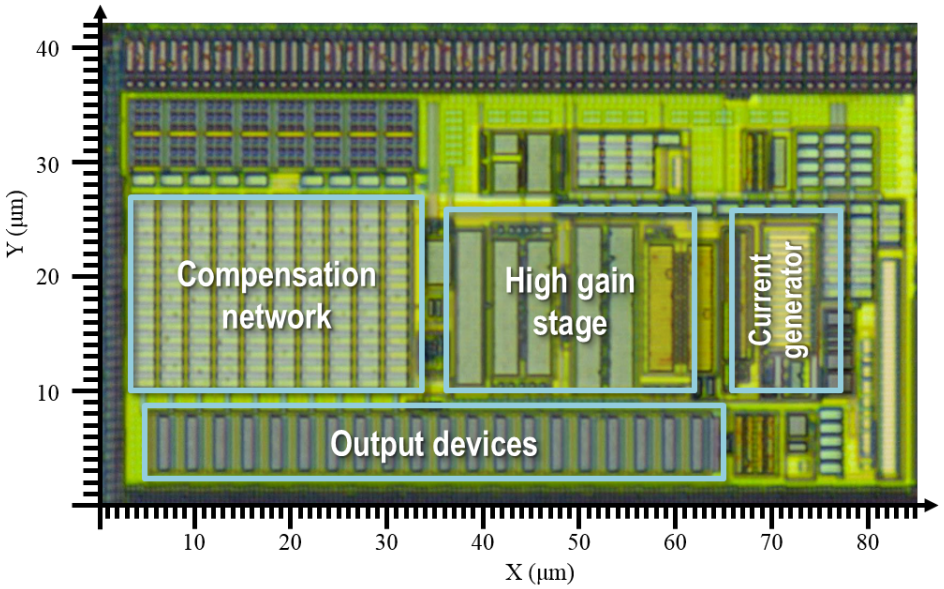


Figure 6.17: Die microphotograph of 28 nm ultra-small LDO.

regulators described in [66]–[68], [88].

6.3 Summary

A distributed power delivery system with six ultra-small fully integrated low-dropout regulators is described in this chapter. The system is fabricated in a 28 nm CMOS process and exhibits a fast transient response with excellent load regulation

under PVT and current sharing variations. An adaptive bias technique is used to enhance the transient performance and increase the power efficiency by, respectively, boosting and decreasing the bias current. A voltage droop of less than 10% and a current efficiency of 99.49% are measured. In addition, an adaptive compensation network is employed within the power delivery system that allows the co-design of a system of distributed parallel LDO regulators, yielding a stable system response within $-25\,^{\circ}\mathrm{C}$ to $105\,^{\circ}\mathrm{C}$ and 10% voltage variations. The system is believed to be the first successful silicon demonstration of stable parallel analog LDO regulators. Each of the LDO regulators within the adaptive networks and bias current generator occupies $85\,\mu\mathrm{m} \times 42\,\mu\mathrm{m} = 0.003\,57\,\mathrm{mm}^2$. No off-chip capacitors are required.

Chapter 7

Digitally Controlled Pulse Width Modulator for On-Chip Power Management

Voltage controlled oscillators (VCOs) are widely used to generate a switching signal where certain characteristics of this signal can be controlled. These controllable switching signals can be efficiently exploited in different switching and linear regulators to adaptively scale voltages on-chip, providing an important circuit level means for DVS systems. Two types of VCOs are primarily used in high performance integrated circuits; inductor-capacitor (LC) oscillators and ring oscillators. LC oscillators can operate at high frequencies and exhibit superior noise performance. Alternatively, ring oscillators occupy significantly smaller on-chip area with a wider tuning range. Due to these advantages, ring oscillators have found widespread use in modern ICs [82], [161]–[165].

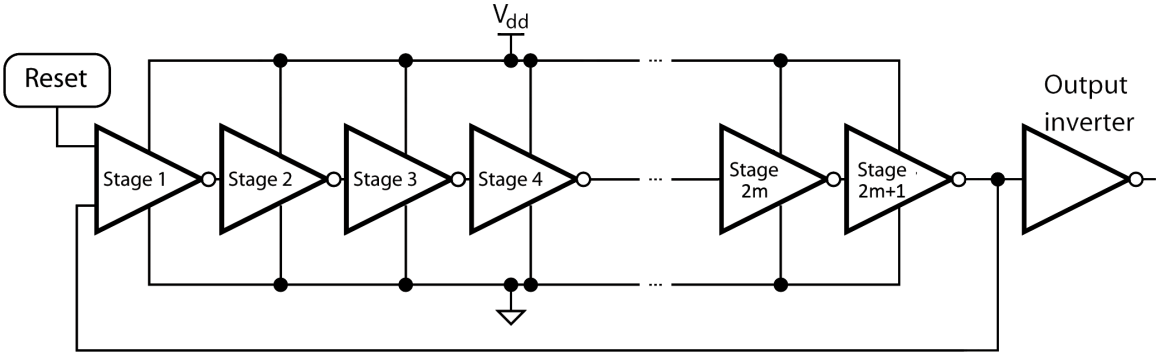


Figure 7.1: Conventional ring oscillator. Note that an odd number of inverters is required for the system to oscillate.

A conventional ring oscillator consists of an odd number of inverters where the output of the last inverter is fed back to the input of the first inverter, as shown in Figure 7.1. The delay provided by each inverter in this chain produces a phase shift in the switching signal. The sum of these individual delays (*i.e.*, phase shifts) and the feedback from the last to the first inverter produces a total phase shift of π that causes the circuit to oscillate. The frequency of this oscillation depends upon the sum of the inverter delays within the chain [166].

The duty cycle of the generated switching signal is typically 50% for conventional ring oscillators where the PMOS and NMOS transistors within the inverters provide the same rise and fall transition times. The duty cycle of a ring oscillator can be tuned by controlling the transition time of the inverters within the ring oscillator. Header and footer circuits are widely used to control the current supplied to the PMOS and NMOS transistors within the ring oscillator inverter chain [167]. Although the

header and footer circuits are typically used to control the frequency, these circuits can also control the duty cycle of a ring oscillator.

In this chapter, a digitally controlled pulse width modulator (PWM) comprised of a header circuit, ring oscillator, and duty cycle to voltage (DC2V) converter is described [165], [168]. The duty cycle of the PWM is determined from proposed closed-form expressions, yielding a simple dependence on the header current. The high accuracy of the proposed expressions is confirmed by simulation results. The header circuit controls the amount of current delivered to the PMOS transistors within the ring oscillator. Contrary to conventional header circuits, where the header is connected to each of the inverters within the ring oscillator chain, the proposed header circuit is connected to every other inverter stage to dynamically control the pulse width of the output signal. This header circuit provides a high granularity duty cycle control with a step size of 2% of the period. An additional header circuit regulates the supply current delivered to the remaining inverter stages, providing improved control while maintaining a constant switching frequency. Additionally, a DC2V converter, based on the frequency to voltage converter proposed in [169], maintains the accuracy of the PWM under process, voltage, and temperature (PVT) variations. Under PVT variations, the maximum change in duty cycle is less than 2.7% of the period.

Owing to the small on-chip area, fast control circuitry, high accuracy under PVT variations, and dynamic duty cycle and frequency control governed by accurate closed-form expression, the proposed PWM is an effective circuit to dynamically change the duty cycle of the input switching signal for on-chip voltage regulators. This circuit enables high granularity dynamic voltage scaling (DVS) at run time and reduces the response time from milliseconds to nanoseconds.

The remaining part of the chapter is organized as follows. The proposed PWM architecture is described in Section 7.1, where the working principle of the header circuitry and DC2V converter is explained, and the analytic expressions for the proposed PWM timing parameters are provided. In Section 9.2, the functionality and accuracy of the proposed circuit under PVT variations are validated with predictive technology models at the 22-nm technology node. Some concluding remarks are offered in Section 7.3.

7.1 Description of the proposed PWM architecture

A schematic of the proposed PWM is shown in Figure 7.2. A header circuit is connected to the ring oscillator to current starve every other stage in the ring

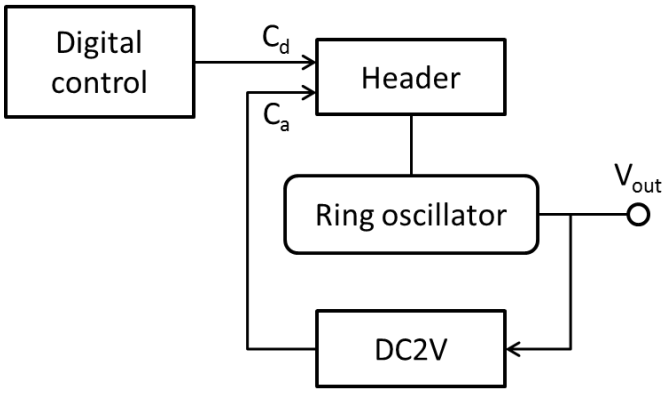


Figure 7.2: Proposed PWM. The header circuitry has two input control signals, digital control (C_d) and analog control (C_a). C_d is used to dynamically change the individual transistors to provide a high granularity duty cycle control whereas C_a maintains a constant current from the header to the ring oscillator under PVT variations.

oscillator chain. Digital control circuitry provides multiple control signals (C_d) to dynamically change the duty cycle, and a DC2V converter ensures the accuracy of the duty cycle under PVT variations by providing an analog signal to the header circuit. The working principles of these circuits are explained in the following subsections.

7.1.1 Header circuitry

An addition based current source, as shown in Figure 7.3, has been proposed in [170]. This circuit is used as a header in [171] to compensate for temperature and process variations by maintaining a constant current to the ring oscillator. Note that this header circuit has one input voltage that controls the gate voltage of M_1 and M_2. Alternatively, the gate voltage of M_3 is controlled by the drain of M_2. The

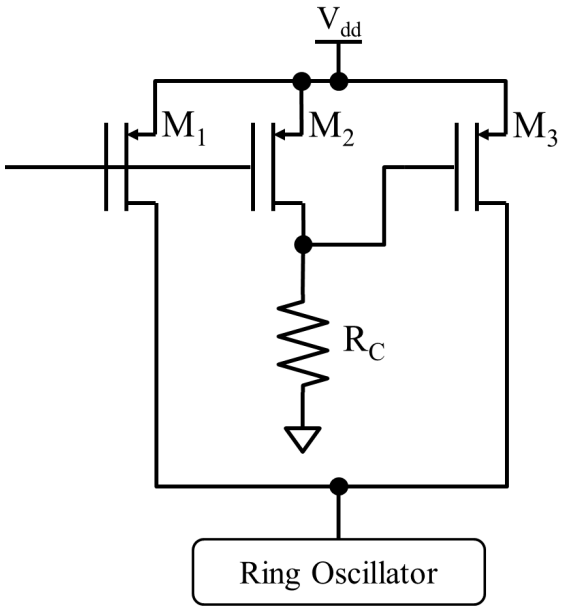


Figure 7.3: Addition based current source used as a header circuit [171].

amount of current flowing through M_1 (and M_2) is therefore inversely proportional to the amount of current through M_3. For example, if current through M_1 (and M_2) increases, the gate voltage of M_3 also increases, decreasing the current through M_3. The sum of these inversely proportional currents is the input current to the ring oscillator, which is approximately constant over a wide range of temperature and process variations [171].

Alternatively, in this chapter, a modified version of this header circuit, as depicted in Figure 7.4, is introduced to control the duty cycle by changing the transition time of the PMOS transistors at every other inverter stage within the ring oscillator. Gates M_1 and M_2 are controlled by the analog signal C_a. As opposed to a single transistor

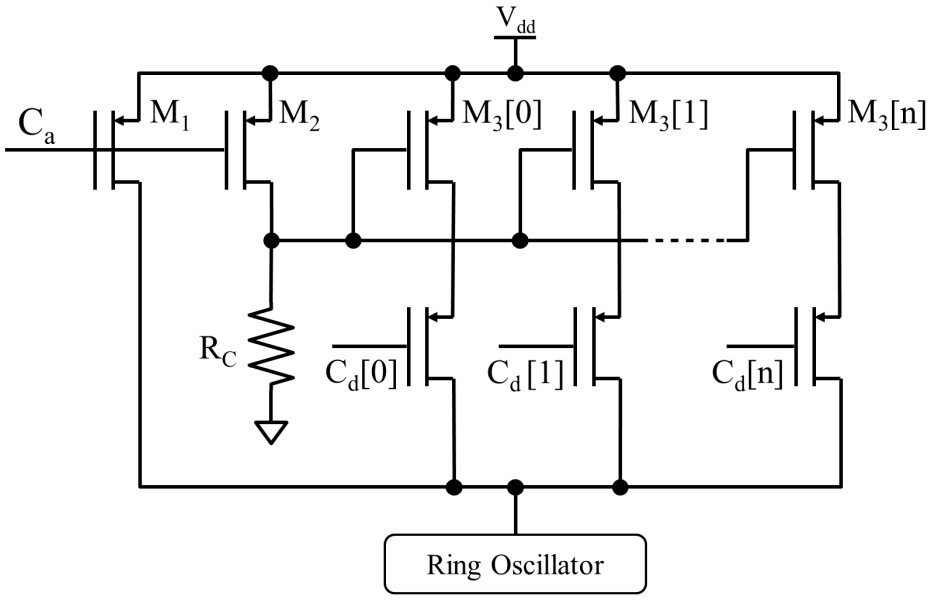


Figure 7.4: Parallel PMOS transistors replace M_3 to improve the granularity of the current control as well as behave as switch transistors to turn on different sections of the header circuitry.

M_3 whose gate is connected to a resistor as shown in Figure 7.3, multiple parallel PMOS transistors $M_3[i]$, $i = 0, \ldots, n$ are added in place of M_3 in the proposed header circuit. The PMOS transistors are designed with increasing device size to provide both increased dynamic range and dynamic control of the duty cycle with 2% increments. All of these transistors have the same gate-to-source voltage, but the voltage at the drain terminals is controlled by other switch transistors. Additional PMOS transistors are connected in series with these transistors as switch transistors. The gate voltage of these switch transistors is controlled by a digital controller that turns on (and off) the individual header stages through control signals $C_d[i]$, $i =$

$0, \ldots, n$. Turning on all of the header stages produces the maximum current to the ring oscillator which in turn minimizes the duty cycle. The variations in the leakage current is more prominent when the device size is small. To mitigate these variations, larger than minimum size transistors are used in sub-65 nm technology nodes [172]. The first two transistors in the header circuit (M_1 and M_2) are therefore comparably large to minimize any mismatches. A minimum channel length of 150 nm is used for these two input transistors as opposed to 40 nm for the remaining transistors.

7.1.2 Duty cycle to voltage converter

The frequency to voltage converter proposed in [169] operates as a DC2V converter. A circuit schematic of this DC2V converter is shown in Figure 7.5. There are primarily three different phases of this circuit. During the first phase, capacitor C_1 is charged through transistor P_1. In the second phase, transistors (*i.e.*, switches) N_2 and N_3 are turned on to allow charge sharing between C_1 and C_2. During the last phase, C_1 is discharged through N_1. The charge time of C_1 depends upon the duty cycle of the input switching signal. A signal with a greater duty cycle causes more charge to accumulate on C_1, increasing the output voltage of the DC2V converter. The proposed DC2V converter controls the bias current from the header circuitry through negative feedback, mitigating PVT variations. Intuitively, when the header

current is reduced, the duty cycle of the ring oscillator is greater, increasing the output voltage of the DC2V converter. As a result, the voltage at the gate of the M_2 transistor increases and the current I_C through resistor decreases, pulling down the gate voltage of the active header stages. Thus, the current flow through the header to the ring oscillator is increased, compensating for the initial reduction in current. A more complete explanation of the working principles of this circuit as well as the logic controller block is available in [169].

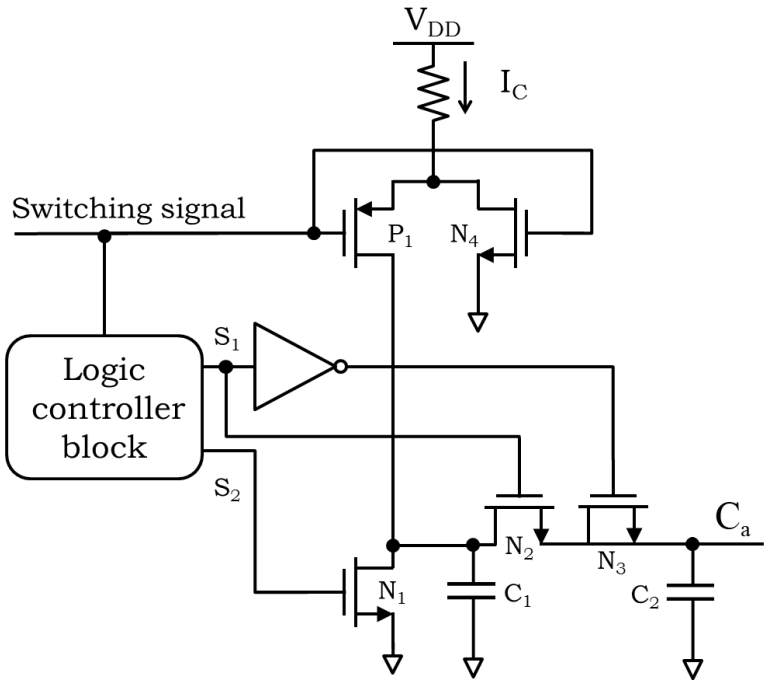


Figure 7.5: Frequency to voltage converter proposed in [169] operating as a duty cycle to voltage converter.

7.1.3 Ring oscillator topology for pulse width modulation

To create a single low-to-high oscillation at the output of the ring oscillator, the signal propagates twice through the entire ring oscillator stages. During the first pass, the PMOS transistor in the odd stages (P_{odd} transistors) and the NMOS transistor in the even stages (N_{even} transistors) are active, contributing to the T_{high} delay of the switching signal at the output of the ring oscillator. Alternatively, during the second round, the PMOS transistor in the even stages (P_{even} transistors) and the NMOS transistor in the odd stages (N_{odd} transistors) are active, determining the T_{low} delay. A periodic signal that switches between zero and 1 volt with duty cycle D and constant frequency $1/P$ is considered in this chapter. The period and duty cycle of a switching signal are defined, respectively, as

$$P \equiv T_{high} + T_{low}, \tag{7.1}$$

$$D \equiv \frac{T_{high}}{T_{high} + T_{low}}. \tag{7.2}$$

All of the MOSFET transistors are designed with similar rise and fall transition times, contributing equally to the high and low portions of the output signal. The half period of a conventional 50% duty cycle ($T_{0,high} = T_{0,low}$) ring oscillator with

$2m + 1$ stages is therefore

$$T_0 = T_{0,high} = T_{0,low} = (2m + 1)\frac{C_G \Delta V_{out}}{I_{ave}}, \tag{7.3}$$

where C_G is the input gate capacitance of the next stage, ΔV_{out} is the voltage change at the output during a single rise/fall transition, and I_{ave} is the average current flowing through a single stage active transistor. The period and duty cycle of a conventional ring oscillator are, respectively, $P_0 = 2T_0$ and $D_0 = 1/2$.

The time required to charge the output capacitance of each ring oscillator stage, $C_G \Delta V_{out}/I_{ave}$, depends directly on the current flowing through the stage, affecting the response of the following stage and the frequency of the switching signal at the output. Lower (higher) than I_{ave} current through all of the P_{odd} and/or N_{even} transistors slows down (speeds up) the response of these stages, increasing (decreasing) the $T_{0,high}$ delay, duty cycle, and period of the switching signal at the output. Alternatively, current starvation (enhancement) of all of the P_{even} and/or N_{even} transistors slows down (speeds up) the response of these stages. The $T_{0,low}$ delay is therefore increased (decreased) in this configuration, decreasing (increasing) the duty cycle and increasing (decreasing) the period of the switching signal at the output. By connecting certain ring oscillator stages to a header circuit (in this case, the odd stages), the duty cycle of the ring oscillator can be controlled through starvation/enhancement

Table 7.1: Timing parameters of the current controlled ring oscillator

Controlled stages	Affected stages	Current at the starved (enhanced) stage	Transition delay at the controlled and affected stages	T_{high}	T_{low}	Duty cycle	Period
P_{odd}	N_{even}	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)	$\Uparrow$ ($\Downarrow$)	Const.	$\Uparrow$ ($\Downarrow$)	$\Uparrow$ ($\Downarrow$)
N_{even}	P_{odd}	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)	$\Uparrow$ ($\Downarrow$)	Const.	$\Uparrow$ ($\Downarrow$)	$\Uparrow$ ($\Downarrow$)
P_{even}	N_{odd}	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)	Const.	$\Uparrow$ ($\Downarrow$)	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)
N_{odd}	P_{even}	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)	Const.	$\Uparrow$ ($\Downarrow$)	$\Downarrow$ ($\Uparrow$)	$\Uparrow$ ($\Downarrow$)

of the current flowing to the ring oscillator stages. The effect of current starvation/enhancement on the ring oscillator timing behavior is illustrated in Table 7.1.

Consider a ring oscillator with $2m+1$ stages with a header connected to $m+1$ P_{odd} transistors, supplying the current $I_{bias} = \alpha I_{ave}$, as shown in Figure 7.6. During the first round, only the biased P_{odd} and the affected N_{even} transistors are active, contributing, respectively, T_{bias} and $T_{bias\_affected}$ delays to the $T_{bias,high} = T_{bias} + T_{bias\_affected}$ delay at the output of the ring oscillator. The delay contribution of the $m+1$ biased stages to the ring oscillator period is

$$T_{bias} = \frac{(m+1)C_G \Delta V_{out}}{I_{bias}} = \frac{m+1}{2m+1}\frac{T_0}{\alpha}. \tag{7.4}$$

Limiting the header current ($\alpha < 1$) to the P_{odd} transistors increases the transition

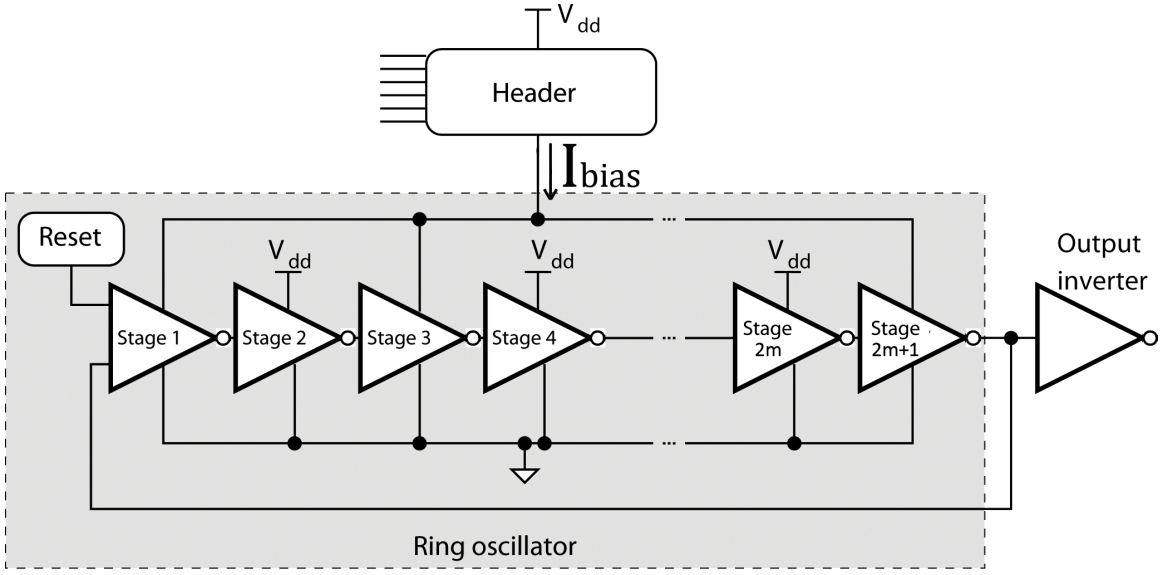


Figure 7.6: Ring oscillator with current controlled P_{odd} transistors.

delay of these stages, slowing the input transition time of the conventionally connected N_{even} transistors. Under these conditions, the conventionally connected N_{even} transistors switch more slowly. Alternatively, in those configurations where the P_{odd} transistors are enhanced ($\alpha > 1$) rather than starved, the input at the driven N_{even} transistors approaches an ideal step input, yielding faster switching of these conventionally connected stages. The delay of a conventional ring oscillator stage driven by a biased stage is inversely proportional to the bias current. The contribution of the m conventionally connected N_{even} transistors to the period of the biased ring oscillator is therefore

$$T_{bias\_affected} = \frac{m}{(2m+1)} \frac{T_0}{\alpha}. \tag{7.5}$$

During the second round, only the P_{even} and N_{odd} transistors are active, contributing to the $T_{bias,low}$ delay at the output of the ring oscillator. These transistors are not biased and are therefore unaffected by the biased stages of the ring oscillator. Thus, the $T_{bias,low}$ delay remains unchanged $T_{bias,low} = T_{0,low} = T_0$, determining the duty cycle of the proposed ring oscillator,

$$D_{bias} = \frac{T_{bias,high}}{T_{bias,high} + T_{bias,low}} = \frac{1}{1+\alpha}. \tag{7.6}$$

The period of the proposed ring oscillator is therefore

$$P_{bias} = T_{bias,high} + T_{bias,low} = T_0(1 + \frac{1}{\alpha}) = \frac{T_0}{1 - D_{bias}}. \tag{7.7}$$

The duty cycle of a biased ring oscillator is a function of the bias parameter $\alpha = I_{bias}/I_{ave}$ and does not depend on the number of stages $2m + 1$. For $\alpha = 1$, a duty cycle of 50% in (7.6) corresponds to a duty cycle of a conventional ring oscillator with balanced rise and fall times. Alternatively, the theoretical 100% duty cycle limit is achieved as $\alpha \to 0$.

The proposed approach permits configuring a ring oscillator with a wide range of duty cycles. The period of a biased ring oscillator, however, depends on α (see (7.7)) and varies with the bias current. Thus, α is constrained by the minimum and

maximum period $T_0 < P_{min} \leq P_{bias} \leq P_{max}$,

$$\frac{T_0}{P_{max} - T_0} \leq \alpha \leq \frac{T_0}{P_{min} - T_0}. \tag{7.8}$$

Note that the frequency of the switching signal generated at the output of the ring oscillator changes while varying the duty cycle of the signal. An improved version of the aforementioned ring oscillator with two header circuits is proposed in the next section to maintain a constant frequency under varying duty cycle ratios.

7.1.4 Ring oscillator topology for pulse width modulation with constant frequency

The duty cycle of a switching signal can be controlled by changing the current sourced to the odd (or even) stages of a conventional ring oscillator, as demonstrated in Section 7.1.3. The period of the switching signal in the proposed topology, however, scales with the duty cycle (7.7), affecting the operational frequency of the ring oscillator. To provide a wide range of duty cycle while maintaining a constant frequency, an additional level of control over the timing parameters of the ring oscillator is required.

Consider a ring oscillator with $2m + 1$ stages and two headers H_A and H_B that

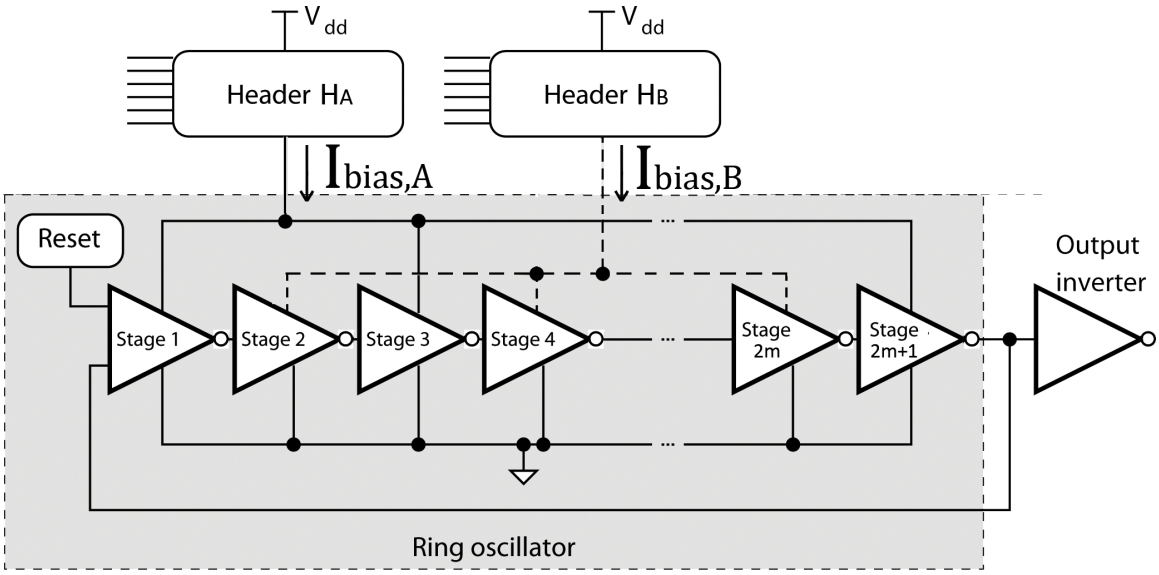


Figure 7.7: Ring oscillator with current controlled P_{odd} and P_{even} transistors.

supply, respectively, the current $I_{bias,A} = \alpha I_{ave}$ to the $m+1$ P_{odd} transistors and $I_{bias,B} = \beta I_{ave}$ to the m P_{even} transistors, as shown in Figure 7.7. The currents flowing through the P_{odd} and P_{even} transistors affect, respectively, the operating speed of the N_{even} and N_{odd} transistors, as described in Section 7.1.3. Alternatively, the P_{odd} and N_{even} transistors are active during the first pass through the ring oscillator independent of the P_{even} and N_{odd} transistors that are active during the second pass. Thus, the timing parameters of the ring oscillator shown in Figure 7.7 are similar to the parameters used in Section 7.1.3, yielding the duty cycle of the ring oscillator,

$$D_{bias} = \frac{1/\alpha}{1/\alpha + 1/\beta} = \frac{1}{1+\alpha/\beta}, \tag{7.9}$$

and period

$$P_{bias} = T_0(1/\alpha + 1/\beta). \tag{7.10}$$

To maintain a constant period, the constraint $P^{(2)}_{bias} = 2T_0$ is used in (7.10), yielding $\beta = \alpha/(2\alpha - 1)$ and therefore

$$D_{bias}\,(P = 2T_0) = \frac{1}{2\alpha}. \tag{7.11}$$

Substituting the period, (7.10), and duty cycle, (7.11), of the controlled ring oscillator in Figure 7.7, and the half period of a conventional ring oscillator with a 50% duty cycle, (7.3), the expressions for the currents $I_{bias,A}$ and $I_{bias,B}$ shown in Figure 7.7 are, respectively,

$$I_{bias,A} = \frac{I_{ave}}{2D}, \tag{7.12}$$

$$I_{bias,B} = I_{bias,A} \cdot \frac{D}{1-D} = I_{ave} \cdot \frac{1}{2(1-D)}. \tag{7.13}$$

Thus, to design a switching signal with a specific duty cycle D and period P, the proposed ring oscillator topology shown in Figure 7.7 should be used with the bias currents $I_{bias,A}$ and $I_{bias,B}$ described by, respectively, (7.12) and (7.13). The currents $I_{bias,A}$ and $I_{bias,B}$ are generated independently, and produce a variation insensitive duty cycle and frequency with properly compensated currents. A constant duty

cycle under PVT variations is therefore a useful indicator for PVT mitigation in the proposed PWM.

7.2 Simulation results

A seven stage ring oscillator is described in this chapter to provide a switching signal with a wide range of duty cycles. The proposed circuit is designed in a 22 nm CMOS predictive technology model (PTM) [173]. Certain parameters in the technology model file are modified based on [174] to include process corners such as typical-typical (TT), slow-slow (SS), fast-fast (FF), fast-slow (FS), and slow-fast (SF). Simulation results characterizing the accuracy of the proposed PWM are shown in Section 7.2.1 for different duty cycle ratios under PVT variations. The effect of the bias current on the duty cycle of the ring oscillator output is discussed in Section 7.2.3 without constraints on the period of the output signal and in Section 7.2.2 under a constant period constraint.

7.2.1 Proposed pulse width modulator under PVT variations

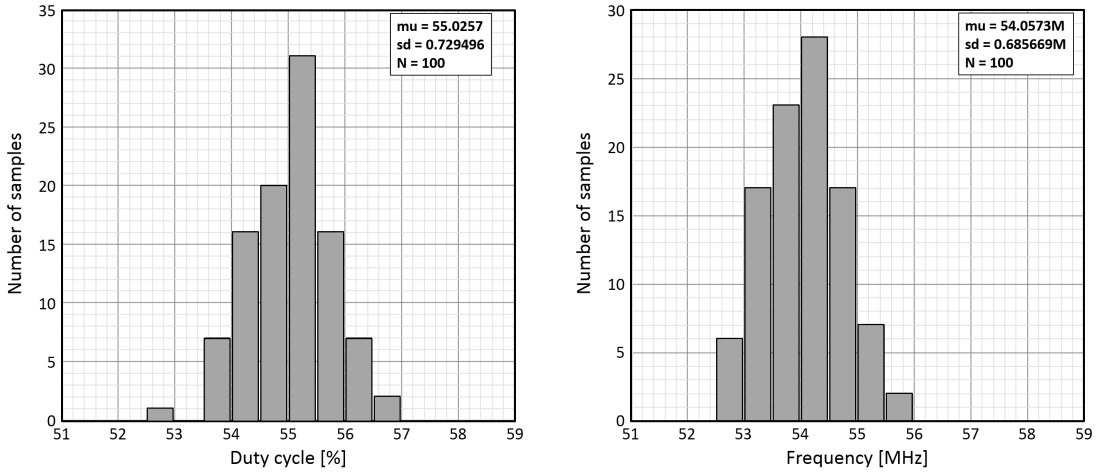
To evaluate the effect of PVT variations on the proposed PWM, the current flowing through the P_{odd} transistors in the first, third, fifth, and seventh stages is controlled by the header circuit, as shown in Figure 7.6 for $m = 3$. The remaining PMOS and NMOS transistors in this section are conventionally connected directly to, respectively, V_{dd} and ground. The supply voltage varies ±5% from the nominal 0.95 V and the temperature varies from 27°C to 80°C. The simulations have been performed for TT, SS, FF, FS, and SF process corners for the 22 nm predictive CMOS model [173]. The corners (TT, FF, SS, FS, and SF) have been generated by modifying the related parameters in the model files [174]. The per cent deviation for different duty cycle ratios is listed in Table 7.2. The deviation of the duty cycle under PVT variations is less than 2.7% of the targeted duty cycle.

The Monte Carlo simulations that consider process and mismatch variations are shown in Figure 7.8 for a duty cycle of 55% and frequency of 55 MHz, yielding a standard variation of, respectively, 0.73% and 0.69 MHz.

Transistors with smaller dimensions are more sensitive to PVT variations [175], [176] and exhibit greater leakage current variations [172] than the wider transistors. The narrower transistors within the header circuitry turn on if a switching signal

Table 7.2: Change in the duty cycle of the proposed PWM under PVT variations for the 22 nm predictive CMOS model.

V_{dd}	Process	Temperature	Duty cycle			
			55%	65%	75%	85%
1.0	TT	27	55.03	64.79	74.58	85.79
1.0	TT	80	55.13	65.00	74.02	85.08
1.0	FF	27	55.01	64.86	74.67	85.83
1.0	FF	80	55.29	65.36	74.01	84.08
1.0	SS	27	55.17	65.01	74.87	85.75
1.0	SS	80	55.15	64.88	74.57	85.07
1.0	FS	27	55.15	65.34	75.60	87.06
1.0	FS	80	55.23	65.49	75.71	86.64
1.0	SF	27	55.51	65.37	74.52	84.36
1.0	SF	80	55.51	65.25	74.01	82.85
0.9	TT	27	55.10	65.00	74.77	86.02
0.9	TT	80	55.09	64.92	74.58	85.76
0.9	FF	27	55.00	64.90	74.66	86.28
0.9	FF	80	55.10	65.06	74.74	85.57
0.9	SS	27	55.30	65.34	75.21	86.08
0.9	SS	80	55.21	65.03	74.73	85.44
0.9	FS	27	55.26	65.81	76.18	87.40
0.9	FS	80	55.25	65.68	75.97	87.16
0.9	SF	27	55.61	65.40	74.38	84.26
0.9	SF	80	55.49	65.07	73.88	83.54
Maximum variations [%]			± 0.55	± 0.78	± 1.53	± 2.68



(a) (b)

Figure 7.8: Monte Carlo simulation of (a) duty cycle, and (b) frequency distribution.

with a greater duty cycle is required. The effect of PVT variations is therefore more prominent on those signals with a wider duty cycle. This trend can be observed in Table 7.2, where the deviation for signals with a 50% duty cycle is smaller than for those signals with a 90% duty cycle.

7.2.2 Duty cycle controlled pulse width modulator

The accuracy of the analytic expressions of the duty cycle presented in Section 7.1.3 is evaluated in this section for a 25% to 90% range of duty cycle. The circuit shown in Figure 7.6 is used with an ideal current source replacing the header. The current I_{bias} flowing through the P_{odd} transistors is therefore controlled by an ideal

current source. The rise time at the output of the P_{odd} transistors degrades with a longer charge time, increasing the duty cycle of the switching signal at the output of the ring oscillator.

The expressions for the duty cycle in (7.6) are verified with simulations for bias currents between 2 and 50 μA, as shown in Figure 7.9. Note the good agreement between the theoretical and simulation results ($error < 4.4\%$). When the current through the controlled stages is neither starved nor enhanced, the simulated circuit oscillates with a 50% duty cycle, yielding $\alpha = 1$ where $I_{bias} = I_{ave} = 19\ \mu A$. Using the analytic expression in (7.6), the duty cycle can be tuned with a digitally programmable control block.

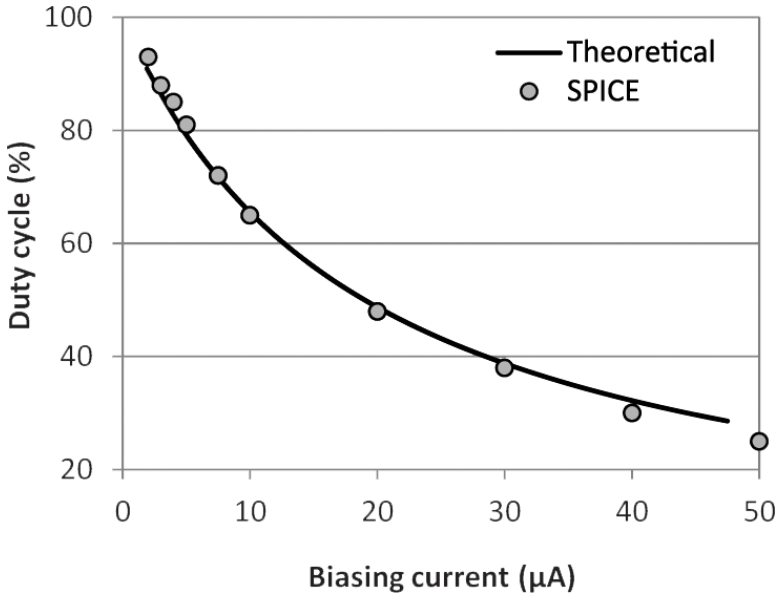
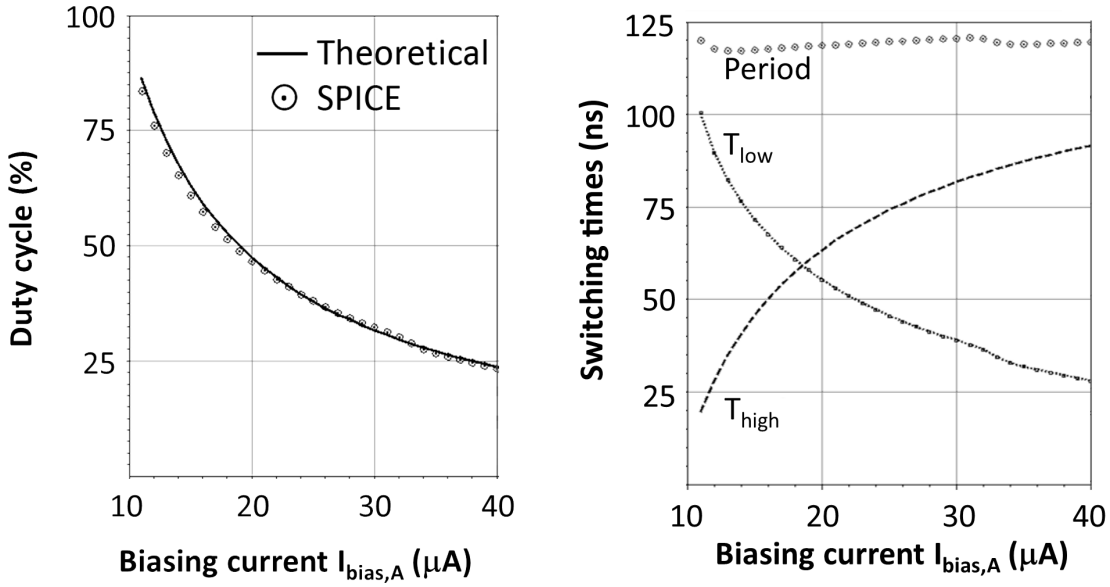


Figure 7.9: Duty cycle varies between 25% and 90% when the header current changes from 50 to 2 μA ($error < 4.4\%$).

7.2.3 Duty cycle and frequency controlled pulse width modulator

The accuracy of the analytic expressions for the duty cycle under the constant frequency constraint (see Section 7.1.4) is evaluated in this section at 8.33 MHz for a 25% to 90% range of duty cycle. The current supply to the ring oscillator is controlled with two headers, exhibiting a frequency of 8.33 MHz $\pm$ 1.25 % for all values of duty cycle. The circuit shown in Figure 7.7 is modeled by ideal current sources replacing the headers H_A and H_B. The currents $I_{bias,A}$ and $I_{bias,B}$ flowing, respectively, through the P_{odd} and P_{even} transistors are assumed to be controlled by ideal current sources. Intuitively, the rise time at the output node of the P_{odd} transistors increases with a longer charge time, increasing the duty cycle of the switching signal. To mitigate the effect of the longer $T_{bias,high}$ delay on the period of the switching signal, the rise time at the output of the P_{even} transistors, based on (7.13), is decreased. The analytic expression for the duty cycle and period, respectively, (7.10) and (7.11), is verified by the simulations for $I_{bias,A}$ currents between 11 and 40 µA, as shown in Figure 7.10. Note the good agreement between the theoretical and simulation results ($error < 3.1\%$).



(a) (b)

Figure 7.10: Header current $I_{bias,A}$ changes from 40 to 11 µA. (a) Duty cycle varies between 25% and 90% ($error < 3.1\%$). (b) Period remains approximately constant ($error < 1.25\%$).

7.3 Summary

A digitally controlled PWM with a wide pulse width ranging from 25% to 90% is proposed in this chapter. An enhanced header circuit is also presented to provide a greater range of header current. The header circuit is connected to every other stage of the ring oscillator to significantly improve the dynamic range of the pulse width. The parallel configuration of the M_3 transistors within the header circuit controls the duty cycle with high granularity. To efficiently control both the duty cycle and the

period of the switching signal, a more advanced version of the PWM is also proposed. Every other stage in this topology is controlled by a header circuit. An additional header circuit, however, is connected to the remaining ring oscillator stages to provide enhanced control over the frequency of the switching signal. A DC2V converter samples the duty cycle of the output signal and generates an analog voltage to control the header current. The PVT variations are compensated by the feedback loop generated by the DC2V converter. Under PVT variations, deviations in the pulse width are less than 2.7% of the switching signal period. Both the duty cycle and the period of the proposed circuit are analytically determined as a function of the header current, simplifying the control over the PWM timing parameters. The accuracy of the analytic expressions describing the duty cycle is evaluated with simulation results, yielding less than 3.1% and 4.4% error, respectively, with and without the controlled frequency for the pulse width modulator. A constant frequency with less than 1.25% variation is reported for different values of the duty cycle. The proposed pulse width modulator provides a means for dynamically changing the voltage in adaptive systems using fast control circuitry, providing high accuracy under PVT variations and dynamic duty cycle and period control.

Chapter 8

Stability in Distributed Power Delivery Networks

While the quality of a power supply can be efficiently addressed with a point-of-load (POL) power delivery system [69], [70], the complexity of a distributed POL power supply system is a significant design issue [164]. To cope with design complexity in complex analog circuits, automated modeling, optimization, and synthesis techniques are typically considered [177]. Automated tools for analog circuits with hundreds of components have recently become commercially available [177]. To automate the design of a power delivery system, accurate methods to evaluate performance metrics (e.g., quality of transient response, stability, and power) are required. A distributed system with multiple LDO regulators delivering power to a single grid exhibits degraded stability due to complex interactions among the LDO regulators,

power distribution network, and current loads. Thus, the stability of these parallel connected voltage regulators is a primary performance concern and should be accurately evaluated.

The stability of a single closed loop system is traditionally determined by the difference between the phase and 180° (phase margin), measured in degrees between the output and input of the open loop system. In systems with multiple dependent loops, the open loop approach is, however, not practical because there is no straightforward method to identify the unstable loop. In this work, an alternative passivity-based stability criterion (PBSC) is proposed. This criterion imposes a passivity condition on the power grid impedance. Based on the proposed criterion, accurate system requirements for exponential and marginal stability in distributed power delivery systems are provided. Automating the design flow of a power delivery system based on the proposed stability approach is demonstrated with a typical parametric circuit performance modeling technique [177]. A distributed power delivery system based on the proposed stability criterion has been fabricated in an advanced 28 nm CMOS technology, and exhibits stable voltage regulation over a wide range of temperature, voltage, and load sharing conditions.

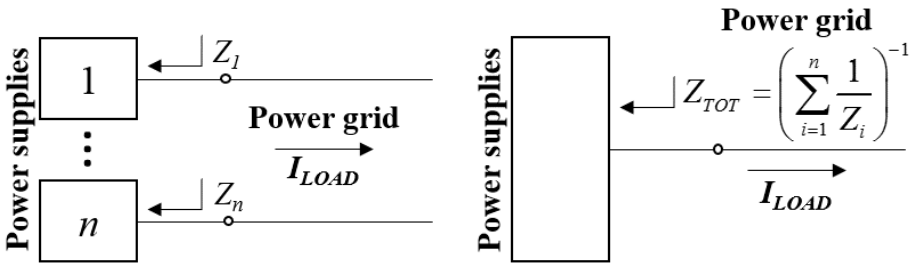
The rest of the chapter is organized as follows. The stability criterion is introduced in Section 8.1. A distributed power delivery system is evaluated in Section 8.2

based on the proposed stability criterion. Simulation results exhibit strong correlation between the stability of a distributed power delivery system and the proposed PBSC criterion. Automated design with the proposed stability criterion is described in Section 8.3. Experimental results for a distributed power delivery system with six on-chip LDOs are presented in Section 8.4. The chapter is concluded in Section 8.5.

8.1 Passivity-based stability of distributed power delivery systems

Understanding the effects of the frequency domain parameters on the time domain characteristics provides significant insight into the analysis and transient behavior of complex systems [178]–[182]. Traditionally, the phase margin of the output response ($\angle V_{OUT}/V_{IN}|_{V_{OUT}=V_{IN}}$) is used to determine the stability of a single LDO regulator. Similarly, a straightforward criterion is required for determining the stability of a distributed power delivery system.

A distributed power delivery system with $n \geq 2$ power supplies driving a single power grid is depicted in Figure 8.1(a). In this distributed system, the power supplies can be combined into a single power delivery system, yielding an equivalent single port network, as shown in Figure 8.1(b). Note that the output impedance of the



(a) (b)

Figure 8.1: Power delivery system (a) with $n \geq 2$ distributed power supplies, and (b) reduced single port network.

equivalent single port network, Z_{TOT} in Figure 8.1(b), is the parallel combination of all of the output impedances $Z_i, i = 1, \ldots, N$ of the individual power supplies shown in Figure 8.1(a). The output impedance of a distributed power delivery system is, therefore, straightforward to evaluate based on the individual output impedance of the parallel connected components. Alternatively, there is no straightforward method to identify the single loop that causes the instability in a system with multiple interacting feedback paths. The open loop transfer function, traditionally used to determine the stability of a lumped power delivery system, cannot be applied in a distributed power delivery system with multiple control loops [183]. A criterion for evaluating the stability of a multi-feedback path system composed of distributed power regulators is therefore needed.

Sufficient conditions for a stable distributed power delivery system are described

in this section. These conditions are based on the observation, proven in [184], that a linear, time-invariant (LTI) system is stable when coupled to an arbitrary passive environment if and only if the driving point impedance is a passive system. Thus, a distributed power delivery system is stable if and only if the equivalent output impedance Z_{TOT} satisfies passivity requirements. The passivity of an LTI system is described here in terms of practical frequency domain parameters.

An LTI system is passive if the system can only absorb energy, yielding, in mathematical terms [185],

$$\int_{-\infty}^{T} v(t)i(t)dt \geq 0, \forall\ T, \tag{8.1}$$

where $v(t)$ and $i(t)$ are, respectively, the voltage across the system and current flowing through the system. The total energy delivered to a passive system is determined from (8.1) based on the Parseval Theorem, exhibiting, for all positive currents,

$$\frac{1}{\pi}\int_{0}^{+\infty} Re[Z(j\omega)]|I(j\omega)|^2 d\omega \geq 0, \tag{8.2}$$

where $Z(s) = V(s)/I(s)$ is the system impedance, and $V(s)$ and $I(s)$ are, respectively, the phasor voltage and current of the system. The passivity condition based on (8.2), $\{Re[Z(\sigma + j\omega)] \geq 0, \forall \sigma > 0\}$, can be simplified, yielding the following sufficient conditions for passivity of an LTI system: $Z(s)$ has no right half plane (RHP)

poles, and the phase of $Z(s)$ is within the $(-90°, +90°)$ range $\forall\ \omega$ [186].

A distributed system is, therefore, exponentially stable (converges within an exponential envelope) if and only if the impedance of the system satisfies these passivity requirements, marginally stable (oscillates with constant amplitude) if and only if the voltage and current phasors are shifted by precisely 90°, and unstable otherwise. The phase of the output impedance is an efficient alternative to determine the stability of these distributed systems, since the traditional phase margin approach is not practical due to the multiple control loops.

8.2 Distributed power delivery systems

A distributed power delivery system is modeled and analyzed in this section based on the proposed passivity criterion. Specific circuit level requirements are determined to maintain overall system stability. The proposed small signal model is described in subsection 8.2.1. The passivity analysis is presented in subsection 8.2.2.

8.2.1 Distributed power delivery model

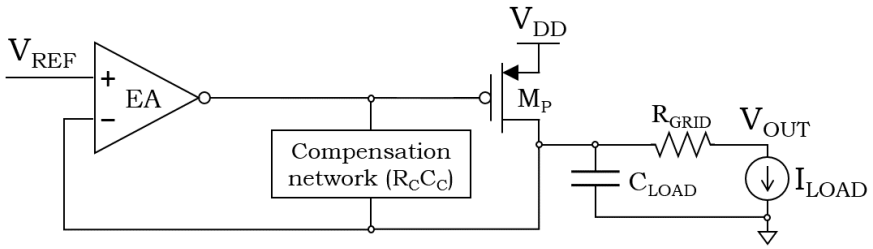
A primary parameter in determining the stability of an individual LDO is the phase margin which is a function of the gain and phase of the transfer function. Each system pole and zero introduce a phase shift of, respectively, −90° and +90°. For

example, a single pole LDO exhibits a phase margin of at least 90° and is always stable. Alternatively, a feedback based LDO regulator with two poles exhibits two negative phase shifts of 90° and potentially a step response that is only marginally stable with no phase margin. Thus, the first two uncompensated poles are the most significant poles in determining the phase margin and stability of an LDO regulator. To improve the stability and step response of an LDO regulator, RC compensation [66], [143]–[146], [187], which introduces a zero into the transfer function, is often used, virtually canceling the effect of a pole. A third order model with three poles and a single zero that cancels one of the poles is therefore a practical choice for analyzing the stability of an LDO regulator. A typical LDO regulator composed of an error amplifier (EA), output device (M_P), and compensation network $R_C C_C$ is depicted in Figure 8.2(a). A third order small signal model of the circuit is described in Chapter 6 and is shown in Figure 8.2(b). The output impedance of the LDO regulator, shown in Figure 8.2, is

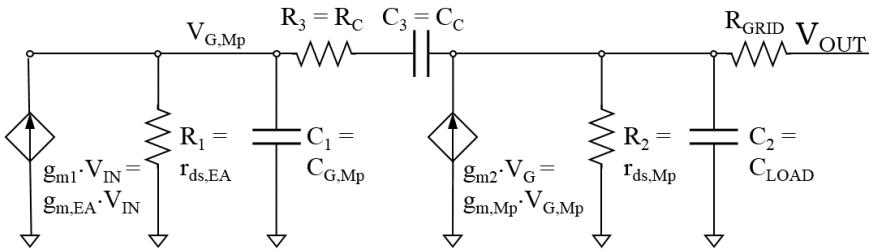
$$\begin{aligned} Z_{OUT}(s) &= \frac{R_2}{1+A_0} \cdot \frac{N(s)}{D(s)} + R_{GRID} \approx \frac{R_2}{A_0} \cdot \frac{N(s)}{D(s)} + R_{GRID} \\ &= \frac{R_2}{A_0} \cdot \frac{1 + N_1 s + N_2 s^2}{1 + \frac{D_1}{A_0} s + \frac{D_2}{A_0} s^2 + \frac{D_3}{A_0} s^3} + R_{GRID}, \end{aligned} \tag{8.3}$$

where

$$A_0 = -g_{m1} g_{m2} R_1 R_2, \tag{8.4}$$



(a)



(b)

Figure 8.2: Low-dropout regulator with $R_C C_C$ compensation, (a) topology, and (b) small signal linear model.

$$N_1 = R_1C_1 + R_3C_3 + R_1C_3, \tag{8.5}$$

$$N_2 = R_1R_3C_1C_3, \tag{8.6}$$

$$D_1 = R_1C_1 + R_2C_2 + R_3C_3 + R_1C_3 + R_2C_3(1 + g_{m2}R_1) + A_0 \cdot (R_3 - \frac{1}{g_{m2}})C_3, \tag{8.7}$$

$$D_2 = R_1R_2C_1C_2 + R_1C_1R_3C_3 + R_2C_2R_3C_3 + R_1R_2(C_1 + C_2)C_3, \tag{8.8}$$

$$D_3 = R_1R_2R_3C_1C_2C_3. \tag{8.9}$$

This output impedance exhibits three poles (p_1, p_2, and p_3) and three zeros (z_1, z_2, and z_3) and can be written as a rational function,

$$Z_{OUT} = \frac{R_2 + A_0 R_{GRID}}{A_0} \cdot \frac{(1+\frac{s}{z_1})(1+\frac{s}{z_2})(1+\frac{s}{z_3})}{(1+\frac{s}{p_1})(1+\frac{s}{p_2})(1+\frac{s}{p_3})}. \tag{8.10}$$

The current load of each regulator is a strong function of the specific location of the distributed LDO regulators and activity of the loads in the power delivery system. Certain small signal parameters are affected by the DC load current, changing the stability characteristics of each LDO regulator. For example, the output resistance ($R_2 = r_{ds,M_P}$ in Figure 8.2(b)) and transconductance ($g_{m2} = g_{m,M_P}$ in Figure 8.2(b)) of an LDO output device, respectively, decrease and increase with increasing load current, shifting the zeros and poles of individual regulators, affecting the stability of the overall system. Alternatively, other parameters that determine the operation of the error amplifier and compensation network (e.g., $R_1 = r_{ds,EA}$, $C_1 = C_{G,M_P}$, $R_3 = R_C$, and $C_3 = C_C$) are practically independent of the DC load current. Note that the zeros of the output impedance $N(s)$ are dominated by the load independent components (R_1, R_3, C_1, and C_3). Alternatively, the denominator $D(s)$ of Z_{OUT} is a strong function of the load dependent components, R_2 and g_{m2}. The zero frequencies in the proposed model are therefore less affected by load variations than the pole frequencies. For simplicity, the zeros of the output impedance Z_{OUT} are assumed to

be independent of the DC load current.

The output impedance of parallel connected voltage regulators is a primary factor in determining the stability of a distributed power delivery system, and is a strong function of the poles and zeros of the individual LDO regulators. To maintain a stable distributed system, the effect of the poles and zeros in (8.10) on the relative location of the poles and zeros in the distributed system need to be determined.

In a distributed power delivery system, the total current load is shared among all of the power supplies. Voltage regulators in close proximity with the current load supply the greatest portion of the total current, which can be significantly higher than the average current supplied by a single regulator [69]. A distributed power delivery system that addresses this issue of load sharing variations is shown in Figure 8.3. The small signal model shown in Figure 8.2(b) is assumed for each individual LDO regulator. The total load current I_{LOAD} in the proposed model is distributed based on the individual power grid resistors, $R_{GRID}^{(i)}, i = 1, \ldots, n$. The total output impedance of the system depicted in Figure 8.3 is

$$\begin{aligned} Z_{OUT}^{TOT} = Z_{OUT}^{(1)} || \ldots || Z_{OUT}^{(n)} &= \left[\sum_{i=1}^{n} \frac{A_0^{(i)}}{R_2^{(i)} + A_0^{(i)} R_{GRID}^{(i)}} \cdot \frac{D(s)^{(i)}}{N(s)^{(i)}} \right]^{-1} \\ &= N(s) \cdot \left[\sum_{i=1}^{n} \frac{A_0^{(i)} D(s)^{(i)}}{R_2^{(i)} + A_0^{(i)} R_{GRID}^{(i)}} \right]^{-1}, \end{aligned} \tag{8.11}$$

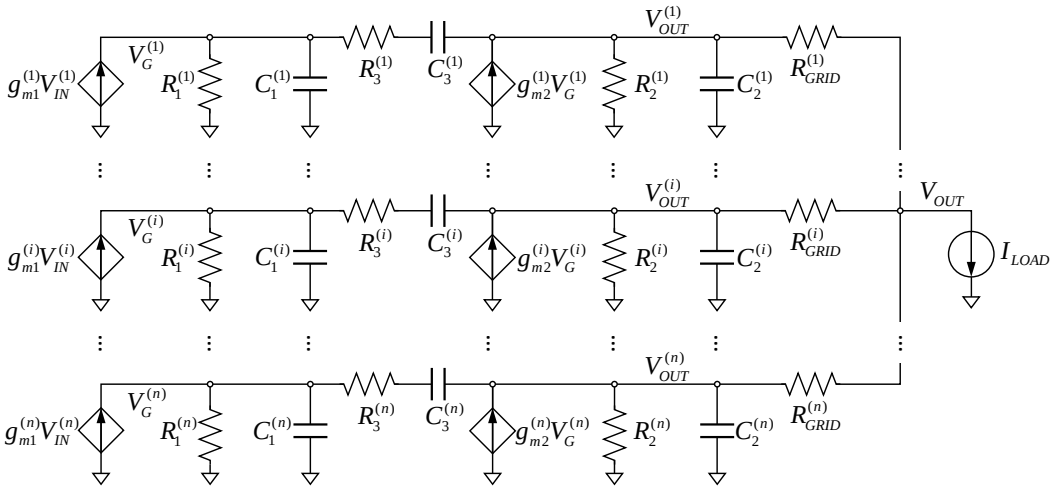


Figure 8.3: Distributed power delivery system with n LDO regulators driving a common power grid.

where the zeros are independent of load current variations ($N(s)^{(i)} = N(s)\ \forall\ i$). Alternatively, the poles in $D(s)^{(i)}$ are shifted and $A_0^{(i)}/(R_2^{(i)} + A_0^{(i)}R_{GRID}^{(i)})$ changes under load current variations, affecting both the root locations and amplitude of the summed polynomials in (8.11).

Intuitively, the summation of n horizontally shifted third order polynomials yields a third order polynomial when the roots are sufficiently spread in frequency. The magnitude of the new polynomial is the sum of the magnitude of the individual polynomials, and the roots are a weighted average of the horizontal frequency shifts. To better understand this concept, consider two horizontally shifted third order polynomials, $D(s)^{(a)}$ and $D(s)^{(b)}$, as shown in Figure 8.4. The sum of the polynomials at $p_1^{(a)}$ and $p_1^{(b)}$ is, respectively, positive and negative. The first root of the new

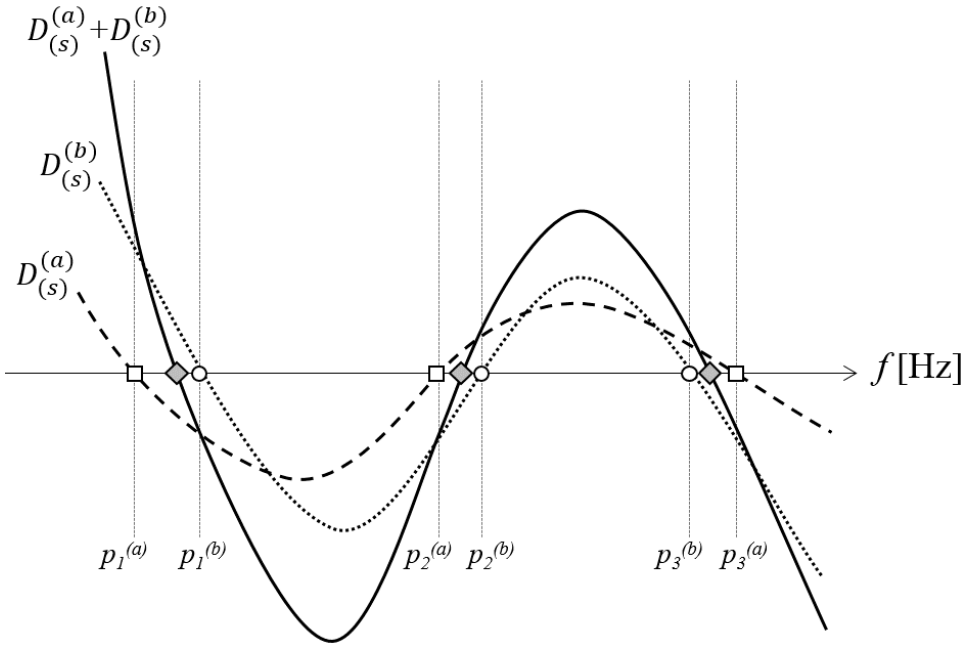


Figure 8.4: Summation of two horizontally shifted polynomials.

polynomial p_1^{ave} is therefore determined by the weighted average of the original roots, $p_1^{(a)}$ and $p_1^{(b)}$. Similarly, the second (p_2^{ave}) and third (p_3^{ave}) roots of the summation are a weighted average of the original shifted roots. The expression in (8.11) can therefore be rewritten as

$$Z_{OUT}^{TOT} = \left(\sum_{i=1}^{n} \frac{A_0^{(i)}}{R_2^{(i)} + A_0^{(i)} R_{GRID}^{(i)}} \right)^{-1} \cdot \frac{N(s)}{D^{ave}(s)}, \tag{8.12}$$

where $D(s)^{ave} = \left(1 + \frac{s}{p_1^{ave}}\right)\left(1 + \frac{s}{p_2^{ave}}\right)\left(1 + \frac{s}{p_3^{ave}}\right)$, and $\min_i\{p_k^{(i)}\} < p_k^{ave} < \max_i\{p_k^{(i)}\}$.

The total impedance of the distributed power delivery system shown in Figure 8.3 is therefore a rational third order function. The poles of the distributed system

are limited by the frequency range of the individual LDO poles ($\min_i\{p_k^{(i)}\}$ and $\max_i\{p_k^{(i)}\}$), exhibiting no RHP poles for individually stable LDO regulators. To maintain stability in a distributed power delivery system with n LDO regulators, the phase of $Z_{OUT}^{TOT}(s)$ must be within the $(-90°, +90°)$ range $\forall\ s$.

To demonstrate the concept of a stable distributed power delivery system based on the passivity-based stability criterion, a distributed power delivery system with $n = 6$ LDO regulators is evaluated. A model of the power delivery system with six LDO regulators and a distributed power delivery network from Chapter 6 is used in this chapter as shown in Figure 8.5. Each LDO in a distributed power delivery system is loaded by an effective current which is a weighted sum of all of the transient currents within the system. The effect of transient current switching on the voltage regulation of a distributed system is therefore similar to a single LDO system. Alternatively, each LDO regulator in a power distribution system contributes differently to the voltage regulation based on the position of the active current loads and the consumed current. The stability of a distributed system is therefore a strong function of the local current shared among the distributed regulators. Current sharing is therefore a primary concern in a distributed power delivery system.

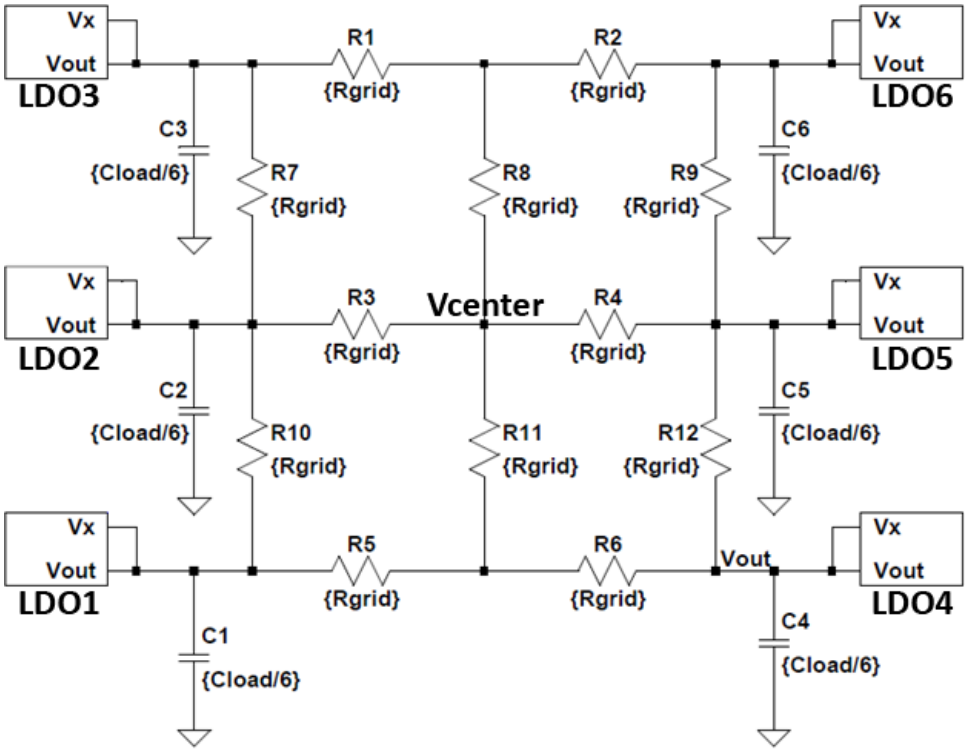
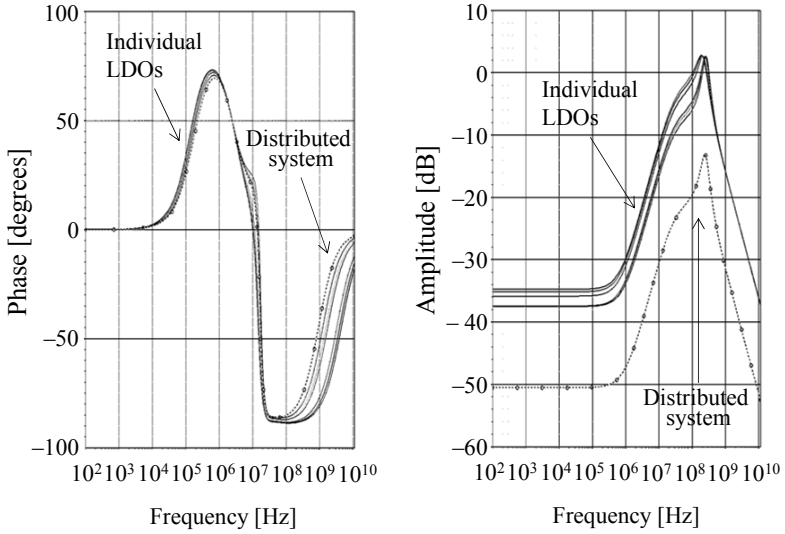


Figure 8.5: Model of distributed LDO and power distribution network.

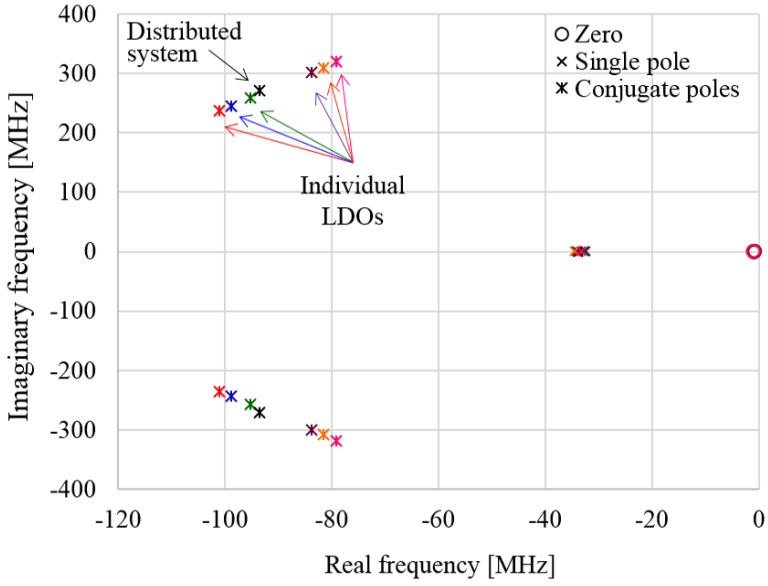
8.2.2 Passivity analysis

The stability of the proposed power delivery system is demonstrated on an example system assuming a total current load of 300 mA. Load sharing among the LDO regulators in the system exhibits a wide range of LDO currents (between 20 mA and 100 mA for an individual LDO regulator). The LDO in close proximity with the current load supplies the largest portion (100 mA) of the total current requirements, which is higher by a factor of two than the average current load (52 mA) supplied by a single LDO. Alternatively, remote LDOs supply significantly less current (down

to 20 mA), only half of the average LDO load current. The output impedance of the system under this load sharing scenario is evaluated for each of the LDO regulators and the combined distributed power delivery system. The phase, magnitude, poles, and zeros within the range of interest are shown in Figure 8.6, demonstrating the concepts of dense zero frequencies and pole frequency averaging used in (8.12). Based on the pole frequency averaging concept shown in (8.12), a distributed system with individually stable LDO regulators *under all feasible load currents* exhibits no RHP poles. To demonstrate the effect of the phase of the output impedance on the stability of a distributed system, the transient response and output impedance phase $\angle Z_{OUT}$ of the distributed system with six LDO regulators is shown in Figure 8.7. In agreement with the passivity-based stability criterion, the output response diverges (oscillates with increasing amplitude), and converges within an exponential envelope for, respectively, $|\angle Z_{OUT}| > 90°$ and $|\angle Z_{OUT}| < 90°$. Note that the system with $C_C = 0.5\,\text{pF}$ and $\max_{\forall f}\{\angle Z_{OUT}\} = 89°$ slowly converges to the steady-state solution, exhibiting an underdamped response inappropriate for voltage regulation in power delivery systems. Alternatively, a system with $C_C = 5\,\text{pF}$ and $\max_{\forall f}\{\angle Z_{OUT}\} = 70°$ exhibits an overdamped response with a significant stability margin. A strong correlation therefore exists between the phase shift of the output voltage and load current, and the effective stability margin of the system. Based on this observation, the phase

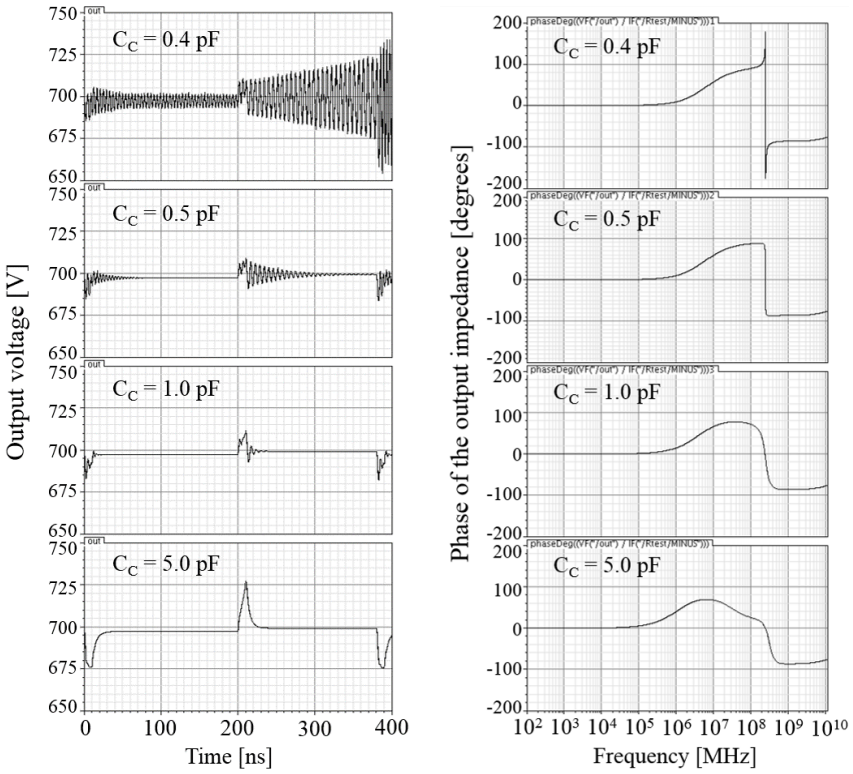


(a) (b)



(c)

Figure 8.6: Output impedance of individual LDO regulators loaded by different currents (between 20 mA and 100 mA) and the combined system output impedance, (a) phase $\angle Z_{OUT}$), (b) amplitude $|Z_{OUT}|$, and (c) poles and zeros.



(a) (b)

Figure 8.7: Output response of a distributed power delivery system with different compensation ($C_C = 0.4pF$, $C_C = 0.5pF$, $C_C = 1pF$, and $C_C = 5pF$), showing the correlation between the (a) transient response, and (b) phase of the output impedance.

margin of the output impedance for a distributed power delivery system is

$$PM(Z_{out}) = 90^\circ - \max_{\forall f}\{\angle Z_{OUT}\}. \tag{8.13}$$

A distributed power delivery system is therefore unstable, stable, or marginally stable if the phase margin of the output impedance is, respectively, negative, positive, or zero. A safe phase margin of the output impedance should be determined based on specific design criteria to avoid excessively underdamped and overdamped voltage regulation systems.

8.3 Parametric circuit performance modeling technique

Existing automated techniques for analog circuit design are based on numerical optimization and evaluation engines [177]. Parametric performance modeling characterizes the performance of an analog circuit (e.g., gain, bandwidth (BW), slew rate (SR), or phase margin (PM)) based on circuit design variables (e.g., device sizes and biasing) [177]. The performance of an individual power supply is typically determined by a set of parameters, such as the DC gain, phase margin, offset, slew rate, and power. Alternatively, a distributed power delivery system should be evaluated based on both the individual power supply performance and additional performance metrics of the combined system, such as the *phase margin of the output impedance*. To reduce the design complexity of modern distributed power delivery systems, the

proposed passivity-based stability criterion should be integrated within existing design automation techniques. A typical automated flow of optimization-based analog circuit sizing is shown in Figure 8.8 [177]. An automated flow for designing a stable

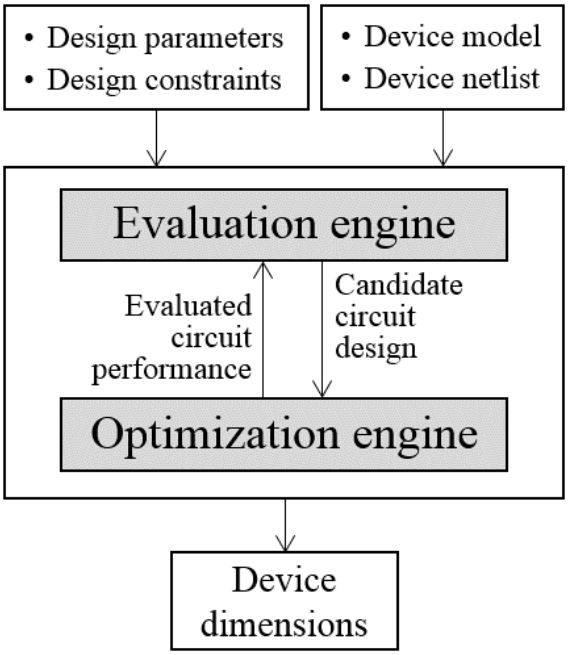


Figure 8.8: Conventional design flow for an analog circuit.

distributed power delivery system is shown in Figure 8.9. The first stage of the proposed flow is based on a typical parametric performance modeling technique [177]. During this stage, an LDO regulator is synthesized based on the specific LDO topology and design objectives. The output of the first stage is used during the second stage to determine the number and location of the parallel connected power supplies in the proposed distributed power delivery system. During this second stage,

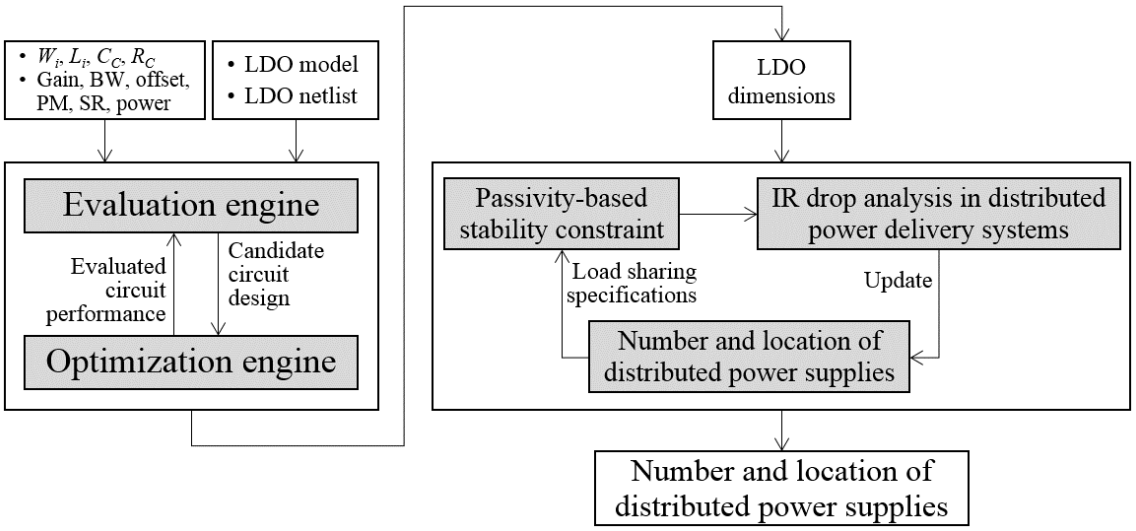


Figure 8.9: Automated PBSC-based design flow for a distributed power delivery system.

a distributed power delivery system with a different number and location of voltage regulators is iteratively evaluated based on the proposed passivity-based stability criterion and distributed power supply placement algorithms [69], [107]. During each iteration, the worst case load sharing scenario is determined for the specific power delivery system. The passivity-based stability of the overall distributed system is evaluated based on the individual current loads. The number and location of the power supplies are updated if required. Finally, the number and location of the parallel connected power supplies is determined that satisfies the quality of power and stability requirements of the distributed power delivery system.

8.4 Experimental results

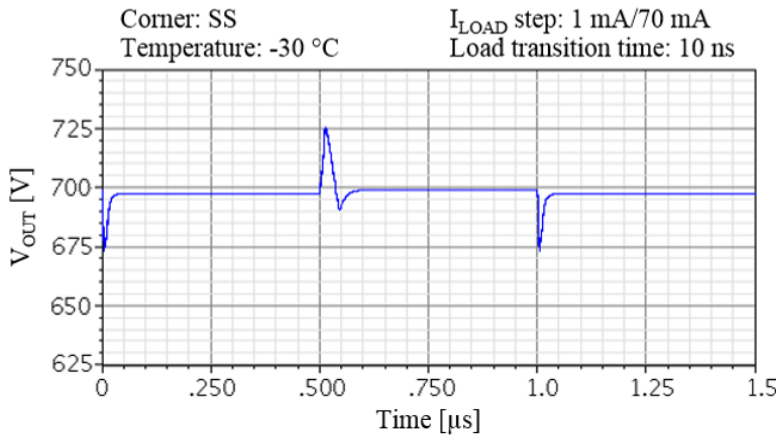
A power delivery system with six LDO regulators has been designed based on the proposed passivity-based stability criterion and fabricated in an advanced 28 nm CMOS technology. The regulators are distributed around the perimeter of the on-chip circuits, as shown in Figure 8.5, regulating the power grid at 0.7 volts with an input voltage of 1.0 volt and maximum load current of 750 mA. The on-chip regulators simultaneously drive a power network, delivering current to the on-chip integrated circuits. To evaluate the stability of the proposed distributed system under aggressive load sharing conditions, a current generating circuit is placed in close proximity to each LDO regulator. These current loads can be activated separately or simultaneously, exhibiting different current loads at individual LDO regulators. All of the measurements are performed on the LDO regulators within the distributed power delivery system.

Load sharing variations are a primary concern for maintaining stability in distributed power delivery systems. Feasible current loads for the proposed power delivery system are based on the system specifications. When all of the integrated loads are activated simultaneously, the proposed distributed system is loaded with a maximum current of $6 \times 125\,\text{mA} = 750\,\text{mA}$. Assuming a uniform distribution of load

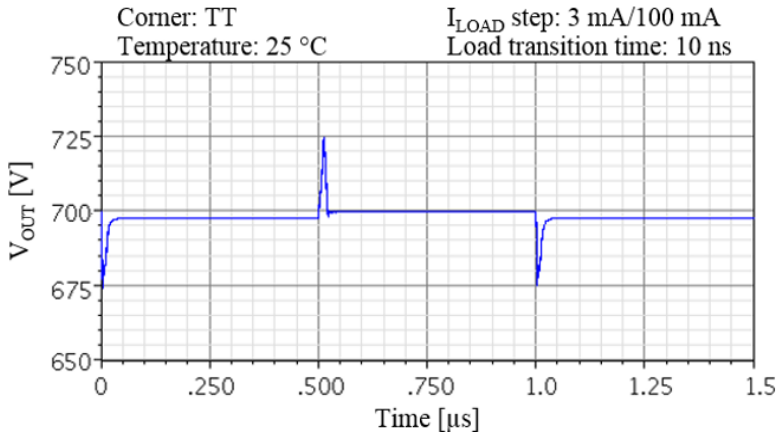
circuits, the maximum load current is evenly distributed among the voltage regulators. Alternatively, currents lower than the maximum load current can be drawn by those circuits located in close proximity to a specific regulator and farther from the other regulators, exhibiting significant load sharing variations. In this work, the distributed LDO regulators simultaneously supply a range of currents between 8.67 mA and 73.5 mA, while maintaining a stable power grid at 0.7 volts. The individual LDO regulators exhibit a stable response and sufficient phase margin under a wide range of feasible load currents. The transient response of the regulator is shown in Figure 8.10 for different load steps, process corners, and temperatures, exhibiting a stable response over a wide range of design parameters.

Modern integrated circuits exhibit aggressive transient characteristics and are expected to mitigate significant PVT variations. To illustrate the mitigation of load sharing variations in a complex power delivery system, the output of the individual load circuits is switched from 8.67 mA to 73.5 mA at different times within a 10 ns period, simultaneously generating different load currents and switching activities at the LDO regulators.

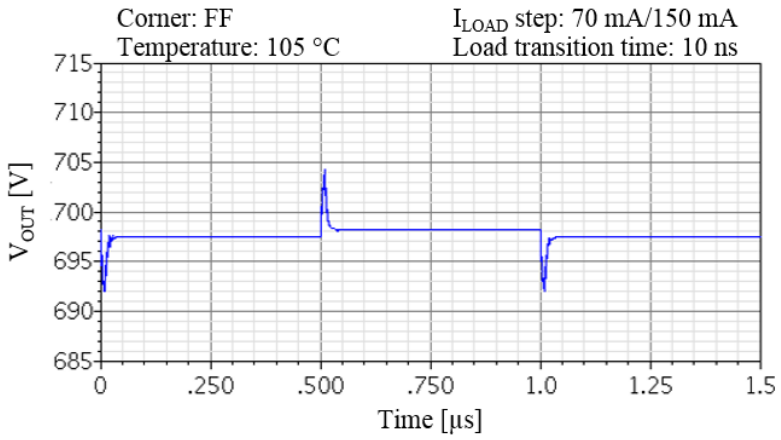
To evaluate the effect of load sharing variations on the stability of a power delivery system, the system is measured for PVT variations. The total load current of the distributed power delivery system with six LDO regulators is stepped from



(a)



(b)

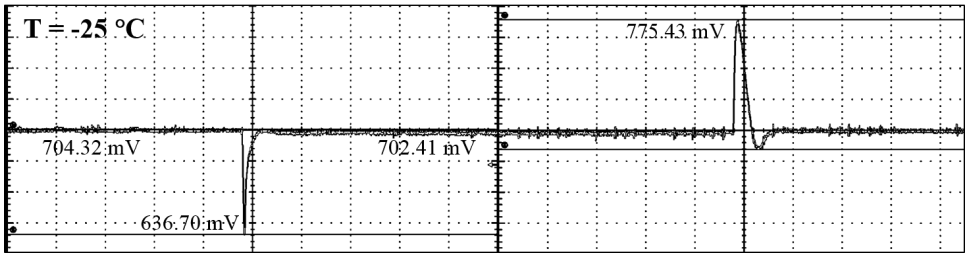


(c)

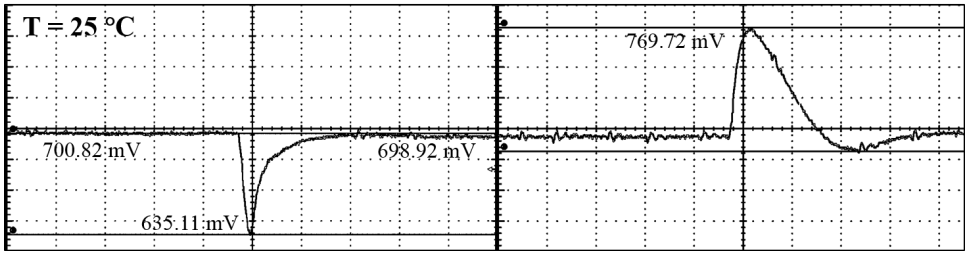
Figure 8.10: Output voltage of a single LDO regulator at (a) slow corner and −30 °C, (b) typical corner and 25 °C, and (c) fast corner and 105 °C.

52 mA to 441 mA in 10 ns. Due to load sharing and PVT variations, the load current of a single LDO regulator significantly increases or decreases as compared to the nominal current, as described in Section 8.2. The measured transient response of the experimental distributed system for nominal input and output voltages of, respectively, 1.0 volt and 0.7 volts is illustrated in Figure 8.11 for −25 °C, 25 °C, and 125 °C. To evaluate the proposed system under line variations, the output is tested at 25 °C under ±10% input voltage variations. The measured transient response is illustrated in Figure 8.12. For all types of variations, the distributed power delivery system exhibits a stable response over a wide range of temperatures with less than 10% voltage droop.

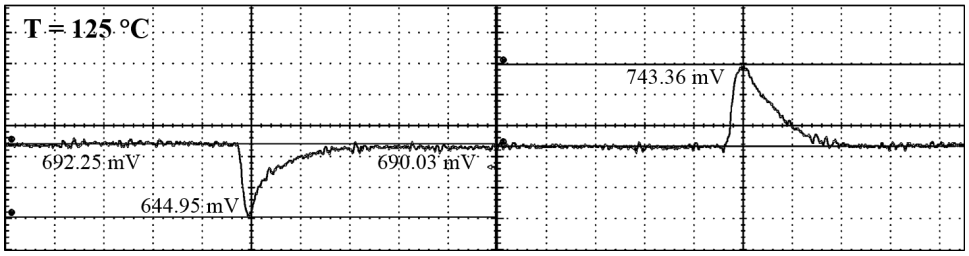
To demonstrate the stability of the distributed power delivery system under maximum load currents and fast load transitions, the load current of the system is stepped from 52 mA to 788 mA in 5 ns at 25 °C. The measured transient response for nominal input and output voltages of, respectively, 1.0 volt and 0.7 volts is illustrated in Figure 8.13, exhibiting a stable response and voltage droop of 0.1 volts. The response time of the system based on the equivalent parasitic capacitance of the load circuit ($C_{Load} = 472\,\text{pF}$), maximum load current ($I_{Load,MAX} = 788\,\text{mA}$), and voltage droop ($\Delta V_{OUT} = 100\,\text{mV}$) is evaluated as $T_R = C_{TOT} \cdot I_{Load,MAX}/\Delta V_{OUT} = 0.064\,\text{ns}$. Based on these experimental results, the system of six parallel LDO regulators yields a



(a)

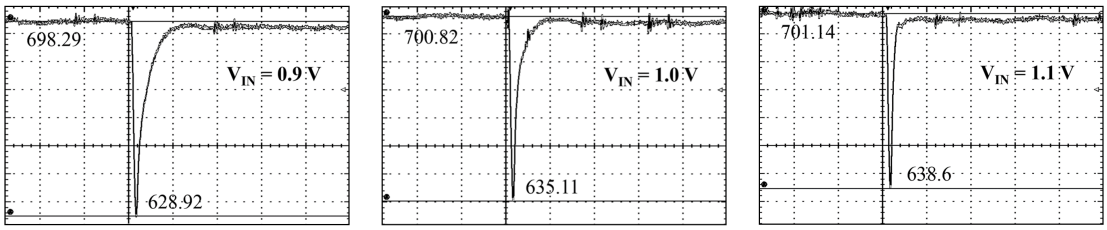


(b)



(c)

Figure 8.11: Transient step response at the load for $V_{IN} = 1$ volt, $V_{OUT} = 0.7$ volts, and a load current step from 52 mA to 441 mA in 10 ns, measured at (a) $T = -25\,^{\circ}\mathrm{C}$, (b) $T = 25\,^{\circ}\mathrm{C}$, and (c) $T = 125\,^{\circ}\mathrm{C}$.



(a) (b) (c)

Figure 8.12: Transient step response at the load at a temperature of 25 °C for V_{OUT} = 0.7 volts and a load current step from 52 mA to 441 mA in 10 ns, measured for (a) V_{IN} = 0.9 volts, (b) V_{IN} = 1.0 volt, and (c) V_{IN} = 1.1 volts.

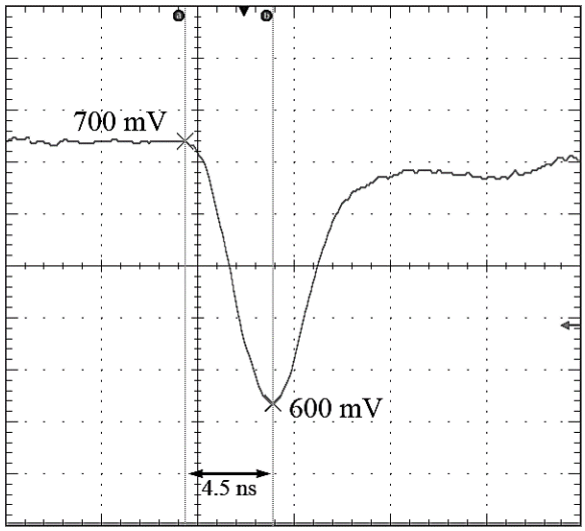


Figure 8.13: Measured transient response for a load current step from 52 mA to 788 mA in 5 ns.

stable response over a wide range of load and PVT variations.

8.5 Summary

Distributed on-chip power regulation and delivery are necessary for delivering high quality power to modern high performance integrated circuits. Significant load sharing and PVT variations however pose stability challenges on the co-design of these multiple on-chip voltage regulators. Thus, the design complexity of the parallel voltage regulators driving the same power grid is traded off for high power quality. To design a stable closed loop regulator, sufficient phase margin in the open loop transfer function is required. Phase margin is therefore a sufficient parameter for determining the stability of a single LDO. Evaluating the open loop characteristics is however not practical with parallel LDO regulators because of the multiple regulation loops. Evaluation of the stability of a distributed power delivery system is therefore not possible with the traditional phase margin criterion.

An alternative passivity-based stability criterion is proposed in this chapter for evaluating the stability of parallel voltage regulators driving a single power grid. Based on this criterion, a distributed power delivery system is stable if and only if the total output impedance of the parallel connected LDOs exhibits no right half plane poles and a phase between $-90°$ and $+90°$. Similar to a single voltage regulator, the

phase margin of the output impedance (the difference between the maximum phase and 90°) should be used to determine the stability of a distributed power delivery system.

A practical third order model of a typical LDO regulator is used to demonstrate the strong correlation between the proposed PBSC method and the stability of a distributed power delivery system. Feasible load sharing variations are evaluated based on system specifications, and upper and lower limits for load sharing variations are determined. Each of the LDO regulators is designed with sufficient phase margin to deliver stable power while satisfying the required load sharing limitations. A distributed power delivery system with six ultra-small fully integrated low-dropout regulators is evaluated to demonstrate the proposed PBSC concept. The system, fabricated in a 28 nm CMOS process, exhibits a stable response with excellent load regulation under PVT and current sharing variations. This power delivery system is believed to be the first successful silicon demonstration of stable parallel analog LDO regulators. The proposed passivity-based stability criterion is shown to be a simple and efficient method to design stable distributed power delivery systems.

Chapter 9

Power Network-on-Chip for Scalable Power Delivery

To facilitate the integration of diverse functions, architectural, circuit, device, and material level power delivery solutions are required. Per core dynamic voltage and frequency scaling is a primary concern for efficiently managing a power budget, and requires the on-chip integration of compact controllers within hundreds of power domains and thousands of cores, further increasing the design complexity of these power delivery systems. While in-package and on-chip power integration has recently became a primary concern [46], [47], research remains focused on developing more compact and efficient power supplies. A methodology to design and manage in-package and on-chip power has to date not been a topic of emphasis. Thus, power delivery in modern ICs is currently dominated by *ad hoc* approaches. With the increasing number of power domains, the greater granularity of the on-chip supply

voltages, and domain adaptive power requirements, the design of the power delivery process has greatly increased in complexity, and is impractical without a systematic methodology. The primary objective here is to provide a systematic methodology for distributed on-chip power delivery and management.

The concept of a power network on-chip (PNoC) is introduced here as a vehicle for distributed on-chip power management. The analogy between a NoC and PNoC is illustrated in Figure 9.1 with simplified NoC and PNoC models. Similar to a network-on-chip, a PNoC decreases the design complexity of power delivery systems, while enhancing the control of the quality of power (QoP) and dynamic voltage scaling (DVS), and providing a scalable platform for efficient power management.

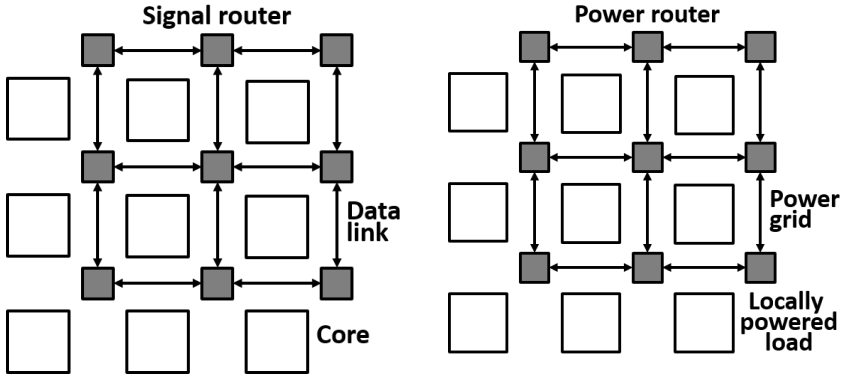


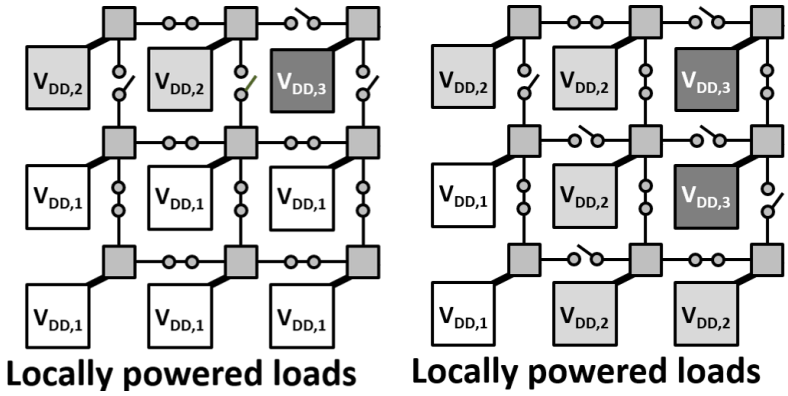
Figure 9.1: On-chip networks based on the approach of separation of functionality, (a) network-on-chip (NoC), and (b) power network-on-chip (PNoC).

The rest of the chapter is organized as follows. The principles of the PNoC

design methodology are described in Section 9.1. In Section 9.2, the performance of the PNoC architecture is compared with existing approaches based on the evaluation of several test cases. Design and performance issues of the PNoC architecture are also discussed. Some concluding remarks are offered in Section 9.3.

9.1 Power network-on-chip architecture

The key concept in systemizing the design of power delivery is to convert the power off-chip, in-package, and/or on-chip with multiple power efficient but large switching power supplies, deliver the power to on-chip voltage clusters, and regulate the power with hundreds of linear low dropout (LDO) regulators at the point-of-load [70]. A power network-on-chip is proposed here as a novel systematic solution to on-chip power delivery that leverages distributed point-of-load power delivery within a fine grained power management framework. The PNoC architecture is a mesh of power routers and locally powered loads, as depicted in Figure 9.1. The power routers are connected through power switches, distributing current to those local loads with similar voltage requirements. A PNoC is illustrated in Figure 9.2 for a single voltage cluster with nine locally powered loads and three different supply voltages, $V_{DD,1}$, $V_{DD,2}$, and $V_{DD,3}$. The power network configuration is shown at two different times, t_1 and t_2.



(a) (b)

Figure 9.2: On-chip power network with multiple locally powered loads and three supply voltage levels (a) PNoC configuration at time t_1, and (b) PNoC configuration at time t_2.

The concept of a power network-on-chip is proposed here to virtually manage the power in SoCs through specialized power routers, switches, and programmable control logic, while supporting scalable power delivery in heterogeneous ICs. A PNoC is comprised of physical links and routers that provide both virtual and physical power routing. This system senses the voltages and currents throughout the system, and manages the POL regulators through power switches. Based on the sensed voltages and currents, a programmable unit makes real-time decisions to apply a new set of configurations to the routers per time slot, dynamically managing the on-chip power delivery process. Novel algorithms are required to dynamically customize the power delivery policies through a specialized microcontroller that routes the power.

These algorithms satisfy real-time power and performance requirements.

A PNoC composed of power routers connected to global power grids and locally powered loads is illustrated in Figure 9.3. Global power from the converters is man-

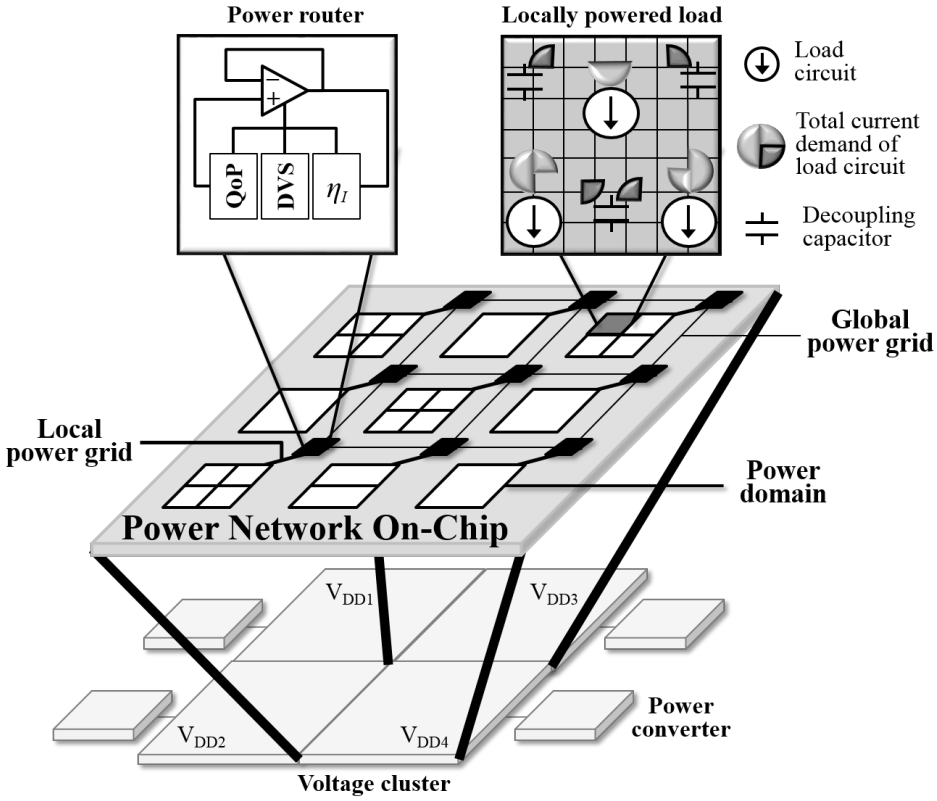


Figure 9.3: On-chip power network with routers distributing the current over the power grid to the local loads.

aged by the power routers, delivered to individual power domains, and regulated within the locally powered loads. These locally powered loads combine all of the current loads located within a specific on-chip power domain with the decoupling capacitors that supply the local current demand within that region. To support DVS,

power switches in the proposed PNoC are dynamically controlled, (dis)connecting power routers within individual voltage clusters in real time. Current loads powered at similar voltage levels therefore draw current from all of the connected power routers, lessening temporary current variations. Similar to a mesh based clock distribution network [188], the shared power supply lessens the effect of the on-chip parasitic impedances, enhancing the voltage regulation and quality of power.

The power routers, local current loads, and power grids are described in the following subsections. Different PNoC topologies and specific design objectives are also considered.

9.1.1 Power routers

The efficient management of the energy budget is dynamically maintained by the power routers. Each power domain is controlled by a single power router. A router topology ranges from a simple linear voltage regulator, shown in Figure 9.4(a), to a complex power delivery system, as depicted in Figure 9.4(b), with sensors, dynamically adaptable power supplies, switches, and a microcontroller. The proposed structure features real-time voltage/frequency scaling, adaptable energy allocation, and precise control over the on-chip QoP. With the PNoC routers the power is managed locally based on specific local current and voltage demands, decreasing the

dependency on remotely located loads and power supplies. Thus, the scalability of power delivery process is enhanced with PNoC approach.

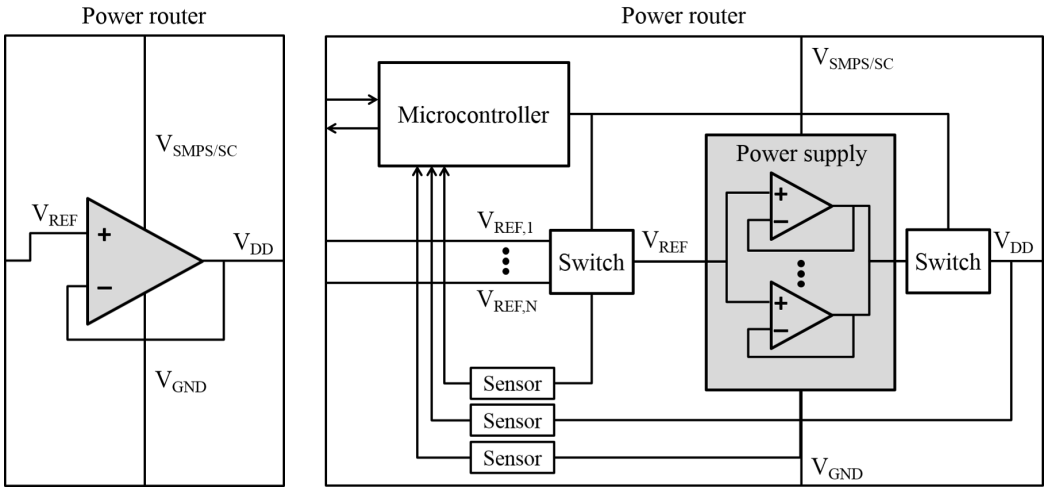


Figure 9.4: Power routers for PNoC a) Simple topology with linear voltage regulator. b) Advanced topology with dynamically adaptable voltage regulator and microcontroller.

9.1.2 Locally powered loads

Locally powered loads with different current demands and power budgets can be efficiently managed with this proposed PNoC. The local power grids provide a specific voltage to the nearby load circuits. The highly complex interactions among the multiple power supplies, decoupling capacitors, and load circuits need to be considered, where the interactions among the nearby components are typically more

significant. The effective region for a point-of-load power supply is the overlap of the effective regions of the surrounding decoupling capacitors [69], [189]. Loads within the same effective region are combined into a single equivalent locally powered load regulated by a dedicated LDO. All of the LDO regulators within a power domain are controlled by a single power router. A model and closed-form expressions of the interactions among the power supplies, decoupling capacitors, and current loads are required to efficiently partition an IC with billions of loads into power domains and locally powered loads.

9.1.3 Power grid

A shared power grid with multiple LDO sources and dedicated power grids each driven by a separate LDO are considered. Dedicated power grids require fine grain distribution of the local power current. Alternatively, a shared power grid with multiple LDO sources is prone to stability issues due to current sharing, and process, voltage, and temperature variations. Specialized adaptive mechanisms are included within the power routers to stabilize a power delivery system that includes a multi-source shared power grid.

9.2 Case study

To evaluate the performance of the power router, a PNoC with four power routers is considered, supplying power to four power domains. IBM power grid benchmarks [141] model the behavior of the individual power domains. To simulate a dynamic power supply in PNoC, the original IBM voltage profiles have been scaled to generate the target power supplies between 0.5 volts and 0.8 volts. Target voltage profiles with four voltage levels (0.8 volts, 0.75 volts, 0.7 volts, and 0.65 volts) within a PNoC are illustrated in Figure 9.5. The number of power domains with each of the four supply voltages changes dynamically based on the transient power requirements of the power domains.

A schematic of a PNoC with four power domains (I, II, III, and IV) and four power routers (PR_I, PR_{II}, PR_{III}, and PR_{IV}) is presented in Figure 9.6. Each of the power routers is composed of an LDO with four switch controlled reference voltages to support dynamic voltage scaling. In addition, the power routers feature adaptive RC compensation and current boost networks controlled by load sensors to provide quality of power control and optimization, as shown in Figure 9.7. The adaptive RC compensation network is comprised of a capacitive block connected in series with two resistive blocks, all digitally controlled. These RC impedances are digitally configured to stabilize the LDO within the power routers under a wide range of

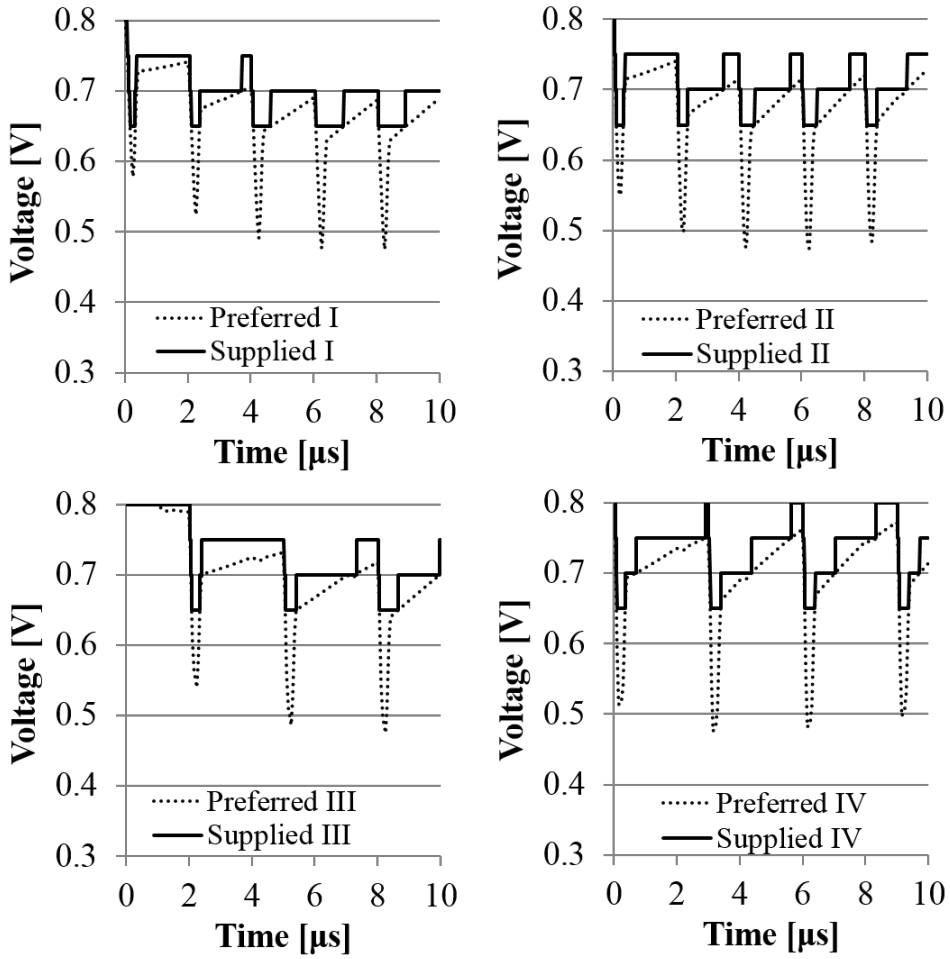


Figure 9.5: Preferred and supplied voltage levels in PNoC with four power domains.

process variations. The proposed current boost circuit is composed of a sensor block that follows the output voltage at the drain of transistor M_P (see Figure 9.7), and a current boost block that controls the current through the differential pair within the LDO. When a high slew rate transition at the output of the LDO occurs, the boost mode is activated, raising the tail current of the LDO differential pair. Alternatively,

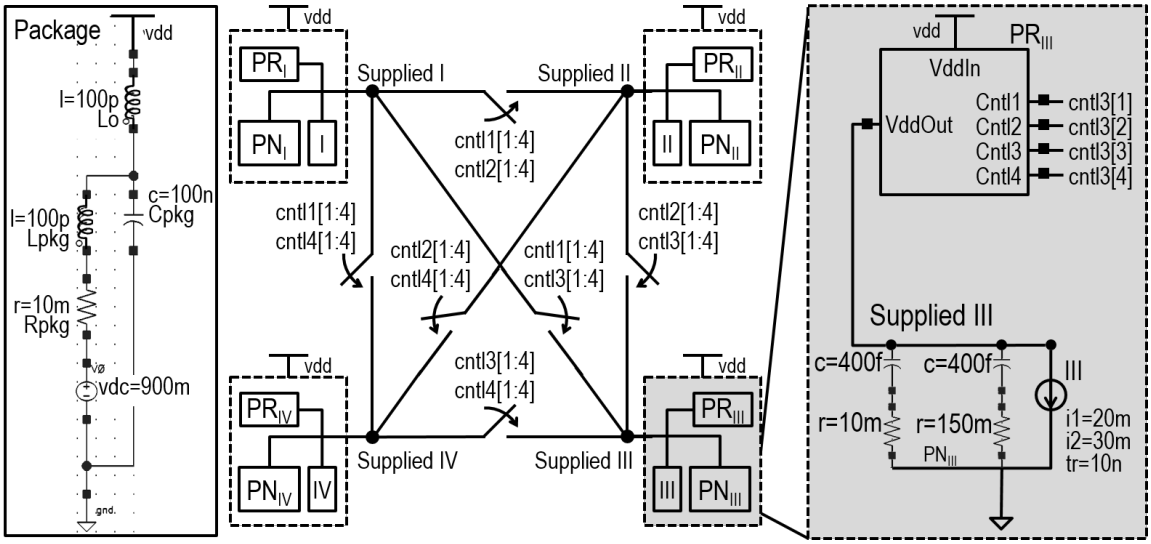


Figure 9.6: Proposed PNoC with four power domains and four power routers connected with control switches.

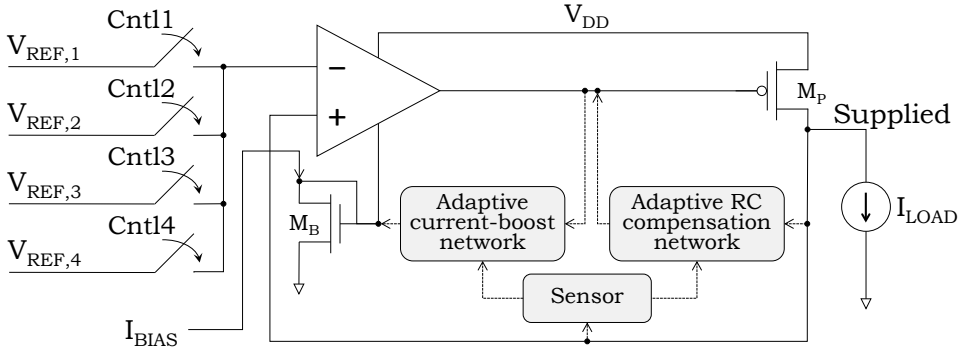


Figure 9.7: Power router with voltage regulator, load sensor, and adaptive networks.

during regular mode, no additional current flows into the differential pair, enhancing the power efficiency of the LDO. Detailed description of the suggestive configurations of the load sensor and adaptive networks are provided in Chapter 6.

The power routers are connected with controlled switches to mitigate load transitions in domains with similar supply voltages. To model the RLC parasitic impedances of the package and power network, non-ideal LDO input and output impedances (PN_I, PN_{II}, PN_{III}, and PN_{IV}) are considered, as shown in Figure 9.6. The load current characteristics are listed in Table 9.1 for each of the four domains. The proposed

Table 9.1: Output load in a PNoC with four power domains.

Domain	I	II	III	IV
Minimum output current [mA]	10	75	20	20
Maximum output current [mA]	50	10	30	20
Output transition time [ns]	50	50	10	N/A

PNoC is simulated in SPICE. Simulation results are presented in Figure 9.8, exhibiting a maximum error of 0.35%, 2.0%, and 2.7% for, respectively, the steady state dropout voltage, load regulation due to the output current switching, and load regulation due to dynamic PNoC reconfigurations. Good correlation with the required power supply in Figure 9.5 is demonstrated. The power savings in each of the power domains range between 21.0% to 31.6% as compared to a system without dynamic voltage scaling. These average power savings show that the PNoC architecture can control the power supplies in real-time, optimizing the power efficiency of the overall power delivery system [190]–[192].

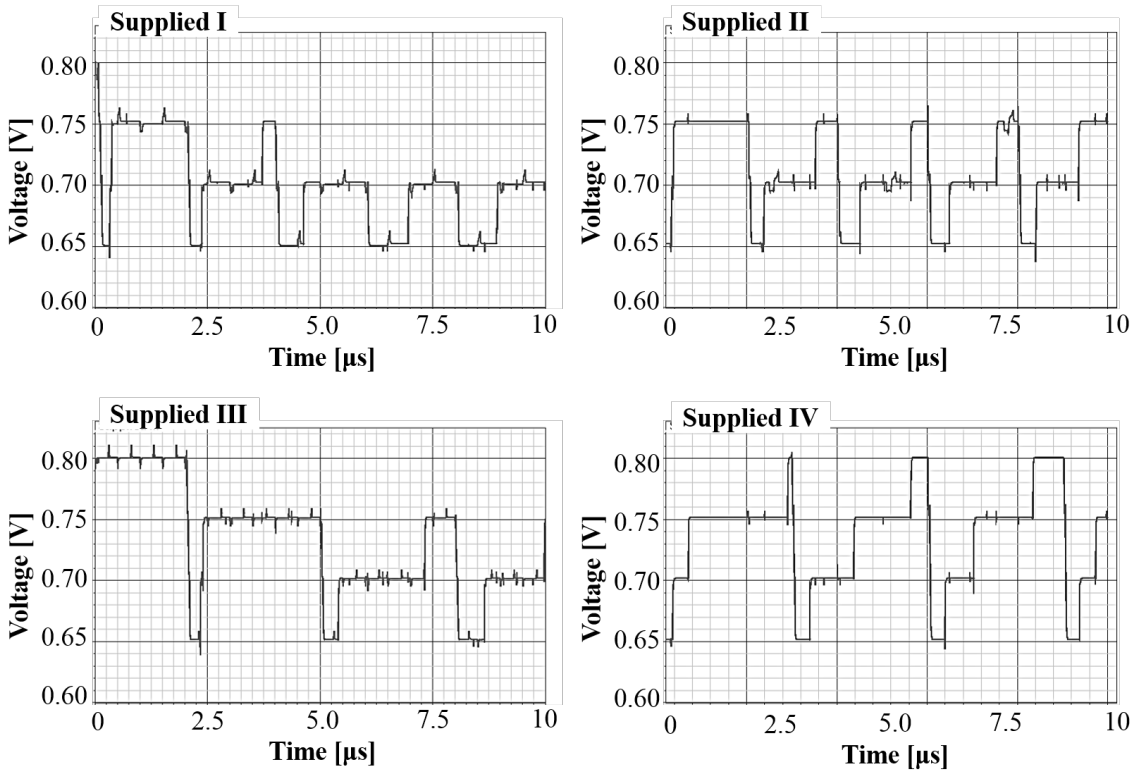


Figure 9.8: Voltage levels in PNoC with four power domains.

9.3 Summary

To address the issues of power delivery complexity and quality of power, a power delivery system should provide a scalable modular architecture that supports integration of additional functional blocks and power features (e.g., DVS, adaptive RC compensation, and efficiency optimization with adaptive current boost) without requiring the re-design of the power delivery system. The architecture should also support heterogeneous circuits and technologies.

The concept and architecture of an on-chip power network is described in this chapter. The primary objectives of an on-chip PNoC based power delivery system are 1) to exploit the concept of on-chip networks for systematic power delivery in SoCs to reduce design complexity while increasing scalability, 2) to provide a methodology that separates power conversion and regulation for efficiently enhancing the quality of power, 3) to enable the application of local power routing through a specialized microcontroller for on-chip power management, and 4) to utilize small area power supplies as point-of-load voltage regulators.

Chapter 10

Future Work

Delivering high quality, dynamically controlled power to support overall power efficient and portable systems is a critical challenge in *heterogeneous* systems-on-chip. Efficiently providing multiple point-of-load on-chip regulated voltages requires fundamental changes to the power delivery and management process. In MtM systems, SoC platforms will consist of hundreds of heterogeneous hardware agents (application processors, DSP cores, fixed function accelerators, sensors, and actuators), each with unique power, performance, and quality of service (QoS) requirements. To support DVS and DVFS, scalable power delivery systems are required that are capable of *dynamically* redistributing high quality regulated power within billions of loads to multiple functionally diverse power domains. To satisfy power limits at each evolving technology node, the increasing availability of "dark silicon" should be exploited

[193]. To switch transistors with greater granularity in real-time, a distributed architecture coupled with a dynamically controlled power delivery and management methodology is required.

To cope with the design complexity of distributed on-chip power delivery, a novel and scalable power delivery platform, PNoC, has been introduced that relies on the reuse of components, architectures, methodologies, and applications. Two different research directions based on topics presented in this dissertation are proposed in this chapter for further investigation. A distributed architecture for a power network on-chip is introduced in Section 10.1. Security issues in modern ICs and a secure PNoC solution are discussed in Section 10.2. A summary is offered in Section 10.3.

10.1 Power network on-chip for distributed power management

Centralized dynamic power management is commonly used to increase power efficiency in modern heterogeneous systems. A typical power delivery scheme is illustrated in Figure 10.1 for an SoC with multiple power domains within each of n individual modules. In this configuration, the power is managed by a single device, exhibiting a centralized power management system. Centralized power delivery

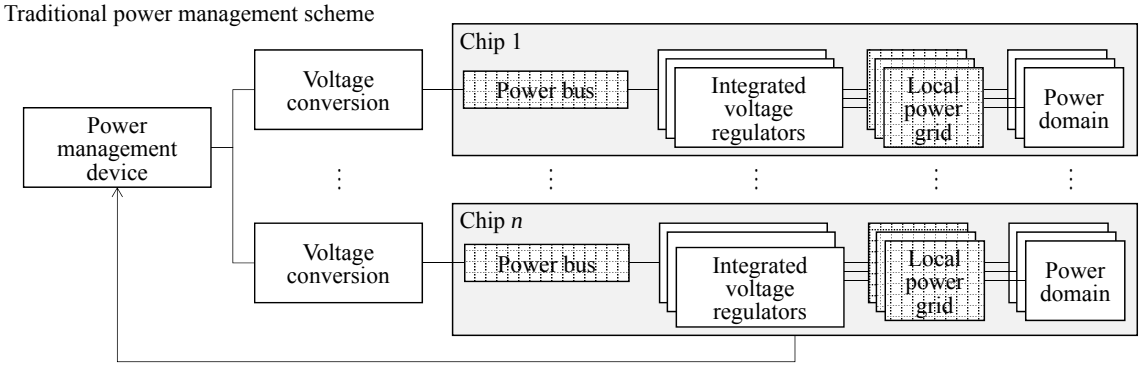


Figure 10.1: Centralized power management of an SoC with n modules.

systems are often characterized by a long feedback path from the individual power domains to a single power management controller, dissipating extra power due to unnecessary data transport, and limiting real-time control over the energy budget due to the slow feedback response. In addition, a centralized approach places all of the power management functions in one or a few locations, affecting system scalability.

The concept of a power network on-chip is introduced in Chapter 9 as a systematic and scalable methodology for on-chip power management. A distributed power delivery system based on the PNoC approach is shown in Figure 10.2. In this scheme, the energy budget is managed locally with intelligent distributed power routers, considering essential information such as timing slacks, voltages, currents, and temperatures that are sensed within the individual power domains. Real-time fine grain power management is possible by shortening the power domain-to-router feedback path (as compared to the centralized approach). Power management adjustments

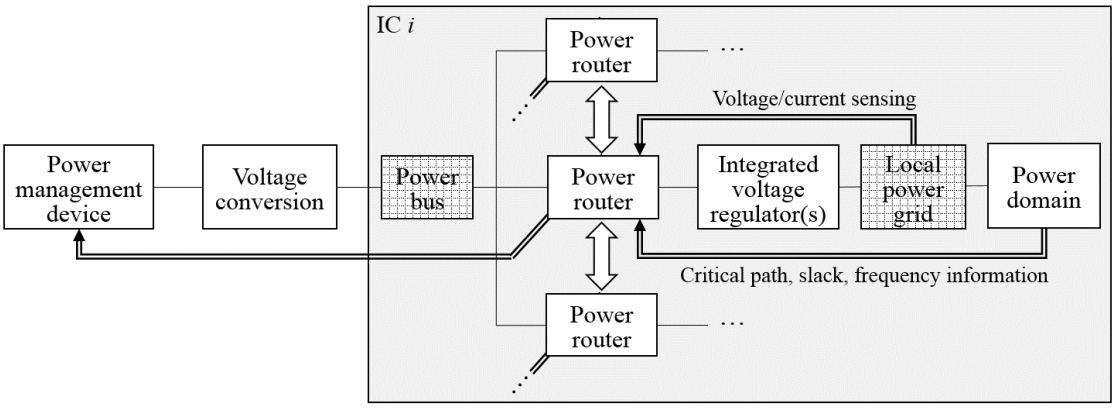


Figure 10.2: Distributed PNoC scheme for power management of a single IC.

should therefore redistribute the overall energy budget in a near real-time manner by communicating local power management decisions among the individual power routers within a PNoC. A centralized power management mechanism can also be integrated to support predetermined DVS and DVFS operations. Novel algorithms and controllers will need to be developed to coordinate these management functions among the intelligent, distributed, and centralized power management mechanisms.

10.2 Secure power network on-chip

With existing techniques for leakage current analysis, the security of the on-chip information has become a growing concern. Non-invasive methods, such as side channel power attacks [194]–[196], are widely used to decode encrypted information based on the co-analysis of stored data and leakage current profiles. Alternatively, invasive

power attacks (e.g., battery draining by sleep deprivation) [197] can be initiated to reduce the performance and/or lifetime of computing devices. Several methods to mitigate power analysis attacks have recently been proposed at the architectural, algorithmic, and circuit levels [198]–[200]. A common countermeasure for power analysis is masking the effective power consumption of an IC by injecting additional power or modifying power profiles with DVS techniques [198]. These approaches however degrade the overall power efficiency of the system. Alternatively, securing information and performance with traditional algorithmic encryption techniques lack scalability and often do not exhibit sufficient security [199]. To efficiently mitigate power analysis attacks, specialized design flows and circuits that disrupt the correlation between the internal leakage power and overall power dissipation need to be developed [200]. Accurate evaluation of power efficiency and security levels are required. Secure on-chip power networks will need to be developed to efficiently manage power in modern ICs.

10.3 Summary

To cope with the increasing design complexity of heterogeneous, distributed, and dynamically controllable power delivery systems while enhancing the scalability of both static and dynamic power delivery systems, a systematization of the power

delivery and management process is required. The concept of intelligent on-chip networks for delivering power has been suggested as future research that addresses the objectives of scalable, systematic power management in high performance heterogeneous systems. Integration of emerging techniques for mitigating power attacks within the proposed PNoC platform has also been proposed as possible future research problems.

A secure power management system based on a distributed PNoC architecture has been suggested as possible long term research. To exploit the concept of on-chip networks for a secure and systematic power delivery and management process while reducing design complexity and increasing scalability, three primary components are required: (1) a systematic methodology that exploits the distributed nature of the proposed PNoC architecture to provide real-time adaptive and efficient control over the quality of power, (2) algorithmic solutions for distributed power management, and (3) specialized circuits and software/firmware architectural solutions to mitigate power attacks in performance sensitive systems.

Chapter 11

Conclusions

The continued advance of society and emerging market segments require functionally diverse semiconductors. In these modern heterogeneous systems, functionally diverse circuits are integrated on-chip, requiring a wide range of high quality DC voltages. In addition, as the need for portable, high performance ICs increases, intelligent management of the energy budget becomes a primary concern. The future of heterogeneous, high performance systems is strongly dependent upon the power delivery system and deeply affected by the quality of on-chip power, availability of fine grain dynamically controlled voltage levels, and the ability to manage power in real-time. On-chip integration of several power supplies is no longer sufficient to address all of these power delivery challenges. To satisfy evolving power delivery requirements, the classical approach for power generation and distribution has changed over the last decade. Power generation circuits are physically closer to the

loads to provide enhanced control over the quality of delivered power. Heterogeneous power delivery systems with different types of off-chip, in-package, and distributed on-chip power converters must be considered. Software and firmware solutions are necessary to address the high design complexity of these hierarchical nonlinear power delivery systems with thousands of power delivery components and billions of loads. Dynamic voltage and frequency scaling increases energy efficiency over time. Fine grain power management schemes optimize the delivery of power in terms of quality, area, efficiency, and design complexity. An effective power delivery solution should provide a systematic methodology, scalable distributed architectures, algorithms for distributed power management, and specialized circuits.

A platform for scalable power delivery and management has been proposed in this dissertation. The key concept of this platform is to manage the overall energy budget with fine grain distributed on-chip power networks, providing local feedback paths from the billions of loads to multiple, locally intelligent power routers. The proposed PNoC approach addresses the issues of design complexity and scalability by providing a modular architecture that supports integration of functional blocks and power features without requiring re-design of the power delivery system. Architectural, algorithmic, and circuit level requirements for intelligent PNoC-based power delivery are necessary, and some possible solutions are described in this dissertation.

A power delivery topology and related algorithms to co-design power supplies within the PNoC framework are proposed. Circuit and design level solutions that address the challenges of distributed on-chip power delivery and intelligent power management such as on-chip area limitations, dynamic power control, and stability are also demonstrated. Integrating emerging technologies (e.g., mitigation of invasive and non-invasive power attacks) within the proposed PNoC framework is discussed.

A heterogeneous power delivery topology is proposed in this dissertation as an architecture for delivering power in PNoC systems. While optimizing the characteristics of the individual power supplies is important, optimizing system-wide power efficiency is critical in high performance ICs. To maintain high quality on-chip power, the power should be converted and regulated within a hierarchical structure composed of different types of power supplies. To convert power with minimum power losses while avoiding area consuming on-chip passive components, power efficient switching mode power supplies should be placed off-chip or in-package. Area efficient linear regulators should be distributed on-chip to deliver and locally control the load circuitry, while reducing power losses from the low voltages dropped across the linear regulators.

Exhaustive *ad hoc* approaches for co-designing power supplies at different levels of hierarchy within a power delivery system exhibit significant design complexity and

are computationally impractical in DVS/DVFS systems with hundreds to thousands of power domains. A computationally efficient methodology to co-design switching converters and on-chip linear regulators within a heterogeneous system is proposed, achieving high quality power and efficiency within limited on-chip area. Dynamically clustering a heterogeneous power delivery system is demonstrated on a suite of IBM benchmark circuits, exhibiting a computationally and power efficient alternative to existing *ad hoc* methodologies that employ either switching or linear on-chip power supplies.

Advanced circuit solutions, such as specialized power routers, switches, programmable control logic, and ultra-small power supplies, are required to provide efficient high quality dynamically managed power within the PNoC framework. To demonstrate the feasibility of the proposed concept of distributed, dynamically controllable, power delivery systems, several power delivery circuits have been proposed. To provide a circuit level means for dynamically scaling the voltage in adaptive systems, a digitally controlled pulse width modulator has been developed, including closed-form duty cycle expressions, and validated under PVT variations. Another key component of distributed power delivery systems is an ultra-small power efficient linear regulator. An ultra-small power efficient linear low dropout regulator has been designed, manufactured, and tested in a 28 nm CMOS process, exhibiting excellent

load regulation under PVT variations.

An important challenge in the realization of distributed power delivery systems is the stability of multi-feedback structures. A distributed system with multiple parallel connected power supplies and dependent feedback paths may exhibit degraded stability due to complex interactions among the power supplies, power distribution network, and current loads. To provide a stable distributed power delivery system, a passivity-based stability criterion has been proposed. Based on this criterion, a distributed power delivery system is stable if and only if the output impedance of the parallel connected power supplies exhibits no right half plane poles and a phase between -90° and $+90^{\circ}$. This criterion can be used to analyze the stability of complex distributed power delivery systems and integrated within design automation environments. To demonstrate these distributed design concepts and techniques, a distributed power delivery system with six identical ultra-small fully integrated low-dropout regulators has been designed and fabricated in a 28 nm CMOS process. This system of distributed parallel LDO regulators exhibits a stable system response over a wide range of temperatures and voltage variations. The system is believed to be the first successful silicon demonstration of stable parallel analog linear regulators.

To conclude, a fine grain power management framework has been proposed to dynamically control the power delivery process while optimally allocating power

among different power domains at run time. As opposed to traditional practices where power is generated with either efficient switching or compact linear power regulators, the proposed distributed heterogeneous approach enables system-wide fine grain optimization of the overall energy budget and efficiency, while satisfying on-chip area constraints. To cope with emerging challenges, such as quality of power, system-wide power efficiency, real-time distributed power management, and design complexity, a variety of circuits, algorithms, and architectural level techniques has been proposed within the power networks on-chip framework. The development of these new capabilities will fundamentally change the way power is delivered and managed, simultaneously enhancing the quality of power, performance, and scalability in high complexity, high performance integrated circuits.

Bibliography

[1] J. Bardeen and W. H. Brattain, "The Transistor, a Semi-Conductor Triode," *Physics Review*, Vol. 74, No. 2, pp. 230–231, July 1948.

[2] W. F. Brinkman, D. E. Haggan, and W. W. Troutman, "A History of the Invention of the Transistor and Where It Will Lead Us," *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 12, pp. 1858–1865, December 1997.

[3] Nobel Media AB, "Nobel Prize," 2013. [Online]. Available from: http://www.nobelprize.org.

[4] J. E. Lilienfeld, *Method and Apparatus for Controlling Electric Currents*, U.S. Patent No. 1,745,175, October 1926.

[5] O. Heil, *Improvements in or Relating to Electrical Amplifiers and Other Control Arrangements andd Devices*, British Patent No. 439,457, March 1935.

[6] J. A. Hoerni, *Method of Manufacturing Semiconductor Devices*, U.S. Patent No. 3,025,589, May 1959.

[7] J. A. Hoerni, "Planar Silicon Transistors and Diodes," *IEEE Transactions on Electron Devices*, Vol. 8, No. 2, pp. 655–657, March 1961.

[8] V. F. Pavlidis, I. Savidis, and E. G. Friedman, "Clock Distribution Architectures for 3-D SOI Integrated Circuits," *Proceedings of the IEEE International SOI Conference*, pp. 111–112, October 2008.

[9] G. E. Moore, "Cramming More Components onto Integrated Circuits," *Electronics Magazine*, Vol. 38, No. 8, pp. 114–117, April 1965.

[10] Semiconductor Industry Association, *The International Technology Roadmap for Semiconductors*, California, 2012.

[11] W. Arden, M. Brillout, P. Cogez, M. Graef, B. Huizing, and R. Mahnkopf, "More-than-Moore," 2010, [Online]. Available from: http://www.itrs.net.

[12] P. Cauvet, S. Bernard, and M. Renovell, "System-in-Package, a Combination of Challenges and Solutions," *Proceedings of the IEEE VLSI Test Symposium*, pp. 193–199, May 2007.

[13] W. Wolf, A. A. Jerraya, and G. Martin, "Multiprocessor System-on-Chip (MPSoC) Technology," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 27, No. 10, pp. 1701–1713, October 2008.

[14] V. F. Pavlidis and E. G. Friedman, "Interconnect-Based Design Methodologies for Three-Dimensional Integrated Circuits," *Proceedings of the IEEE*, Vol. 97, pp. 123–140, January 2009.

[15] V. F. Pavlidis and E. G. Friedman, *Three-Dimensional Integrated Circuit Design*, Morgan Kaufmann, 2009.

[16] E. G. Friedman, "Clock Distribution Networks in Synchronous Digital Integrated Circuits," *Proceedings of the IEEE*, Vol. 89, No. 5, pp. 665–692, May 2001.

[17] I. Vaisband, E. G. Friedman, R. Ginosar, and A. Kolodny, "Low Power Clock Network Design," *Journal of Low Power Electronics and Applications*, Vol. 1, No. 1, pp. 219–246, June 2011.

[18] I. Vaisband, E. Friedman, R. Ginosar, and A. Kolodny, "Energy Metrics for Power Efficient Crosslink and Mesh Topologies," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1656–1659, Mat 2012.

[19] D. F. Wendel *et al.*, "POWER7, a Highly Parallel, Scalable Multi-Core High End Server Processor," *IEEE Journal of Solid-State Circuits*, Vol. 46, No. 1, pp. 145–161, January 2011.

[20] S. Kumar, A. Jantsch, J.-P. Soininen, M. Forsell, M. Millberg, J. Oberg, K. Tiensyrja, and A. Hemani, "A Network on Chip Architecture and Design Methodology," *Proceedings of the IEEE Computer Society Annual Symposium on VLSI*, pp. 105–112, April 2002.

[21] Intel Corporation, *4004: Single Chip 4-Bit P-Channel Microprocessor*, March 1987.

[22] J. D. Warnock *et al.*, "The Circuit and Physical Design of the POWER4 Microprocessor," *IBM Journal of Research and Development*, Vol. 46, No. 1, pp. 27–51, January 2002.

[23] Y. Kikuchi *et al.*, "A 40 nm 222 mW H.264 Full-HD Decoding, 25 Power Domains, 14-Core Application Processor With x512b Stacked DRAM," *IEEE Journal of Solid-State Circuits*, Vol. 46, No. 1, pp. 32–41, January 2011.

[24] E. Salman and E. G. Friedman, *High Performance Integrated Circuit Design*, McGraw-Hill, 2012.

[25] "International Technology Roadmap for Semiconductors.," Available online: www.itrs.net/Links/2011ITRS/2011Tables/PIDS_2011Tables.xlsx.

[26] Intel Corporation, *8008: 8-Bit Parallel Central Processor Unit*, November 1972.

[27] Intel Corporation, *8080A/8080A-1/8080A-2: 8-Bit N-Channel Microprocessor*, November 1986.

[28] Intel Corporation, *8085AH/8085AH-2/8085AH-1: 8-Bit HMOS Microprocessors*, September 1987.

[29] Intel Corporation, *8086: 16-Bit HMOS Microprocessors*, September 1990.

[30] Intel Corporation, *80C286/883: High Performance Microprocessor with Memory Management and Protection*, September 1988.

[31] Intel Corporation, *Intel386 EX Embedded Microprocessor*, October 1998.

[32] Intel Corporation, *Embedded Intel486 SX Processor*, August 2004.

[33] Intel Corporation, *Pentium Processor at iCOMP Intex 610-75MHz*, June 1997.

[34] Intel Corporation, *Pentium Pro Processor with 1 MB L2 Cache at 200 MHz*, January 1998.

[35] Intel Corporation, *Pentium II Processor at 233MHz, 266MHz, 300MHz and 333MHz*, January 1998.

[36] Intel Corporation, *Pentium III Processor for the SC242 at 450MHz to 1.0GHz*, July 2002.

[37] Intel Corporation, *Pentium IV Processor in the 423-Pin Package at 1.30, 1.40, 1.50, 1.60, 1.70, 1.80, 1.90, 2GHz*, August 2001.

[38] Intel Corporation, *Pentium IV Processor on 90nm Process*, February 2005.

[39] Intel Corporation, *Intel Core i7-900 Desktop Processor Extreme Edition Series on 32-nm Process*, July 2010.

[40] T. López, R. Elferich, and E. Alarcón, *Voltage Regulators for Next Generation Microprocessors*, Springer, 2010.

[41] V. Kursun and E. G. Friedman, *Multi-Voltage CMOS Circuit Design*, Wiley & Sons, 2006.

[42] F. Waldron, J. Slowey, A. Alderman, B. Narveson, and S. C. O'Mathuna, "Technology Roadmapping for Power Supply in Package (PSiP) and Power Supply on Chip (PwrSoC)," *Proceedings of the IEEE International Applied Power Electronics Conference and Exposition*, pp. 525–532, February 2010.

[43] N. Wang, J. Hannon, R. Foley, K. Mccarthy, T. O'Donnell, K. Rodgers, F. Waldron, and C. O. Mathuuna, "Integrated Magnetics on Silicon for Power Supply in Package (PSiP) and Power Supply On Chip (PwrSoC)," *Proceedings of the Electronics System-Integration Technology Conference*, pp. 1–6, September 2010.

[44] T. M. Andersen, C. M. Zingerli, F. Krismer, J. W. Kolar, and C. O'Mathuna, "Inductor Optimization Procedure for Power Supply in Package and Power Supply on Chip," *Proceedings of the IEEE Energy Conversion Congress and Exposition*, pp. 1320–1327, September 2011.

[45] J. Lu, H. Jia, X. Wang, K. Padmanabhan, W. G. Hurley, and Z. J. Shen, "Modeling, Design, and Characterization of Multiturn Bondwire Inductors with Ferrite Epoxy Glob Cores for Power Supply System-on-Chip or System-in-Package Applications," *IEEE Transactions on Power Electronics*, Vol. 25, No. 8, pp. 2010–2017, August 2010.

[46] J. Kim *et al.*, "Chip-Package Hierarchical Power Distribution Network Modeling and Analysis Based on a Segmentation Method," *IEEE Transactions on Advanced Packaging*, Vol. 33, No. 3, pp. 647–659, August 2010.

[47] P. Hazucha *et al.*, "A 233-MHz 80%-87% Efficient Four-Phase DC-DC Converter Utilizing Air-Core Inductors on Package," *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 4, pp. 838–845, April 2005.

[48] C. O'Mathuna, N. Wang, S. Kulkarni, and S. Roy, "Review of Integrated Magnetics for Power Supply on Chip (PwrSoC)," *IEEE Transactions on Power Electronics*, Vol. 27, No. 1, pp. 4799–4816, November 2012.

[49] K. Onizuka, K. Inagaki, H. Kawaguchi, M. Takamiya, and T. Sakurai, "Stacked-Chip Implementation of On-Chip Buck Converter for Distributed Power Supply System in SiPs," *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 11, pp. 2404–2410, November 2007.

[50] S. C. O. Mathuna, T. O'Donnell, N. Wang, and K. Rinne, "Magnetics on Silicon: An Enabling Technology for Power Supply On Chip," *IEEE Transactions on Power Electronics*, Vol. 20, No. 3, pp. 585–592, May 2005.

[51] D. Yeh, L.-S. Peh, S. Borkar, J. Darringer, A. Agarwal, and W.-M. Hwu, "Thousand-Core Chips [Roundtable]," *IEEE Design and Test of Computers*, Vol. 25, No. 3, pp. 272–278, May/June 2008.

[52] S. Y. Borkar, "Thousand Core Chips - A Technology Perspective," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 746–749, June 2007.

[53] D. N. Truong, W. H. Cheng, T. Mohsenin, Z. Yu, A. T. Jacobson, G. Landge, M. J. Meeuwsen, C. Watnik, A. T. Tran, Z. Xiao, E. W. Work, J. W. Webb, P. V. Mejia, and B. M. Baas, "A 167-Processor Computational Platform in 65 nm CMOS," *IEEE Journal of Solid-State Circuits*, Vol. 44, No. 4, pp. 1130–1144, April 2009.

[54] T. Hattori *et al.*, "A Power Management Scheme Controlling 20 Power Domains for a Single-Chip Mobile Processor," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 542–543, February 2006.

[55] Y. Kanno *et al.*, "Hierarchical Power Distribution with 20 Power Domains in 90-nm Low-Power Multi-CPU Processor," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 540–541, February 2006.

[56] R. W. Erickson and D. Maksimovic, *Fundamentals of Power Electronics*, Kluwer Academic Publishers, 2001.

[57] M. Swaminathan, J. Kim, I. Novak, and J. P. Libous, "Power Distribution Networks for System-on-Package: Status and Challenges," *IEEE Transactions on Advanced Packaging*, Vol. 27, No. 2, pp. 286–300, May 2004.

[58] G. Patounakis, Y. W. Li, and K. L. Shepard, "A Fully Integrated On-Chip DC-DC Conversion and Power Management System," *IEEE Journal of Solid-State Circuits*, Vol. 39, No. 3, pp. 443–451, March 2004.

[59] J. Rosenfeld and E. G. Friedman, "Quasi-Resonant Interconnects: A Low Power, Low Latency Design Methodology," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 17, No. 2, pp. 181–193, February 2009.

[60] G. Palumbo and D. Pappalardo, "Charge Pump Circuits: An Overview on Design Strategies and Topologies," *IEEE Circuits and Systems Magazine*, Vol. 10, No. 1, pp. 31–45, March 2010.

[61] C. Jia, H. Chen, M. Liu, C. Zhang, and Z. Wang, "Integrated Power Management Circuit for Piezoelectronic Generator in Wireless Monitoring System of Orthopaedic Implants," *IET Circuits, Devices and Systems*, Vol. 2, No. 6, pp. 485–494, December 2008.

[62] C. Hong, L. Ming, H. Wenhan, C. Yi, J. Chen, Z. Chun, and W. Zihua, "Low-Power Circuits for the Bidirectional Wireless Monitoring System of the Orthopedic Implants," *IEEE Journal on Biomedical Circuits and Systems*, Vol. 3, No. 6, pp. 437–443, December 2009.

[63] A. Cabrini, A. Fantini, and G. Torell, "High-Efficiency Regulated Charge Pump for Non-Volatile Memories," *Proceedings of the IEEE International Conference on Electronics, Circuits and Systems*, pp. 720–723, December 2006.

[64] Q. Fan, X. Fu, P. Niu, G. Yang, and T. Gao, "A Novel Low Voltage and High Speed CMOS Charge Pump Circuit," *Proceeding of the IEEE International Conference on Signal Processing Systems*, pp. V3–389–V3–391, July 2010.

[65] I. Vaisband, M. Saadat, and B. Murmann, "A Closed-Loop Reconfigurable Switched-Capacitor DC-DC Converter for Sub-mW Energy Harvesting Applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 62, No. 2, pp. 385–394, February 2015.

[66] J. Guo and K. N. Leung, "A 6-μW Chip-Area-Efficient Output-Capacitorless LDO in 90-nm CMOS Technology," *IEEE Journal of Solid-State Circuits*, Vol. 45, No. 9, pp. 1896–1905, September 2010.

[67] Y.-H. Lam and W.-H. Ki, "A 0.9 V 0.35 μm Adaptively Biased CMOS LDO Regulator with Fast Transient Response," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 442–626, February 2008.

[68] M. Al-Shyoukh, H. Lee, and R. Perez, "A Transient-Enhanced Low-Quiescent Current Low-Dropout Regulator with Buffer Impedance Attenuation," *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 8, pp. 1732–1742, August 2007.

[69] S. Kose and E. G. Friedman, "Distributed On-Chip Power Delivery," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 2, No. 4, pp. 704–713, December 2012.

[70] I. Vaisband and E. G. Friedman, "Heterogeneous Methodology for Energy Efficient Distribution of On-Chip Power Supplies," *IEEE Transactions on Power Electronics*, Vol. 28, No. 9, pp. 4267–4280, September 2013.

[71] S. Kose, S. Tam, S. Pinzon, B. McDermott, and E. G. Friedman, "An Area Efficient On-Chip Hybrid Voltage Regulator,," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 398–403, March 2012.

[72] J. Gjanci and M. H. Chowdhury, "A Hybrid Scheme for On-Chip Voltage Regulation in System-On-a-Chip (SOC)," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 19, No. 11, pp. 1949–1959, November 2011.

[73] Barry Boehm, "Spiral Development: Experience, Principles, and Refinements Spiral Development Workshop," July 2000.

[74] Semiconductor Industry Association, California, *The International Technology Roadmap for Semiconductors*, 2007.

[75] L. D. Smith, "Decoupling capacitor calculations for CMOS circuits," *IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 101–105, November 1994.

[76] C. R. Paul, "Effectiveness of Multiple Decoupling Capacitors," *IEEE Microwave and Wireless Components Letters*, Vol. 34, No. 2, pp. 130–133, May 1992.

[77] T. Roy, L. Smith, and J. Prymak, "ESR and ESL of Ceramic Capacitor Applied to Decoupling Applications," *IEEE Topical Meeting on Electrical Performance of Electronic Packaging*, pp. 213–216, October 1998.

[78] L. K. Wang and H. H. Chen, "On-Chip Decoupling Capacitor Design to Reduce Switching-Noise-Induced Instability in CMOS/SOI VLSI," *Proceedings of the IEEE International SOI Conference*, pp. 100–101, October 1995.

[79] H. H. Chen and S. E. Schuster, "On-Chip Decoupling Capacitor Optimization for High-Performance VLSI Design," *Proceedings of the IEEE International Symposium on VLSI Technology, Systems, and Applications*, pp. 99–103, May/June 1995.

[80] P. Larsson, "Resonance and Damping in CMOS Circuits with On-Chip Decoupling Capacitance," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 45, No. 8, pp. 849–858, August 1998.

[81] M. D. Pant, P. Pant, and D. S. Wills, "On-Chip Decoupling Capacitor Optimization Using Architectural Level Prediction," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 10, No. 3, pp. 319–326, June 2002.

[82] S. Kose, S. Tam, S. Pinzon, B. McDermott, and E. G. Friedman, "Active Filter Based Hybrid On-Chip DC-DC Converters for Point-of-Load Voltage Regulation," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 21, No. 4, pp. 680–691, April 2013.

[83] S. Kose and E. G Friedman, "Distributed Power Network Co-Design with On-Chip Power Supplies and Decoupling Capacitors," *Proceedings of the Workshop on System Level Interconnect Prediction*, June 2011.

[84] S. Kose and E. G Friedman, "Simultaneous Co-Design of Distributed On-Chip Power Supplies and Decoupling Capacitors," *Proceedings of the IEEE International SOC Conference*, pp. 15–18, September 2010.

[85] V. Kursun, S. G. Narendra, V. K. De, and E. G. Friedman, "Analysis of Buck Converters for On-Chip Integration With a Dual Supply Voltage Microprocessor," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 11, No. 3, pp. 514–522, June 2003.

[86] V. Kursun, S. G. Narendra, V. K. De, and E. G. Friedman, "High Input Voltage Step-Down DC-DC Converters For Integration in a Low Voltage CMOS Process," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 517–521, March 2004.

[87] T. Y. Man, K. N. Leung, C. Y. Leung, P. K. T. Mok, and M. Chan, "Development of Single-Transistor-Control LDO Based on Flipped Voltage Follower for SoC," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 55, No. 5, pp. 1392–1401, June 2008.

[88] P. Hazucha *et al.*, "Area-Efficient Linear Regulator with Ultra-Fast Load Regulation," *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 4, pp. 933–940, April 2005.

[89] M. Wens and M. S. J. Steyaert, "A Fully Integrated CMOS 800-mW Fourphase Semiconstant ON/OFF-Time Step-Down Converter," *IEEE Transactions on Power Electronics*, Vol. 26, No. 2, pp. 326–333, February 2011.

[90] H. Nam, Y. Ahn, and J. Roh, "5-V Buck Converter Using 3.3-V Standard CMOS Process With Adaptive Power Transistor Driver Increasing Efficiency and Maximum Load Capacity," *IEEE Transactions on Power Electronics*, Vol. 27, No. 1, pp. 463–471, January 2012.

[91] L. Wang, Y. Pei, X. Yang, Y. Qin, and Z. Wang, "Improving Light and Intermediate Load Efficiencies of Buck Converters With Planar Nonlinear Inductors and Variable On Time Control," *IEEE Transactions on Power Electronics*, Vol. 27, No. 1, pp. 342–353, January 2012.

[92] W. Yan, W. Li, and R. Liu, "A Noise-Shaped Buck DCDC Converter with Improved Light-Load Efficiency and Fast Transient Response," *IEEE Transactions on Power Electronics*, Vol. 26, No. 12, pp. 3908–3924, December 2011.

[93] H. Jia, J. Lu, X. Wang, K. Padmanabhan, and Z. J. Shen, "Integration of a Monolithic Buck Converter Power IC and Bondwire Inductors with Ferrite Epoxy Glob Cores," *IEEE Transactions on Power Electronics*, Vol. 26, No. 6, pp. 1627–1630, June 2011.

[94] Y.-H. Lee, S.-C. Huang, S.-W. Wang, W.-C. Wu, P.-C. Huang, H.-H. Ho, Y.-T. Lai, and K.-H. Chen, "Power-Tracking Embedded BuckBoost Converter with Fast Dynamic Voltage Scaling for the SoC System," *IEEE Transactions on Power Electronics*, Vol. 27, No. 3, pp. 1271–1282, March 2012.

[95] Y. Ahn, H. Nam, and J. Roh, "A 50-MHz Fully Integrated Low-Swing Buck Converter Using Packaging Inductors," *IEEE Transactions on Power Electronics*, Vol. 27, No. 10, pp. 4347–4356, October 2012.

[96] M. Bathily, B. Allard, and F. Hasbani, "A 200-MHz Integrated Buck Converter with Resonant Gate Drivers for an RF Power Amplifier," *IEEE Transactions on Power Electronics*, Vol. 27, No. 2, pp. 610–613, February 2012.

[97] Y. Ramadass, A. Fayed, and A. Chandrakasan, "A Fully-Integrated Switched-Capacitor Step-Down DC-DC Converter with Digital Capacitance Modulation in 45 nm CMOS," *IEEE Journal of Solid-State Circuits*, Vol. 45, No. 12, pp. 2557–2565, December 2010.

[98] H.-P. Le, S. R. Sanders, and E. Alon, "Design Techniques for Fully Integrated Switched-Capacitor DC-DC Converters," *IEEE Journal of Solid-State Circuits*, Vol. 46, No. 9, pp. 2120–2131, September 2011.

[99] S. S. Sapatnekar and H. Su, "Analysis and Optimization of Power Grids," *IEEE Design and Test*, Vol. 20, No. 3, pp. 904–912, May 2003.

[100] J. N. Kozhaya, S. R. Nassif, and F. N. Najm, "A Multigrid-Like Technique for Power Grid Analysis," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 10, pp. 1148–1160, October 2002.

[101] M. Zhao, R. V. Panda, S. S. Sapatnekar, and D. Blaauw, "Hierarchical Analysis of Power Distribution Networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 21, No. 2, pp. 159–168, February 2002.

[102] P. G. Doyle and J. L. Snell, *Random Walks and Electrical Networks*, Mathematical Association of America, 1984.

[103] H. Qian, S. R. Nassif, and S. S. Sapatnekar, "Power Grid Analysis Using Random Walks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 8, pp. 1204–1224, August 2005.

[104] H. Qian and S. S. Sapatnekar, "Hierarchical Random-Walk Algorithms for Power Grid Analysis," *Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference*, pp. 499–504, January 2004.

[105] H. Qian, S. R. Nassif, and S. S. Sapatnekar, "Random Walks in a Supply Network," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 93–98, June 2003.

[106] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 1984.

[107] S. Kose and E. G Friedman, "Efficient Algorithms for Fast IR Drop Analysis Exploiting Locality," *Integration, The VLSI Journal*, Vol. 45, No. 2, pp. 149–161, March 2012.

[108] S. Kose and E. G. Friedman, "Fast Algorithms for IR Voltage Drop Analysis Exploiting Locality," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 996–1001, June 2011.

[109] Z. Zeng and P. Li, "Locality-Driven Parallel Power Grid Optimization," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 28, No. 8, pp. 1190–1200, August 2009.

[110] B. Yan, S. X.-D. Tan, G. Chen, and L. Wu, "Modeling and Simulation for On-Chip Power Grid Networks by Locally Dominant Krylov Subspace Method," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 744–749, November 2008.

[111] P. Li, "Variational Analysis of Large Power Grids by Exploring Statistical Sampling Sharing and Spatial Locality," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 645–651, November 2005.

[112] E. Chiprout, "Fast Flip-Chip Power Grid Analysis Via Locality and Grid Shells," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 485–488, November 2004.

[113] S. Pant, D. Blaauw, and E. Chiprout, "Power Grid Physics and Implications for CAD," *IEEE Design and Test of Computers*, Vol. 24, No. 3, pp. 246–254, May 2007.

[114] K. Wang and M. Marek-Sadowska, "On-Chip Power-Supply Network Optimization using Multigrid-Based Technique," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 24, No. 3, pp. 407–417, March 2005.

[115] X.-D. S. Tan and C.-J. R. Shi, "Fast Power/Ground Network Optimization Based on Equivalent Circuit Modeling," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 550–554, June 2001.

[116] M. Popovich, E. G. Friedman, R. M. Secareanu, and O. L. Hartin, "Efficient Placement of Distributed On-Chip Decoupling Capacitors in Nanoscale ICs," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 811–816, November 2007.

[117] M. Popovich, E. G. Friedman, M. Sotman, and A. Kolodny, "Efficient Distributed On-Chip Decoupling Capacitors for Nanoscale ICs," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 16, No. 7, pp. 1717–1721, July 2008.

[118] Z. Zeng, X. Ye, Z. Feng, and P. Li, "Tradeoff Analysis and Optimization of Power Delivery Networks with On-Chip Voltage Regulation," *Proceedings of the IEEE/ACM Design Automation Conference*, pp. 831–836, June 2010.

[119] M. S. Daskin, *Network and Discrete Location: Models, Algorithms, and Applications*, John Wiley and Sons, 1995.

[120] Z. Drezner and H. Hamacher, *Facility Location: Applications and Theory*, Springer, 2002.

[121] R. Z. Farahani, M. S. Seifi, and N. Asgari, "Multiple Criteria Facility Location Problems: A Survey," *Applied Mathematical Modelling*, Vol. 34, No. 7, pp. 1689–1709, October 2010.

[122] N. Viswanathan *et al.*, "The ISPD-2011 Routability-Driven Placement Contest and Benchmark Suite," *Proceedings of the ACM International Symposium on Physical Design*, pp. 141–146, March 2011.

[123] V. Vorperian, "Simplified Analysis of PWM Converters Using Model of PWM Switch. II. Discontinuous Conduction Mode," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 26, No. 3, pp. 497–505, May 1990.

[124] V. Vorperian, "Simplified Analysis of PWM Converters Using Model of PWM Switch. Continuous Conduction Mode," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 26, No. 3, pp. 490–496, May 1990.

[125] F. Wei and A. Fayed, "A Feasibility Study of High-Frequency Buck Regulators in Nanometer CMOS Technologies," *Proceedings of the IEEE Dallas Workshop on Circuits and Systems*, pp. 1–4, October 2009.

[126] Vishay, "Passive Components," Available online: http://www.vishay.com/power-ics/step-down-regulators.

[127] Johanson Technology, "Integrated Passive Components.," Available online: http://www.johansontechnology.com/images/stories/catalog/JTI_CAT_2012.pdf.

[128] D. Lu and C. P. Wong, *Materials for Advanced Packaging*, Springer, 2008.

[129] S. Bhunia and S. Mukhopadhyay, *Low-Power Variation-Tolerant Design in Nanometer Silicon*, Springer, 2011.

[130] R. G. Dreslinski, M. Wieckowski, D. Blaauw, D. Sylvester, and T. Mudge, "Near-Threshold Computing: Reclaiming Moores Law Through Energy Efficient Integrated Circuits," *Proceedings of the IEEE*, Vol. 98, No. 2, pp. 253–266, February 2010.

[131] K. Ishida, Y. Ryu, X. Zhang, P.-H. Chen, K. Watanabe, M. Takamiya, and T. Sakurai, "0.5-V Input Digital LDO with 98.7% Current Efficiency and 2.7-A Quiescent Current in 65nm CMOS," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 1–4, September 2010.

[132] K. Hirairi, Y. Okuma, H. Fuketa, T. Yasufuku, M. Takamiya, M. Nomura, H. Shinohara, and T. Sakurai, "13% Power Reduction in 16b Integer Unit in 40nm CMOS by Adaptive Power Supply Voltage Control with Parity-Based Error Prediction and Detection (PEPD) and Fully Integrated Digital LDO," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 486–488, February 2012.

[133] B. Amelifard and M. Pedram, "Optimal Design of the Power-Delivery Network for Multiple Voltage-Island System-on-Chips," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 28, No. 6, pp. 888–900, June 2009.

[134] I. Vaisband and E. G. Friedman, "Energy Efficient Adaptive Clustering of On-Chip Power Delivery Systems," *Integration, The VLSI Journal*, Vol. 48, pp. 1–9, January 2015.

[135] I. Vaisband and E. G. Friedman, "Computationally Efficient Clustering of Power Supplies in Heterogeneous Real Time Systems," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 1628–1631, June 2014.

[136] Yongtae Kim and Peng Li, "An Ultra-Low Voltage Digitally Controlled Low-Dropout Regulator with Digital Background Calibration," *Proceedings of the IEEE International Symposium on Quality Electronic Design*, pp. 151–158, March 2012.

[137] Y. Okuma *et al.*, "0.5-V Input Digital LDO with 98.7% Current Efficiency and 2.7-μA Quiescent Current in 65nm CMOS," *Proceedings of the IEEE Custom Integrated Circuits Conference*, pp. 1–4, September 2010.

[138] M. Ho, K. N. Leung, and K.-L. Mac, "A Low-Power Fast-Transient 90-nm Low-Dropout Regulator With Multiple Small-Gain Stages," *IEEE Journal of Solid-State Circuits*, Vol. 45, No. 11, pp. 2466–2475, November 2010.

[139] Y. Xiong, S. Sun, H. Jia, P. Shea, and Z. J. Shen, "New Physical Insights on Power MOSFET Switching Losses," *IEEE Transactions on Power Electronics*, Vol. 24, No. 2, pp. 525–531, February 2009.

[140] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer, 1999.

[141] S. R. Nassif, "Power Grid Analysis Benchmarks," *Proceedings of the IEEE/ACM Asia and South Pacific Design Automation Conference*, pp. 376–381, January 2008.

[142] M. G. Degrauwe, J. Rijmenants, E. A. Vittoz, and H. J. de Man, "Adaptive Biasing CMOS Amplifiers," *IEEE Journal of Solid-State Circuits*, Vol. 17, No. 3, pp. 522–528, June 1982.

[143] W.-J. Huang and S.-I. Liu, "Capacitor-Free Low Dropout Regulators using Nested Miller Compensation with Active Resistor and 1-Bit Programmable Capacitor Array," *IET Electronics Letters*, Vol. 2, No. 3, pp. 306–316, June 2008.

[144] K. N. Leung and P. K. T. Mok, "A Capacitor-Free CMOS Low-Dropout Regulator with Damping-Factor-Control Frequency Compensation," *IEEE Journal of Solid-State Circuits*, Vol. 38, No. 10, pp. 1691–1702, October 2003.

[145] C. K. Chava and J. Silva-Martinez, "A Frequency Compensation Scheme for LDO Voltage Regulators," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 51, No. 6, pp. 1041–1050, June 2004.

[146] X. Fan, C. Mishra, and E. S.-Sinencio, "Single Miller Capacitor Frequency Compensation Technique for Low-Power Multistage Amplifiers," *IEEE Journal of Solid-State Circuits*, Vol. 40, No. 3, pp. 584–592, March 2005.

[147] R. J. Milliken, J. S.-Martinez, and E. S.-Sinencio, "Full On-Chip CMOS Low-Dropout Voltage Regulator," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 54, No. 9, pp. 1879–1890, September 2007.

[148] T. Y. Man, P. K. T. Mok, and M. Chan, "A High Slew-Rate Push-Pull Output Amplifier for Low-Quiescent Current Low-Dropout Regulators with Transient-Response Improvement," *IEEE Transactions on Circuits and Systems II: Express Briefs*, Vol. 54, No. 9, pp. 755–759, September 2007.

[149] K. N. Leung, Y. S. Ng, K. Y. Yim, and P. Y. Or, "An Adaptive Current-Boosting Voltage Buffer for Low-Power Low Dropout Regulators," *Proceeding of the IEEE Conference on Electron Devices and Solid-State Circuits*, pp. 485–488, December 2007.

[150] M. El-Nozahi, A. Amer, J. Torres, K. Entesari, and E. Sanchez Sinencio, "High PSR Low Drop-Out Regulator with Feedforward Ripple Cancellation Technique," *IEEE Journal of Solid-State Circuits*, Vol. 45, No. 3, pp. 565–577, March 2010.

[151] P. Y. Or and K. N. Leung, "An Output-Capacitorless Low-Dropout Regulator with Direct Voltage-Spike Detection," *IEEE Journal of Solid-State Circuits*, Vol. 45, No. 2, pp. 458–466, February 2010.

[152] G. A. Rincon-Mora and P. E. Allen, "A Low-Voltage, Low Quiescent Current, Low Drop-Out Regulator," *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 1, pp. 36–44, January 1998.

[153] Y.-H. Lee, S.-Y. Peng, C.-C. Chiu, A.C.-H. Wu, K.-H. Chen, Y.-H. Lin, S.-W. Wang, T.-Y. Tsai, C.-C. Huang, and and C.-C. Lee, "A Low Quiescent Current Asynchronous Digital-LDO with PLL-Modulated Fast-DVS Power Management in 40 nm SoC for MIPS Performance Improvement," *IEEE Journal of Solid-State Circuits*, Vol. 48, No. 4, pp. 1018–1030, April 2013.

[154] P. Li, "Design and Analysis of IC Power Delivery," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 664–666, November 2012.

[155] S. Lai, B. Yan, and P. Li, "Stability Assurance and Design Optimization of Large Power Delivery Networks with Multiple On-Chip Voltage Regulators," *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 247–254, November 2012.

[156] S. Lai and P. Li, "A Fully On-Chip Area-Efficient CMOS Low-Dropout Regulator with Load Regulation," *Analog Integrated Circuits and Signal Processing*, Vol. 72, No. 2, pp. 925–1030, February 2012.

[157] S. Lai, B. Yan, and P. Li, "Localized Stability Checking and Design of IC Power Delivery With Distributed Voltage Regulators," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 32, No. 9, pp. 1321–1334, September 2013.

[158] A. J. D'Souza, R. Singh, J. R. Prabhu, G. Chowdary, A. Seedher, S. Somayajula, N. R. Nalam, L. Cimaz, S. Le Coq, P. Kallam, S. Sundar, S. Cheng, S. Tumati, and W. Huang, "A Fully Integrated Power-Management Solution for a 65nm CMOS Cellular Handset Chip," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 382–384, February 2011.

[159] J. F. Bulzacchelli, Z. Toprak-Deniz, T. M. Rasmus, J. A. Iadanza, W. L. Bucossi, S. Kim, R. Blanco, C. E. Cox, M. Chhabra, C. D. LeBlanc, C. L. Trudeau, and D. J. Friedman, "Dual-Loop System of Distributed Microregulators With High DC Accuracy, Load Response Time Below 500 ps, and 85-mV Dropout Voltage," *IEEE Journal of Solid-State Circuits*, Vol. 47, No. 4, pp. 863–874, April 2012.

[160] F. Lima, A. Geraldes, T. Marques, J. N. Ramalho, and P. Casimiro, "Embedded CMOS Distributed Voltage Regulator for Large Core Loads," *Proceedings of the IEEE European Solid-State Circuits Conference*, pp. 521–524, September 2003.

[161] B. Razavi, *Design of Analog CMOS Integrated Circuits*, McGraw-Hill, 2001.

[162] B. H. Calhoun and A. Chandrakasan, "Ultra-Dynamic Voltage Scaling Using Sub-Threshold Operation and Local Voltage Dithering in 90nm CMOS," *Proceedings of the IEEE International Solid-State Circuits Conference*, pp. 300–599, February 2005.

[163] T. D. Burd, T. A. Pering, A. J. Stratakos, and R. W. Brodersen, "A Dynamic Voltage Scaled Microprocessor System," *IEEE Journal of Solid-State Circuits*, Vol. 35, No. 11, pp. 1571–1580, November 2000.

[164] R. Jakushokas, M. Popovich, A. V. Mezhiba, S. Kose, and E. G. Friedman, *Power Distribution Networks with On-Chip Decoupling Capacitors, Second Edition*, Springer, 2011.

[165] S. Kose, I. Vaisband, and E. G. Friedman, "Digitally Controlled Wide Range Pulse Width Modulator for On-Chip Power Supplies," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 2251–2254, May 2013.

[166] S. Docking and M. Sachdev, "A Method to Derive an Equation for the Oscillation Frequency of a Ring Oscillator," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 50, No. 2, pp. 259–264, February 2003.

[167] W. Kolodziejski, S. Kuta, and J. Jasielski, "Current Controlled Delay Line Elements' Improvement Study," *Proceedings of the International Conference on Signals and Electronic Systems*, pp. 1–4, September 2012.

[168] I. Vaisband, M. Azhar, E. G. Friedman, and S. Kose, "Digitally Controlled Pulse Width Modulator for On-Chip Power Management," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 22, No. 12, pp. 2527–2534, January 2014.

[169] A. Djemouai, M. Sawan, and M. Slamani, "High Performance Integrated CMOS Frequency to Voltage Converter," *Proceedings of the International Conference on Microelectronics*, pp. 63–66, December 1998.

[170] A. M. Pappu, X. Zhang, A. V. Harrison, and A. B. Apsel, "Process-Invariant Current Source Design: Methodology and Examples," *IEEE Journal of Solid-State Circuits*, Vol. 42, No. 10, pp. 2293–2302, October 2007.

[171] X. Zhang and A. B. Apsel, "A Low-Power, Process and Temperature Compensated Ring Oscillator with Addition Based Current Source," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 58, No. 5, pp. 868–878, May 2011.

[172] M. Anis and M. H. Aburahma, "Leakage Current Variability in Nanometer Technologies," *Proceedings of the International Workshop on System-on-Chip for Real-Time Applications*, pp. 60–63, July 2005.

[173] NIMO Group, "Predictive Technology Model (PTM)," 2008. [Online]. Available from: http://www.eas.asu.edu/∼ptm, Arizona State University.

[174] Y. Cao, *Predictive Technology Model for Robust Nanoelectronic Design*, Springer, 2011.

[175] Y. Kishiwada, S. Ueda, Y. Miyawaki, and T. Matusoka, "Process Variation Compensation with Effective Gate-Width Tuning for Low-Voltage CMOS Digital Circuits," *Proceedings of the IEEE International Meeting for Future of Electron Devices*, pp. 1–2, May 2012.

[176] M. S. Gupta, J. A. Rivers, P. Bose, G.-Y. Wei, and D. Brooks, "Tribeca: Design for PVT Variations with Local Recovery and Fine-Grained Adaptation," *Proceedings of the Annual IEEE/ACM International Symposium on Microarchitecture*, pp. 435–446, December 2009.

[177] R. A. Rutenbar, G. G. E. Gielen, and J. Roychowdhury, "Hierarchical Modeling, Optimization, and Synthesis for System-Level Analog and RF Designs," *Proceedings of the IEEE*, Vol. 95, No. 3, pp. 640–669, March 2007.

[178] J. H. Mulligan Jr., "The Effect of Pole and Zero Locations on the Transient Response of Linear Dynamic Systems," *Proceedings of the Institute of Radio Engineers*, Vol. 37, No. 5, pp. 516–529, May 1949.

[179] J. J. Kelly, M. S. Ghausi, and J. H. Mulligan Jr., "On the Analysis of Composite Lumped Distributed Systems," *Elsevier Journal of Solid-State Electronics*, Vol. 284, No. 3, pp. 170–192, September 1967.

[180] A. Riccobono and E. Santi, "A Novel Passivity-Based Stability Criterion (PBSC) for Switching Converter DC Distribution Systems," *Proceedings of the IEEE International Applied Power Electronics Conference and Exposition*, pp. 2560–2567, February 2012.

[181] J. C. West and J. Potts, "A Simple Connection Between Closed-Loop Transient Response and Open-Loop Frequency Response," *Proceedings of the IEE - Part II: Power Engineering*, Vol. 100, No. 75, pp. 201–212, June 1953.

[182] J. Wagner and G. Stolovitzky, "Stability and Time-Delay Modeling of Negative Feedback Loops," *Proceedings of the IEEE*, Vol. 96, No. 8, pp. 1398–1410, August 2008.

[183] D. A. Johns and K. Martin, *Analog Integrated Circuit Design*, Wiley, 1997.

[184] J. E. Colgate, *The Control of Dynamically Interacting Systems*, Ph.D. Thesis, Massachusetts Institute of Technology, August 1988.

[185] J. L. Wyatt, L. O. Jr. Chua, J. Gannett, I. Goknar, and D. Green, "Energy Concepts in the State-Space Theory of Nonlinear n-Ports: Part I-Passivity," *IEEE Transactions on Circuits and Systems*, Vol. 28, No. 1, pp. 48–61, January 1981.

[186] O. Brune, *Synthesis of a Finite Two-Terminal Network whose Driving-Point Impedance is a Prescribed Function of Frequency*, Ph.D. Thesis, Massachusetts Institute of Technology, August 1931.

[187] X. Peng, S. Willy, L. Hou, J. Wang, and W. Wu, "Impedance Adapting Compensation for Low-Power Multistage Amplifiers," *IEEE Journal of Solid-State Circuits*, Vol. 46, No. 2, pp. 445–451, February 2011.

[188] P. J. Restle *et al.*, "A Clock Distribution Network for Microprocessors," *IEEE Journal of Solid-State Circuits*, Vol. 36, No. 5, pp. 792–799, May 2001.

[189] M. Popovich, M. Sotman, A. Kolodny, and E. G. Friedman, "Effective Radii of On-Chip Decoupling Capacitors," *IEEE Transactions on Very Large Scale Integration (VLSI) Circuits*, Vol. 16, No. 7, pp. 894–907, July 2008.

[190] I. Vaisband and E. G. Friedman, *Power Network On-Chip for Scalable Power Delivery*, United States Patent (pending).

[191] I. Vaisband and E. G. Friedman, "Power Network On-Chip for Scalable Power Delivery," *Proceedings of the Workshop on System Level Interconnect Prediction*, pp. 1–5, June 2014.

[192] I. Vaisband and E. G. Friedman, "Dynamic Power Management with Power Network-on-Chip," *Proceeding of the IEEE International Conference on Circuits and Systems*, June 2014.

[193] H. Esmaeilzadeh, E. Blem, R. St. Amant, K. Sankaralingam, and D. Burger, "Dark Silicon and the End of Multicore Scaling," *Proceedings of the ACM International Symposium on Computer Architecture*, pp. 365–376, June 2011.

[194] P. Kocher, J. Jaffe, and B. Jun, "Differential Power Analysis," *Proceedings of the International Conference on Advances in Cryptology*, pp. 388–397, August 1999.

[195] M. Alioto, L. Giancane, G. Scotti, and A. Trifiletti, "Leakage Power Analysis Attacks: A Novel Class of Attacks to Nanometer Cryptographic Circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 57, No. 2, pp. 355–367, February 2010.

[196] L. Lin and W. Burleson, "Leakage-Based Differential Power Analysis (LDPA) on Sub-90nm CMOS Cryptosystems," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 252–255, May 2008.

[197] T. Martin, M. Hsiao, D. S. Ha, and J. Krishnaswami, "Denial-Of-Service Attacks on Battery-Powered Mobile Computers," *Proceedings of the IEEE Annual Conference on Pervasive Computing and Communications*, pp. 309–318, March 2004.

[198] H. Vahedi, R. Muresan, and S. Gregori, "On-Chip Current Flattening Circuit with Dynamic Voltage Scaling," *Proceedings of the IEEE International Symposium on Circuits and Systems*, pp. 4277–4280, May 2006.

[199] K. Tiri and I. Verbauwhede, "A Digital Design Flow for Secure Integrated Circuits," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 25, No. 7, pp. 1197–1208, July 2006.

[200] O. A. Uzun and S. Kose, "Converter-Gating: A Power Efficient and Secure On-Chip Power Delivery System," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 4, No. 2, pp. 169–179, June 2014.