

# Charge Pump Circuit Design

- ✓ Architecture
- ✓ Concept to Implementation
- ✓ Operation and Analysis

Feng Pan  
Tapan Samaddar

---

## Charge Pump Circuit Design

---

# Charge Pump Circuit Design

**Feng Pan  
Tapan Samaddar**

## ABOUT THE AUTHORS

FENG PAN is doing circuit design at SanDisk Corporation, where he is involved in defining the architecture and implementing charge pumps and related high-voltage circuits over several generations of NAND Flash memory products. He is a holder of 18 U.S. design patents. Mr. Pan previously worked at AMD, where he contributed to the company's NOR Flash memory chip designs. He holds an MS degree from Stanford University and a BS degree from U.C. Berkeley.

TAPAN SAMADDAR is a design engineer at SanDisk Corporation, where he leads the company's high-voltage circuit designs. Previously, he was employed by T-RAM, Inc., where he was involved in the design of high-speed SRAM-compatible memory chips. He is a holder of several U.S. design patents. Mr. Samaddar has also worked on high-density NOR Flash memory at Atmel Corporation, and on high-speed cache memory designs at ST Microelectronics.

**McGraw-Hill**

New York Chicago San Francisco Lisbon London Madrid  
Mexico City Milan New Delhi San Juan Seoul  
Singapore Sydney Toronto

Copyright © 2006 by The McGraw-Hill Companies, Inc. All rights reserved. Printed in the United States of America. Except as permitted under the United States Copyright Act of 1976, no part of this publication may be reproduced or distributed in any form or by any means, or stored in a data base or retrieval system, without the prior written permission of the publisher.

# Charge Pump Circuit Design

1 2 3 4 5 6 7 8 9 0 DOC/DOC 019876

ISBN 0-07-147045-X

The sponsoring editor for this book was Wendy Rinaldi, the production supervisor was Jean Bodeaux, the editorial supervisor was Jody McKenzie, and the project manager was Samik Roy Chowdhury (Sam). It was set in New Century Schoolbook by International Typesetting and Composition. The art director for the cover was Margaret Webster-Shapiro.

Printed and bound by RR Donnelley.

McGraw-Hill books are available at special quantity discounts to use as premiums and sales promotions, or for use in corporate training programs. For more information, please write to the Director of Special Sales, McGraw-Hill Professional, Two Penn Plaza, New York, NY 10121-2298. Or contact your local bookstore.

Information contained in this work has been obtained by The McGraw-Hill Companies, Inc. ("McGraw-Hill") from sources believed to be reliable. However, neither McGraw-Hill nor its authors guarantee the accuracy or completeness of any information published herein and neither McGraw-Hill nor its authors shall be responsible for any errors, omissions, or damages arising out of use of this information. This work is published with the understanding that McGraw-Hill and its authors are supplying information but are not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought.

## Contents

|                                                             |           |
|-------------------------------------------------------------|-----------|
| Preface                                                     | ix        |
| Acknowledgments                                             | xi        |
| Introduction                                                | xiii      |
| <b>Chapter 1. History of the High-Voltage Charge Pump</b>   | <b>1</b>  |
| 1.1 Using a Transformer to Generate High Voltages           | 1         |
| 1.2 The Cockcroft-Walton High-voltage Charge Pump           | 3         |
| 1.3 The Dickson Charge Pump                                 | 5         |
| 1.3.1 The body effect                                       | 6         |
| 1.3.2 Implication of body effect on the Dickson charge pump | 7         |
| 1.4 Better Solutions                                        | 8         |
| <b>Chapter 2. Basic MOS Device Physics</b>                  | <b>11</b> |
| 2.1 The P-N Junction                                        | 12        |
| 2.1.1 Reverse bias                                          | 13        |
| 2.1.2 Forward bias                                          | 14        |
| 2.1.3 P-N junction diode characteristics                    | 16        |
| 2.2 The MOS Capacitor                                       | 17        |
| 2.2.1 The flat band condition                               | 17        |
| 2.2.2 Accumulation                                          | 18        |
| 2.2.3 Depletion                                             | 19        |
| 2.2.4 Weak inversion                                        | 21        |
| 2.2.5 Strong inversion                                      | 21        |
| 2.3 Capacitance Variation of a MOS                          | 22        |
| 2.4 The Threshold Voltage                                   | 23        |
| 2.5 Metal-Oxide Field-Effect Transistor                     | 27        |
| 2.5.1 Device operation                                      | 27        |
| 2.5.2 Second-order effects for MOSFET operation             | 32        |
| 2.6 Latch-up in CMOS Technology                             | 33        |
| 2.7 Merits of PMOS Versus NMOS in Circuit Design            | 35        |
| 2.8 The MOSFET Model                                        | 36        |
| 2.9 SPICE Simulation Convergence                            | 38        |
| 2.10 Conclusion                                             | 40        |

|                                                          |            |
|----------------------------------------------------------|------------|
| <b>Chapter 3. Basic Operation of a Charge Pump</b>       | <b>41</b>  |
| 3.1 Charge Pump System                                   | 41         |
| 3.2 Basic Concept: The Bucket Capacitor Model            | 43         |
| 3.3 The Dickson Charge Pump                              | 45         |
| 3.4 Dynamic Analysis of the Charge Pump                  | 49         |
| 3.4.1 The body effect revisited                          | 54         |
| 3.5 Dynamic Analysis of the Charge Pump with Body Effect | 55         |
| 3.6 Conclusion                                           | 57         |
| <b>Chapter 4. Charge Pump Design Criteria</b>            | <b>59</b>  |
| 4.1 Technology                                           | 60         |
| 4.1.1 System supply voltage                              | 61         |
| 4.1.2 Silicon dioxide ( $\text{SiO}_2$ )                 | 66         |
| 4.1.3 Resistor                                           | 70         |
| 4.1.4 Transistor specification                           | 72         |
| 4.2 Specification                                        | 75         |
| 4.2.1 Output load characteristics                        | 75         |
| 4.2.2 Pump output voltage                                | 77         |
| 4.2.3 Pump output current                                | 80         |
| 4.2.4 Ripple on regulated output voltage                 | 82         |
| 4.2.5 Pump regulation                                    | 85         |
| 4.2.6 Capacitive divider                                 | 85         |
| 4.2.7 Resistive divider                                  | 87         |
| 4.2.8 MOSFET biased type regulator                       | 88         |
| 4.2.9 Which scheme should be used in design?             | 89         |
| 4.3 Pump Power Consumption                               | 90         |
| 4.4 Die Size of the Charge Pump                          | 90         |
| 4.5 Conclusion                                           | 91         |
| <b>Chapter 5. How to Design a Basic Charge Pump</b>      | <b>93</b>  |
| 5.1 Charge Pump Specifications                           | 94         |
| 5.1.1 Output voltage                                     | 95         |
| 5.1.2 Current drivability                                | 95         |
| 5.1.3 Output ramp-up time and recovery time              | 98         |
| 5.1.4 Power consumption                                  | 99         |
| 5.2 Pump Clock Source                                    | 100        |
| 5.3 Regulator Design                                     | 102        |
| 5.4 Non-overlapping Clock Generator                      | 102        |
| 5.5 Cross-coupled Voltage Doubler Design                 | 104        |
| 5.6 Logical Effort for Clock Buffer Sizing               | 106        |
| 5.7 Parasitic R and C                                    | 114        |
| 5.8 Power Bus and Bower Bus Capacitance                  | 116        |
| 5.9 Conclusion                                           | 117        |
| <b>Chapter 6. Designing a Better Charge Pump</b>         | <b>119</b> |
| 6.1 Parameters Associated with Pump Performance          | 120        |
| 6.1.1 Charge transfer point of view                      | 121        |
| 6.1.2 Voltage point of view                              | 125        |

|                                                                                 |            |
|---------------------------------------------------------------------------------|------------|
| 6.2 How to Improve Charge Pump Efficiency                                       | 128        |
| 6.2.1 V <sub>t</sub> cancellation scheme                                        | 129        |
| 6.2.2 Pump design using high-amplitude pump clocks                              | 130        |
| 6.3 Regulation of the Pump                                                      | 132        |
| 6.3.1 Resistive divider versus capacitive divider                               | 132        |
| 6.3.2 Regulation controls                                                       | 135        |
| 6.3.3 Noise control for regulation                                              | 136        |
| 6.4 Power Consumption versus Pump Performance                                   | 136        |
| 6.5 Charge Pump Area Efficiency                                                 | 141        |
| 6.6 Layout Requirements for Pump Design                                         | 142        |
| 6.6.1 Parasitic capacitance                                                     | 143        |
| 6.6.2 Miller effect (parasitic capacitance)                                     | 145        |
| 6.6.3 Junction capacitance                                                      | 146        |
| 6.6.4 Improving layout efficiency per unit area                                 | 148        |
| 6.6.5 Well resistance                                                           | 150        |
| 6.6.6 Critical signal width                                                     | 153        |
| 6.6.7 Clock buffering                                                           | 155        |
| 6.6.8 Power bus and decoupling capacitance                                      | 158        |
| 6.7 Conclusion                                                                  | 161        |
| <b>Chapter 7. Different Charge Pump Architectures</b>                           | <b>163</b> |
| 7.1 The 2-Phase Positive Charge Pump—Revisited                                  | 163        |
| 7.2 The 4-Phase Positive Charge Pump                                            | 166        |
| 7.3 The Modified 2-Phase Positive Charge Pump with Doubled Pump Clock Amplitude | 178        |
| 7.4 The Static CTS Charge Pump                                                  | 183        |
| 7.5 The Positive Charge Pump with Very High Amplitude Pump Clocks               | 185        |
| 7.6 The 2-Phase Negative Charge Pump                                            | 186        |
| 7.7 The 2-Phase Negative Charge Pump with Triple Well Technology                | 188        |
| 7.8 Conclusion                                                                  | 190        |
| <b>Chapter 8. Future Design References</b>                                      | <b>193</b> |
| 8.1 Area Versus Performance                                                     | 193        |
| 8.1.1 Output performance versus pump clock frequency                            | 193        |
| 8.1.2 Output performance versus pump clock amplitude                            | 196        |
| 8.2.3 Output performance versus number of pump stages                           | 196        |
| 8.2 Power Consumption                                                           | 198        |
| 8.2.1 Power consumption versus pump clock frequency                             | 198        |
| 8.2.2 Power consumption versus number of pump stages                            | 199        |
| 8.2.3 Power consumption versus pump clock amplitude                             | 200        |
| 8.3 Noise Controls for the Charge Pump                                          | 202        |
| 8.3.1 Noise versus filtering capacitance                                        | 202        |
| 8.3.2 Noise versus balance of pump power                                        | 204        |
| 8.4 Off-Chip Charge Pump                                                        | 207        |
| 8.4.1 Off-chip capacitive charge pump                                           | 208        |
| 8.4.2 Off-chip inductive charge pump                                            | 211        |
| 8.5 Conclusion                                                                  | 213        |

|                                                                                          |     |
|------------------------------------------------------------------------------------------|-----|
| Chapter 9. A Practical Charge Pump Design Example and Analysis                           | 215 |
| 9.1 Design Specification                                                                 | 217 |
| 9.2 Design Steps                                                                         | 218 |
| 9.2.1 Step 1: Determine $N$ , the initial number of stages                               | 218 |
| 9.2.2 Step 2: Determine $f$ , the initial pump operating frequency                       | 219 |
| 9.2.3 Step 3: Determine $C$ , the initial size of the pump capacitor                     | 219 |
| 9.2.4 Step 4: Determine the diode $W/L$ , the initial size of the diode-connected MOSFET | 220 |
| 9.3 Initial Simulation and Analysis                                                      | 220 |
| 9.4 Pump Performance Characterization                                                    | 228 |
| 9.4.1 Pump efficiency calculation                                                        | 228 |
| 9.4.2 Pump $I$ - $V$ characteristics                                                     | 231 |
| 9.4.3 Pump output current versus pump clock frequency                                    | 232 |
| 9.4.4 Pump output current versus MOSFET sizes                                            | 233 |
| 9.4.5 Pump output current versus clock driver size                                       | 234 |
| 9.4.6 Design summary                                                                     | 235 |
| 9.5 Conclusion                                                                           | 236 |

|       |     |
|-------|-----|
| Index | 239 |
|-------|-----|

## Preface

Charge pumps are finding increased attention and novel diversified usage in the new era of nanometer-generation chips used in different systems, specifically those incorporating nonvolatile memory. Many of the present and future nanometer-generation chips' performance depend heavily on the ability to efficiently generate high voltages on-chip while meeting stringent power and area requirements. And yet, charge pump design, being purely analog in nature and involving high voltage, needs meticulous design techniques, intensive semiconductor device analysis, careful design layout planning, and accurate parasitic extraction process to produce excellent results in real implementation on silicon.

This book is a product of our years of quest for a practical book on charge pump circuit design. Having made significant contributions in different successful projects at various companies, we have always felt the need for a book on the topic of charge pump design. From our early days we constantly felt the challenge of working on a subject where there are no books and our sources of information were limited to only a few pages of description on various text books and different IEEE-published papers and journals. Most of these documents, while giving us skeletal ideas about the basic architecture and enhancements of charge pumps, did nothing to guide us intricately through different design conceptions and implementation processes. Charge pump, being a pure analog design, carries an inherent risk of unanticipated effects, which when overlooked can significantly reduce circuit performance or cripple the circuit operation.

Both of us have individually looked for books or materials on this topic and asked colleagues and veterans for any published materials, but we soon found out that everyone wished there was a book on this particular topic. By writing this book we have tried to bring our combined personal design experiences along with the knowledge assimilated from our study of different journals and research papers. This book covers the basics of charge pump circuits in detail and provides a thorough mathematical derivation and analysis of charge pump operation. It also strives

to explain the different aspects for an excellent charge pump design, and explains each step in detail. Every effort has been made to provide enough hands-on design information, potential pitfalls to avoid, and practical ideas harnessed from our years of designing charge pumps, which are being used in many chips, finding mass scale adoption.

This book assumes a basic knowledge of semiconductor device physics and MOSFET operation and is targeted toward almost every semiconductor chip design engineer who is involved in analog circuit design and memory circuit design. The book takes a relatively novice reader through various aspects and gives sufficient information to enable him or her to complete their design from conception to actual layout implementation. Further audiences include systems designers and board level integrators. Also this book should be essentially helpful to almost all electrical engineering professors and students at all levels.

This book is organized into nine chapters. Chapter 1 starts with a history of charge pump evolution from the Noble award winning work of Cockcroft and Walton to the eventual adaptation of John F. Dickson. Chapter 2 is intended for a quick refresh of basic MOS device physics and different second order effects relevant to charge pump operation. It also discusses SPICE simulators and BSIM models while providing suggestions to avoid the dreaded SPICE convergence issue. Chapter 3 dives straight into the heart of charge pump operation and quantitatively analyzes the pump characteristics. Chapter 4 is where the basic implementation details are discussed along with different charge pump controlling blocks. It analyzes the different pump regulation schemes and quantifies them. Chapter 5, a prelude to Chapter 9, introduces the different parameters for charge pump specifications and discusses various implementation details. Now, once the basic operation of a 2-phase charge pump and its many characteristics and specifications have been understood, Chapter 6 takes it to the next level and discusses how to design a better charge pump. Chapter 7 discusses various charge pump architectures, such as the modified 2-phase charge pump, the 4-phase charge pump, and the CTS charge pump. Chapter 8 provides future design references and discusses different circuit and system effects that affect the performance of the pump. Finally, Chapter 9 provides a practical design example and discusses the influence of different parameters while analyzing the characteristics of a charge pump.

It is believed that if the reader applies appropriate techniques, which have been presented here, he or she should be able to design charge pumps that meet design and performance specifications and produce excellent operating circuits on silicon.

*Feng Pan  
Tapan Samaddar*

## Acknowledgments

I wish to acknowledge all the individuals who provided tremendous help and support. A special thanks go to Wendy Rinaldi and Alex McDonald, acquisitions editors, for believing in us as first-time writers, and guiding and supporting us throughout the entire duration of this project. I am thankful for Samik Roy Chowdhury (Sam) and his team at International Typesetting and Composition for all their assistance and hard work in making this timely production.

My co-author, Tapan Samaddar, for sparking the idea to write this book on charge pump design, and for his inspiring work in researching and completing this book.

My wife, Judy, for her patience and understanding. During the last year while she was pregnant, I had to work many late nights and weekends on this book. She encouraged me and supported me in many ways. I am very grateful to her. My lovely daughter, Tiffany, who was born in early April this year, for her trust and understanding in her daddy, and for forgiving her daddy for not being able to give her time. I am indebted to her. I would also like to thank my parents and my brother for their guidance and support throughout my life.

Many thanks to those who supported me in the past to make this book production possible.

*Feng Pan*

Writing this book began with a lot of energy and excitement. However, after about six months of persistent writing, drawing, and revising, while meeting tough deadlines, schedules, and surprises at the office, we reached midway in the book and immediately began to feel the streaks of insanity, realizing that the book will never finish on time. It was only through the continued efforts, support, and help of many outstanding individuals that we were able to complete this book.

A special thank you goes to Wendy Rinaldi and Alex McDonald, who have given us much needed guidance, helped encourage the progress, and collated the manuscript. Great appreciation is also given to Sam and his team, who did a marvelous job editing the manuscript, shaped it into a book, and helped finish everything on time.

My loving wife, Somali, has made many contributions, from typing characters and drawing figures to finding numerous errors and raising questions that made me reexamine my own understanding. I am very grateful to her for her support. I would also like to thank my mother and my father for providing me with emotional support and for all the help they have given me throughout the years.

Did I forget anyone? Thanks to all of you who inspired and provided support to write this book.

*Tapan Samaddar*

## Introduction

The approaching nanometer generation of large-scale integrated (LSI) circuits requires power-supply voltages of less than 2 V to enable low-power operation and increased battery life. In addition, because low-power technology has become a primary target for the mainstream LSI designs to meet with the nomadic computing era, oxide thickness, transistor dimensions, and voltage-scaling approaches that are most effective for low power are accelerating and spreading at an unprecedented pace. Therefore, low-voltage circuit technologies for processors, memory, and analog circuits are intensively being investigated.

Many of the system blocks—such as EEPROMs, Flash memories, power management blocks, audio and video codecs, image sensor circuits, and displays—require internal voltages higher than the system supply voltage. This internal high-voltage supply needs to be generated in-system or on-chip. The traditional approach of switch-capacitor circuits or inductor-based linear regulators consumes too much power and silicon area to justify today's shrinking needs. An on-chip charge pump design provides an excellent solution and eliminates the need for an inductor. Having no inductor alleviates any potential electromagnetic-interference concerns that could have an impact on sensitive RF receivers or wireless chipsets. Another advantage is to reduce the cost of using discrete off-chip components. The charge pump solution eliminates the need for DC/DC boost converters and expensive low-profile inductors that are required to meet the size limitations of handheld devices and cellphones.

The advent of personal information devices, digital cameras, and MP3 players has fuelled a boom for nonvolatile memories, particularly Flash memories, because of their high-density, moderate power consumption and high endurance for mechanical shock and vibration. Solid state memories, such as Flash, contain no moving parts and allow for easy and fast data storage. EEPROMs and Flash memory have been some of the biggest drivers to create better and efficient charge pumps. Even though the supply voltage is decreasing, a Flash memory will still need a high internal programming/erase voltage, up to 30 V, regardless of the

power supply trend, and this voltage also needs to be controlled precisely to achieve a narrow deviation in the threshold voltages of its memory-cell transistors.

A familiar problem in system engineering is the subsystem whose power requirements are not met by the main power supply. In such cases, the available supply rails are not directly usable, nor is the direct use of battery voltage (when available) always an option. Lack of space can prevent inclusion of the optimal number of batteries, and in other cases the gradually declining voltage of a discharging battery is not acceptable for the application. Voltage converters can generate the desired voltage levels, and charge pumps are often the best choice for these applications requiring some combination of low power, simplicity, and low cost. Charge pumps are easy to use, because they require no expensive inductors or additional discrete components. Further, charge pumps can be the only option for certain applications, such as those in telecom applications, which require generating +5 V from the available -48 V.

With the increasing popularity of color LCD displays in cellphones, PDAs, and digital cameras, white LEDs are becoming popular illumination sources. Whereas monochrome displays can use colored light sources, such as electroluminescent backlights or colored LEDs, color displays require a white light source to properly display color. The ubiquitous red and green LEDs have a typical forward-voltage drop of about 1.6 V to 2.4 V and can be driven by a simple battery pack. White LEDs, however, typically have a forward-voltage drop of 3 V to 4 V and are more likely to need a separate power supply. The traditional direct LED drivers and inductor-based boost converters have their own limitations. Charge pumps are increasingly finding ground in these applications due to their better performance—they provide the smallest and lowest-cost solution because they rely only on small capacitors for high voltage generation. As video features become more integrated into mobile phone use, improvements in power consumption for LCD backlighting are essential for the maintenance and improvement of overall battery life. Chip vendors competing for a share of this red-hot white LED driver market are pitching advanced charge pump sources that deliver greater backlighting power for the color displays in larger and more complex portable/wireless devices.

Therefore, much of the present and future chip performance will depend heavily on the ability to efficiently generate high voltages on-chip while meeting stringent power and area requirements. And yet, charge pumps, being purely analog in nature and involving high voltage, need meticulous design techniques, intensive semiconductor device analysis, careful design layout planning, and an accurate parasitic extraction process to produce excellent results in real implementation on silicon.

To summarize, charge pumps are finding increased attention and novel diversified usage in the new era of nanometer-generation chips used in

different systems. This book strives to explain the different aspects for an excellent charge pump design, and explains each step in detail. It is full of extra hands-on design information, potential pitfalls to avoid, and practical ideas harnessed from the authors' years of charge pump design experience, which are currently being used in many chips, finding mass scale adoption.

## Basic MOS Device Physics

In today's evolving chip-manufacturing industry, a solid understanding of semiconductor device physics is essential for the design and implementation of circuits that are analog in nature. In general, all circuits, whether a simple half adder or an operational amplifier, are analog in nature, while the digital abstraction is only an abstraction as long as certain design criteria are met. For its operation, an analog circuit not only relies on basic device operation, but its performance also is dictated by many second-order effects.<sup>1</sup> Because the circuit designer needs to make decisions about which effects need to be considered and which effects can be ignored, a thorough insight into the device physics is important. In this chapter, we will study the basics of MOSFETs (metal oxide semiconductor field-effect transistors) at an elementary level, covering the necessary topics to understand basic circuit operations. We will also talk about secondary effects of MOSFET in detail, specifically those are important for charge pump operation.

The forward- and reverse-biased p-n junctions form the basics of semiconductor technology, and the properties of forward- and reverse-biased junctions have an important influence on the characteristics of many semiconductor devices. For example, reverse-biased p-n junctions, formed by creating an n<sup>+</sup> junction or an n-well in the p-substrate, exist in almost every integrated circuit. These junctions contribute as voltage-dependent parasitic capacitances and leakage current components. Further, a number of important characteristics of the active devices, such as the basic operation of the MOS device, breakdown voltage, and output resistance, depend directly on the properties of the depletion region of the p-n junction. To address its importance and usage at many different places, particularly for understanding charge pump operation, we provide a brief analysis of the forward and reverse-biased p-n junction diode device characteristics first.

## 2.1 The P-N Junction

N-type silicon has many mobile electrons, whereas p-type silicon has many mobile holes. When these two types of semiconductors are brought together in contact, two events occur at the interface due to the diffusions of majority carriers:

1. Electrons migrate from n-type material into p-type material, leaving behind uncompensated donors (+ ions).
2. Holes pour out of p-type material and move toward n-type material, leaving behind uncompensated acceptors (– ions).

The electrons and holes then recombine near the interface, producing a region depleted of free carriers, leaving behind regions with only fixed charges (ionized donors and acceptors).

The diffusion process cannot continue indefinitely because the depleted region creates an electric field whose direction opposes the diffusion of majority carriers (electrons in n-type and holes in p-type). The electric field will sweep minority carriers (holes in n-type and electrons in p-type) across the junction so that there is a drift current of electrons from the p-type side toward the n-type side and of holes from the n-type side toward the p-type side, which is in the opposite direction to the diffusion current. The junction field builds up until these two current flows are equal and equilibrium is established (no net current flow), as shown in Figure 2-1. This process also gives rise to a parasitic (depletion) capacitance.

The energy band diagram for a p-n junction with zero voltage bias can be seen in Figure 2-2.

The zero bias built-in potential can be expressed as

$$\Phi_0 = V_t \ln \left( \frac{N_A N_D}{n_i^2} \right) \quad (2-1)$$

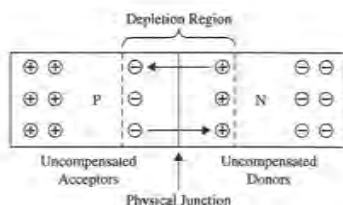


Figure 2-1 P-N junction diode.

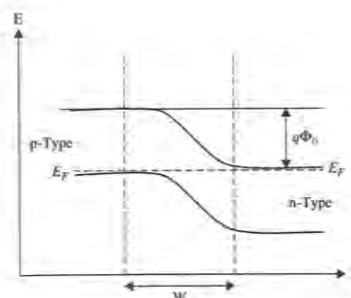


Figure 2-2 The energy band diagram for a p-n junction at 0 V bias.

where  $V_t = \frac{kT}{q}$  is the thermal equivalent voltage, which is close to 26 mV at room temperature,  $N_A$  and  $N_D$  are the doping concentrations of the n-type and p-type semiconductors, respectively, and  $n_i$  is the intrinsic carrier concentration of silicon, which is about  $1.45 \times 10^{10}$  atoms/cm<sup>3</sup>.

The depletion capacitance  $C_j$  of the p-n junction diode can be expressed as

$$C_j = \frac{C_{j0}}{\left[ 1 - \left( \frac{V_d}{\Phi_0} \right) \right]^m} \quad (2-2)$$

where  $V_d$  is forward bias voltage applied across the junction,  $C_{j0}$  is the zero-bias capacitance (that is, the effective capacitance when  $V_d = 0$  V), and  $m$  is the grading profile, which usually varies from 0.2 to 0.5.

### 2.1.1 Reverse bias

If we apply a bias,  $V_B < 0$ , by connecting the positive terminal of a DC source to the n-side and the negative terminal to the p-side, the process causes electrons from the n-side to become attracted toward the positive terminal, and vice versa. This changes the depth of the depletion area, which becomes wider, and the potential barrier grows higher.

As can be seen from the energy band diagram in Figure 2-3, the built-in potential is augmented by the applied reverse bias,  $V_B$ , and the total voltage across the junction becomes  $\Phi_0 + V_B$ . The barrier is so high

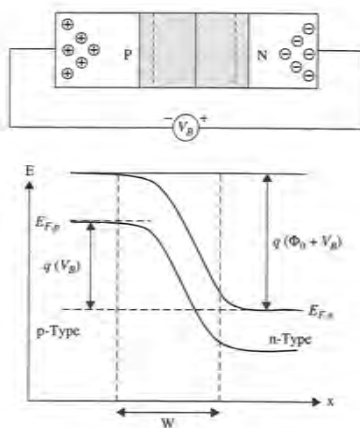


Figure 2-3 P-N junction diode at reverse bias.

that few electrons can cross from the n-type to the p-type region, thus reducing the diffusion current almost to zero.

Nonetheless, enough electron/hole pairs are generated in the depletion region, and these cause a small drift current as they are swept across the depletion region.

The drift current is relatively insensitive to the height of the potential barrier because it is created by random electron/hole pair generation and all the minority carriers generated may diffuse to the depletion region and be swept across it, whatever the size of reverse potential.

### 2.1.2 Forward bias

Next, if we apply a bias,  $V_B > 0$ , by connecting the positive terminal of a DC source to the p-side and the negative terminal to the n-side, the barrier height will be reduced to  $\Phi_0 - V_B$  and the depletion region will narrow, as shown in Figure 2-4.

Consequently, majority carriers are able to surmount the potential barrier much more easily than in the equilibrium case so that the diffusion current becomes much larger than the drift current. Once the excess holes cross into the n-type region, they recombine with the electrons. This process is the same for electrons crossing into the p-type region.

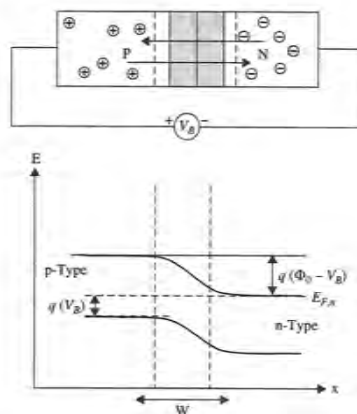


Figure 2-4 P-N junction diode at forward bias.

In equilibrium, as holes diffuse away, they must be met by a constant supply of electrons with which they recombine. Therefore, the current must be supplied at a rate that equals the concentration of holes at the edge of the depletion region. Thus, the current due to hole injection can be expressed as follows:

$$J_p = q \frac{D_p}{L_p} p_{n0} \left( e^{V_B/V_i} - 1 \right) \quad (2-3)$$

Similarly, the current due to electron injection can be expressed as follows:

$$J_n = q \frac{D_n}{L_n} n_{p0} \left( e^{V_B/V_i} - 1 \right) \quad (2-4)$$

The total net current is the sum of the above two equations, 2-3 and 2-4. Hence

$$I = A(J_p + J_n) \quad (2-5)$$

Substituting equations 2-3 and 2-4 in equation 2-5 we get

$$I = qA \left( \frac{D_p}{L_p} p_{n0} + \frac{D_n}{L_n} n_{p0} \right) \left( e^{V/V_t} - 1 \right) \quad \text{or} \quad I = I_s \left( e^{V/V_t} - 1 \right) \quad (2-5a)$$

where  $I_s$  is known as the reverse leakage current

$$I_s = qA \left( \frac{D_p}{L_p} p_{n0} + \frac{D_n}{L_n} n_{p0} \right) \quad (2-6)$$

As the equation indicates, this current is very strong temperature and bias dependent. We can also see that for forward bias (positive  $V$ ) the net current increases exponentially with voltage, whereas for reverse bias (negative  $V$ ) the current is essentially constant and equal to  $-I_s$ .

### 2.1.3 P-N junction diode characteristics

Figure 2-5 shows the I-V characteristic of a p-n junction diode. The diode's I-V characteristic can be approximated by two regions of operation. Below a certain bias voltage, the depletion layer has significant width, and the diode can be thought of as an open (nonconductive) circuit. As the bias voltage is increased, at certain conditions the diode will become conductive and allow charges to flow through from the p-side to the n-side, at which point it can be thought of as a closed switch with almost

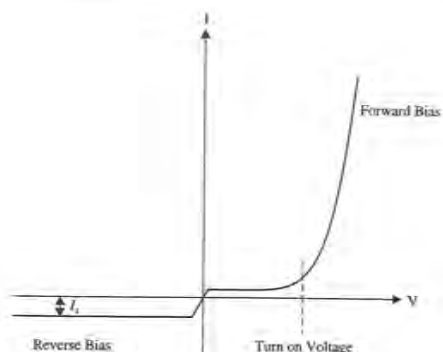


Figure 2-5 I-V characteristics of a p-n junction diode.

zero impedance. Actually, because the transfer function is exponential, a little increase in forward voltage bias causes a large conduction of current. The positive bias voltage above which the diode becomes conducting is actually the built-in potential that needs to be overcome for conduction to occur and it is generally close to 0.6 V.

During the reverse bias, the current through the device is very low (in the nA range) for all reverse voltages in general. Special diodes, such as avalanche and zener diodes, operate in the reverse bias region, such that the reverse voltage is "clamped" to a known value (called "zener voltage") and is often used for voltage regulation.

## 2.2 The MOS Capacitor

Before we begin our discussion of the MOSFET, it is essential to achieve a satisfactory understanding of the MOS capacitor fundamentals. Consider the n-type MOSFET in capacitor<sup>2-5</sup> configuration in Figure 2-6. The drain and source are grounded, and the gate is connected to a variable-voltage source. The  $p^+$  substrate connection is also grounded. When the gate bias  $V_{gs} = 0$  V, there is no potential difference across the gate and the substrate and hence there are no stored charges across the two conducting plates. The MOSFET can be considered to be in a steady state at this time.

### 2.2.1 The flat band condition

As shown in Figure 2-7, in an ideal MOS capacitor, the metal work function,  $\Phi_m$ , is equal to the semiconductor work function,  $\Phi_s$ . Therefore, the Fermi level of the semiconductor,  $E_{FS}$ , is aligned with the Fermi level of the gate,  $E_{FG}$ . There is no band bending in any region of the MOS capacitor. This assumes that the gate dielectric does not have any trapped charges and the semiconductor is uniformly doped.

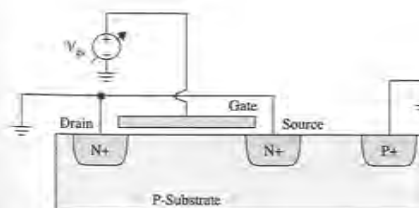


Figure 2-6 MOS transistor structure.

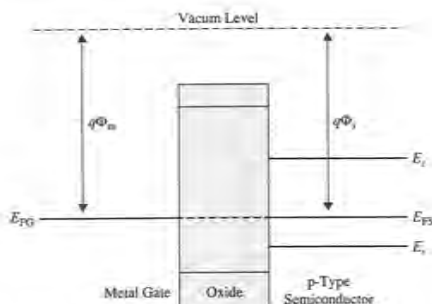


Figure 2-7 Flat band diagram at steady state.

### 2.2.2 Accumulation

When the gate bias,  $V_{gs}$ , is negative, the gate acquires a negative charge. The source of this negative charge is electrons supplied by the voltage source. Because charge neutrality is always maintained across the MOS capacitor, a net positive charge must be available in the silicon substrate to counterbalance the negative charge stored on the gate. This is achieved by an accumulation of majority carrier holes under the gate, as shown in Figure 2-8. (Note: Because the silicon is p-type, the majority carriers are holes.) This condition, during which the majority carrier concentration is greater near the  $S_i - SiO_2$  interface, compared to the substrate bulk, is called *accumulation*.

During a negative gate bias, the gate's Fermi level is increased with respect to the Fermi level of the silicon substrate by an amount equal

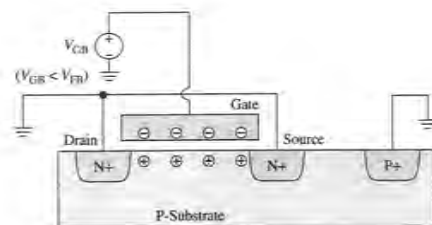


Figure 2-8 MOS transistor in accumulation.

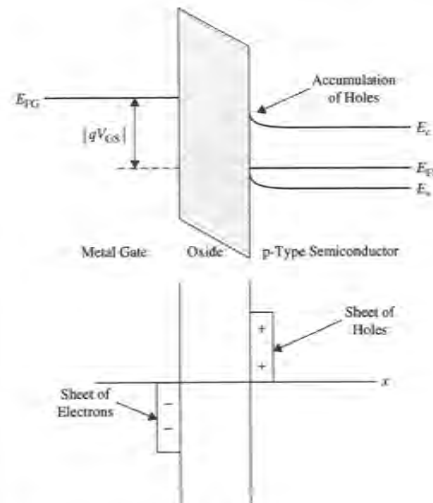


Figure 2-9 Fermi level at accumulation.

to  $qV_{gs}$  as shown in Figure 2-9. The accumulation of holes under the gate dielectric causes the energy bands in the silicon substrate to bend upward, and this brings the valence band closer to the Fermi level. Yet, during this whole process, the Fermi level in the substrate is not changed because there is no net flow of current from the gate to the substrate.

### 2.2.3 Depletion

As shown in Figure 2-10, as  $V_{gs}$  is increased from negative to slightly positive (not negative enough to attract a lot of holes and not negative enough to attract a lot of electrons), the surface under the gate is depleted—i.e., the holes under the gate are pushed away, leaving behind ionized, negatively charged acceptor atoms, thus creating a depletion region. The charge in the depletion region, rising from the ionized acceptor atoms, is exactly equal to the charge on the gate in order to preserve charge neutrality.

As opposed to the accumulation phenomenon, with a positive gate bias, the Fermi level of the gate is lowered with respect to the Fermi level of

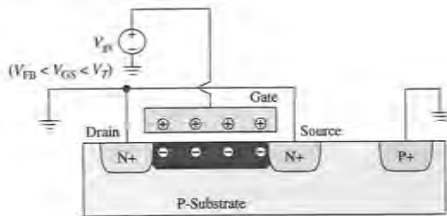


Figure 2-10 MOS transistor in depletion.

the substrate, during the depletion phase, as shown in Figure 2-11. The energy bands bend downward, resulting in a positive surface potential. Under the influence of the positive gate bias, hole depletion under the gate causes the valence band to move away from the Fermi level. As the gate potential is increased slowly, the depletion will slowly cause a band bending at the surface, such that at a particular time the intrinsic level will coincide with the Fermi level, and the surface will resemble an intrinsic material.

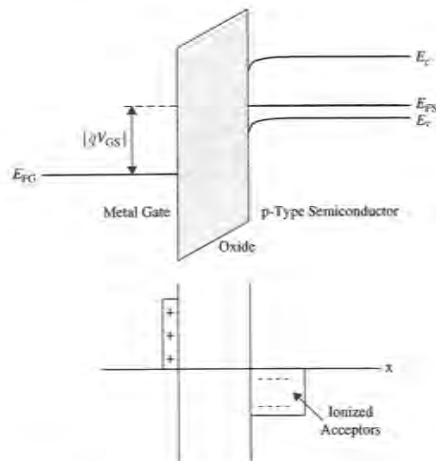


Figure 2-11 Fermi level at depletion.

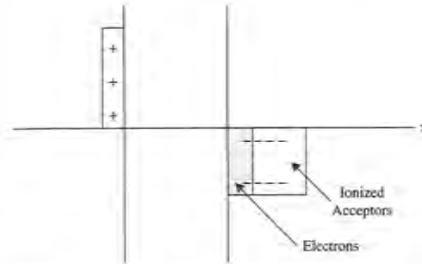


Figure 2-12 Fermi level at weak inversion.

#### 2.2.4 Weak inversion

When the gate potential reaches above a certain threshold, as shown in Figure 2-12, the region under the gate oxide begins to attract more electrons. With electrons accumulated near the surface of the substrate, the surface begins to change gradually from intrinsic type to n-type. Because this makes the surface of opposite polarity, the region under the gate is termed "inverted." The negative charge in the semiconductor is composed of ionized acceptor atoms in the depletion region and free electrons in the inversion layer. At this onset of inversion, the electron concentration at the surface is still less than the hole concentration in the neutral bulk. Thus, this condition is referred to as "weak inversion" because the surface under the oxide is not heavily  $n^+$ .

#### 2.2.5 Strong inversion

When  $V_{gs}$  is sufficiently large, such that a large number of electrons are attracted under the gate, the surface of p-substrate is said to be "strongly inverted" (that is, no longer p-type, as shown in Figure 2-13).

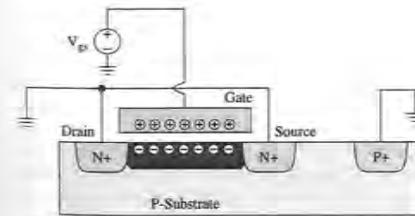


Figure 2-13 MOS transistor in strong inversion.

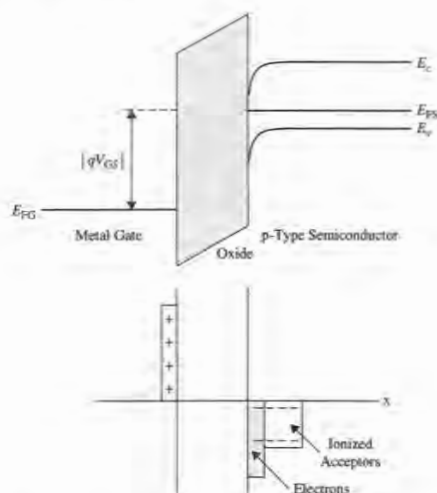


Figure 2-14 Fermi level at strong inversion.

With the increased gate bias, the band bending increases, as shown in Figure 2-14. The depletion region becomes deeper and the electron concentration in the inversion layer increases. When the electron concentration under the gate approaches the hole concentration in the bulk, a strong inversion layer is said to be formed. The surface potential required to achieve strong inversion is defined as the threshold voltage of the MOSFET transistor.

### 2.3 Capacitance Variation of a MOS

To understand the C-V curve of n-type MOSFET transistor shown in Figure 2-15, let us start with a very negative gate-source voltage bias. The negative potential on the gate attracts a lot of holes from the substrate to under the gate. During this condition, the MOSFET is operating in the strong accumulation region and can be viewed as a capacitor having unit capacitance of  $C_{ox}$  because the whole capacitance is essentially formed between the gate and the substrate, and is separated by gate oxide.

It is important to understand that during this time, the capacitance between gate and substrate runs through the large parasitic

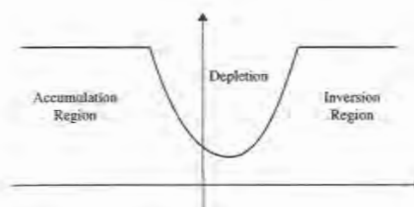


Figure 2-15 C-V curve of n-type MOSFET transistor.

resistance of the substrate, and hence this configuration may not have equivalent unit capacitance of  $C_{ox}$  when using the MOSFET at a high frequency.

Next, as  $V_{gs}$  rises, the depletion region starts to form under the gate oxide. The concentration of the holes under the gate starts to reduce, and the MOSFET slowly enters into the weak inversion region. Under this condition, the equivalent MOSFET capacitance consists of the gate oxide capacitance in serial with the capacitance from the depletion layer in the substrate. The formation of these serial capacitances causes a reduction in the total equivalent gate to substrate capacitance for the MOSFET.

Finally, as  $V_{gs}$  is raised further and above MOSFET threshold voltage  $V_t$ , the MOSFET goes from the weak inversion region into the strong inversion region. The substrate under the gate attracts enough electrons to sustain a channel between the source/drain of MOSFET transistor. The equivalent unit MOSFET capacitance once again becomes  $C_{ox}$ . This is the best biasing condition for a MOSFET to act as a capacitor.

### 2.4 The Threshold Voltage

The first parameter of interest that characterizes the switching behavior of the MOSFET device is the threshold voltage,  $V_t$ .<sup>2-5</sup> This is defined as  $V_{gs}$  voltage at which a MOSFET starts to conduct between source and drain of the transistor. We can graph the relative conduction against the difference in gate-to-source voltage in terms of the source-to-drain current ( $I_{ds}$ ) versus the different gate-to-source voltage ( $V_{gs}$ ). These graphs for a fixed drain-source voltage,  $V_{ds}$ , are shown in Figure 2-16. It is also possible to make n-channel MOSFET transistors to conduct when the gate voltage is equal to the source voltage (i.e.,  $V_{gs} = 0$  V), whereas others require a positive difference between the gate and the source voltages to start conduction (negative for PMOS devices). Those devices that are normally non-conducting are classified as "enhancement-mode devices,"

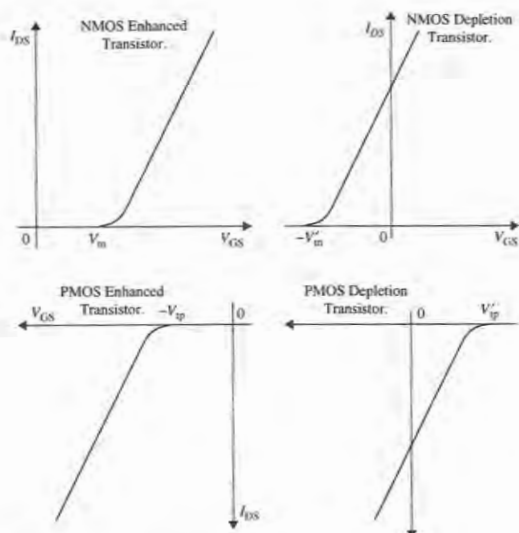


Figure 2-16 Forward bias, reverse bias characteristics of MOSFETs.

whereas those devices that can conduct even with 0 V gate bias are called "depletion-mode devices." The n-channel transistors and p-channel transistors are the dual opposites of each other—that is, the voltage polarities required for correct operation are opposite. The threshold voltages for n-channel and p-channel devices are denoted by  $V_{TN}$  and  $V_{TP}$ , respectively.

The threshold voltage of a MOSFET capacitor can also be realized as the gate voltage,  $V_{GB}$ , required to create and maintain a strong inversion region underneath the gate. Under this condition, the gate can maintain a large concentration of electrons under it and the electrons can flow from the source to the region in substrate underneath the gate and eventually to the drain. Thus, a channel of charge carriers is formed under the gate between source and drain regions when the MOSFET is turned on. During this time, we can also say that the MOSFET is inverted. In reality, the MOSFET turn-on phenomenon is a gradual function of the gate-to-source biasing voltage, making it difficult to determine  $V_t$  precisely.

As previously mentioned, the surface underneath the gate of n-type MOSFET is inverted from p-type to n-type when the applied  $V_{gs}$  is greater than the threshold voltage,  $V_t$ . Under this condition, there is also a depletion region that is formed between the channel and the substrate. The thickness of this depletion region can be expressed as

$$X_d = \sqrt{\frac{2\epsilon_{si}\Phi}{qN_A}} \quad (2-7)$$

where  $N_A$  is the doping density of the uniform doped p-type substrate,  $\epsilon_{si}$  is the permittivity of the silicon, and  $\Phi$  is the potential in the depletion layer at the oxide-silicon interface.  $\Phi$  can be expressed as  $\Phi = |\Phi_S - \Phi_F|$ .

Also,  $\Phi_S$  is the potential, underneath the gate, at the oxide silicon interface, and  $\Phi_F$  is the potential of the p-type substrate, also known as the Fermi level of the p-type substrate.  $\Phi_F$  can be expressed as

$$\Phi_F = -\frac{kT}{q} \ln \left[ \frac{N_A}{n_i} \right] \quad (2-8)$$

where  $n_i$  is the intrinsic carrier concentration of silicon ( $\approx 1.45 \times 10^{16}$  atoms/m<sup>3</sup>). The absence of holes in the depletion region leaves a net negative charge due to the immobile acceptor atoms that remain behind. This negative charge is equal to the charge attracted under the gate.

The charge per unit area in the depletion region can be expressed as  $Q = qN_A X_d$ . Using Equation 2-7, the charge can be expressed as follows:

$$Q = \sqrt{(2qN_A \epsilon_{si} \Phi)} = \sqrt{(2qN_A \epsilon_{si} |\Phi_S - \Phi_F|)} \quad (2-9)$$

If the surface potential,  $\Phi_S$ , is equal to  $\Phi_F$ , then the MOSFET is operating in the accumulation region. As  $V_{gs}$  is increased, the surface potential becomes more positive, and when  $\Phi_S = 0$  V, the surface under the oxide has become depleted. As  $V_{gs}$  is increased further, such that  $\Phi_S = -\Phi_F$ , the surface near substrate is inverted and becomes n-type material—that is, the electron concentration at the semiconductor-oxide interface is equal to the substrate doping concentration. The value of  $V_{gs}$  at which  $\Phi_S = -\Phi_F$  is called the threshold voltage of MOSFET. Hence, during the entire transition, the magnitude of change of  $\Phi_S$  is  $2\Phi_F$ . In the presence of an inversion layer and without any substrate bias, the charge in the depletion region can be expressed as follows:

$$Q_{s0} = \sqrt{(2qN_A \epsilon_{si} |2\Phi_F|)} \quad (2-10)$$

If a substrate bias voltage  $V_{sb}$  is applied between the source and the substrate, the  $V_{gs}$  voltage required to produce the inversion layer becomes  $(2|\Phi_F| + V_{sb})$ , and the charge stored in the depletion region can be expressed as follows:

$$Q_b = \sqrt{(2qN_A \epsilon_{si} (2|\Phi_F| + V_{sb}))} \quad (2-11)$$

During the presence of the inversion layer, a parallel plate capacitance is formed between the gate and the inversion region, with the gate oxide acting as the insulator. This capacitor can be expressed as  $C_{ox} = Q_b / V_{GI}$ , where  $V_{GI}$  is the voltage difference between the gate and the inversion region.

Hence, the  $V_{gs}$  voltage required to produce and sustain the inversion layer, called the "threshold voltage" ( $V_t$ ), can be calculated by adding the following components:

- The contact potential that exists between the gate metal and the silicon  $\Phi_{ms}$ . In fact,  $\Phi_{ms}$  can be expressed as  $\Phi_{ms} = (\Phi_{gate} - \Phi_{oxide}) + (\Phi_{oxide} - \Phi_F) = (\Phi_{gate} - \Phi_F)$ .
- The surface inversion potential ( $-2\Phi_F$  is required to produce the inversion layer).
- The potential  $V_{GI}$  that exists between the gate and the inversion layer.
- The positive charge,  $Q_{ss}$ , that exists at the silicon-oxide interface due to imperfections from the growth of gate oxide or as a result of ion implantations used to adjust the threshold voltage of the MOSFET. This charge must be compensated by the gate-source voltage contribution of  $-Q_{ss} / C_{ox}$ .

Therefore, adding all these components, we can express the threshold voltage as follows:

$$\begin{aligned} V_t &= \Phi_{ms} - 2\Phi_F + \frac{Q_b}{C_{ox}} - \frac{Q_{ss}}{C_{ox}} \\ &= \Phi_{ms} - 2\Phi_F + \frac{Q_{b0} - Q_{ss}}{C_{ox}} - \frac{Q_{b0} - Q_b}{C_{ox}} \end{aligned} \quad (2-12)$$

Using Equation 2-11 and substituting into the preceding equation, we get

$$\begin{aligned} V_t &= \Phi_{ms} - 2\Phi_F + \frac{Q_{b0} - Q_{ss}}{C_{ox}} + \frac{\sqrt{2q\epsilon_{si}N_A}}{C_{ox}} \left[ \sqrt{2\Phi_F + V_{sb}} - \sqrt{2\Phi_F} \right] \\ &= \Phi_{ms} - 2\Phi_F + \frac{Q_{b0} - Q_{ss}}{C_{ox}} + \gamma \left[ \sqrt{2\Phi_F + V_{sb}} - \sqrt{2\Phi_F} \right] \end{aligned} \quad (2-13)$$

where the parameter  $\gamma$  can be defined as the substrate-bias (or body-effect) coefficient.

$$\gamma = \frac{\sqrt{2q\epsilon_{si}N_A}}{C_{ox}}$$

When  $V_{sb} = 0$  V (that is, when the source and substrate are at the same potential), we can simplify the preceding equation as follows:

$$V_{t0} = \Phi_{ms} - 2\Phi_F + \frac{Q_{b0} - Q_{ss}}{C_{ox}} \quad (2-14)$$

Next, by substituting Equation 2-15 into Equation 2-14, we can derive a simpler version for the threshold voltage:

$$V_t = V_{t0} + \gamma \left[ \sqrt{2\Phi_F + V_{sb}} - \sqrt{2\Phi_F} \right] \quad (2-15)$$

In general, the value of  $V_{t0}$  is usually adjusted during fabrication by implanting additional impurities into the active area. As shown in Equation 2-15, the threshold voltage increases when the back bias,  $V_{sb}$ , is applied. A positive bias on the substrate results in a wider depletion region and assists in balancing the gate charge. This causes electron concentration in the inversion layer to decrease. Thus, a higher gate voltage is needed to achieve the onset of inversion resulting in an increase of the threshold voltage. In addition, the doping concentration and oxide thickness can also have impacts on the threshold voltage dependence on back bias. Lower doping concentrations and thinner oxide result in a weaker dependence of back bias on threshold voltage.

## 2.5 Metal-Oxide Field-Effect Transistor

The MOSFET shown in Figure 2-17 is an n-channel MOSFET, in which electrons can flow from source to drain in the channel induced under the gate oxide. Both n-channel and p-channel MOSFETs are extensively used. In fact, CMOS IC technology relies on the ability to use both devices on the same chip.

### 2.5.1 Device operation

To understand the MOS device operation, consider the setup shown in Figure 2-18. The nodes of an enhancement-type NMOS transistor have been connected through different voltage sources and ammeters to apply voltages and monitor the device current. Typically,  $V_s = V_B$  and  $V_D > V_s$ . For simplicity, we can assume that the body and the

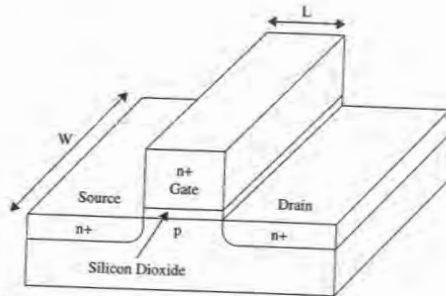


Figure 2-17 MOSFET structure.

source terminals are tied to the ground (i.e.,  $V_{SB} = 0$  V). These yield the following:

$$V_{GS} = V_{GB} - V_{SB} = V_G$$

$$V_{DS} = V_{DB} - V_{SB} = V_D$$

**Cut-off region of MOSFET.** When the source and bulk of the n-type MOSFET are biased at 0 V, with a small positive voltage is applied on the drain and 0 V is applied to the gate—i.e.,  $V_{DS} > 0$  and  $V_{GS} < 0$ —the drain is a reverse-biased p-n junction. Conduction band electrons in the source region encounter a potential barrier determined by the built-in potential of the source junction. Assuming this is enhancement-type MOSFET with  $V_t > 0$ , electrons cannot enter into the channel region and, hence, no current flows from the source to the drain. This is referred to as the cut-off state.

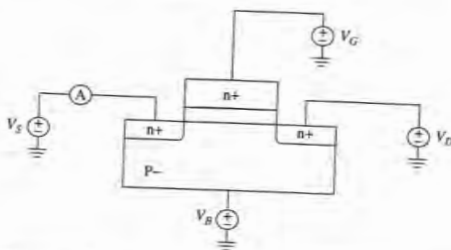


Figure 2-18 Experimental setup for studying MOSFET characteristics.

**Linear region of MOSFET.** As  $V_{GS}$  is increased higher than  $V_t$ , with band bending in the channel region ( $\psi_s > 0$ ), it brings the conduction band in the channel region closer to the conduction band in the source region, reducing the height of the potential barrier of electrons. Electrons can now enter the channel, and the current flow from source to drain can be established.

In the low-drain bias regime, the drain-to-source current increases almost linearly with drain bias voltage. Indeed, here the channel resembles an ideal resistor obeying Ohm's law. The channel resistance is determined by the electron concentration in the channel, which is a function of the gate bias. Therefore, the channel acts like a voltage-controlled resistor whose resistance is determined by the applied gate bias. As the gate bias is increased, the slope of the I-V characteristic gradually increases due to the increasing conductivity of the channel. We obtain different slopes for different gate biases. This region where the channel behaves like a resistor is referred to as the linear region of MOSFET. The drain-to-source current in the linear regime is given by

$$I_{DS,lin} = \frac{W}{L} \mu C'_{ox} \left[ (V_{GS} - V_t) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad (2-16)$$

where  $V_t$  is the threshold voltage,  $C'_{ox}$  is the gate capacitance per unit area, and  $\mu$  is the effective channel mobility, which differs from bulk mobility. We will deal with the concept of effective channel mobility later.

Threshold voltage in the preceding equation is defined as

$$V_t = V_{FB} + 2\Phi_F + \gamma \sqrt{2\Phi_F + V_{SB}} \quad (2-17)$$

For very small  $V_{DS}$ , the second term in the parentheses can be ignored, and the expression for drain current can be simplified to

$$I_{DS,lin} = \frac{W}{L} \mu C'_{ox} (V_{GS} - V_t) V_{DS} \quad (2-18)$$

which defines I-V curve as a straight line with a slope equal to the channel conductance.

$$\sigma_c = \frac{W}{L} \mu C'_{ox} (V_{GS} - V_t) \quad (2-19)$$

The MOSFET bias conditions under different modes of operation is shown in Figure 2-19.

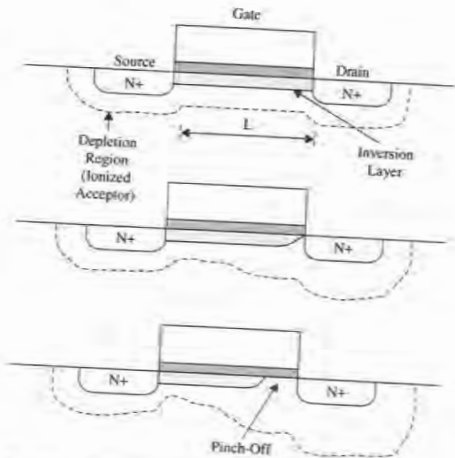


Figure 2-19 MOSFET in different bias conditions.

**Saturation region of MOSFET.** Next, let us consider what happens when  $V_{DS}$  is increased such that  $V_{DS} \geq V_{GS} - V_t$ . Now, because both the gate and the drain are positively biased, the potential difference across the oxide is smaller near the drain end. Again, because the positive charge on the gate is determined by the potential drop across the gate oxide, the gate charge is smaller near the drain end. This implies that the amount of negative charge in the semiconductor needed to preserve charge neutrality will also be smaller near the drain. Consequently, the electron concentration in the inversion layer near the drain end drops when  $V_{DS} \geq V_{GS} - V_t$ . This result is reasonable because we know that the gate-to-channel voltage at which there is no channel (before the onset of a weak inversion) is equal to  $V_t$  by definition of the threshold voltage. Therefore, at the point where the channel pinches off near the drain side, the channel voltage is  $V_{GS} - V_t$ , so the horizontal electric field across the channel, during pinch off, does not depend on  $V_{DS}$  but instead on the voltage across the channel, which is  $V_{GS} - V_t$ . Hence, Equation 2-16 is no longer valid when  $V_{DS} \geq V_{GS} - V_t$ . In this case, the drain current can be derived by substituting  $V_{DS} = V_{GS} - V_t$  in Equation 2-16 as

$$I_{DS,lin} = \frac{W}{L} \mu C_{ox} \left[ (V_{GS} - V_t) V_{DS} - \frac{1}{2} V_{DS}^2 \right] \quad (2-20)$$

As  $V_{DS}$  is increased beyond  $V_{DS,sat}$ , the width of the pinch-off region increases. However, the voltage that drops across the inversion layer remains constant and equal to  $V_{DS,sat}$ . The portion of the drain bias in excess of  $V_{DS,sat}$  appears across the pinch-off region. In a long channel MOSFET, the width of the pinch-off region is small relative to the total length of the channel and the variation can be ignored. Thus, neither the length nor the voltage across the inversion layer changes beyond pinch-off, resulting in a drain current independent of drain bias. Consequently, the drain current saturates. In smaller devices, this assumption falls apart and leads to a phenomenon called "channel length modulation." This can be explained by the fact that as  $V_{DS}$  is increased above  $V_{GS} - V_t$ , the effective channel length of MOSFET is not the physical drawn length " $L$ " anymore because of the presence of a depletion region between the physical pinch-off point in the channel at the drain end and the drain region itself. The effective channel length can be expressed as

$$L_{eff} = L_{drawn} - X_d$$

where  $X_d$  is the depletion width. Hence, rewriting Equation 2-20 gives us the following:

$$I_{DS,sat} = \frac{W}{L_{eff}} \mu C_{ox} (V_{GS} - V_t)^2 \quad (2-21)$$

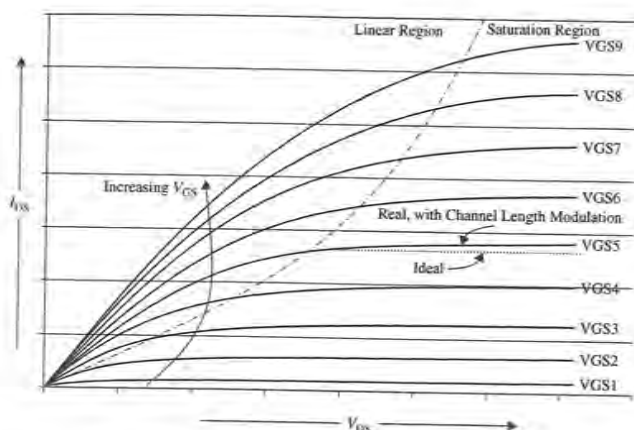
Now, because  $L_{eff}$  or  $X_d$  is a function of  $V_{DS}$ , it follows that  $I_{DS,sat}$  in Equation 2-21 is actually a function of  $V_{DS}$ . In fact, taking channel length modulation into account,  $I_{DS,sat}$  can be expressed as

$$I_{D,sat} = \frac{W}{L} \mu C_{ox} (V_{GS} - V_t)^2 (1 + \lambda V_{DS}) \quad (2-22)$$

where  $\lambda$  is a channel length modulation parameter, whose typical values are between  $0.05 \text{ V}^{-1}$  and  $0.005 \text{ V}^{-1}$ .

From this discussion, it is also evident that the electron distribution is highest near the source and decreases near the drain. To keep a constant current throughout the channel, the electrons travel slower near the source and speed up near the drain. In fact, in the pinch-off region, the electron density is negligibly small. Therefore, in this region, in order to maintain the same current level, the electrons have to travel at much higher speeds to transport the same magnitude of charge.

Figure 2-20 shows  $I_{DS}$  vs.  $V_{DS}$  plot for an enhanced n-type MOSFET. As shown, at low  $V_{DS}$ , the drain current increases almost linearly with  $V_{DS}$ , resulting in a series of straight lines with slopes increasing with  $V_{GS}$ . At high  $V_{DS}$ , the drain current tends toward saturation and becomes almost independent of  $V_{DS}$ .

Figure 2-20  $I_{DS}$  vs.  $V_{DS}$  for MOSFET.

### 2.5.2 Second-order effects for MOSFET operation

Our analysis of the MOSFET characteristics so far has entailed making several simplifying assumptions and analyzing an ideal device behavior. The operation of a real MOSFET may show some deviations from the model described which becomes especially acute as the dimensions of MOSFET reach the nano-meter region. It is essential to understand of some of these effects, also termed as second-order effects, and their impact on the device behavior for the design of charge pumps.<sup>2-5</sup> In this section we will describe two second order effects that will be essential in our subsequent circuit analysis.

**Punch-through.** When the drain of MOSFET is biased at voltage level that is high with respect to the source, the depletion region near the drain could extend to the depletion region near the source side and two of them merge together. It will cause current to flow irrespective of the gate voltage (i.e., even if  $V_{GS}$  is 0 V). This phenomenon is known as the punch-through effect. It is sensitive to the effective channel length of MOSFET device, and the applied voltage levels on source and drain of the devices. Since a high voltage charge pump is built based upon CMOS process, and voltage levels seen on internal nodes are much higher than chip supply voltage, extra attention is needed in dealing with any MOSFET transistors that are connected to the high voltage signals.

**Impact ionization.** As the length of MOSFET transistor is reduced, the electric field near the drain of the transistor will increase (for a fixed-drain voltage) due to saturation. For submicron gate lengths, the electric field near the drain can become so high that electrons are imparted with enough energy to become what is termed "hot." These hot electrons impact the drain, dislodging holes that are then swept toward the negatively charged substrate and appear as the substrate current. This effect is known as "impact ionization." Moreover, the electrons can penetrate the gate oxide, causing a gate current. Over the time this can lead to a degradation of the MOSFET device parameters (threshold voltage, subthreshold current, and transconductance), which in turn can lead to the failure of the circuits eventually. Although the substrate current may be used in a positive manner to estimate the severity of the hot electron effect, it can lead to poor refresh times in dynamic memories, noise in mixed signal systems, and possibly latch-up. Hot holes, on the other hand, do not normally present a problem because of their lower mobility.

### 2.6 Latch-up in CMOS Technology

The device structures present in the standard CMOS technology are inherently composed of parasitic PNP and NPN bipolar structures that can give rise to a combined SCR-type effect, known as "latch-up." Even though this problem occurs more frequently in regions near I/O interface, a circuit designer contemplating a charge pump design should be well aware of the dangerous latch-up effect when both NMOS- and PMOS-type MOSFETs are used in charge pump circuits handling high-voltage swings in every clock cycle.

Figure 2-21 shows the parasitic bipolar MOSFETs that are responsible for latch-up in CMOS technology. Figure 2-22 shows the schematic of the MOSFET parasitics. The emitter, base, and collector of the bipolar npn,  $Q_1$ , is formed by the source of the PMOS, the n-well, and the substrate, respectively. Bipolar npn  $Q_2$ 's collector, base, and emitter are formed by the n-well, substrate, and source of the NMOS transistor. The resistors  $RW_1$  and  $RW_2$  are the resistances of the n-well, and resistors  $RS_1$  and  $RS_2$  are the resistances of the substrate. Capacitors  $C_1$  and  $C_2$  are the parasitic capacitors that are present between the drain, source, and the substrate of the PMOS and NMOS transistors.

In a normal operating region, the bias voltages at nodes  $b_1$  and  $b_2$  are such that the parasitic bipolar devices are non-conducting. Because  $Q_1$  is off, the potential at node  $b_2$  is close to 0 V. Since  $Q_2$  is off too, the potential at node  $b_1$  is close to the supply voltage. Now consider what happens when a spurious positive voltage spike appears at node  $b_2$  through the parasitic capacitor  $C_2$ . This will cause a high positive bias

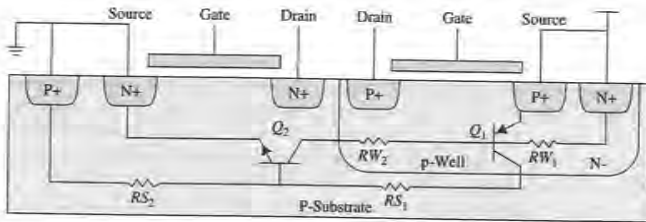


Figure 2-21 Cross-section view of an inverter showing the parasitic components.

at node  $b_2$  and cause  $Q_2$  to turn on. A current through  $Q_2$  will reduce the voltage at node  $b_1$ . Now if the gain of the transistors and the input voltage spike amplitude are high enough while resistances  $RW_1$  and  $RW_2$  are low enough, a turned-on  $Q_2$  will cause  $Q_1$  to turn on. As a consequence,  $Q_1$  will supply the necessary current to sustain the bias voltage at node  $b_2$ . Hence, this will give rise to a self-sustaining mechanism, resulting in a low-resistance path from the power supply to ground. A similar argument can be made for negative-going input pulses at node  $b_1$  through capacitor  $C_1$ .

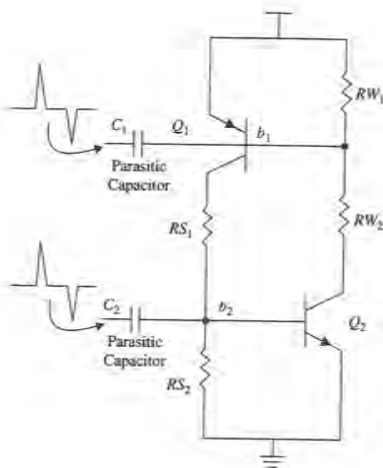


Figure 2-22 Schematic of the parasitic components related to latch-up.

The following common techniques can help to reduce the latch-up problems in real practice:

- **Reduce the parasitic resistances  $RW_1$  and  $RW_2$ .** In general, if these resistances are zero, the parasitic PNP and NPN transistors will never turn on. The values of these resistances are strong functions of the distance between substrate near active area and well contacts, and the total number of the contacts. With plenty of these contacts placed close to the NMOS and PMOS transistors, the value of  $RW_1$  and  $RW_2$  can be kept at a minimum.
- **Reduce the noise amplitude at nodes  $b_1$  and  $b_2$  by reducing the size of capacitors  $C_1$  and  $C_2$ .** Reducing the size of the NMOS and PMOS transistors will reduce the size of the parasitic capacitors  $C_1$  and  $C_2$ , thus lowering the magnitude of the signal fed through.

In general, large MOSFETs, required to drive heavy loads, are especially susceptible to latch-up because of the large drain depletion capacitances. One of the best ways to design latch-up free circuits, especially for charge pumps, is to use only one type of MOSFET in the circuit.

## 2.7 Merits of PMOS Versus NMOS in Circuit Design

The modern CMOS process uses two types of MOSFETs: NMOS and PMOS. Early MOS IC developments focused on creating LSI chips that used either all NMOS or all PMOS devices. (CMOS, which employs both NMOS and PMOS transistors, came later.) PMOS IC devices were developed before NMOS ICs because they were easier to make. NMOS transistors were susceptible to manufacturing contaminations that caused them to malfunction. However, this contamination did not affect PMOS transistors in quite the same way because they use holes instead of electrons as majority carriers. But products built with PMOS transistors were inherently slow because they relied on lower-mobility holes (electron mobility is about two to three times higher than hole mobility) to carry current instead of electrons, which made the industry focus more on building products with NMOS devices.<sup>6</sup>

Yet, with the advent of technology and the established superiority of the CMOS process, both PMOS and NMOS devices are being used in modern-day circuits. However, analog circuits—and particularly the charge pumps—can be created entirely out of either PMOS or NMOS transistors. A detailed analysis will be done in Chapter 8 of this book. In general, besides NMOS being faster, for given dimensions and bias currents, NMOS transistors exhibit higher output resistance, which equates to higher gain in amplifiers. Further, to conduct the same

magnitude of current, an NMOS transistor can be smaller than the PMOS. But PMOS transistors have one formidable advantage: because modern silicon substrate are p-type, a separate n-well is created to build PMOS transistors, which allows for the independent control of a PMOS transistor's body bias without impacting the neighboring transistors. Hence, even though each MOSFET type has its own set of advantages and disadvantages, it depends on the application and the circuit designer to select the right type of devices for producing the most efficient operation.

## 2.8 The MOSFET Model

Almost all electronic circuit simulations performed today are done using the SPICE simulator or different variations of SPICE-based simulators. SPICE, an acronym for Simulation Program with Integrated Circuit Emphasis, is a general-purpose circuit simulation program for nonlinear DC, nonlinear transient, and linear AC analyses. Circuits may contain resistors, capacitors, inductors, mutual inductors, independent voltage and current sources, dependent sources, lossless and lossy transmission lines, switches, uniform distributed RC lines, and the five most common semiconductor devices: diodes, BJTs, JFETs, MESFETs, and MOSFETs. SPICE originated at the EECS Department of the University of California at Berkeley.<sup>7,8</sup> Now, to represent the behavior of transistors and other circuit elements in circuit simulations, SPICE requires an accurate model for each device. Over the last few decades, MOSFET modeling has evolved from basic empirical models to really sophisticated levels for accurately modeling many second-order effects. Throughout this chapter we saw the ideal equations that describe the behavior of MOS transistors. Although these incorporate some nonideal effects, they do not accurately model a specific device for a particular process. This is especially true for devices with ever-shrinking dimensions. In the original implementation of SPICE, as released by U.C. Berkeley, three levels of accuracy were provided for MOSFET models for SPICE simulations; these were denoted as Level 1, Level 2, Level 3. The Level 1 model, also known as the Schichman-Hodges model, uses basic device equations (square law) for  $I_{DS}$  to calculate current flow, device capacitances, and other parameters. However, the Level 1 model eliminates most of the MOSFET's second-order effects, notable among them the carrier velocity saturation effect. The Level 1 model estimates the basic I-V curves of long channel devices with reasonable accuracy, but it still predicts the output impedance of transistors in saturation quite poorly.

The Level 2 model was then developed to represent many higher-order effects. The Level 2 model, also known as the Grove-Frohmman model, uses the 3/2 power equations that result when we include the variation

of threshold voltage across the channel. During the derivation of the MOSFET  $I_{DS}$  equations, it was assumed that the threshold voltage is constant along the channel. This assumption is not correct because the charge in the depletion region under the channel varies according to the local voltage. The Level 2 model addresses this issue, along with other second-order effects, including the effect of velocity saturation. Silicon data indicate that the Level 2 model provides reasonable I-V accuracy for wide short channel devices in the saturation. But simulating with the Level 2 model is computationally intensive because of the existing 3/2 power equations. The Level 3 model was then developed around 1980 with some simplified equations and some empirical constants added to improve accuracy for short channel length devices. It is computationally more efficient, replacing the 3/2-power terms with a first-order Taylor expansion. The drain-induced barrier-lowering effect was also added. The Level 3 model is impressively physical, modeling two-dimensional effects based on junction depth and depletion depths.

The BSIM series models,<sup>7,8</sup> as the MOSFET models originating from U.C. Berkeley are named, were developed as simple physical models with many new "engineering" parameters added to simplify the equations and yet attain better accuracy. These models have enjoyed widespread adoption and have become a standard for defining CMOS processes. BSIM models have evolved, especially over the last decade, from BSIM1, and BSIM2, and BSIM3 to the latest BSIM4 models. In general, the BSIM models have incorporated the following important features:

- Vertical field dependence of carrier mobility
- Carrier velocity saturation
- Drain-induced barrier lowering
- Depletion charge sharing by source and drain
- Nonuniform doping profile for ion-implanted devices
- Channel length modulation
- Subthreshold conduction
- Geometric dependence of electrical parameters
- Gate tunneling model (BSIM4)
- Intrinsic gate resistance model (BSIM4)
- Substrate resistance model (BSIM4)
- Layout-dependent parasitic model (BSIM4)

It is interesting to note that as the BSIM model has evolved, it has also increased the number of device parameters accordingly, as shown in Figure 2-23. Whereas BSIM1 needs about 50 parameters, BSIM2 needs about 70 parameters, and BSIM3 needs about 170 parameters.

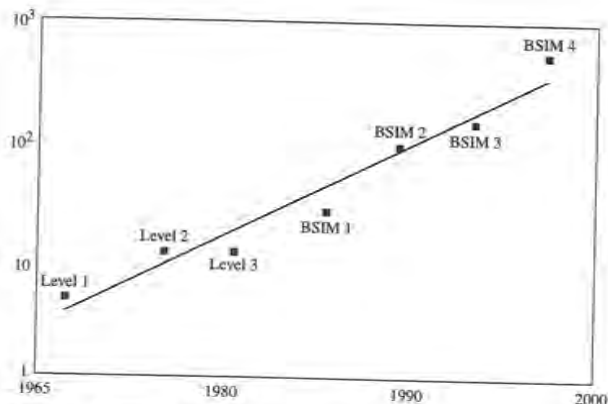


Figure 2-23 MOSFET model parameter variation.

The latest BSIM models strive to deal with frontier effects of the progressing technology, incorporating ultra short channel effects and providing reasonable accuracy for subthreshold and strong inversion operation.

## 2.9 SPICE Simulation Convergence

Normal SPICE-based simulators solve the MOSFET equations for DC and transient analyses through an iterative process.<sup>9,10</sup> However, for a number of reasons, the iterative process does not always converge on a solution, especially for purely analog circuits involving capacitors, as in the case of charge pump circuits, and this becomes a major source of frustration. In the following paragraphs we will discuss ways to mitigate this major issue.

Consider a typical convergence problem. SPICE makes an initial guess at a circuit's node voltages and, using the circuit conductances, calculates the mesh currents. Next, SPICE uses this current information to recalculate the node voltages and starts the cycle over again. The procedure is repeated until the entire node voltages settle to within certain tolerance limits, which can be altered with OPTIONS parameters, such as RELTOL, VNTOL, and ABSTOL.

Now, if the node voltages do not settle down to the tolerance limits within a certain number of iterations, the DC analysis results in an error message, such as "No convergence in DC analysis." SPICE then exits out of the simulation process because an initial stable operating point is required before running a DC or transient analysis.

When transient analysis is run, this iterative process repeats for each individual time step. If the node voltages do not settle down, SPICE reduces the time step and tries again to determine the node voltages. But, if the time step reduces beyond a certain fraction of the total analysis time, SPICE issues the error message "Time step too small" and quits the simulation.

In general, initial DC convergence problems occur because of incorrect initial voltage guesses, model discontinuities, unstable/bistable operating points, or unrealistic circuit impedances. Further model discontinuities or incorrect circuit connections are usually the causes of transient-analysis failure.

To solve the DC convergence problems, any of the following five steps may be used:

- Make sure all the circuit connections are valid. Check for floating nodes, dangling nodes, and incorrect connections.
- Increase the ABSTOL, RELTOL and VNTOL options.

ABSTOL is the absolute current tolerance. Its default value is 1 pA. This means that when the node current reaches within 1 pA of its actual value, SPICE assumes that the current has converged and moves to the next step. VNTOL is the node voltage tolerance. Its default value is 1  $\mu$ V. RELTOL is the relative tolerance parameter, and its default value is 0.1%. RELTOL and VNTOL act together during a simulation. Let us assume during the course of a simulation the actual value of a node voltage is 1 V. The RELTOL parameter will signify a convergence when the node voltage is within 0.1% of 1 V (i.e., within 1 mV of a volt), whereas the VNTOL parameter signifies a convergence when the node voltage is within 1  $\mu$ V of a volt. SPICE will use the larger of the two numbers (in this case, the RELTOL parameter) to flag the convergence.

By increasing the value of these parameters, the convergence problems may be avoided at the cost of simulation accuracy. To help circuit convergence, these parameter values may be reduced as follows:

1. option ABSTOL=1e-6 VNTOL=1e-3 RELTOL=1e-2
2. option ABSTOL=1e-3 VNTOL=5e-3 RELTOL=1e-1

- Use initial conditions by inserting a UIC keyword in the TRAN statement. For example, using ".tran 1n 100n UIC" causes SPICE to completely bypass the DC analysis. Add ".IC v(X)= " initial-condition statements to assist in the initial stages of the transient analysis.

- Ramp up the power supplies. In general, using a DC voltage source for the power supply may cause convergence problems. Instead, ramp up the supply voltage(s) slowly to get around these problems.

For example, the `vvdd vdd 0 pwl 0n 0v 1000n 2.5v` command will ramp up the input  $V_{DD}$  power supply voltage slowly from 0 V to 2.5 V over 1000n seconds.

Further, convergence during DC sweep can also be helped by avoiding absolute power supply boundaries. Hence, sweeping a voltage source from 0 V to 3 V may prove troublesome, but sweeping it from 0.1 V to 2.95 V may work.

- Reduce the rise/fall time of the input stimuli. During transient analysis, if the circuit uses external input signals, pay attention to the rise/fall times of these input stimuli. Applying signals with sharp rise/fall times may flag "Time step too small" errors. A simple solution is to experiment with longer rise/fall times for the input signals.

## 2.10 Conclusion

The main purpose of this chapter was to quickly review the basic p-n junction diode, MOSFET capacitor, and MOSFET characteristics and operation to allow us to gain a better perspective into the operations of the charge pump circuits in the coming chapters. This chapter also touched on MOSFET SPICE models and convergence issues during SPICE simulation, which will always be present during charge pump circuit simulations.

## References

1. Pelliconi, R., I. David, B. Andrea, P. Marco, and L.R. Pier. "Power Efficient Charge Pump in Deep Submicron Standard CMOS Technology." *Solid-State Circuits Conference*, 2001. ESSCIRC 2001. *Proceedings of the 27th European*.
2. Razavi, B. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, New York, 2001.
3. Weste, H.E.N. and K. Eshraghian. *Principles of CMOS VLSI Design: A Systems Perspective*, Second Edition. Addison-Wesley, 1994.
4. Baker, J.R., W.H. Li, and E.D. Boyce. *CMOS Circuit Design, Layout, and Simulation*. IEEE Press, New York, 1998.
5. Gray, R.P., J.P. Hurst, H.S. Lewis, and G.R. Meyer. *Analysis and Design of Analog Integrated Circuits*, Fourth Edition. John Wiley & Sons, 2001.
6. Encore desktop calculator at HP. <http://www.hp9825.com/html/prologues.html>.
7. BSIM SPICE-related documents. <[http://www-device.eecs.berkeley.edu/~bsim3/bsim4\\_intro.html](http://www-device.eecs.berkeley.edu/~bsim3/bsim4_intro.html)>.
8. "Engineering BSIM for the Nano-Technology Era and Beyond." *Fifth International Conference on Modeling and Simulation of Microsystems*.
9. Quarles, T.L. "Analysis of Reference and Convergence Issues for Circuit Simulation." University of California, Berkeley. ERL Memo M89/42.
10. Hymowitz, Charles. "Step-by-step procedures help you solve Spice convergence problems." *EDN design feature*. March 3, 1994.

# Basic Operation of a Charge Pump

## 3.1 Charge Pump System

Charge pumps have been used to generate high voltages for many applications, such as EEPROMs and Flash memories for programming and erasing of the floating-gate. In general, a charge pump is a closed-loop system because the pump output is usually regulated at a predetermined level. To achieve the regulation, the pump needs to be kept turned on, as long as the output voltage is lower than the regulation level. If the pump output level reaches or exceeds the regulation level, the extra charge from pump output must be shunted away or the pump must be completely shut off. Since the system is relying on feedback control to maintain the regulation, obviously there is output noise near regulation level.

A basic block diagram of the charge pump's operation is shown in Figure 3-1. Assuming the pump output voltage is starting from zero at the beginning of operation, the output of the regulator is high (meaning that the pump output voltage is lower than the regulation level). Feedback signal from the regulator would enable pumping clocks to be applied to the charge pump. Charge is transferred from stage to stage and the potential energy of the charge is elevated as it propagates through stages. The output of the charge pump will start to ramp up once the charge reaches the output. During this time the regulator is enabled to constantly sample the output voltage and compare it with the reference voltage. When the output reaches the regulation level, the regulator senses it and deactivates the pumping clock. The charge pump is shut off.

If the output load is purely capacitive in nature, the pump output voltage should remain unchanged for a long period of time and in principle the pump does not need to be turned on again once regulation level

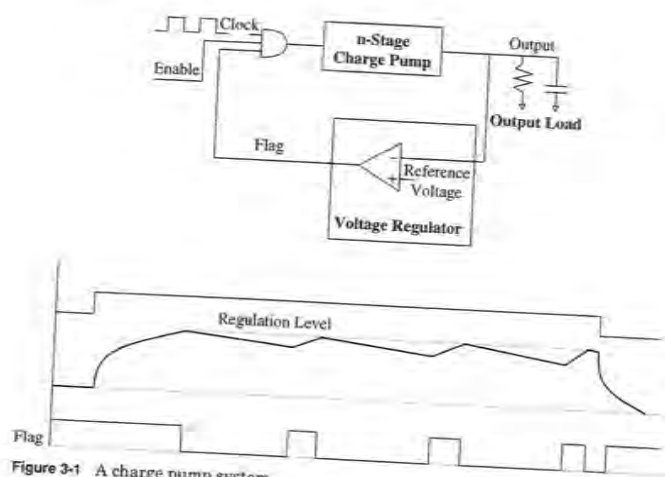


Figure 3-1 A charge pump system.

is reached. In reality, though, this scenario seldom happens. Capacitive coupling from nearby signals, the DC current attributing from different output leakage components, DC current dissipation in the regulator, and myriad other components will obviously cause the output of the charge pump to be discharged over time. Depending on the pump's feedback response time and other implementations, the regulator will eventually detect the deviation of output voltage from the regulation level and enable the charge pump. Since the output is closer to the original preset level, the pump should be able to replenish the lost charges faster and reach regulation level only in a short amount of time. As shown in Figure 3-1, this process will introduce output voltage overshoots and undershoots near regulation level. If the amplitude of noise is large, it may be detrimental to the functionality and device-related reliability of the circuit. Many techniques have been proposed to reduce the output noise for different applications, such as operation-dependent power balancing between pump and output load scheme, and the voltage-controlled oscillator (VCO) approach.

The operation-dependent power balancing between pump and output load scheme relies on matching pump output strength with the loading circuit power consumption at all time. The charge pump may have different output power specifications in many different operations. Even within the same operation the loading circuit may have different current loads over time. Noise occurs if there is mismatch between the

delivered power from the pump and the consumed power from the load. If pump strength can be adjusted based on operation, then the output noise can be minimized. Pump strength can be adjusted as required by varying clock frequency, pump clock amplitude, sizing of effective boosting capacitance, etc., the second approach uses a voltage-controlled oscillator instead of a steady clock generator. The clock frequency of the VCO will be high when the pump is initially starting up. Then this clock frequency will slowly decrease in an almost linear fashion as the output voltage approaches the regulation level and will almost stop when the value is reached. The latter approach is more complex to implement and not suitable for all applications.

At the end of the high-voltage operations, it is required to shut off the charge pump, discharge the high-voltage potential at the output and all internal nodes, and bring them down to normal chip supply levels. This shut-off and discharging process should be carefully managed. If the pump circuit, including the regulator, is shut off before discharging, it may create unwarranted high-voltage stress on some circuits. Further, discharging the high voltage to internal chip supplies may create high-ground bounce or high- $V_{DD}$  bounce without careful timing controls. The noise could create problems for other circuits still in operation. In general, it is better to discharge the internal high-voltage nodes during a pre-defined period for high-voltage recovery purposes only.

Most of the techniques and processes discussed so far will be explained in detail in the following chapters. Throughout this chapter, we will focus on understanding the basics of charge pump operation. We will start with the bucket capacitor model and then discuss the operation of the Dickson charge pump and the different second-order effects, using equations to reinforce the concepts.

### 3.2 Basic Concept: The Bucket Capacitor Model

Charge pumps are circuits that generate a voltage level higher than the chip supply voltage from which they operate. To see how this is possible, consider the simple circuit shown in Figure 3-2, which consists of a single capacitor and three switches.<sup>1</sup>

During clock phase,  $\phi$ , switches  $S_1$  and  $S_3$  are closed and the capacitor is charged to the supply voltage,  $V_{DD}$ . Next, in the second clock phase,  $\phi_b$ , switches  $S_1$  and  $S_3$  are opened and switch  $S_2$  is closed. The bottom plate of the capacitor assumes a potential of  $V_{DD}$ , while the capacitor maintains its charge of  $Q = C \times V_{DD}$  from the previous phase. This means the following during  $\phi_b$ :

$$Q = C(V_{out} - V_{DD}) = C \times V_{DD} \text{ i.e. } V_{out} = 2 \times V_{DD} \quad (3-1)$$

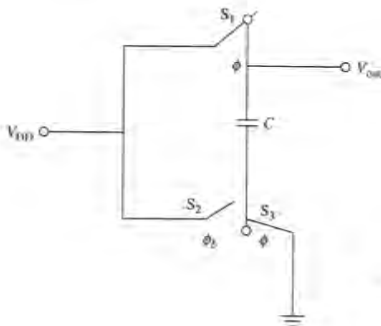


Figure 3-2 Simple voltage doubler.

Thus, in the absence of a DC load, an output voltage can be generated that is twice the input supply voltage. In order to accommodate a load at the output, the circuit should be modified by adding an output capacitance, as shown in Figure 3-3.

In this case, the ideal output voltage is given by

$$V_{out} = \frac{C}{C + C_{out}} \times 2 \times V_{DD} \quad (3-2)$$

Needless to say, the presence of the output capacitor  $C_{out}$  will reduce the output voltage, depending on the output load capacitor. If a load,  $R_L$ , is present, a ripple voltage,  $V_R$ , will also be generated at the output.

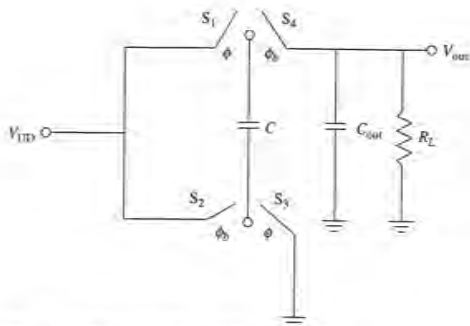


Figure 3-3 Voltage doubler with a load.

The ripple voltage can be reduced by making  $C_{out}$  sufficiently large so that  $V_R$  is negligible compared to  $V_{out}$ , but in doing so the output voltage will also be reduced. It is important to note that the switches mentioned in this circuit can actually be implemented using n-type MOSFETs or p-type MOSFETs, and the capacitor can be a standard metal-metal capacitance or realized using the gate oxide of the MOSFET, with its source-drain substrate connected together to form one end and the gate forming the other end. It is also important to pay considerable attention to the type of MOSFETs used in circuits involving high voltages. We will discuss this point later. Next, we will see a practical method of generating high voltage using the Dickson charge pump.

### 3.3 The Dickson Charge Pump

The operation of the Dickson pump circuit<sup>2,3</sup> is illustrated in Figure 3-4, which shows the typical voltage waveforms in an n-stage multiplier. As you can see,  $\phi$  and  $\phi_b$  are two out-of-phase clocks with amplitude  $V_\phi$  and are capacitively coupled to alternate nodes along the diode chain. The two clocks increase the potential voltage of charge at subsequent nodes by pumping packets of charge along the diode chain as the coupling capacitors are successively charged and discharged during each half of the clock cycle.

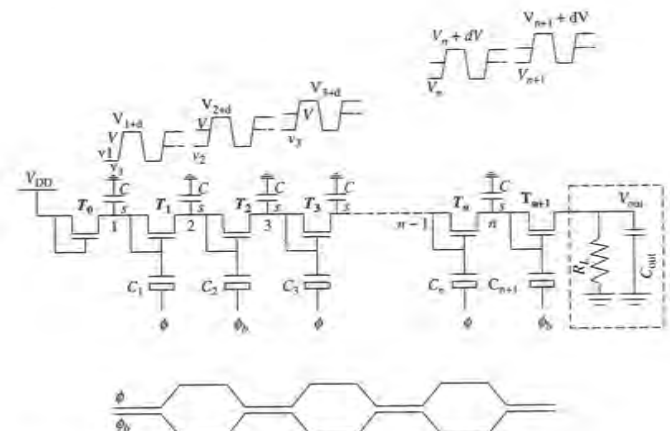


Figure 3-4 Internal waveforms of a 2-phase charge pump.

As shown in Figure 3-4, the difference between the voltages of the  $n$ th and  $(n+1)$ th nodes at the end of each pumping cycle is given by

$$\Delta V = V_{n+1} - V_n = V'_\phi - V_D \quad (3-3)$$

where  $V'_\phi$  is the voltage swing at each node due to capacitive coupling from the clock,  $V_D$  is the forward bias diode voltage, and  $V_L$  is the voltage by which the capacitors are discharged down when the multiplier is supplying an output current ( $I_{out}$ ).<sup>4</sup> For the clock coupling capacitance ( $C$ ) and stray capacitance ( $C_s$ ) at each node, the voltage gain can be represented as

$$V'_\phi = \left( \frac{C}{C+C_s} \right) V_\phi \quad (3-4)$$

When the clock  $\phi$  is low and  $\phi_b$  is high, MOSFET  $T_0$  conducts while MOSFET  $T_1$  is off. When  $T_0$  is on, the voltage at node 1 settles at  $V_{DD} - V_D$ . Next, when the clock  $\phi$  goes high, the voltage at node 1 becomes

$$V_1 = V_{DD} + (V'_\phi - V_D) \quad (3-5)$$

During the time  $\phi_b$  is low and assuming the clock period is sufficiently long, MOSFET  $T_1$  will conduct until the voltage at node 2 charges up to

$$V_2 = V_{DD} + (V'_\phi - V_D) - V_D \quad (3-6)$$

Next, when the clock  $\phi$  goes low and  $\phi_b$  is high, the voltage at node 2 becomes

$$V'_2 = V_{DD} + 2(V'_\phi - V_D) \quad (3-7)$$

Hence, for  $N$  stages, the voltage at node " $n$ " will be  $V_n = V_{DD} + N \cdot (V'_\phi - V_D)$ . MOSFET  $T_{n+1}$  forms an isolating diode, forcing the output voltage to be

$$V_{out} = V_{DD} + N(V'_\phi - V_D) - V_D \quad (3-8)$$

Equating the value of  $V'_\phi$  from Equation 3-2 we get the following:

$$V_{out} = V_{DD} + N \left[ \left( \frac{C}{C+C_s} \right) V_\phi - V_D \right] - V_D \quad (3-9)$$

This equation shows the output voltage in an ideal situation when the pump is not delivering any output load current. Because the pump will

be connected to an output load, which could be drawing a load current, the output voltage will not remain at the voltage expressed in Equation 3-9. Assuming  $V_L$  as the voltage drop per stage for supplying the average load current, the charge pumped by each diode per clock cycle is  $(C+C_s)V_L$ . The current supplied by the pump, at a clock frequency  $f$ , is given by

$$I_{out} = f(C+C_s)V_L \quad (3-10)$$

Rewriting Equation 3-10, the output voltage for  $N$  stages is reduced by an amount

$$\frac{N \times I_{out}}{(C+C_s)f}$$

Rewriting Equation 3-9 by incorporating the effects of the load current, the output voltage becomes

$$V_{out} = V_{DD} + N \left[ \left( \frac{C}{C+C_s} \right) V_\phi - V_D - \frac{I_{out}}{(C+C_s)f} \right] - V_D \quad (3-11)$$

where  $V_{out}$  is the maximum peak output voltage at the last stage. Next, when  $\phi_b$  goes low, MOSFET  $T_{n+1}$  turns off and the output load resistance,  $R_L$ , will discharge the load capacitor. Then, when  $\phi_b$  goes high, MOSFET  $T_{n+1}$  will conduct and the output will charge up. This will give rise to a ripple at the output, defined as  $V_R$ . Because  $C_{out}$  is sufficiently large,  $V_R$  will be small compared to  $V_{out}$ .  $V_R$  can be expressed as

$$V_R = \frac{I_{out}}{f \times C_{out}} = \frac{V_{out}}{f \times R_L \times C_{out}} \quad (3-12)$$

The output ripple voltage is the noise to the loading circuits. Large amplitude of ripple could affect the operation of the chip, such as PLL, EEPROM, or flash memory applications. As can be observed from Equation 3-12, the ripple voltage can be reduced by increasing the clock frequency  $f$  or by increasing the output load capacitance. Each of these methods has its own limitations and drawbacks. Frequency cannot be blindly increased without decreasing the pump's efficiency after a certain limit. Also, increasing the load capacitance will reduce the output ramp-up time adversely.

From Equation 3-11, it can be seen that the charge pump will work provided that

$$N \left[ \left( \frac{C}{C+C_s} \right) V_\phi - V_D - \frac{I_{out}}{(C+C_s)f} \right] - V_D > 0$$

which simplifies as follows:

$$\left[ \left( \frac{C}{C+C_s} \right) V_\phi - V_D - \frac{I_{out}}{(C+C_s)f} \right] > 0 \quad (3-13)$$

Now Equation 3-11 can be rewritten as

$$V_{out} = V_\phi - I_{out} R_S \quad (3-14)$$

where

$$V_\phi = V_{DD} + N \left[ \left( \frac{C}{C+C_s} \right) V_\phi - V_D \right] - V_D$$

and

$$R_S = \frac{N}{(C+C_s)f} \quad (3-14a)$$

Here,  $V_\phi$  and  $R_S$  are the open-circuit output voltage and output series resistance of the charge pump, respectively.<sup>8</sup> Therefore, Equation 3-14 can be expressed as a simple equivalent circuit of the charge pump, as shown in Figure 3-5.

In deriving this model for the charge pump, it has so far been assumed that the capacitors are fully charged and discharged with the diode connected MOSFET cutoff voltage,  $V_D$ . In practice, this is not the case due to the nonlinear voltage-current characteristics and internal series resistance,  $R_D$ , of the MOSFETs. This results in a residual voltage in addition to  $V_D$  remaining across the diodes at the end of each cycle, causing the multiplier output series resistance,  $R_S$ , to increase in a nonlinear manner with load current. However, by making  $R_D$  sufficiently small, the increase of  $R_S$  due to this effect can be minimized. Proper care should be taken to size the MOSFETs by keeping minimum permissible channel length and allowing sufficient MOSFET width.

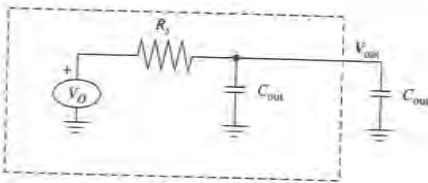


Figure 3-5 Equivalent model of the Dickson charge pump.

### 3.4 Dynamic Analysis of the Charge Pump

Because the internal nodes of the charge pump are always in a state of flux, we need to make some simple assumptions to quantify its dynamic operation.<sup>6</sup> Assume that the charge pump has slowly risen to the maximum voltage and is maintaining the output voltage constant. Under this scenario, the charge transferred from one capacitor to the next equals the charge transferred to the output. We also need to make the following assumptions:

- The parasitic capacitance at each stage is negligibly small.
- The clock cycle time is sufficiently large, compared to all the RC time constant in the circuit.

Next, consider the simplified circuit shown in Figure 3-6.

Assuming a complete charge transfer, the charge stored at node 1 at time  $T_1$  is

$$Q_1 = C(V_{DD} - V_i) \quad (3-15)$$

Now the positive edge of the clock  $\phi$  injects a charge at node 1, which causes  $M_1$  to conduct and transfer charges to node 2. At the end of time  $T_2$ , assuming the charge transfer is complete, the potential at node 2 will rise to

$$V_2 = (Q_2 + q_{in}) / C \quad (3-16)$$

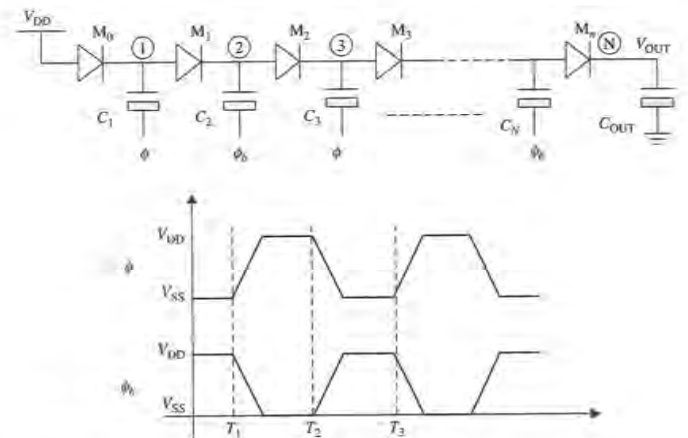


Figure 3-6 A 2-phase charge pump.

where  $Q_2$  equals the existing charge at node 2, and  $q_{inj}$  is the part of the charge transferred to node 2, after being injected by the clock  $\phi$  at node 1. However,

$$V_2 = V_1 - V_t \quad (3-17)$$

where  $V_1$  is the new potential at node 1 at time  $T_2$ , or

$$V_1 = V_{DD} + (Q_1 - q_{inj})/C \quad (3-18)$$

Hence, substituting Equation 3-18 and Equation 3-16 into Equation 3-17, we get

$$V_2 = \left( Q_1 - \frac{(Q_1 - q_{inj})}{C} \right) - V_t = \frac{(Q_2 + q_{inj})}{C}$$

or

$$2Q_1 = Q_2 + 2q_{inj}$$

or

$$Q_2 = 2C(V_{DD} - V_t) - 2q_{inj} \quad (3-19)$$

Again, at time  $T_1$ , the potential at node 3 can be expressed as

$$\frac{Q_3}{C} = V_{DD} + \frac{Q_2}{C} - V_t$$

Substituting Equation 3-19 in the preceding equation, we get the following:

$$Q_3 = 3C(V_{DD} - V_t) - 2q_{inj}$$

The preceding sequence of charges/stages can be expressed as

$$Q(2n-1) = (2n-1)C(V_{DD} - V_t) - 2(n-1)q_{inj} \quad (3-20)$$

when  $n$  is even, and as

$$Q(2n) = 2nC(V_{DD} - V_t) - 2nq_{inj} \quad (3-21)$$

when  $n$  is odd. Here,  $1 < n < N/2$ . Note that the first stage can be defined by  $n = 0$ , but it is not in the domain of the preceding equations.

Further,  $n = 1$  means the physical second stage, and  $n = 2$  means the physical third stage, and so on.

Now the voltage equation of the last stage can also be represented as

$$V_{out} = V_{DD} + \frac{Q(N)}{C} - V_t$$

or

$$Q(N) = C(V_{out} - V_{DD} + V_t) \quad (3-22)$$

Assuming an even number of stages, Equation 3-21 can be expressed as

$$q_{inj} = \frac{Q(2n)}{2n} - C(V_{DD} - V_t)$$

Incorporating Equation 3-22 in Equation 3-21, the final output charge at stage  $N$  can be expressed as

$$q_{inj} = \frac{C}{N} [(N+1)(V_{DD} - V_t) - V_{out}] \quad (3-23)$$

Equations 3-20 and 3-21 can also be expressed in terms of  $V_{out}$  and  $q_{inj}$ . Substituting Equation 3-23 in Equation 3-21, we get the following:

$$Q(2n) = 2kC(V_{DD} - V_t) - \frac{2kC}{N}(N+1)(V_{DD} - V_t) + \frac{2kC}{N}V_{out}$$

$$Q(2n) = \frac{2kC}{N}(V_{out} - V_{DD} - V_t)$$

Similarly, substituting Equation 3-23 in Equation 3-20, we get

$$Q(2n-1) = \frac{2(k-1)}{N}C(V_{out} - V_{DD} - V_t) + C(V_{DD} - V_t)$$

As shown in Figure 3-7, the last stage holds the maximum amount of charge. Further, the charge stored in each stage does not increase linearly. This can be quickly explained by the basic operation of the Dickson pump.

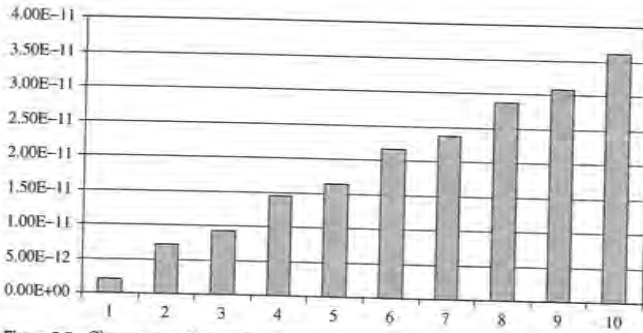


Figure 3-7 Charge stored at each internal stage.

The charge supplied by the power supply, every cycle, in a steady state is equal to the charge transferred to the capacitor,  $C_1$ , through the diode  $M_0$ , plus the charge transferred by the clocks  $\phi$  and  $\phi_b$  through capacitors  $C_1$  to  $C_N$  from one stage to the next.

Solving the equations for the total charge delivered from the power supply and the charge delivered at the output, it can be shown that

$$Q_{VDD}(t) = (N+1) \times C_{load} \times [V_{out}(t) - V_{DD} - V_t] \quad (3-24)$$

where  $C_{load} = C_{out} + C_{pump}$ ,  $C_{pump} \sim NC/3$  (assuming  $N > 4$ ), and  $C_{pump}$  is the inherent capacitance of the charge pump.

Finally, Tanzawa and Tanaka<sup>5</sup> have shown that the rise time,  $T_r$ , for the output voltage to charge up from initial state ( $V_{DD} - V_t$ ) to a final voltage,  $V_{fin}$ , can be expressed as

$$T_r = \frac{\ln \left[ 1 - \frac{V_{fin} - (V_{DD} - V_t)}{N(V_{DD} - V_t)} \right]}{\ln \left[ \frac{1}{1 + \frac{C}{N(C_{load})}} \right]} T_{osc} \quad (3-25)$$

Thus, the average output current during time  $T_r$  can be expressed as

$$I_{out} = \frac{C_{load} [V_{fin} - (V_{DD} - V_t)]}{T_r} \quad (3-26)$$

and the average power supply current consumption during  $T_r$  can be derived from Equation 3-24 as

$$I_{VDD} = \frac{Q_{VDD}(t)}{T_r} = \frac{(N+1)C_{load}(V_{fin} - V_{DD} - V_t)}{T_r} \quad (3-27)$$

where  $C_{load} = C_{out} + C_{pump}$ .

For charge pumps requiring high-output voltages, the number of stages is essentially high ( $N > 8$ , or if  $N = 8$ , the charge pump may be operating with a boosted clock scheme, such as  $2-4V_{DD}$ ).

Hence, to ensure good output current,  $N(V_{DD} - V_t) \gg V_{out}$ . Further, the output load  $C_{load} \gg C$ .

Equation 3-25 can also be expressed as

$$T_r = \frac{\ln \left[ 1 - \frac{V_{fin} - (V_{DD} - V_t)}{N(V_{DD} - V_t)} \right]}{\beta} \quad (3-27a)$$

where

$$\beta = \frac{1}{T_{osc}} \ln \left[ \frac{1}{1 + \frac{C}{N(C_{load})}} \right]$$

is a constant.

Now, a continuous function of the form  $\ln(x)$  can be expanded easily in terms of a Taylor series. In the current case, even though the output voltage,  $V_{out}$ , will be a staircase function as it reaches its destination voltage,  $V_{fin}$ , it can be assumed to be a smooth function because the rise time is very large compared to the clock cycle and the output load capacitance is generally very high compared to the pump capacitor. Hence, simplifying Equation 3-27a using a Taylor series expansion for logarithmic functions, Papaix and Daga<sup>6</sup> have simplified the rise time as follows:

$$T_r = \frac{V_{fin} - (V_{DD} - V_t)}{(V_{DD} - V_t)} \left( \frac{1}{3} + \frac{C_{load}}{NC} \right) T_{osc} \quad (3-28)$$

The power efficiency can be calculated like this:

$$\xi = \frac{V_{out} I_{out}}{V_{DD} I_{VDD}} = \frac{V_{out} C_{load} [V_{fin} - (V_{DD} - V_t)]}{V_{DD} (N+1) C_{load} (V_{fin} - V_{DD} - V_t)} = \frac{V_{out}}{(N+1) V_{DD}} \quad (3-29)$$

This shows that the number of stages only determines the power efficiency when  $V_{out}$  and  $V_{DD}$  are fixed. Further, the rise time,  $T_r$ , is proportional to  $C_{load}/f_{osc}$ ,  $1/N$ ,  $1/V_{DD}$ , and  $1/C$ . In general, the  $W/L$  dimensions of the diodes are not parameters in these equations because the equations were derived assuming a complete charge transfer, every clock cycle. However, in reality, the  $W/L$  ratio of the diodes cannot always be very high, because this will increase the layout area, and the frequency of operation is not always tuned for complete charge transfer. Hence, this incomplete charge transfer scenario will cause an anomaly in the rise time between the expected result from Equation 3-28 and the actual simulation results. The exact derivation of the charge pump's output characteristics during incomplete charge transfer is beyond the scope of this book. For initial study purposes, the reader is requested to design and simulate the charge pump with a relaxed clock period to ensure complete charge transfer. After that he/she can reduce the clock period and study the new effects.

The preceding derivations allow us to synthesize an equivalent model of the charge pump, as shown in Figure 3-8.

From Equation 3-24, and referring to Figure 3-8,  $V_{max}$  equals  $(N+1)(V_{DD} - V_t)$ ,  $R_{eqv}$  equals  $N/C$ , and  $C_{eqv}$  equals  $NC/3$  (assuming  $N > 4$ ).

### 3.4.1 The body effect revisited

In the previous chapter we saw that the threshold voltage of an NMOS transistor can be represented as

$$V_t = V_{t0} + \gamma(\sqrt{\phi_s + V_{sb}} - \sqrt{\phi_s}) \quad (3-30)$$

where

- $\phi_s$  equals the surface potential at threshold and is represented by  $\phi_s = 2V_t \ln \frac{N_A}{n_i}$ .

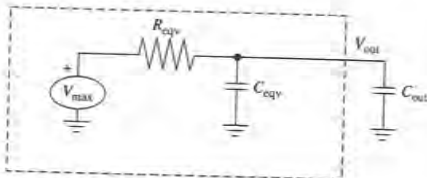


Figure 3-8 Equivalent model of the Dickson pump.

- $\gamma$  equals the body effect coefficient and is represented by

$$\gamma = \frac{t_{ox}}{\epsilon_{ox}} \sqrt{2q\epsilon_s N_A} = \frac{\sqrt{2q\epsilon_s N_A}}{C_{ox}}$$

- $V_{t0}$  equals the zero-bias threshold voltage.
- $V_{sb}$  is the source-to-body voltage bias.

As you can see from the preceding equation, for a particular process corner,  $V_{t0}$ ,  $\gamma$  and  $\phi_s$  are constants. Hence, as the source voltage of an NMOS MOSFET increases, the threshold voltage of the MOSFET also rises, which results in decreased  $I_{ds}$ , so less charge transfer takes place.<sup>4,7</sup>

### 3.5 Dynamic Analysis of the Charge Pump with Body Effect

The derivation for Equation 3-25 assumes a complete charge transfer and is implemented with diodes. In CMOS technology, because diodes are generally implemented with MOSFETs, as shown in Figure 3-9, with the drain and gate tied together, implementing a Dickson charge pump with a series of NMOS transistors will introduce body effect—the  $V_t$  of the MOSFETs will increase as the source voltage increases. Hence, Equation 3-25 needs to be readdressed, incorporating the effect of  $V_t$  variation.

In general, for an NMOS with a zero back bias, when the drain and gate are tied together, the source voltage can be expressed as

$$V_s = V_D - V_{t0}$$

Taking body effect into account, this expression can be modified as  $V_s = \alpha(V_D - V_{t0})$ , where  $\alpha$  is the body effect factor.<sup>3</sup> Therefore,

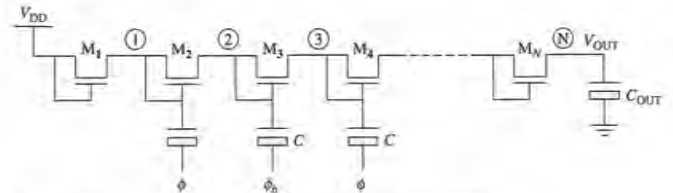


Figure 3-9 Dickson charge pump implemented using MOSFETs.

Equation 3-15 can be expressed as  $Q_1 = \alpha C(V_{DD} - V_{t0})$ . Equations 3-22, 3-23, and 3-25 can also be restated as follows:

$$Q(2n) = \sum_{j=1}^{j=2n} \alpha^j C(V_{DD} - V_t) - \sum_{j=1}^{j=2n} \alpha^{(j-1)} q_{inj} \quad (3-31)$$

$$Q(2n-1) = \sum_{j=1}^{j=2n-1} \alpha^j [C(V_{DD} - V_t) - q_{inj}] - \alpha^{2(n-1)} q_{inj}$$

$$Q(N) = C \left( \frac{V_{out}}{\alpha} - V_s \right) \quad (3-32)$$

Using Equations 3-31 and 3-32, the output charge  $q_{inj}$  can be expressed as

$$q_{inj} = \frac{C \left[ \sum_{j=1}^{N+1} \alpha^j (V_{DD} - V_t) - V_{out} \right]}{\sum_{j=1}^N \alpha^j} \quad (3-33)$$

which was derived by Kanawaja et al., and the output rise time,  $T_r$ , from  $(V_{DD} - V_t)$  to  $V_{fin}$  can be derived as

$$T_r = \frac{\ln \left[ \frac{V_{fin} - (V_{DD} - V_t)}{(V_{DD} - V_t) \left( \sum_{j=1}^{j=N+1} \alpha^j - 1 \right)} \right]}{\ln \left[ \frac{1}{1 + \frac{C}{N(C_{load})}} \right]} T_{osc} \quad (3-34)$$

As a test, you can see that if there is no body effect—i.e., if  $\alpha = 1$ ,  $V_s = V_{th} - V_{t0}$ —then the preceding equation reduces to Equation 3-25.

### 3.6 Conclusion

We have analyzed the Dickson charge pump operation and quantified its operation while deriving its output voltage, current, and rise time characteristics. We also saw that the charge pump can be represented as a simple voltage source with an internal serial impedance. As the number of stages of the charge pump is increased, the pump is able to deliver higher output voltages, but the internal series resistance also increases, thus decreasing the available output current. In the subsequent chapters, we will build on this understanding of the Dickson charge pump<sup>9</sup> and scrutinize different factors affecting the charge pump's performance and efficiency. We will also examine different varieties of charge pumps designed to increase efficiency and get around the crippling high-threshold voltage problem, such as the 4-phase charge pump, CTC charge pump scheme, and charge pumps designed with PMOS transistors.

### References

1. Pylarinos, L. Charge Pumps: An Overview. <http://www.eecg.toronto.edu/~kphang/ece1371/chargepumps.pdf>.
2. Dickson, J. "On-chip High-Voltage Generation in NMOS Integrated Circuits Using an Improved Voltage Multiplier Technique." *IEEE Journal of Solid-State Circuits*, Vol. 11, No. 6, pp. 374–378, June 1976.
3. Witters, J.S., G. Groeseneken, and H.E. Maes. "Analysis and modeling of on-chip high-voltage generator circuits for use in EEPROM circuits." *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 5, pp. 1372–1380, October 1989.
4. San, H., H. Kobayashi, T. Myono, T. Iijima, and N. Kuroiwa. "Highly-Efficient Low-Voltage-Operation Charge Pump Circuits Using Bootstrapped Gate Transfer Switches." *IEEE Japan Transactions of EIS*, Vol. 120-C.
5. Tanzawa T. and T. Tanaka. "A dynamic analysis of the Dickson charge pump circuit." *IEEE Journal of Solid-State Circuits*, Vol. 32, No. 8, August 1997.
6. Papaix, Caroline and Jean-Michel Daga. High Voltage Generation for Low Power Large VDD Range Non Volatile Memories. PATMOS 2001. <http://patmos2001.eivd.ch>.
7. Lin, H., J. Lu, and T.Y. Lin. "A new 4-phase charge pump without body effects for low supply voltages." Asia-Pacific Conference on ASICs, August 6-8, 2002.
8. Khouri, O. et al. "Low Output Resistance Charge Pump for Flash Memory Programming." International Workshop on Memory Technology, Design, and Testing. MTDT 2001.
9. Kim, et al. High voltage generating charge pump circuit. U.S. patent number 6,661,682.

## Charge Pump Design Criteria

A charge pump is a common system, the construction of which is based upon simple CMOS transistors and capacitors. Relying on charge conservation theory and the capacitive coupling method, the charge pump generates the final output voltage, which can be either higher or lower than the given chip supply voltages.

Figure 4-1 shows a generic high-voltage charge pump block and the output path network. There are seven components: a high-voltage charge pump, a noise-filtering capacitance ( $C_{\text{filter}}$ ), a decoupling capacitance ( $C_{\text{decoupling}}$ ), a resistor ( $R_{\text{regulation}}$ ) from the resistor divider circuit used in high-voltage regulation path, a high-voltage switch used to connect and disconnect the pump output to the loading circuits, a capacitive output loading ( $C_{\text{load}}$ ), and a current loading ( $I_{\text{load}}$ ). The charge pump is designed to bring its output to the desired voltage level within a fixed amount of time. The load current must be supplied at designed regulation levels.

Many criteria should be considered before the actual design work can start. Normally designers should consider at least two things: the available technology and the product specification. Technology involves the wafer processing for the chips. Models and parameters such as capacitance, sheet resistance, interconnection, and transistors are set by the technology and are used by designers in circuit design. The product specification contains the requirements the final product must achieve. Specifications such as power-consumption requirements, the pump-regulation level, the power efficiency of the design and die size, and so on, are the targets the circuit design has to meet on silicon. The following sections discuss the different major aspects that must be understood before starting the charge pump design.

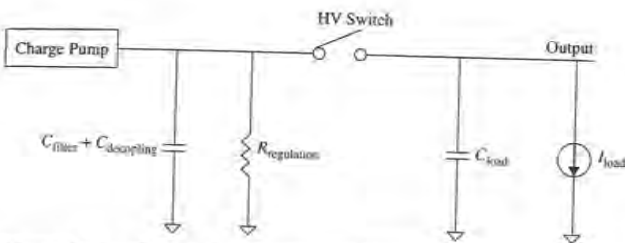


Figure 4-1 Generic high-voltage charge pump block and the output path network.

#### 4.1 Technology

At any time one to many technologies may exist to facilitate the design of the charge pump. Some technologies may be more advanced at a given time, and the final product can be more advanced at a given time and cheaper in terms of manufacturing cost. It is important to know what is available to designers and the pros and cons for each technology. More advanced technologies may not have the best economic return in the actual production.

Looking back in history, it is clear that technology evolves over time to meet the needs of a design. With more applications being introduced and more demands being created, better technologies are developed to allow more advanced designs to be feasible at lower costs. As mentioned in Chapter 1, the very first charge pump was proposed by Swiss physicist Heinrich Greinacher in 1919. By cascading more than one diode capacitor in series, a voltage greater than the supply voltage can be generated. Later, this technique was used by John Douglas Cockcroft and Ernest Thomas Sinton Walton to generate voltage potentials of more than 800,000 volts in their particle accelerator. This technique was again adopted by John F. Dickson, who proposed the technique of implementing the charge pump on a silicon chip.

Gradually charge pumps became widely used on chips, such as EPROM, EEPROM, Flash memory, PLL design, and so on. They help generate a voltage source that is higher in potential than the available chip power supplies. For example, EPROM/EEPROM uses high voltages potentials to electrically program, or electrically erase, the bits in the nonvolatile memories. The high-voltage potential required is generally higher than the chip supply voltage given. Before 1990, the needed high-voltage potential could not be generated on chips. In order to program or erase the EPROM/EEPROM memories, an external EPROM/EEPROM programmer was used. The desired high-voltage potential has to be generated off-chip. Due to this constraint, the nonvolatile memory chips have to be inserted by

the programmer to be programmed or be erased. Later, they are inserted back into the system. It is inconvenient to use nonvolatile memory in this fashion at system level. With the introduction of Flash EEPROM memory in the early 1990s, the high-voltage potentials could be generated on the chip itself. There is no need to use external power supplies to supply high voltages any more. Flash memory chips can be programmed directly or erased without the need for external high voltage sources. "In-system programming" is the term for on-chip programming or erasure.

In the early days, the high-voltage potentials needed by EPROM/EEPROM<sup>1-3</sup> were near 10~12 V. Later, as the chips migrated from the desktop to handheld devices, the power supply used in the system was scaled down to reduce the power consumption. At the same time, the off-chip high-voltage generation was migrated onto the chip itself. Now the charge pump is an essential component in many chip designs. Without the advance of technology, it is impossible to imagine the proliferation of cell phones, digital cameras, and MP3 players. They all rely on the nonvolatile memories in the system to store critical information. The charge pumps are the essential building blocks that make those applications feasible.

Technology determines the architecture and circuit design. To design a charge pump, designers need to understand the system supply voltage, oxide, and resistor and transistor specifications. All of these play a crucial role in determining circuit performance, total die size, and the power consumption of the design for the final products.

##### 4.1.1 System supply voltage

The power supply for the chip comes from the system board where it is being soldered. The absolute levels of supply voltages are fixed by the system specification. For example, if the chip was designed for boot information storage on the PC main board, the supply voltage would be around 5 V. If the chip was designed to be used in a handheld device, such as a cellphone or PDA, the available supply voltage would be around 1.8 V or 3 V from the batteries. The charge pump uses the given chip power supplies as its charge source. It also uses the chip power supplies as the drives for all supporting pump circuits, such as clock generation and clock buffers. After the initial charge enters into the pump stages, its potential energy is elevated by the work of capacitive coupling. In each clock cycle, the change of potential energy depends on the clock driver strength, boosting capacitance sizes, and parasitic loadings on internal nodes. Ideally, if there is no loss of charge during transfer, and if all circuits are operating at 100% efficiency, then all the power consumed from the chip power supply by the charge pump should be delivered completely to the output of the charge pump. If the pump output power consumption is fixed, then Equation 4-1 should

hold true at all times for the same pump circuit operating at different power supply voltages.

$$\begin{aligned} P_1 &= V_1 \times I_1 \\ P_2 &= V_2 \times I_2 \\ P_1 &= P_2 \Rightarrow V_2 \times I_2 = V_1 \times I_1 \end{aligned} \quad (4-1)$$

For example, in an ideal case, a 5 V charge pump, which consumes 10 mA of  $I_{cc}$  current, has a total power consumption of 50 mW. The output of a charge pump can deliver 50 mW of power at its regulation level. For an equivalent 3 V design to have 50 mW output power capability, the charge pump should consume 16.7 mA of  $I_{cc}$  current from 3 V power supply. The total power delivered to the output of the charge pump is still 50 mW. In reality, charge loss occurs during transfer due to parasitic loadings and due to the threshold of NMOS diodes. Also, the circuit efficiency is not 100% and all peripheral supporting circuits consume power.

Figure 4-2 shows the ideal relationship between the  $I_{cc}$  of a pump versus the chip supply voltage for an ideal pump with 100% circuit efficiency. If the pump output power requirement is not changed, the total power

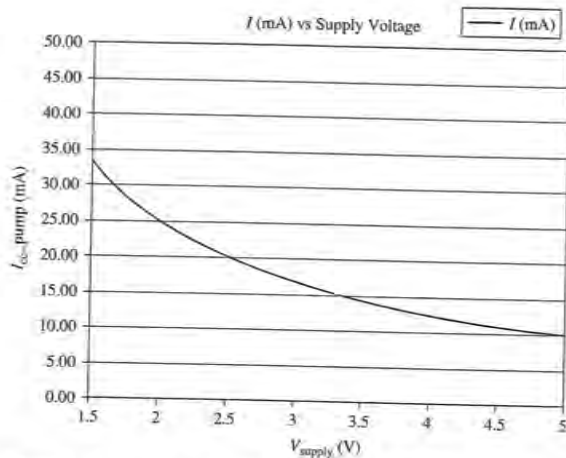


Figure 4-2  $I_{cc}$  of pump vs. chip supply voltage in an ideal case.

consumed from the chip power supply should not change. Stated another way, the product of  $I_{cc}$  and  $V_{cc}$  should be constant. In reality, Equation 4-1 will not hold because the power supply voltage varies for the identical pump design. There is the loss of charge due to parasitic capacitive loadings on each stage, and there is the loss of charge due to the threshold of NMOS diodes at each pump stage. Peripheral supporting circuits will consume some amount of power in addition to the actual pump. In addition, the circuit efficiency is reduced dramatically as the supply voltage is scaled down.

As the chip power supply voltage is scaled down, as shown in Figure 4-2, the  $I_{cc}$  current would increase inversely proportionally. In reality, the efficiency of the transistor and the efficiency of the capacitance would drop as the driving voltage is scaled down. Current drivability would decrease as  $V_{gs}$  and  $V_{ds}$  are reduced. The threshold voltage of diode-connected NMOS becomes a more significant factor to impede the transfer of charge between stages. In order to meet the same pump output performance at a lower chip power supply voltage, the sizes of transistor and capacitance have to be increased further rather than proportionally with scaling of supply voltage to compensate for the reduction of supply voltage.

In Figure 4-3, a realistic  $I_{cc}$  versus  $V_{supply}$  curve is plotted on top of the ideal curve discussed in Figure 4-2. As the chip supply voltage is scaled down, the pump has to consume more current and more power in order to meet the same pump output performance. The actual trendlines head upward away from the ideal line as the chip power supply goes lower

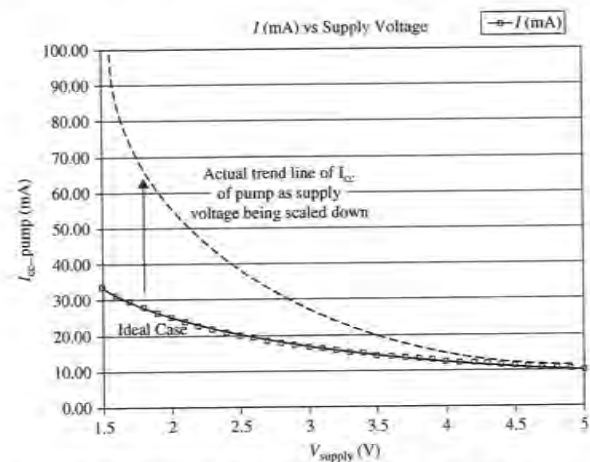


Figure 4-3  $I_{cc}$  of pump vs. chip supply voltage in a realistic case.

and lower. The power efficiency of the charge pump worsens with the scaling of the chip supply voltage. The phenomenon of increased power consumption with the scaling down of the chip power supply has a great impact in terms of the overall chip power specification.

For ASIC chips, the scaling of technology allows the scaling down of oxide thickness as well as the scaling down of the transistor dimension. More transistors can be packed into the same layout area on the silicon. As a consequence, thinner oxide allows the chip power supply voltage to be scaled down to prevent oxide reliability issues. The direct benefit is that the total power consumption for the same number of transistor counts is reduced.

Figure 4-4 shows a simple logic circuit with one inverter driving another inverter. Node *In* switches from high to low, and node *Out* is charged up by the chip supply voltage  $V_{\text{supply}}$ .

Equation 4-2 represents the total loading capacitance seen by the driving inverter. The gate oxide thickness is assumed to be identical for both NMOS and PMOS transistors. The total loading capacitance is the product of unit gate capacitance  $C_{\text{ox}}$  and the total transistor gate area, as shown in Equation 4-2. As technology is scaled down, the oxide thickness is also scaled down and  $C_{\text{ox}}$  increases. At mean time, the dimensions of the transistors are scaled down too and this allows the total gate area to reduce. From Equation 4-2,  $C_{\text{total}}$  is unlikely to vary strongly with the scaling of technology.

$$C_{\text{load}} = C_{\text{ox}}(W_p L_p + W_n L_n) \quad (4-2)$$

$V_{\text{supply}}$  is scaled down as the gate oxide thickness is scaled down to prevent any oxide reliability issue. Equation 4-3 represents the total charge consumed for one clock cycle for this simple logic gate. Because the loading capacitance is assumed to remain unchanged with the scaling of technology,  $Q_{\text{total}}$  is scaled down with the technology, as shown in

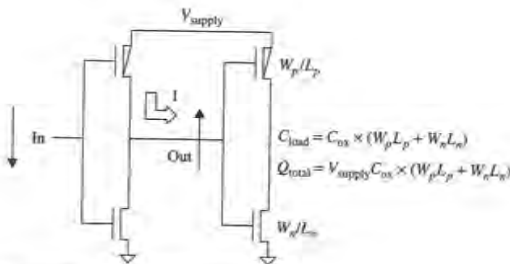


Figure 4-4 Charging of an inverter.

Equation 4-3. If the transistor count does not increase, the total power consumption of the same ASIC chip design should be scaled down with the technology. However, this phenomenon is unlikely to occur because the transistor count is usually increased to include more functions on-chip in new designs. The trend of total power consumption of ASIC chips would stay flat or go up against the technology scaling trend.

$$Q_{\text{total}} = V_{\text{supply}} \times C_{\text{load}} = V_{\text{supply}} C_{\text{ox}}(W_p L_p + W_n L_n) \quad (4-3)$$

For chips (including on-chip charge pumps) to generate high-voltage potentials, it is a different story regarding power consumption. Consider the following equations:

$$Q_{\text{ramp}} = \frac{V_{\text{regulation}} \times C_{\text{load}}}{\# \text{ cycles}} \quad (4-4)$$

$$Q_{\text{regulation}} = I_{\text{load}}(t) \times t_{\text{cycle}} \quad (4-5)$$

$$Q_{\text{cycle}} = Q_{\text{ramp}} + Q_{\text{regulation}} \quad (4-6)$$

$$Q_{\text{cycle}} = \frac{V_{\text{regulation}} \times C_{\text{load}}}{\# \text{ cycles}} + I_{\text{load}}(t) \times t_{\text{cycle}}$$

Equation 4-6 denotes the charge per cycle that needs to be delivered to the output of the charge pump as a function of time ( $t$ ). The first term in Equation 4-6 is represented by Equation 4-4. It denotes the minimum average charge transferred per clock cycle to charge up the output capacitive loading within the specified ramp-up time requirement. If the architecture does not change,  $C_{\text{load}}$  is unlikely to be changed with the same design.  $V_{\text{regulation}}$  is a technical parameter specified by design requirements. It is also unlikely to vary too much from generation to generation. The second term in Equation 4-6 is represented by Equation 4-5. It denotes the additional charge consumed by the current load circuits connected to the output of the charge pump. This term is unlikely to be changed unless the architecture or the loading circuit is changed. Because Equation 4-4 and Equation 4-5 are not likely to vary too much with the technology, the total power consumption required on the output of the charge pump is not likely to be scaled down with the scaling of technology. As a consequence, the total power consumption of the entire charge pump is not likely to be scaled down either. This phenomenon is fundamentally different from most ASIC designs.

Chips have a power budget determined by system design. As the power supply drops, to meet the same output performance for the charge pump (from Figure 4-3 and Equation 4-6), the  $I_{cc}$  current of the chip must increase at a higher rate than the rate of voltage supply scaling. It is important for designers to understand what circuits are being designed in the first place; then they can correctly define the chip specification for overall power consumption and chip performance. Another bad impact due to the lowering chip power supply implies that a larger capacitive area is needed to transfer more charge within each clock cycle to meet the same charge pump performance. If the performance of the pump is not allowed to be scaled down with the lowering power supply, the  $I_{cc}$  current needs to increase dramatically, as shown in Figure 4-3. In order to conduct a larger current than designs from previous generation, the overall pumping capacitance must increase at least proportionally to meet the output current required. Larger capacitance translates to a larger layout area, a larger clock driver size, more buffering stages for pump clock drivers, a longer route for the signal to travel, and larger parasitic loadings on all internal nodes of charge pump. All these have a negative effect on the performance of the pump. To further compensate for these losses, the size of the pump must be increased further. Overall, these negative impacts do not favor the charge pump design at all.

Another impact related to chip power supply is power bus planning, which could indirectly lower the supply voltage near the circuits even further. Even with the given chip supplies' specified system, tremendous effort is required to work out the details of power bus planning. Insufficient or bad power bus planning on-chip can cause the internal power supplies to have significant offsets from the voltage levels given the power supply pins. On the system level, bad board design and incorrect noise filtering on the power bus could cause additional voltage drops on the power supplies before even reaching the power supply pins of the chips.

Charge pump designers need to have enough noise margin estimated for the power supplies. Larger noise margins will not kill the designs. Of course, too much noise margin built into a design will increase the die size and total power consumption. Derivative projects can always improve based on silicon debugging from the mother chip. An underestimate of the power margin could kill the design and might require an expensive full-mask revision as well as delay production.

#### 4.1.2 Silicon dioxide ( $\text{SiO}_2$ )

Silicon dioxide ( $\text{SiO}_2$ ) is one of most important building blocks for semiconductor devices. It is used not only for isolation purpose, such as gate oxide or field oxide, but also as an active circuit component in many

analog circuits. In pump design, the charge pump needs to transfer the charge from lower potentials to higher potentials. The common practice of changing potential energy for charge is based upon charge conservation theory and capacitive coupling technique. The pumping capacitance per stage is determined by the pump output ramp-up speed and the pump output power consumption. Efficient unit area capacitance and a reduction of the total layout area for the pump capacitance required are both crucial in charge pump design.

Let us review the capacitance at a basic level to see how it is used to meet the pump design requirements. As shown in Figure 4-5, two conducting plates, separated by some dielectric material, form the capacitor. In Figure 4-5, the dielectric material is assumed to be  $\text{SiO}_2$ .

Plate A is connected to the anode of the battery, and plate B is connected to the cathode of the battery. The potential voltage of the battery is  $V$ . As the positive charge is accumulated on the top plate, the negative charge is attracted toward the bottom plate. Because two plates are separated by  $\text{SiO}_2$ , no conduction can occur between nodes A and B. The electric field,  $E$ , is built up and points from plate A to plate B.

The calculation for the parallel plate capacitance is given in Equation 4-7 for the sake of this discussion.  $\epsilon_0$  is defined as the permittivity of the dielectric material.  $\text{SiO}_2$  is one of the most commonly used dielectric materials for semiconductor processing.  $\epsilon_r$  is the relative permittivity of the air. Also, *area* is the total effective overlapping area between the two parallel plates, and *thickness* is the thickness of the dielectric material, or the spacing between the two parallel plates. From Equation 4-7, if the dielectric material is fixed, the total capacitance ( $C_{\text{total}}$ ) is proportional to the area of overlapping capacitance, and is inversely proportional to the distance separating those two plates.

$$C_{\text{total}} = \epsilon_0 \epsilon_r \frac{\text{area}}{\text{thickness}} \quad (4-7)$$

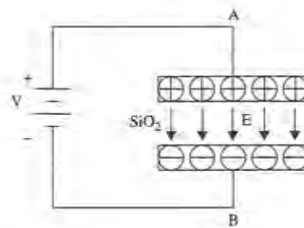


Figure 4-5 Capacitance connected to a battery.

In actual design, the gate capacitance of a MOSFET is usually chosen as the boosting capacitance because of its maximum capacitive efficiency over area. Two of the conducting plates are normally made of polysilicon and the channel formed underneath the gate. In between the two plates is  $\text{SiO}_2$ , which separates the two conducting nodes. This gate capacitor is commonly used. The dielectric material could be silicon dioxide, nitride, ONO, and so on. The MOSFET transistor has to be biased in either the accumulation region or the inversion region to maximize its equivalent gate capacitance equal to the gate oxide capacitance. The gate oxide is chosen as the boosting capacitance because its thickness is approximately  $1/10 \sim 1/20$  of the thickness of the field oxide. It is the thinnest dielectric layer that can be found on silicon. It is also well controlled in terms of the silicon process. For the same silicon area, the gate capacitance generates the largest unit capacitance over an area compared with other types of capacitance.

Figure 4-6 shows the generic representation of MOSFET gate capacitance as well as various components related to gate capacitance. The polysilicon gate is one of the terminals for the capacitor. The other terminal is the connection for the source/channel/drain of the NMOS transistor. It is ideal to have MOSFET gate capacitance identically to  $C_{ox}$ , with no variation in term of biasing voltages. Also, the design should be simple. However, this capacitance does vary with the biasing conditions of the MOSFET. Figure 4-7 shows the gate capacitance of NMOS versus  $V_{GB}$  biasing. The C-V curve of NMOS gate capacitance can be divided into three regions, as shown in Equations 4-8 through 4-10.

$V_{GB} > V_{Tn}$  defines the inversion region.

$$C_{gate} = C_{ox} \quad (4-8)$$

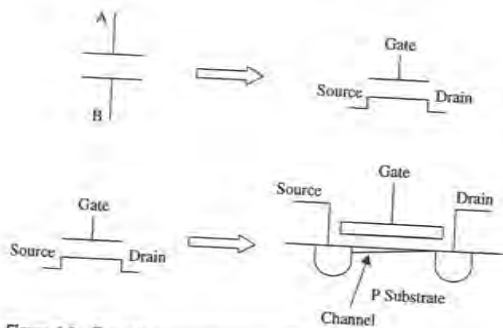


Figure 4-6 Generic representation of MOSFET gate capacitance.

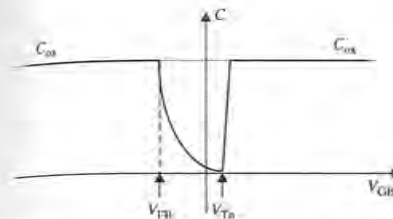


Figure 4-7 C-V curve of NMOS gate capacitance.

$V_{Tn} > V_{GB} > V_{FB}$  defines the depletion region.

$$C_{gate} = \frac{C_{dep} C_{ox}}{C_{dep} + C_{ox}} = \frac{C_{ox}}{1 + \frac{\epsilon_s t_{ox}}{\epsilon_{ox} \lambda_{dep}}} \quad (4-9)$$

$V_{FB} > V_{GB}$  defines accumulation region.

$$C_{gate} = C_{ox} \quad (4-10)$$

As the biasing conditions change for NMOS, the equivalent gate capacitance varies between inversion, depletion, and accumulation regions. It is important to choose the right devices and operate them in the targeted regions. Otherwise, the gate capacitance cannot be simply treated as  $C_{ox}$  in design.

The capacitance is inversely proportional to oxide thickness. The thinner the gate oxide is, the greater the gate capacitance per unit layout area. Stated another way, over the same silicon area, more charges can be stored across the capacitance with thinner dielectric material in between. Chip design demands this unit capacitance over area to be as high as possible to reduce the overall layout area. However, the oxide thickness (or the dielectric thickness) cannot be scaled down forever. The breakdown voltage of the dielectric material constrains the design. Oxide breakdown has always been a serious reliability concern in the semiconductor industry because of the continuous pressure toward smaller and thinner devices. There are three reasons to drive down the gate oxide thickness. The first reason is that smaller device dimensions can be achieved. This allows more and more devices to be packed into the same area compared with previous technologies. The second reason is that the current drivability is increased with the same device dimension. The third reason is that shrinking down the oxide thickness allows

the power supply voltage to be reduced. Less power consumption should be expected over the same transistor count.

Higher electrical fields in the oxide increase the tunneling of carriers from the channel into the oxide.<sup>4</sup> These carriers slowly degrade the quality of the oxide and lead, over time, to failure of the oxide. This effect is referred to as *time dependent destructive breakdown* (TDDDB). The thinnest oxide layers today are already less than 50 angstroms thick. An oxide layer can break down instantaneously at  $8-11 \times 10^6$  V per centimeter of thickness, or 8–11 V per angstrom of thickness. Based on the voltage levels required, the oxide thickness should be carefully designed for high-voltage devices to avoid any reliability issues.

#### 4.1.3 Resistor

The resistor is another critical component in charge pump design. Important characteristics of a resistor are sheet resistance of material, temperature coefficient in design, junction leakage and bias-dependant of resistance.

The effect of the resistor could be good or bad, depending on the application. Most of the time, resistance causes a bad effect in circuit designs. The *RC* delay, the first order delay metric of a circuit, of any wiring is one of the main causes for the slowdown for signal transitions. The charge pump relies on capacitive coupling to transfer charge and to elevate the potential energy of the charge. Because pump clocks are operating at a relative high frequency to drive large boosting capacitance, the metal layers used by pump clocks contribute to the *RC* delay of clock signals. As the charge is transferred from stage to stage, the interconnecting metal layers between the stages cause extra *RC* delay, thus impeding the transfer. The effect of resistance mentioned here is an example causing bad charge pump performance.

Sometimes the circuit designs do use resistors to perform important functions. For example, the simple delay element used in design can be purely a *RC* delay circuit. In terms of pump regulations, the resistor divider scheme is commonly used. In analog circuit design, the resistor also plays many important roles.

The first parameter related to the resistor should be the sheet resistance. Figure 4-8 shows the cross-section view of a conducting wire. The width of the wire is  $W$ . The thickness is  $t$ , and the length of the wire is  $L$ . The total resistance is given in Equation 4-11. The sheet resistance (or the resistance per unit length) of the material is described in Equation 4-12. Choosing the right type of material with the right range of sheet resistance is very important to design. The sheet resistance varies with different material. It is process-dependent. Metal layers have very small sheet resistance. WELL material has very high sheet resistance. Sometimes, if a design needs mega-ohm resistance, it does

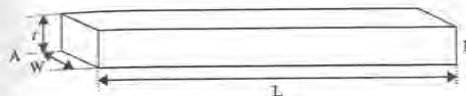


Figure 4-8 Sheet resistance of a conducting material.

not make sense to choose metal layer for the resistor. The layout could explode due to the metal layout's small sheet resistance. On the other hand, if 1 ohm of resistance is needed, it does not make sense to use WELL resistance. The discrepancy due to width and length could cause a much larger discrepancy in terms of the final resistance value.

$$R_{\text{total}} = \frac{\rho L}{tW} = \left( \frac{\rho}{tW} \right) L = R_{\text{sheet}} L \quad (4-11)$$

$$R_{\text{sheet}} = \frac{\rho}{tW} \quad (4-12)$$

The second parameter associated with the resistor is the temperature coefficient of the material. As temperature varies, the effective resistance will vary, as shown in Equation 4-13.  $R_0$  is the initial resistance at temperature  $T_0$ , which is normally set at 27°C.  $\alpha$  is the temperature coefficient of the conducting material. The degree of variation is material-dependent. In charge pump design, if the resistor divider scheme is used for pump regulation, the regulation level of high-voltage output is immune to the temperature variation. However, the DC current consumption through the resistor network will vary as the temperature changes. Sometimes the output of the charge pump goes through a low-pass filter before connecting to the loading circuits. *RC* delay would vary with the temperature. For many analog circuit designs, such as band gap reference, reference current generation, and so on, the temperature coefficients of resistors need to be studied.

$$R(t) = R_0 [1 + \alpha(t - T_0)] \quad (4-13)$$

The third parameter associated with resistance is the junction leakage current of the resistor. Resistor types, such as metal resistors and ploy resistors, have no leakage to neighboring dielectric materials. Other type of materials, such as  $n^+$  resistors,  $p^+$  resistors, and NWELL resistors, all have some amount of leakage through the junction. This current could be represented by the reverse-biased diode leakage. What is the concern of current leakage through the resistor? It is all about power consumption of the charge pump. A charge pump design-type of circuit

is not power efficient. To deliver 50  $\mu\text{A}$  of current on the output at a specified regulation level, the pump itself could consume 5 mA of current from its power supply. As the level of regulation increases, or as the power supply voltage scales down, this power efficiency is reduced even further. Minimizing any unwanted current on the output of the pump is becoming a crucial issue. Even if no physical resistor is intended to be connected to the output of the pump, all the source/drain junctions of the transistors and the WELLS connected to the output of the pump act as sources of leakage. They are an addition to the current load on the output of charge pump, and they tend to cause the pump to burn extra power and increase the size of the pump design.

The fourth parameter associated with the resistor concerns bias-dependent resistance. This parameter is mainly related to  $n^+$  resistors,  $p^+$  resistors, and WELL resistors. As the potential voltage changes across the surface of the resistive material, the depletion region underneath changes with the biasing voltage. As a consequence, the sheet resistance may not be the resistance expected from the design manual. This variation could affect pump output regulation levels, timing delay, and so on. Understanding the effect allows the countermeasure to be built into the actual circuits.

#### 4.1.4 Transistor specification

Transistors are the fundamental building blocks for CMOS circuits. Many device characteristics are important to the design, such as  $I_{ds}$ , threshold voltage  $V_t$ , and body bias effect,  $W_{max}$  and  $L_{min}$ , of transistors, junction breakdown voltage, gate oxide breakdown voltage and snapback effect, etc.<sup>5,6</sup>

For high-voltage charge pump design, transistors can be divided into two main categories: low-voltage transistors for logic functions and high-voltage transistors used in high-voltage paths. Low-voltage transistors are mainly used for logic operations in the peripheral of the charge pump. Functional blocks, such as clock generation, the delay element, and buffers are normally implemented by low-voltage type transistors. The potential voltages used by those blocks are typically chip supply voltages or potential voltages being regulated down from those sources. They are no different from the devices used for other logic operations. High-voltage transistors are used for connections in the high-voltage path, not only in the charge pump itself, but also all the peripheral circuits that experience high-voltage potential in operations. The device characteristics of the normal low-voltage transistors, such as  $I_{ds}$ , threshold voltage  $V_t$ , body bias effect, and so on, are equally important to high-voltage transistors. Due to high-voltage potentials, other aspects need to be specifically watched. Source/drain punchthrough issues, junction breakdown, oxide breakdown, and the snapback effect are all critical in dealing with high-voltage devices.

High-voltage transistors need to conduct current and transfer charges.  $I_{ds}$  is an important parameter that can be used to characterize transistors.  $I_{ds}$  in the linear region is expressed in Equation 4-14, and  $I_{ds}$  in the saturation region is expressed in Equation 4-15. Current drivability is important in terms of charge pump performance, such as the precharging of internal nodes and the charge transferring between stages. The higher the current drivability, the faster the charge can be delivered. Internal nodes can reach equilibrium faster, and faster clock frequency is applicable.

$$I_{ds} = \mu_n c_{ox} \left( \frac{W}{L} \right) \left[ (V_{gs} - V_t) V_{ds} - V_{ds}^2 / 2 \right] \quad (\text{Linear region}) \quad (4-14)$$

$$I_{ds} = \frac{1}{2} \mu_n c_{ox} \left( \frac{W}{L} \right) (V_{gs} - V_t)^2 \quad (\text{Saturation region}) \quad (4-15)$$

The threshold voltage  $V_t$  of a high-voltage MOSFET is very crucial. It affects the charge transfer, leakage, and size of the charge pump. The effect of body bias further compounds the problem even more severely as the source bias rises. Equation 4-16 measures the  $V_t$  of a MOSFET with source bias  $V_{sb}$ . The threshold of an NMOS device prevents the charge from being fully transferred between the source and drain of the device. At each pump stage, the efficiency of the charge pump is reduced due to its inability to transfer all the charge. A five-stage charge pump with 3 V power supply and 1 V threshold voltage has only 12.5% output power efficiency in an ideal case. If an eight-stage charge pump is used under similar conditions, the output power efficiency is only 2.4%. These cases do not consider the variation in threshold voltage. If we consider the body bias effect on  $V_t$ , later stages would suffer an even bigger voltage drop between source and drain nodes than that of the earlier stages. For charge pump design, it is preferable to have a high-voltage device with a low  $V_t$  and the minimum body bias. However, the threshold voltage cannot be too low due to concerns explained later in this chapter regarding device leakage and punch-through issues.

$$V_t = V_{t0} + \gamma \left( \sqrt{2|\Phi_F| + V_{SB}} - \sqrt{2|\Phi_F|} \right) \quad (4-16)$$

The dimensions of high-voltage transistors can impact the layout size of charge pumps. In terms of the transistor width, the maximum width ( $W_{max}$ ) allowed is determined by process and device modeling. Devices wider than  $W_{max}$  are not allowed by layout rules. Modeling of devices in test chips has a limited dimension. It can cause incorrect projection in simulation for a device that is significantly wider than those used on

the test chip for device modeling. A wider device must be broken down into a device with smaller, multiple fingers, as shown in Figure 4-9. The extra spacing rules, contact rules, and routing of the wire will cause the size of the layout to increase. In terms of the length of the high-voltage transistors, the minimum value of channel length ( $L_{min}$ ) is specified. In the discussion of the  $I_{ds}$  of high-voltage transistors in Equation 4-14 and Equation 4-15, it is understood that decreasing  $L$  enables better current drivability for transistors with the same width. This is what designers wish for. However, for high-voltage devices,  $L_{min}$  cannot be minimized forever. The punch-through effect is a critical concern for high-voltage design. As the voltage across the junction increases, the width of the depletion region near the source/drain increases under bias. If the channel is too short, the depletion layers under the source and drain regions could merge together. This will form a continuous depletion region and allow current to flow between source and drain with low gate bias, such as  $V_{gs} < V_t$ . There is rapid increase of  $I_{ds}$  with increasing drain voltage. If this situation occurs, the source and drain of the MOSFET can be considered to be shorted together without the control of gate voltage.  $L_{min}$  is required for high-voltage transistor design rules under specified biasing conditions.

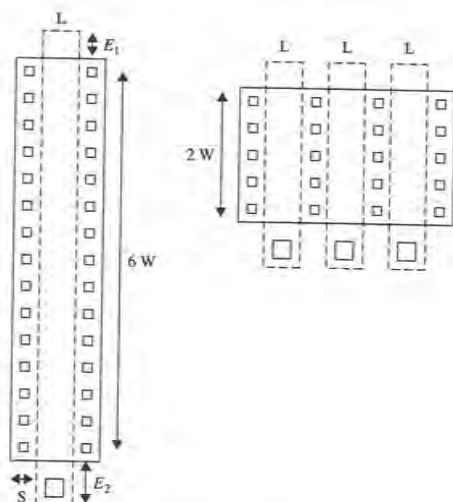


Figure 4-9 Breaking down a wider device into multiple fingers.

For high-voltage transistors, it is possible that either the source or the drain (or both) will connect to some high-voltage potentials. For high-voltage NMOS, the bulk is  $p^-$ , and the source/drain is  $n^+$ .  $p^-$  and  $n^+$  form a simple diode. If the reverse-biased voltage reaches a critical level, there will be a large increase in reverse-biased current as a result of junction breakdown. Once this happens, the transistor could be damaged. This critical level is called *junction breakdown voltage*. In designing high-voltage circuits, high voltage at all nodes should be checked carefully so as not to exceed this critical level.

Oxide breakdown of high-voltage transistor is covered in the section titled "Silicon Dioxide ( $\text{SiO}_2$ ).". Higher fields in the oxide increase the tunneling of carriers from the channel into the oxide. These carriers slowly degrade the quality of the oxide and over time lead to failure of the oxide. This effect is known as time-dependent destructive breakdown (TDDDB). Similar to the check for junction breakdown, all gate voltages in a high-voltage path should be checked so as not to exceed the gate oxide breakdown voltage.

Snapback is another phenomenon that can take place within a MOSFET transistor. When the carrier is hot, a high-energy particle could trigger snapback if the field across the drain region is sufficiently high. Snapback occurs when the parasitic bipolar transistor underneath MOSFET transistor is triggered. If the parasitic BJT (Bipolar Junction Transistor) is turned on, the result is a very high current between the drain and source region of the transistor. A special protection scheme should be used to turn on a MOSFET transistor if its drain is exposed to very high voltages.

## 4.2 Specification

Now that we have looked at the technology issues related to pump design, we will discuss the design specifications for a charge pump.

### 4.2.1 Output load characteristics

The charge pump is designed to provide high-voltage potential to the loading circuits. The characteristics of a loading circuit can dramatically affect the charge pump implementation. The load of charge pumps can generally be divided into two categories: capacitive load and current load. Figure 4-10(a) shows the generic representation of the capacitive load to a charge pump. As long as the load does not draw any DC current during the regulation phase of operation, it can be simplified into a capacitor. Figure 4-10(b) shows the response of pump output voltage over time for a capacitive load.  $T_{regulation}$  is the duration of time for the charge pump output to reach its regulation level. Before the pump output voltage reaches regulation, the required pump output current is

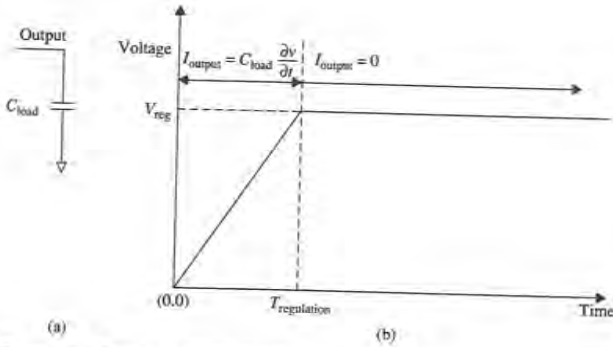


Figure 4-10 Capacitive load to the charge pump.

used to charge up the capacitive load, as shown in Equation 4-17. After the pump output reaches the regulation, no more capacitive charging current is needed. The pump output current in this region is described by Equation 4-18.

At any given time, the minimum current a charge pump design with a capacitive load should deliver to the output is the MAXIMUM function of Equation 4-17 and Equation 4-18. This is rewritten in Equation 4-19.

$$I_{\text{output}} = C_{\text{load}} \frac{\partial V}{\partial t} \quad t \leq T_{\text{regulation}} \quad (4-17)$$

$$I_{\text{output}} = 0 \quad t > T_{\text{regulation}} \quad (4-18)$$

$$I_{\text{output}} = \text{MAX} \left[ C_{\text{load}} \frac{\partial V}{\partial t}, 0 \right] = C_{\text{load}} \frac{\partial V}{\partial t} \quad (4-19)$$

Figure 4-11(a) shows the generic representation of the current load to a charge pump. As long as the load draws DC current during the regulation phase, the load can be considered a current load. Figure 4-11(b) shows the response of the pump over time for the current load.

Before the pump output voltage reaches regulation level, the load current consumed is shown in Equation 4-20. The first term is the average current needed to charge up the capacitive loading, and the second term is the current consumed by the DC current load (in general, due to a resistor divider type regulator) at any given time. After pump output reaches the regulation level, there is no more capacitive current for the capacitive load. The only component left is the current load at the

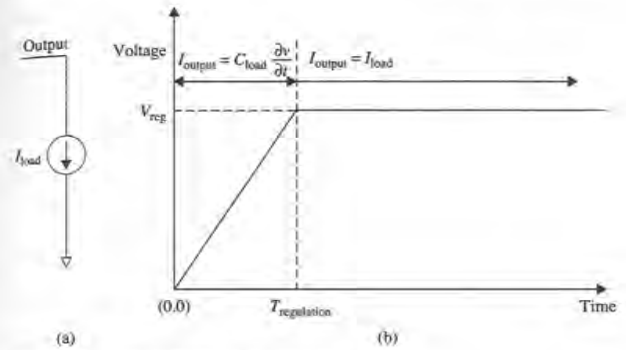


Figure 4-11 Current load to a charge pump.

given time shown in Equation 4-21. At any given time, the charge pump should meet the minimum current requirement delivered to its output. This is the MAXIMUM function of Equation 4-20 and Equation 4-21, and is rewritten in Equation 4-22.

$$I_{\text{output}} = C_{\text{load}} \frac{\partial V}{\partial t} + I_{\text{load}}(t) \quad t \leq T_{\text{regulation}} \quad (4-20)$$

$$I_{\text{output}} = I_{\text{load}}(t) \quad t > T_{\text{regulation}} \quad (4-21)$$

$$I_{\text{output}} = \text{MAX} \left[ C_{\text{load}} \frac{\partial V}{\partial t} + I_{\text{load}}(t), I_{\text{load}}(t) \right] \quad (4-22)$$

Comparing the capacitive load and the current load for the charge pump, the main difference is the DC load current  $I_{\text{output}} = I_{\text{load}}(t)$  during the regulation phase. It implies there is power consumption by the charge pump during regulation. In the capacitive load case, after the load capacitance is charged up to regulation level, no more current is needed on the load, and the charge pump can be shut off to cut off power consumption. In current load cases,  $I_{\text{load}}(t)$  is the component that always exists on the output of charge pump. As a consequence, the pump needs to supply current at all time, so the charge pump has to operate continuously, and so provides the power consumption of the whole circuit.

#### 4.2.2 Pump output voltage

Normally, the power supply is defined as the source voltage that can provides a certain amount of current at a given voltage level. The maximum

power that can be delivered by the supply is defined by the design specification. If the loading current exceeds this maximum requirement, the supply voltage will drop below its regulation level. The charge pump is designed to provide current/power at high-voltage potentials. Although the current is drawn from the output of the charge pump, eventually all the charge and power have to come from the given chip supply.

The performance of a charge pump can be broken down into two parts:

- **Pump output ramp-up speed** This is the pump output recovery speed if more capacitive load being connected in the middle of regulations.
- **Pump output power capability** This is the maximum current a charge pump can deliver while maintaining the output-regulated voltage levels.

Figure 4-12 shows the waveform of output voltage over time. It can be divided into four regions. The first region is for charge pump output to ramp up from its initial level to the final regulation level,  $V_{reg}$ . During this period, the load capacitance is charged up. The second region is a very short regulation phase. Then some capacitive load is connected to the output. Charge sharing will bring down the voltage at the output. Immediately, the pump responds and tries to recover back to  $V_{reg}$ . The last region is when the pump output goes back into regulation once again. This is the first plot designers need to understand. It carries a lot of design specifications that determine the final implementation of a charge pump. The generic high-voltage path from Figure 4-1 has been duplicated in Figure 4-13 for better illustrative purposes. The output waveform of Figure 4-13 would be exactly the same as the one shown in Figure 4-12.

Looking at Figure 4-12, we can extract several key design parameters from the plot. First, this pump is regulated at a level of  $V_{reg}$ . This is

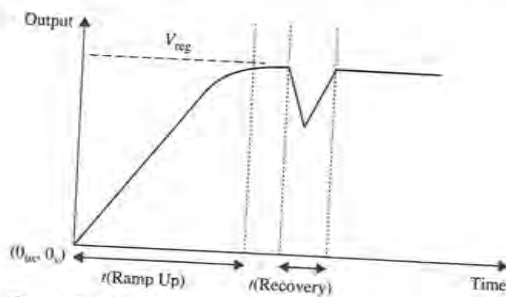


Figure 4-12 Charge pump output voltage versus time.

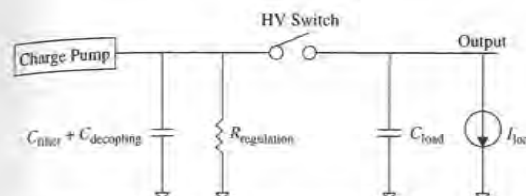


Figure 4-13 Generic high-voltage path with charge pump.

the targeted regulation level in this design. Second, it would take the pump a certain amount of time to ramp up the output node before it reaches regulation level. The time it takes is  $t_{rampup}$ , which is the amount of time to charge up  $C_{filter}$  and  $C_{decoupling}$  to  $V_{reg}$  in Figure 4-13. Third, as Figure 4-13 shows, when the HV switch connects capacitive load  $C_{load}$  to the pump output, charge sharing will immediately bring down the output voltage of the charge pump. It would take the pump some amount of time to recover back to its original regulation level. This recovery time is specified as  $t_{recovery}$ .

How do these parameters relate to the charge pump design? Pump output ramp-up (or recovery) speed is critical to the chip's performance. For example, a Flash memory chip has program and erase timing specifications. Usually for programming a byte or a page of information, the maximum programming pulse width is fixed based on program speed. To meet the speed requirement, within one program pulse, voltage on the wordline has to ramp up and stay at the regulation level for some minimum amount of time. Failing to do so causes inefficient programming and pushes out the actual program speed performance. Ramp-up of the pump output voltage requires the pump to charge up all capacitive loadings within the given amount of time. This is one of the design specifications.

Equation 4-23 calculates the total charge that needs to be transferred to the charge pump output to charge up the capacitive load; Equation 4-24 shows within one clock cycle how much charge should be transferred from stage to stage. This is proportional to the boosting capacitance size, and it is also proportional to the boosting clock voltage amplitude. The threshold of the diode-connected transistor inhibits the full transfer of the charge between stages. Equation 4-25 shows the average current the pump has to supply to pump output to meet the  $t_{rampup}$  requirement. The shorter the ramp-up time required, the more pump current is needed, and the larger capacitance required per stage.

$$Q_{rampup} = (C_{filter} + C_{decoupling}) \times V_{reg} \quad (4-23)$$

$$Q_{\text{cycle}} = C_{\text{boost}} \times (V_{\text{boost}} - V_t) \quad (4-24)$$

$$I_{\text{rampup}} = Q_{\text{rampup}} / t_{\text{rampup}} \quad (4-25)$$

The requirement in the recovery phase is similar to that of the ramp-up of the charge pump. It is assumed that the voltage across  $C_{\text{load}}$  is 0 V before the HV switch closes.

Because  $C_{\text{filter}}$  and  $C_{\text{decoupling}}$  have been charged up, the only capacitive component that needs to be charged up is  $C_{\text{load}}$ . Equation 4-26 through Equation 4-28 for the recovery phase are equivalent to Equation 4-23 through Equation 4-25 for the ramp-up phase. The charge pump has to meet the requirements of both cases.

$$C_{\text{filter}} Q_{\text{recovery}} = C_{\text{load}} \times V_{\text{reg}} \quad (4-26)$$

$$Q_{\text{cycle}} = C_{\text{boost}} \times (V_{\text{boost}} - V_{\text{th}}) \quad (4-27)$$

$$I_{\text{recovery}} = Q_{\text{recovery}} / t_{\text{recovery}} \quad (4-28)$$

#### 4.2.3 Pump output current

At fixed output voltage level, the pump needs to deliver current to the load, in addition to the regulation current if the regulation scheme demands DC current.

Figure 4-14 describes the I-V curve of a charge pump. This curve is also called the load line of the charge pump. The pump output current and the corresponding output voltage are plotted. Each point on the curve is the actual operating point of the pump. It represents how much the maximum output current pump is capable of delivering given the regulated voltage, or with the fixed current consumed on the pump output, the maximum regulation level at which the output of charge pump can be maintained. For example, node  $(V_1, I_1)$  indicates that if the output of the pump was regulated at  $V_1$ , the maximum current pump output that could be delivered is  $I_1$ . With any load current less than or equal to  $I_1$ , the output voltage can be regulated at  $V_1$ . However, if a current larger than  $I_1$  is demanded on the output of the pump, such as  $I_2$  ( $I_2 > I_1$ ), the pump output could not maintain its regulation at  $V_1$ . In Figure 4-14, it is easy to see that  $V_2$  is the maximum voltage level the charge pump can regulate with a load current of  $I_2$ .  $V_2$  is less than  $V_1$ .

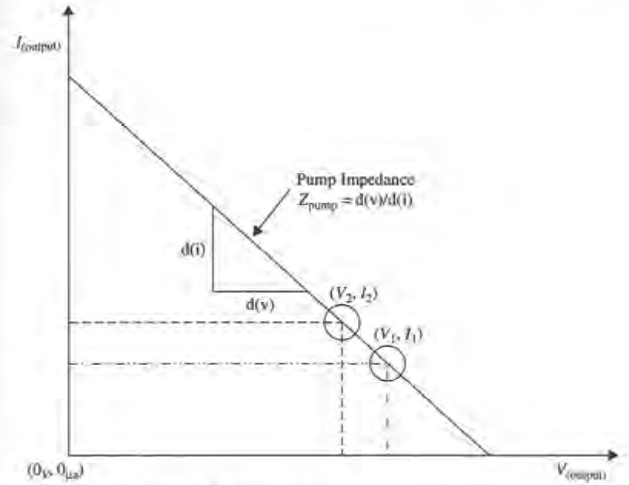


Figure 4-14 I-V curve of a charge pump.

In order to maintain the required output current, the boosting capacitance and the boosting voltage have to meet the requirement shown in Equation 4-29.

$$I_{\text{out}} = \frac{\partial [C_{\text{stage}} \times \{V_{\text{clock}} - V_t\}]}{\partial [T_{\text{clock\_period}}]} \quad (4-29)$$

In practice, the charge transferred per clock cycle from stage to stage has to be greater than or equal to the total charge consumption on the output per clock cycle.

Another important parameter can be derived from Figure 4-14—the slope of the I-V curve. The inverse slope of the I-V curve was described as the impedance of the charge pump. This impedance is related to many parameters in pump design, such as the pump clock frequency, the size of the pump-boosting capacitance, the boosting voltage per stage, the threshold voltage of the diode, the parasitic capacitance within each stage, and so on. The steeper the slope is, the better the current drivability.

Figure 4-15 shows the I-V curve of the charge pump with regulation. The design is regulated at  $V_{\text{reg}}$ . The maximum the pump can supply

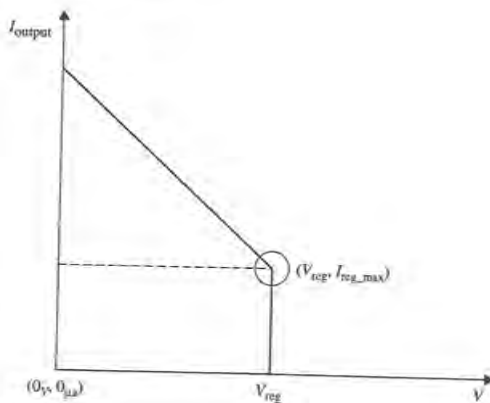


Figure 4-15 I-V curve of a charge pump with regulation.

at  $V_{reg}$  is  $I_{reg\_max}$ . If the load current exceeds this absolute value, the design would fall out of regulation. The new operation point can be found by traversing up the I-V curve based upon load current. The pump design is intended to increase  $I_{reg\_max}$  to be as large as possible.

#### 4.2.4 Ripple on regulated output voltage

With the regulation applied to the pump output, the output would have ripple (or noise) at the regulation level. The waveform of the ripple is shown in Figure 4-16. This noise is usually due to the finite gain of the amplifier used for regulation sensing as well as the delay in feedback control.

In Figure 4-17, the amplitude of ripple near regulation level is  $\Delta V$ .  $t_1$  depends on the time to discharge below regulation level, plus the loop delay of the feedback signal and the response of the control circuit.  $t_2$  depends on the characteristic of pump capacitance, pumping clock frequency, feedback loop delay, response of the control circuit, decoupling capacitance, load capacitance, and so on.

Noise reduction is a critical requirement in many applications. Because the regulation is part of the analog circuit design, definitely from the circuit point of view, detailed analysis can be done to improve the gain of the amplifier, to improve the phase margin, and so on. Two unique aspects should be emphasized here that are specific to charge pump noise reduction. One aspect concerns the balancing of power between the output of the pump and that of the loading circuits. The other aspect concerns the capacitive noise-filtering technique.

First, let us analyze the power that can be supplied by the charge pump and the power consumed by the loading circuits.

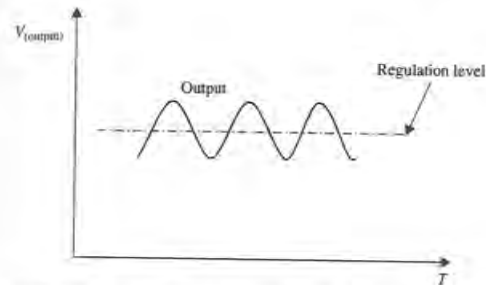


Figure 4-16 Ripple on regulated pump output voltage.

Figure 4-18 shows a generic representation of all the current sources connecting the output of the pump. In the regulation phase, Kirchoff's current law is observed at all times, as shown in Equation 4-30.

$$I_{output} = I_{load} + I_{shunt} + I_{capacitance} \quad (4-30)$$

If at any given time  $I_{load}$  and  $I_{shunt}$  can be maintained to match  $I_{output}$ , and if  $I_{capacitance}$  is equal to 0, then the output voltage would not have any noise because the output power of the pump is balanced by the power being consumed by regulation and load circuits.

However, this situation is rare. In many cases,  $I_{load}$  is relatively stable.  $I_{shunt}$  and  $I_{capacitance}$  need to balance the difference  $\Delta I = I_{output} - I_{load}$ . Due to the delay of feedback response time,  $I_{capacitance}$  needs to absorb the difference  $\Delta I$  right away. The effect is shown for  $\Delta V$  on the output voltage of the charge pump. If  $I_{shunt}$  finally responds, it will contribute to another part of the  $\Delta V$  change. Ideally,  $I_{output}$  should be designed in regulation phase to be near the range of  $I_{load}$ . Any variations in

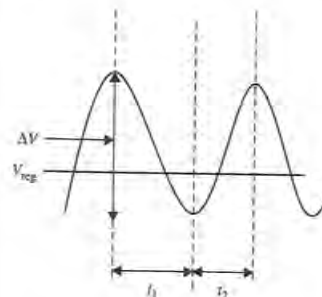


Figure 4-17 Amplitude of ripple.

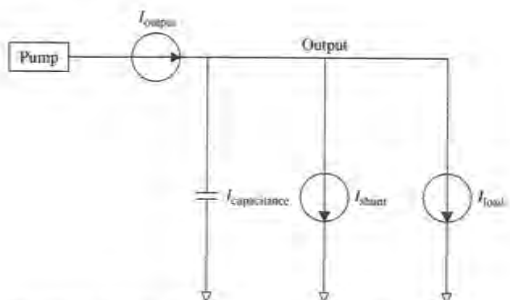


Figure 4-18 Dissection of current consumption on the pump output.

$I_{\text{capacitance}}$  and  $I_{\text{shunt}}$  could be a smaller percentage of  $I_{\text{output}}$ . This allows the potential noise on the output to be contained within a limited range at first.

In designing the charge pump strength, special design techniques can be applied to vary the output power strength based on the operation demands.<sup>7</sup> For example, in the ramp-up phase, regulation phase, or recovery phase, the output capability could vary to just to meet the demand of the load current. With this approach, the root cause of noise is addressed at first instead of the outcome of the noise due to the mismatching of the powers.

Also, capacitive noise filtering can be a very powerful technique. As shown in Equation 4-30, before  $I_{\text{shunt}}$  can respond,  $I_{\text{capacitance}}$  will be the only component that can respond instantaneously. The direct effect is shown in  $\Delta V$  on the pump output voltage.  $\Delta V$  has a direct relation to the total filtering capacitance size, as shown in Equation 4-31.

$$\Delta V = \frac{\Delta Q}{\Delta t_{\text{response}}} = \frac{(I_{\text{output}} - I_{\text{load}} - I_{\text{shunt}})}{\Delta t_{\text{response}}} \quad (4-31)$$

When the filtering capacitance size is increased,  $\Delta V$  would be reduced proportionally from the calculation in Equation 4-31. The need of  $I_{\text{shunt}}$  to reduce output noise in time is becoming less critical. The overall noise on the pump output would be reduced. The drawbacks of this approach include the increasing of the pump ramp-up speed and the possible increasing of the die size. More capacitance on the output node means longer settle time for output. Increasing the filtering capacitance would also increase the layout size. The correct sizing of decoupling capacitance for filtering is important in order to deal with both concerns.

#### 4.2.5 Pump regulation

Regulation of the charge pump is another important concern. Without regulation, the output of the charge pump cannot be determined. Depending on the needs of the application, the implementation of the regulation scheme can be quite different. A common approach of regulation involves some means to sample the pump output voltage and compare the sampled voltage with a known reference voltage through an amplifier. The comparison result is amplified and is fed back to control the pump or supporting circuits in order to maintain the output at regulation level. Many schemes can be used for pump regulation control.

#### 4.2.6 Capacitive divider

One type of commonly used pump regulation scheme is based on the capacitive divider method. This scheme is shown in Figure 4-19.  $C_1$  and  $C_2$  are two known capacitances. Before regulation is enabled, Node *div* has an additional device (not shown in the figure) to initialize the value. It is common to initialize *div* to be 0 V if positive voltage is being regulated on the output.

Once the regulation starts, the initialization devices are shut off. Conservation of the charge is observed on node *div* at all times. Of course, in this discussion, the leakage current on *div* is ignored. After the initialization, node *div* would be coupled up based on capacitive coupling from the pump output. This is shown in Equation 4-32 and Equation 4-33.

$$V_{\text{div}} C_2 + C_1 (V_{\text{div}} - V_{\text{output}}) = 0 \times (C_1 + C_2) = 0 \quad (4-32)$$

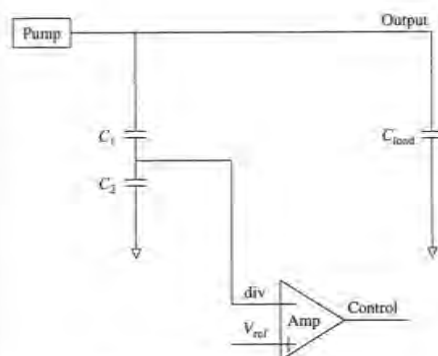


Figure 4-19 Capacitive divider feedback control.

$$V_{div} = V_{output} \times \frac{C_1}{(C_1 + C_2)} \quad (4-33)$$

Equation 4-32 shows the conservation of the charge on node *div* before and after output is being regulated. Equation 4-33 shows how the coupled voltage from  $V_{output}$  to *div* can be calculated. In regulation, the divided voltage on node *div* should be equal to  $V_{ref}$ , the reference voltage in regulation, as shown in Equation 4-34.

$$V_{div} = V_{ref} \quad (4-34)$$

The final regulation level of the pump output can be calculated based on Equation 4-35. The advantages of this approach are that the feedback speed is very quick and the capacitors do not take any DC current from the output of the charge pump when it is being regulated. As shown earlier, any reduction of the pump output current would allow the pump to be sized smaller, with less overall power consumption. The capacitor divider approach fits into this category.

$$\begin{aligned} V_{ref} &= V_{output} \times \frac{C_1}{(C_1 + C_2)} \\ V_{output} &= V_{ref} \times \frac{(C_1 + C_2)}{C_1} \\ V_{output} &= V_{ref} \times \left[ 1 + \frac{C_2}{C_1} \right] \end{aligned} \quad (4-35)$$

There are indeed a few disadvantages to using capacitance for feedback control. First, the capacitance value may shift with the process or biasing voltage. If capacitance is built based on oxide, the oxide thickness could vary from die to die, and from wafer to wafer. If capacitance is based on MOSFET gate capacitance, different biasing voltages across MOSFET would cause gate capacitance to go through different regions. In this case, the value of gate capacitance would change. Second, the parasitic capacitance from wiring and connection could change the actual values of  $C_1$  and  $C_2$  from design. Any deviation will cause the final regulated output to be shifted from the target values. Third, the capacitive divider is based on charge conservation on the middle node in between two capacitances. This middle node needs to be initialized to a certain value through devices. Any leakage currents through the junction diode or subthreshold current would make

conservation of the charge invalid and cause the regulation level on output to be off. The maximum regulation time allowed is based on the tolerance of the regulation level. Operations exceeding this maximum regulation requirement should always refresh the capacitor divider circuits.

#### 4.2.7 Resistive divider

Another type of commonly used regulation scheme is based on resistor divider feedback control. Resistors in series divide the high output voltage, resulting in a lower voltage that can be safely applied to low-voltage transistors in an amplifier.  $V_{div}$  is compared with the reference voltage and generates the feedback control signals. This scheme is shown in Figure 4-20. At any given time,  $V_{div}$  observes the relation given in Equation 4-36. In the regulation phase, the divided voltage  $V_{div}$  should be equal to  $V_{ref}$ . The level of pump output,  $V_{output}$ , in regulation can be calculated based on Equation 4-37.

$$V_{div} = V_{output} \times \frac{R_2}{(R_1 + R_2)} \quad (4-36)$$

$$V_{div} = V_{ref}$$

$$V_{ref} = V_{output} \times \frac{R_2}{(R_1 + R_2)}$$

$$V_{output} = V_{ref} \times \frac{(R_1 + R_2)}{R_2} \quad (4-37)$$

$$V_{output} = V_{ref} \times \left[ 1 + \frac{R_1}{R_2} \right]$$

There are advantages and disadvantages to using a resistor divider in a feedback scheme.

The first advantage of the resistive feedback regulation scheme is that the divided voltage for comparison is based on resistor ratio. This divided voltage does not change if the absolute value of the resistor changes due to doping variation, diffusion variation, and so on. The second advantage of the resistor divider scheme is that the temperature coefficients of resistors are cancelled in Equation 4-36 and Equation 4-37. The temperature coefficient of  $V_{div}$  would not have the components from the resistor divider. The third advantage is that this scheme can be easily implemented without too much worry concerning

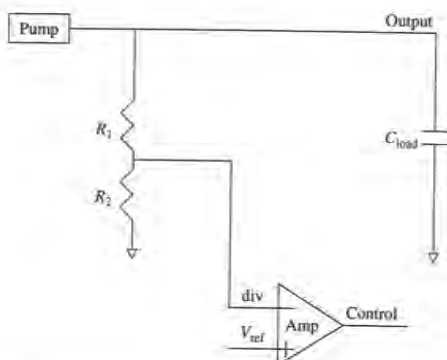


Figure 4-20 Resistive divider feedback control.

parasitic resistance. As long as the sheet resistance of the resistors is large, the impact from parasitic resistance due to metal and contacts is relatively small.

Now the designer should also be aware of some inherent disadvantage associated with the resistor divider type regulator. The first disadvantage of the resistor divider feedback scheme is the power consumption. With resistors connected to the output of the charge pump, it will always consume DC current in regulation. As mentioned earlier, charge pump power efficiency is not very good. If the output current is divided for regulation purposes, the remaining current that can be supplied to the load circuit is reduced. The size of the charge pump may need to be increased to compensate for this extra DC regulation current. Normally the resistance of the divider circuit has to be sized up to allow minimum current consumption while not hurting the layout area due to larger resistance. The second disadvantage is that the resistor divider network is slow in terms of delay in feedback response. The feedback path is based on an RC network. All resistors have some parasitic capacitance associated with them. Because resistance needs to be sized up to reduce unwanted DC current being wasted, the RC delay through the feedback path is normally very large. That is why the resistor divider feedback scheme is slower in response time compared with that of the capacitive divider scheme and could contribute to large output noise in regulation.

#### 4.2.8 MOSFET biased type regulator

A MOSFET biased type regulator can generally be used for pumps generating relatively lower voltages, and it is best suited for negative voltage

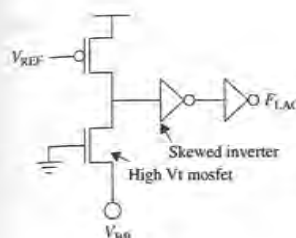


Figure 4-21 A simple biased MOSFET type regulator.

charge pumps. Consider the circuit in Figure 4-21. A high  $V_t$  NMOS device's source is tied to the output of the pump generating a negative output voltage. As  $V_{BB}$  starts to go low and reaches a preset voltage, the skewed inverter turns on and causes the output to toggle. A controlled  $V_{ref}$  voltage source can modify the pump regulation voltage.

#### 4.2.9 Which scheme should be used in design?

Choosing which scheme is best suited for a specific design depends on the requirement of the design. The direct impact from regulation schemes involves regulation speed and the noise on the pump output in regulation.

For example, suppose a charge pump design has the following requirements: to ramp up the output load in  $0.3 \mu s$ , to supply  $1 \text{ mA}$  of current at  $5 \text{ V}$ , and to have only  $100 \text{ mV}$  ripple on the output node in regulation. The capacitive divider feedback scheme may be a good candidate in terms of regulation speed and noise control. A capacitive divider can sample the output voltage almost instantaneously. The feedback signal may control a shunt current path to channel any extra current away from the output. If output current being delivered is lower than  $1 \text{ mA}$ , the feedback path can disable the shunt current path. The pump would be fully on because a constant current of  $1 \text{ mA}$  is required at the output.

As another example, let us say the charge pump has to ramp up the output voltage to  $10 \text{ V}$  in  $5 \mu s$  and does not need to supply any DC current in regulation. In this case, both capacitive and resistive approaches can be chosen to sample and stop the pump when it hits the  $10 \text{ V}$  regulation level. If a capacitive divider is used, there is a timing concern because intermediate nodes of capacitors would have junctions of transistors. The leakage current over long periods of time can cause the regulation to deviate from the target level. Another issue arises when the accuracy is affected due to extra parasitic capacitances. The size of regulation capacitance needs to be relatively large to minimize this effect. Using the resistive divider scheme in this case will not have

the same concerns. All errors due to parasitic capacitance and leakage current do not exist.

### 4.3 Pump Power Consumption

Power consumption of the pump itself is another critical specification that needs to be understood. While transferring the charge from input to output, the charge pump has to consume power in clock switching and capacitive boosting. Sometimes the charge pump only delivers a very small fraction of the power to the output; the total amount of power consumption by itself could be 20X or more. For example, let us say a charge pump with a 3.0 V power supply is capable of delivering 100  $\mu$ A at 10 V to the output by design requirement. To be able to generate this output power, the charge pump in general may be consuming 2 ~ 3 mA of current from the chip power supply. The power consumption specification affects many design considerations, such as charge pump performance, loading circuit, power supply voltage, device threshold voltages, architecture, pump regulation, and so on. It should be studied carefully in the early stages of design to have a realistic power budget on the final chip.

### 4.4 Die Size of the Charge Pump

Every circuit being put on silicon will take some amount of die area. A charge pump especially consumes area in many cases. As the technology is scaled down, the power supply voltage is scaled down, too. For the pump design, this trend can cause serious impact to the sizing of the charge pump.

As analyzed at the beginning of this chapter, the supply voltage is an important design factor. The area of a pump with a 3 V power supply would be very different from a pump with a 1.8 V power supply while providing the same pump performance. The layout area of the 1.8 V design can be 5X or more than that of the 3 V design. Even if technology is not scaled down, the application will require a die size reduction to maintain profitability. Under this scenario, every circuit being laid out on chip has to be as small as possible. Based on the optimization theory, if every circuit could be optimized to the minimum size, the total chip area summed should be the smallest in size. Charge pumps need to be optimized.

The size of a charge pump on silicon is determined by many parameters, such as performance requirement, device threshold voltage, architecture, regulation level, and so on. Optimizing the sizing requires the optimization of every parameter to deliver the best sizing, performance, and power.

### 4.5 Conclusion

This chapter covered some of the fundamental design criteria for charge pump implementation. They are essential to help designers understand the basic requirements, and the capability and the limitations of devices. With all these constraints in mind, circuit designers can have a global view of how the charge pump should be designed as a whole system. It allows most suitable pump architecture to be chosen and realistic implementation to be planned in order to achieve the final design goals. This process is key to any successful charge pump design.

### References

1. Umezawa, A., S. Atsumi, M. Kuriyama, H. Banba, Imamiya, K. Naruke, S. Yamada, E. Obj, M. Oshikiri, T. Suzuki, and S. Tanaka, "A 5-V-only operation 0.6- $\mu$ m flash EEPROM with row decoder scheme in triple-well structure," *IEEE Journal of Solid-State Circuits*, Vol. 27, pp. 1540-1546, 1992.
2. Witters, J. S., G. Groeseneken, and H. E. Maes, "Analysis and modeling of onchip high-voltage generator circuits for use in EEPROM circuits," *IEEE Journal of Solid-State Circuits*, Vol. 24, pp. 1372-1380, October 1989.
3. High-Voltage Generator Circuits for Use in EEPROM Circuits. *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 5, October 1989.
4. Oxide Breakdown <http://www.semiconfareast.com/oxidebreakdown.htm>.
5. Razavi, B. *Design of Analog CMOS Integrated Circuits*. McGraw-Hill, New York, 2001.
6. Gray, R.P., J.P. Hurst, H.S. Lewis, and G.R. Meyer. *Analysis and Design of Analog Integrated Circuits*. Fourth Edition. John Wiley & Sons, New York, 2001.
7. Pan. High voltage ripple reduction. U.S. patent 6,734,718.

## How to Design a Basic Charge Pump

Before starting to design a new charge pump, we need to analyze the design requirements, which will in turn help us define the major pump parameters. In general, the advantage gained by optimizing some of the parameters may be detrimental to other ones, and vice versa. For example, increasing the number of pump stages ( $n$ ) will result in a higher output voltage, but it also increases the internal impedance of the pump and thus cuts down the output current drivability. For a particular number of stages, the output drive current can be increased by increasing the size of the capacitors and the width of diode-connected MOSFETs. But the larger geometry MOSFETs and capacitors will also increase the wire routing length of the clock signals and other internal connections, which will increase the resistance and the parasitic capacitances on those nodes. As we saw earlier in Equation 3-11, parasitic resistance and parasitic capacitance play a significant role in determining the output voltage and the charge pump efficiency. Hence, the number of pump stages ( $n$ ) needs to be carefully chosen so as to orchestrate an optimum pump performance.<sup>1</sup>

Take another charge pump specification—the output voltage ramp-up time or the recovery time requirement. In general, charge pumps used in EEPROM and other applications that need a high-output voltage require the regulated output voltage to be available within a stipulated time after the pump is enabled. Further, most of these applications also require that during the time the pump is being enabled, if due to any sudden loading the pump output droops significantly from its steady regulated value, the charge pump should be able to pull up the output voltage back to the regulated level within a defined time, also called the *recovery time*. The factor that directly dictates the output ramp-up time and the recovery time is the

pump's output drive at any given voltage level. As mentioned, increasing the output drive current by sizing up the boosting capacitance and diode-connected devices has a direct effect on the pump's output voltage capability and its output efficiency. Another way to improve charge pump output current drivability is by increasing the pump clock frequency.<sup>2</sup> As could be seen in Equation 3-14a from Chapter 3, increasing the pump frequency can synthetically reduce the internal impedance of the pump. It effectively increases the output current drivability at any given possible regulated voltage level, thus reducing the ramp-up and recovery times. However, it may not be so easy to operate the pump at a very high frequency. In fact, the maximum frequency of pump operation is determined by the minimum time required to transfer charges from one stage to the next, completely or close to 90% of the maximum value. If the pump is operated at a frequency that exceeds the optimum limit, it will cause inefficient charge transfer and therefore poor overall charge pump performance, not to mention higher power consumption  $CV_{DD}^2 f_{osc}$ . Operating the pump at a frequency that is lower than the optimum limit means the charge pump is not being utilized at its full efficiency level. One of the parameters that limits the higher frequency of operation is the stage-to-stage parasitic capacitance. Optimum capacitor and MOSFET sizes and innovative layout techniques must be used to reduce parasitic capacitances.

In this chapter, we will discuss each of the major parameters in detail and consider how to create trade-offs between one and the others. We will also touch on layout implementation, floor planning, clock distribution, and regulator design—all of which help us start designing a charge pump.

### 5.1 Charge Pump Specifications

The charge pump, such the one listed in Table 5-1, should have the specifications available before its initial design. In general, part of the specifications will be derived from the system specifications, in which different

TABLE 5-1 Charge Pump Specifications

| Specifications                   | Name          | Sample Value |
|----------------------------------|---------------|--------------|
| Steady output voltage            | $V_{out}$     | 1.6 V        |
| Output voltage ripple            | $\Delta V$    | 0.1 V        |
| Output current requirements      | $I_{out}$     | 30 $\mu$ A   |
| Output load capacitance          | $C_{load}$    | 20 pF        |
| Output voltage ramp-up time      | $T_{ramp}$    | 10 $\mu$ S   |
| Output voltage recovery time     | $T_{rec}$     | 2 $\mu$ S    |
| Maximum pump layout area         | $Area_{pump}$ | —            |
| Pump power supply voltage        | $V_{DD}$      | 2.5 V        |
| Average pump current consumption | $I_{pump}$    | 1 mA         |

budgets are defined for different blocks of the system. The other part of the specification will come from device and circuit requirements.

#### 5.1.1 Output voltage

The first step in designing the charge pump is to determine the required output voltage. The voltage required may be a simple  $V_{DD} \pm \delta$  to supply the required back-bias voltage for DRAM memories or a very high output voltage, on the order of 25–30 volts, as required in flash memories. The number of stages ( $n$ ), along with the pump clock amplitude, directly dictates the final output voltage. The simple Dickson charge pump's equation can be used to determine the number of stages required to attain the required output voltage:

$$V_{out} = NV_{\phi} + V_{in} - (N+1)V_D \quad (5-1)$$

For example, if the clock amplitude ( $V_{\phi}$ ) is 2.5 V, the input voltage ( $V_{in}$ ) is 2.5 V, and  $V_D$  is 0.2 V (the  $V_t$  of the diode-connected NMOS transistor, with  $V_t$  assumed to be constant for now), then to generate a  $V_{out}$  of, say 16 V, we need at least six stages ( $N = 6$ ).

However, this is not enough. In general, the MOSFET at the sixth stage will exhibit  $V_t$ , which could be 1–2 volts or higher than expected, depending on the MOSFETs' characteristics.<sup>4</sup> Hence, the output voltage that can be delivered may be much less than 16 volts. Another problem is that we have not yet considered the effects of parasitic capacitance, which will consume charge and increase internal impedance. Finally, and most important, because the pump output is very exponential in nature toward its open circuit output voltage, it is never advisable to design the pump such that its required output voltage is very close to its maximum output voltage, in which case the pump will be deemed too weak.<sup>5</sup> Therefore, it is good practice to design a pump in which the required output voltage is about 70%–80% of its maximum output voltage capacity, and to use an active regulator to regulate the charge pump output.

Continuing with this example, we will start with a Dickson charge pump that has eight to nine stages (i.e., a pump that can theoretically deliver the potential output voltage that is about 30%–40% higher than the specification requirement).

#### 5.1.2 Current drivability

The current drivability is the next most important factor to look for. In general, the pump output current decreases linearly with the increased pump output voltage. The power efficiency of the charge pump decreases

with increased pump output voltage. The internal impedance of the charge pump may be represented as follows:

$$R_s = \left( \frac{N}{C + C_s} \right) \times \frac{1}{f} \quad (5-2)$$

To achieve a higher output voltage, the number of pump stages  $N$  needs to be increased, but as we can see from the previous equation, this also increases the effective internal resistance. Therefore, to achieve higher output current at a high pump output voltage, two important factors—the boosting capacitor size and the pump clock frequency—need to be increased. Also, another factor that is not accounted for is the simple equation—the MOSFET threshold voltage,  $V_t$ , as a function of source bias. As the absolute value of the output voltage increases, the sources of diode-connected devices will increase. As more pump stages are added, more  $V_t$  drops through the pump stages would occur. The pump internal impedance will increase. Taking  $V_t$  into account, a modified internal resistance may be represented as

$$R_s = \left( \frac{N}{C + C_s} \right) \times \frac{1}{f} \quad (5-3)$$

where  $R_{vt}$  represents the effective internal pump impedance imposed due to the gradual increase in  $V_t$  of the diode-connected devices along the chain, with the last stage exhibiting the highest absolute  $V_t$ .

Higher output drive current results in the faster output ramp-up and the quicker recovery times. Next, we will calculate the initial value of the pump boosting capacitance,  $C$ , in a quick two-step process:

1. Because the output load capacitance is already known, with the output ramp-up time requirement (defined in the pump specification), we can easily calculate the linear current required to charge up the output node to the regulation level within the given amount time.

Continuing with the previous example, assuming  $V_{out} = 16$  V,  $T_{rampup} = 10$   $\mu$ s, and  $C_{load} = 20$  pF. The linear current required to charge up the pump output can be calculated as follows:

$$i = C_{load} \frac{V_{out}}{T_{rampup}} = 20E^{-12} \frac{16}{10E^{-6}} = 32 \mu A \quad (5-4)$$

Hence, the pump should be able to deliver a steady linear current at a minimum of 32  $\mu$ A. Note that, in general, the output drive current must be a specification for the charge pump.

2. Now, armed with this information, we need to focus our attention on to the last stages of the Dickson charge pump and consider its operation with the two phases of the clock.

At steady state during regulation, assume the pump is delivering a steady current ( $I_{out}$ ) to the output while maintaining a constant  $V_{out}$  (see Figure 5-1). In this situation,  $I_{out}$ , the current consumed on the output matches exactly the current delivered by the pump. Extending this concept one step inside, the charge dumped on node 5, through the diode  $M_5$  from preceding stage, is equal to the charge lost through diode  $M_5$  within every clock cycle. This charge gain and loss is manifested as voltage increase and decrease on node 5.<sup>6</sup>

Assuming a long clock cycle, during the second half of the clock, the voltage drop on node 5 due to the output current  $I_{out}$  is equal to the voltage gain  $\Delta V$  due to the positive phase of  $V_{ob}$ . Assuming  $\Delta V$  is 800 mV

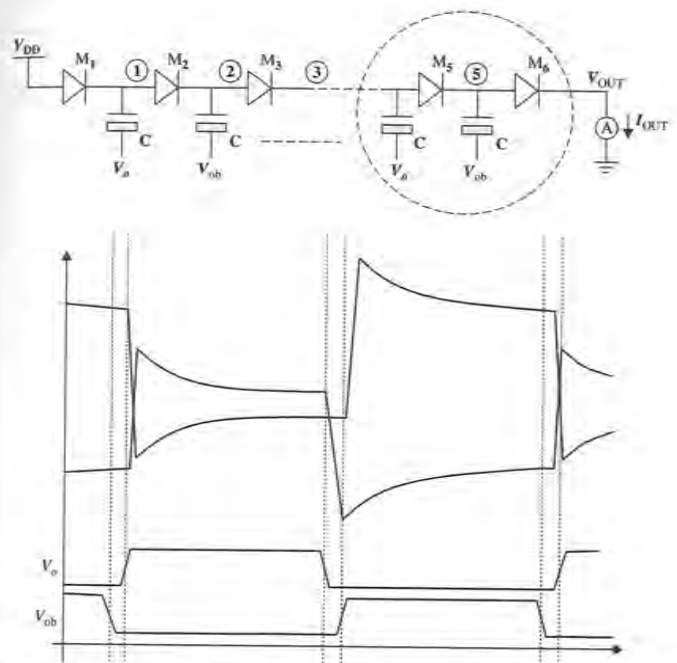


Figure 5-1 Charge pump's internal voltages.

(a typical approximation for the last stage) and the frequency of pump operation is 10 MHz, and because we have already calculated  $I_{out}$  from the previous step, we can use the same equation to determine an initial value of  $C$ :

$$I_{out} = 32 \mu A = C \frac{\Delta V}{\Delta t} = C \frac{800 \times 10^{-3}}{100 \times 10^{-9}} \quad (5-5)$$

$$\Rightarrow C = 4 \text{ pF}$$

Therefore, we can start our simulations with our initial assumption of 4 pF for all pump boosting capacitors. After observing the initial set of simulations, the size of the capacitors can be tweaked to meet different specifications. In general, the initial capacitance assumption is on the higher side and will come down, as the design is being optimized. One more thing to keep in mind while determining the output drive current is the size of the diode-connected MOSFET. In general, the MOSFET width should not be too large because that will increase source/drain parasitic capacitances and will also increase signal routing lengths. But the MOSFET width should not be made too small either because that will limit the MOSFET's  $I_{DS}$ , which in turn will hinder quick charge transfer from one stage to the next. In general, the initial width of the MOSFET can be calculated from the MOSFET's  $I_{DS}$ —versus  $V_{DS}$  curves, with a particular width chosen so that it meets the needed  $I_{out}$  current at saturation. Again, the MOSFET length should be as minimal as possible, but enough to meet all worst-case, device-related reliability parameters.

The current drivability issue is never complete without mentioning the impact of the circuit's capacitor and diode sizes on the overall pump performance. Care must be taken to keep the device sizes smaller, reduce the layout area, and do creative layout floor planning in order to reduce unwanted parasitic.

### 5.1.3 Output ramp-up time and recovery time

The pump can be modeled as a voltage source with an internal impedance of  $R$ , which in turn limits its output current drivability. The output ramp-up time and voltage recovery time are two important benchmarks of the charge pump performance. Most applications require the output to ramp up fast when a particular operation is initiated—an operation that requires the high-voltage output of the pump. As in an EEPROM, the pump voltage is required when a program operation or an erase operation is chosen. Hence, the pump needs to ramp up the output fast enough to complete the required operation in the shortest possible time in order to meet program or erase speed requirements.

The output voltage recovery time is another important measure of the pump's capability. The pump's output current driving capability is inversely proportional to the regulated output voltage. When the pump output is maintaining a steady output, high-voltage state, sudden output charge sharing due to rapid switching activity or spike of load current will cause a droop in the output voltage, which needs to be replenished within the given "recovery time." Designing the pump to operate at 70%–80% of its maximum capacity will generally guarantee this process.

For a pump with " $n$ " number of stages, the output ramp-up time or recovery time can be controlled by the size of the capacitors and the frequency of operation. A larger capacitor will allow more charge to be transferred in each clock cycle and thus cause a faster increase in output voltage. Higher frequency will expedite this process if the pump is capable. With today's process, in which it is possible to push the ring oscillator frequency to the gigahertz range, the frequency of charge operations is not a speed-critical factor. As mentioned earlier, the maximum frequency of charge operation is determined by the pump architecture, the circuit design and the layout-related parasitic. The range of pump frequency is determined by the application and design requirements. Optimization is needed through design iterations. Further, the clock frequency and the pump performance should be verified with parasitics extracted from the layout (i.e., with back annotated simulation results).

### 5.1.4 Power consumption

With the rising demand in handheld devices and a requirement for low-power operation, power consumption (or power efficiency) has been one of the significant factors in charge pump design. Traditionally, charge pumps have been the culprit in the chips that incorporate them because charge pumps exhibit lower efficiency at higher output voltage, which makes them take a significant bite in the overall chip power consumption. In reality, the efficiency for charge pump (generally the ones generating very high voltages) could vary between 5%–50% depending on the particular designs.

Even though increasing the capacitor size or the operation frequency can directly increase the pump performance, the power consumption will create the most significant drawback. As the capacitance ( $C$ ) is increased, the parasitic capacitance ( $C_s$ ) also increases, which will effectively limit the highest frequency of operation and also cause a large  $CV_{DD}^2 f_{osc}$  dissipation. Also, driving this large capacitor requires large drivers, which effectively increase the operating current. There are many design tricks for reducing the  $CV_{DD}^2 f_{osc}$ -related power consumption, such as a multistep frequency control, where the pump is operated at a higher frequency, initially to ramp up to the required level,

and then switched over to a lower frequency to maintain that voltage level under regulation. A more sophisticated method may be to use a VCO-type of oscillator. The oscillator frequency varies as a function of pump output voltage.

A final factor is the contribution of the voltage regulator in power consumption. It is common that resistive voltage dividers are used to sense the output voltage and regulate the charge pump output. To have a fast regulation response time, a lower resistor value may be chosen, which in turn has larger DC regulation current. Therefore, the pump has to deliver this extra current in addition to the load current. In this case, the charge pump will consume more DC power.

## 5.2 Pump Clock Source

The clock source for a charge pump can be derived from the system clock source, if available, or a new clock source may be built to support the pump activities. A simple clock source may be built using a ring oscillator, as shown in Figure 5-2.

It consists of an odd number of inverting gates connected in a ring pattern. The NAND gate has been used to enable the oscillation. The oscillation frequency can be expressed as

$$f = \frac{1}{N(t_{phl} + t_{plh})} \quad (5-6)$$

where  $t_{phl}$  is the signal propagation delay for an input rising edge and  $t_{plh}$  is the signal propagation delay for an input falling edge.  $N$  is the number of stages of the oscillator. For a properly sized gate,  $t_{phl} = t_{plh}$ . Hence, the frequency of oscillation can be expressed as

$$f = \frac{1}{2NT_D} \quad (5-7)$$

where  $T_D$  is the propagation delay, which equals  $t_{phl}$ , which in turn equals  $t_{plh}$ .

Because the ring oscillator needs a higher number of stages to produce a lower frequency, specifically with faster semiconductor processes, a standard practice is to allow the oscillator to operate at a high frequency

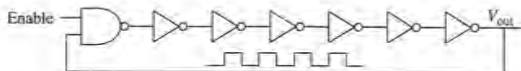


Figure 5-2 A seven-stage ring oscillator.

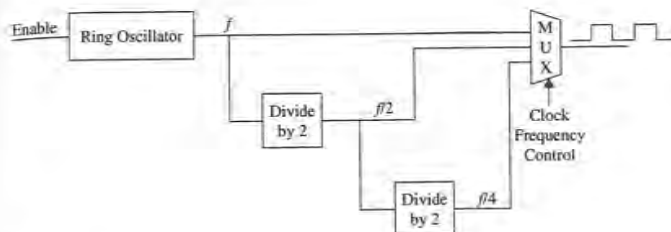


Figure 5-3 Variable frequency clock generator.

and then use a series of dividers to step the frequency down to a required level, as shown in Figure 5-3.

Even though the simple ring oscillator can do the job, it has huge variation—over the process, temperature, and supply voltage variations. The oscillation frequency may vary as much as 30% over the typical operating condition. Keeping the oscillation frequency constant is important because for applications requiring precise pump output voltage, a changing oscillation frequency will cause a change in output voltage noise, power consumption, and output recovery time, etc. To get around this problem, a more-sophisticated, current-controlled ring oscillator may be used.

As shown in Figure 5-4, the current-controlled ring oscillator uses the same inverting gates connected in a ring pattern, with the exception that the current through these inverting gates is controlled by a reference voltage source. This voltage source may actually be derived from a band gap reference voltage generator to produce extra stability across

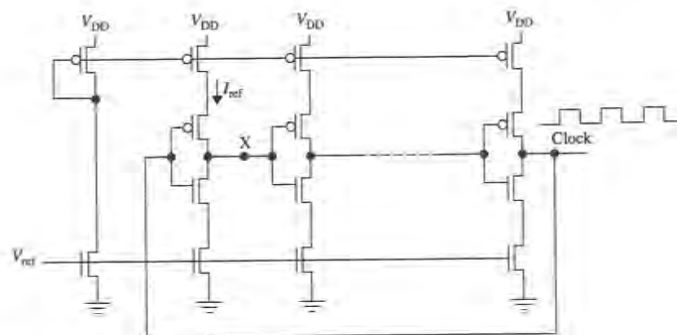


Figure 5-4 Current starved ring oscillator.

temperature and process variations. Assuming the capacitance on node X in Figure 5-4 is  $C_{stage}$ , the frequency of operation can be derived as follows:

$$f = \frac{I_{ref}}{NC_{stage} V_{DD}} \quad (5-8)$$

Another significant advantage of this oscillator is that compared to the simple ring oscillator, this oscillator requires a fewer number of stages to attain the same oscillation frequency, and the oscillation frequency can also be controlled linearly by varying the  $V_{ref}$ .

### 5.3 Regulator Design

Even though some simple low-voltage charge pumps generating voltages in the range of  $2V_{DD}$  or  $V_{DD} + V_t$  may not need a regulator, most of the high-voltage charge pumps almost always need one. Two major types of regulator are generally used: the resistor divider-type regulator, and the capacitor divider-type regulator. Details about these types of regulators were provided in the previous chapter, so we will not discuss them here. Based on the application, the circuit designer should choose the type of regulator that is most suitable for the design.<sup>7-9</sup>

### 5.4 Non-overlapping Clock Generator

In general, the pumps will need a non-overlapping clock for optimum functioning. For a 2-phase Dickson charge pump, the two clock phases may be generated by using one clock source and its inversion as another clock source, as shown in Figure 5-5.

With this type of clock source, consider what happens during time  $t_2$  for a Dickson charge pump in which  $\phi_b$  is connected to stage  $n - 1$  and  $\phi$  is connected to stage  $n$ . During the time  $t_1 - t_2$ , stage  $n - 1$  is charging up stage  $n$ , and if the clock time period is relatively small, the positive going edge of  $\phi$  will cause a quick cutoff of the charge transfer from stage

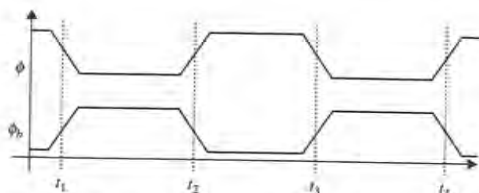


Figure 5-5 A non-overlapping clock.

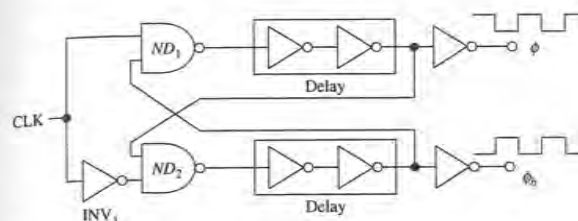


Figure 5-6 A non-overlapping clock generator.

$n - 1$  to stage  $n$ , thus resulting in incomplete charge transfer and, hence, reduced pump performance. To avoid this scenario, a non-overlapping clock is better suited for a Dickson charge pump. A non-overlapping clock can be generated from the original clock source by a simple circuit (see Figure 5-6 and Figure 5-7).

This circuit takes a clock signal and generates a 2-phase non-overlapping clock. The operation is based on the fact that the falling edge of the input clock passes immediately through the NAND gate  $ND_1$ , while the rising edge has first to propagate through the other NAND gate and the cascaded delay element. The resulting signals,  $\phi$  and  $\phi_b$ , have a non-overlapping time equal to the sum of the delays at the NAND gate of the delay element. The delay element is generally made using an even number of inverters. When driving long clock lines, additional buffer stages need to be used to maintain sharp output clock rise and fall times.

Even though the non-overlapping clock generator shown in Figure 5-6 is used extensively in many circuit implementations, it has a small imperfection—the one phase of the clock passes through an extra inverter,  $INV_1$ , and the other phase does not. This problem is generally manifested as a nonsymmetric pulse width of the complementary clocks when the pump is operated at a very high frequency. Even though this is not a problem for a normal Dickson charge pump, it may be a problem for

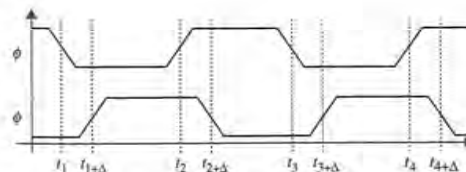


Figure 5-7 A non-overlapping clock generated from the non-overlapping clock generator.

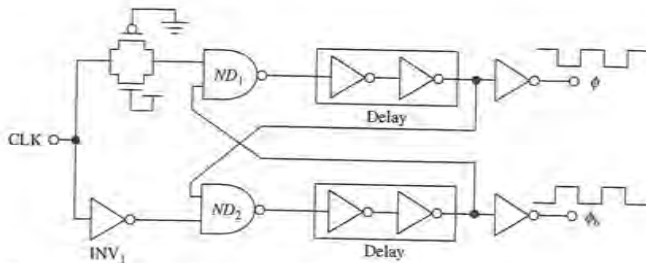


Figure 5-8 A perfected non-overlapping clock generator.

more sophisticated pump designs, such as a 4-phase charge pump, which will be discussed in later chapters. One quick way to fix this issue is to use a passive matching delay element, as shown in Figure 5-8. In this case, the PMOS-NMOS combination pass gate has the same geometry as the PMOS-NMOS MOSFETs in the  $INV_1$  gate to produce symmetric clock output signals.

The initial amount of non-overlap can be kept between 1 ns to 5 ns and later tweaked to provide better results after RC back-annotated simulations are performed. Because the non-overlap time is generally derived from gate delays, this non-overlap will vary with different process, voltage, and temperature conditions and therefore different corner simulations need to be done before optimizing the non-overlap.

### 5.5 Cross-coupled Voltage Doubler Design

During the course of the design of the charge pumps, clocks must be created with amplitudes higher than the power supply voltage of  $V_{DD}$ , such as  $2 \times V_{DD}$ ,  $3 \times V_{DD}$ , and  $4 \times V_{DD}$ . As explained later, using clocks of higher amplitude allows for fewer pump stages to be in serial and thus reduces the internal impedance of the charge pump. This approach makes the pump design much more efficient when the power supplies are scaled down. It also allows the charge pump to be capable of high frequency operation when the dynamic threshold voltage of the diodes is very high at fast clock rate.<sup>3,10-12</sup>

A simple cross-coupled voltage doubler design is shown in Figure 5-9.  $V_o$  and  $V_{ob}$  are the non-overlapping clock inputs, while  $Out$  and  $Out_b$  are the  $2 \times V_{DD}$  amplitude clock outputs. To understand the functionality, let us look at what happens when  $V_o$  goes high. First, it causes MOSFET  $p_{1b}$  to shut off. Then,  $N_{1b}$  turns on, pulling down the output to ground, while MOSFET  $N_{2b}$  turns on, which precharges node  $X_{1b}$  close to  $V_{DD} - V_t$ . In the next clock phase, when clock  $V_o$  goes low, MOSFET  $N_{1b}$  and

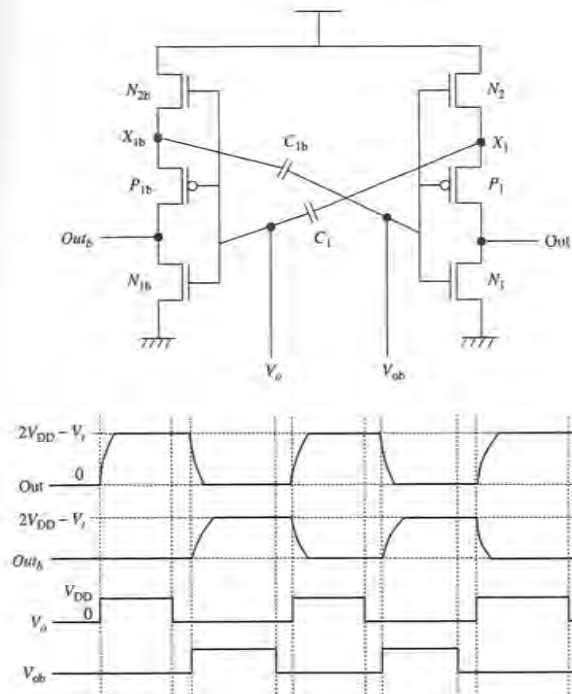


Figure 5-9 A cross-coupled voltage doubler.

$N_{2b}$  turn off and MOSFET  $p_{1b}$  turns on. Next, the clock  $V_{ob}$  goes high, pushing node  $X_{1b}$  to  $2V_{DD} - V_t$ , which is passed through  $p_{1b}$  to the output. A symmetrical operation happens on the right side, resulting in two non-overlapping clocks of higher amplitude. Note that the  $-V_t$  term will always remain, which results in a small inefficiency. Using MOSFETs of lower threshold voltage for  $N_2$  and  $N_{2b}$  will result in improved performance.

If clock amplitudes of higher potential than  $2V_{DD}$  are required, the cross-coupled voltage doubler shown in Figure 5-9 may be cascaded together to produce a clock of amplitude  $3V_{DD} - 2V_t$ , and three such stages may be cascaded together to produce a clock of amplitude  $4V_{DD} - 3V_t$ . Figure 5-10 shows the schematic of two stages cascaded together to produce a  $3V_{DD} - 2V_t$  clock. Even though this process seems simple, the capacitances at nodes  $X_1$  and  $X_{1b}$  need to be kept at a minimum, because

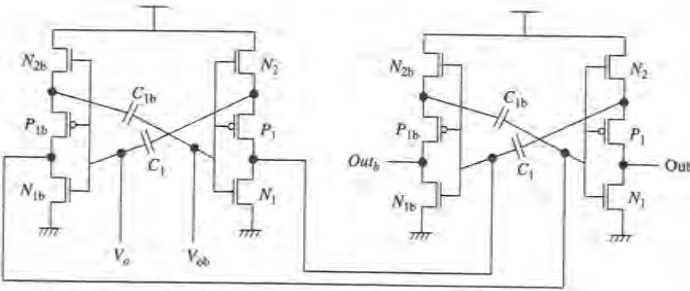


Figure 5-10 A cascaded clock amplitude multiplier.

when the switches from low to high and couples up node  $X_1$  close to the  $2V_{DD}$  level, the charge is actually being shared between capacitor  $C_1$  and the parasitic capacitances. As two stages are cascaded together to produce higher voltages, this charge-sharing problem will really create inefficiency, thus reducing the magnitude of the actual output voltage. Further, NMOS transistors at higher potential will exhibit higher threshold voltage due to high source bias, adding to the circuit's inefficiency. Proper care should be taken to select the actual clock amplitude and the circuit layout.

### 5.6 Logical Effort for Clock Buffer Sizing

Everyone who has gone through a chip design phase has faced the same problem: How to size a signal propagation path through different gates to drive a heavy load or to drive a signal over a long signal line? In other words, how to properly size a series of logic gates to produce the least propagation delay? The *logical effort method*<sup>13</sup> allows us to find optimum solutions for these issues. The logical effort method has the following merits:

- It uses a simple model of delay.
- It allows back-of-the-envelope calculations.
- It helps make rapid comparisons between alternatives.
- It emphasizes remarkable symmetries.

In this chapter, we will quickly detail a simple logical effort methodology and find ways to use logical effort to help make better decisions for clock buffer sizing. You will see that the best performance can be achieved by using optimum fanout for each stage in the logic chain.

You will also see that even though the widely published optimum fanout is "e" ( $\approx 2.71$ ), the real optimum for CMOS circuits ranges from 3.0 to 4.5, depending on the type of gate.

Logical effort expresses delays in a process-independent unit:

$$D_g = D_{abs} / \tau \quad (5-9)$$

Here,  $D_{abs}$  is the actual delay of the gate in units of seconds,  $\tau$  is the process multiplication factor (in seconds), which represents the delay of a minimum inverter driving another minimum inverter in some process.  $\tau \approx 50$  ps for a  $0.8\text{-}\mu\text{m}$  process, and  $\tau \approx 12$  ps for a  $0.18\text{-}\mu\text{m}$  process.  $D_g$  is the absolute process-independent delay.

Next, the gate delay can be expressed as

$$D_g = FG + p_{gate} \quad (5-10)$$

where  $G$  is the logical effort. In general  $G = 1$  for an inverter, by definition. For a 2-input NAND gate  $G = 4/3$  and for a 2-input NOR gate  $G = 5/3$ .

Next,  $F$  is the electrical effort  $= C_{out}/C_{in}$ . The term  $F$  can also be expressed as the ratio of output to input capacitance, which is also termed *fanout*.

Finally,  $p_{gate}$  is the parasitic delay of the gate driving no load.  $p_{gate}$  is set by internal parasitic capacitance.  $p_{gate} = 1$  for an inverter. For a 2-input NAND gate  $p_{gate} = 2$  and for a 2-input NOR gate  $p_{gate} = 2$ .

Logical effort,  $G$ , is defined as the ratio of the input capacitance of a gate to the input capacitance of an inverter delivering the same output current. Figure 5-11 shows the logical effort calculation for the three most common gates. Note that the product of electrical effort and logical effort,  $FG$ , is known as the *stage effort*.

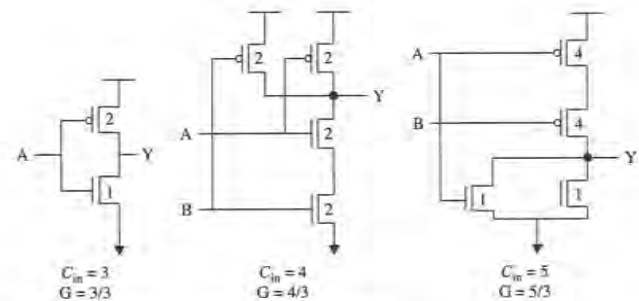


Figure 5-11 Logical effort of common logic gates.

The inverter has a logical effort of 1, by definition. Logical effort for any gate can be determined in the following ways:

- It can be measured from delay versus fanout plots.
- It can be estimated by comparing the gate input area (i.e., by finding the sum of widths, provided the channel lengths are equal) of all the MOSFETs connected to the input and dividing this by the sum of widths of the PMOS and NMOS transistor of an inverter. Figure 5-11 shows the calculation of logical effort by using the gate input area comparison method to calculate the gate delay of an inverter in the following circuit configuration.

Next, let us apply the logical effort concept to find the gate delay of an inverter driving a few gates. By definition, the gate delay will be a function of the fanout, its logical effort, and the intrinsic parasitic capacitance of the gate. Figure 5-12 shows a typical structure.

Using Equation 5-10,  $D_g = FG + p_{\text{gate}}$ , we have the following:

- Logical effort for inverter:  $G = 1$
- Electrical effort:  $F = C_{\text{out}}/C_{\text{in}} = 4$   
= input capacitance of four inverters/input capacitance of one inverter.
- Parasitic delay of an inverter:  $p_{\text{inv}} = 1$

Therefore, gate delay,  $D_g$ , is  $1 \times 4 + 1 = 5$ .

Multiplying  $D_g$  with the process-dependent multiplier,  $\tau$ , will produce the actual gate delay. The process multiplication factor can be derived easily by plotting a delay-vs.-fanout plot for any gate. Figure 5-13 shows such a plot for an inverter with an approximate trendline and the values of the y-intercept and slope automatically calculated by Excel.

Using  $D_{\text{abs}} = \tau(fg + p_{\text{inv}})$  and substituting the values of  $g$ ,  $p_{\text{inv}}$ , and  $f$ , we can calculate the value of  $\tau$  by looking up the values of Figure 5-13, as shown in Table 5-2.

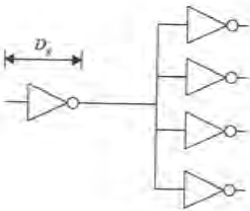


Figure 5-12 A typical circuit setup for calculating gate delay.

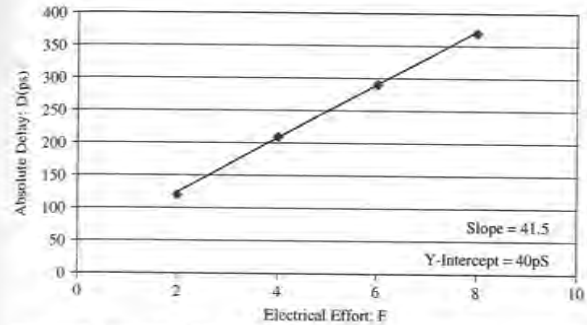


Figure 5-13 Simulated delays of an inverter driving loads with different fanouts for a particular process.

Note that for the inverter,  $g = 1$  and  $p_{\text{inv}} = 1$  by definition. As can be seen from Table 5-2, the average value of  $\tau$  is 41 ps.

For designing large circuits that need several stages of gates or buffers, the total signal path delay can be expressed as

$$D_{\text{tot}} = \sum D_g = \sum fg + \sum p \quad (5-11)$$

Next, it can be proved that the delay is smallest when each stage bears the same stage effort:

$$f_1 g_1 = f_2 g_2 = f_3 g_3 = \dots \quad (5-12)$$

Applying this concept on a series of "n" inverters, the following can be shown:

$$F^n = \frac{C_{\text{load}}}{C_{\text{stage1}}} \quad (5-13)$$

TABLE 5-2 Process Multiplication Factor  $\tau$  Calculation

| $f$ | $g$ | $p_{\text{inv}}$ | $D_{\text{abs}}$ | $\tau = \frac{D_{\text{abs}}}{(f \times g + p_{\text{inv}})}$ |
|-----|-----|------------------|------------------|---------------------------------------------------------------|
| 2   | 1   | 1                | 120              | 40                                                            |
| 4   | 1   | 1                | 210              | 42                                                            |
| 6   | 1   | 1                | 290              | 41.4                                                          |
| 8   | 1   | 1                | 370              | 41.1                                                          |

Hence, taking log on both sides, we have the following:

$$n = \frac{\ln \left[ \frac{C_{\text{load}}}{C_{\text{stage1}}} \right]}{\ln[F]} \quad (5-14)$$

Because each stage bears the same stage effort and assuming all gates are of a similar type (so we can have the same  $p_{\text{gate}}$ ), the total path delay can also be expressed as follows:

$$D_{\text{tot}} = \sum_1^n Dg = \sum_1^n fg + \sum_1^n p = nDg \quad (5-15)$$

Substituting  $n$  from Equation 5-14, we have

$$D_{\text{tot}} = M \frac{Dg}{\ln[F]}$$

$$\text{where } M = \ln \left[ \frac{C_{\text{load}}}{C_{\text{stage1}}} \right]$$

Because  $D_g = FG + p$ , substituting  $D_g$  in the preceding equation we get

$$D_{\text{tot}} = M \frac{(FG + p)}{\ln[F]} \quad (5-16)$$

Next, to find the minimum delay for this path, we should differentiate the preceding equation with respect to  $F$  and equate it to 0.

Using the quotient rule of differentiation, we can write

$$\frac{dD_{\text{tot}}}{dF} = M \left[ \frac{G \ln F - (FG + p) \left( \frac{1}{F} \right)}{(\ln F)^2} \right] = 0$$

$$G \ln F - G = \frac{p}{F} \quad (5-17)$$

$$F [\ln F - 1] = \frac{p}{G}$$

Because for an inverter  $G = 1$ , the preceding equation can be written as

$$F [\ln F - 1] = p \quad (5-18)$$

Neglecting parasitics ( $p = 0$ ), we find the familiar result  $F = 2.718$  (e). For  $p = 1$ , we can solve the equation numerically to get  $F = 3.5$ . Similarly, for a two-input NAND gate and two-input NOR gate, it can be shown that the minimum delay can be achieved by  $F = 3.7$ . Next, we can do a little sensitivity analysis to see how the delay varies if  $F$  is not minimum.

Using

$$D_{\text{tot}} = M \frac{Dg}{\ln[F]} \quad (5-19)$$

and representing the minimum delay as  $\hat{D}_{\text{tot}} = M \frac{D_{\text{min}}}{\ln[F_{\text{min}}]}$ , we can plot

$$\frac{D_{\text{tot}}}{\hat{D}_{\text{tot}}} = \frac{\frac{Dg}{\ln[F]}}{\frac{D_{\text{min}}}{\ln[F_{\text{min}}]}} \quad (5-20)$$

with respect to  $F$ , as shown in Figure 5-14.

Through further analysis, we can see that for different types of standard logic gates, the minimum fanout,  $F_{\text{min}}$ , will vary between  $2.5 < F < 4.5$ . Next, from Figure 5-14, you can see that keeping the fanout per stage  $2.4 < F < 6$  gives a delay within 15% of the optimal delay. In practice, keeping a uniform  $F = 4$  per-stage number allows for

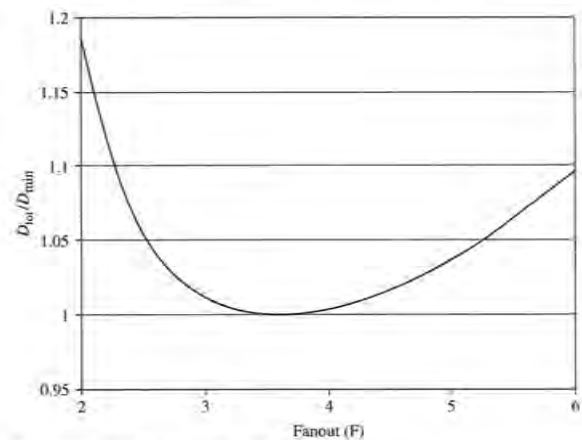


Figure 5-14 Delay sensitivity with fanout variation.

easy calculation and optimum results. An easy way to use the results from logical effort in practice is to first determine the output load capacitance (in terms of gate loading) and then work back by following  $F = 4$  fanout backward through each stage.

Take the following example in Figure 5-15 to determine the best fanout/stage and the number of stages for a clock driver whose final equivalent loading is represented with an inverter of size 960  $\mu$  (PMOS) and 480  $\mu$  (NMOS). Also, for simplicity, let's imagine the initial driver size is 15  $\mu$  (PMOS)/7.5  $\mu$  (NMOS).<sup>14,15</sup>

From the preceding example, we can easily infer the following:

- Signal propagation paths are fastest when fanout is close to 4.
- Path delay is weakly sensitive to the number of stages and gate sizes.
- Using fewer stages does not mean faster propagation.

Keep in mind that logical effort uses a simplistic delay model and that it neglects input rise/fall time effects. In the preceding calculations, interconnect parasitic capacitance loadings were not accounted for. In practice, this will create erroneous results, especially for long interconnecting wires. An easy way to account for parasitic capacitances is to estimate the loading parasitic capacitance and convert it into an equivalent gate capacitance. This should be a one-step process as soon as the parasitic capacitance-to-gate capacitance ratio is determined.

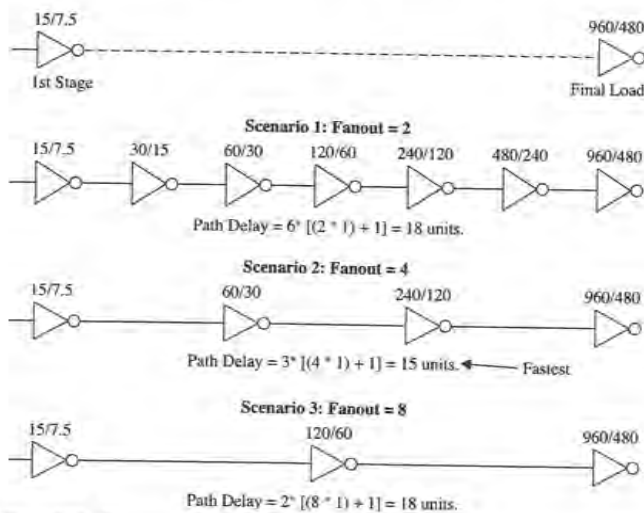


Figure 5-15 Fanout comparison.

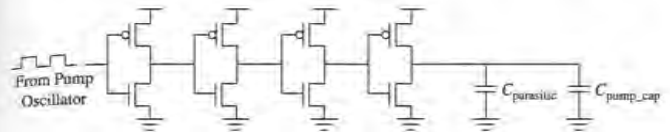


Figure 5-16 A clock signal driving buffer.

Next, connect this equivalent gate capacitance to the output of the driver that is driving the long line (i.e., this driver will see the equivalent gate capacitance as an additional fanout element). This process allows for more accurate estimation of fanout and the total path delay.

A simple clock design can be a series of inverters, sized according to the FO4 ratio, driving the capacitors. Remember that because in most cases the final capacitive load is heavy, an FO4 ratio might not be maintained. Instead, the load may be 10–100 larger than the preceding stage. Nevertheless, a fast ramp up of node  $V_0$   $V_\phi$  is preferred, and extra boost techniques may be applied to hasten it (more about this later).

Even though the simple clock buffer shown in Figure 5-16 is sufficient in many applications, it has an inherent shortcoming for charge pump applications. The clock buffer output voltage,  $V_0$ , is dependent on the supply voltage,  $V_{DD}$ . This means that when the supply voltage variation is on the higher side, the clock amplitude will be higher and the charge pump output ramping will be faster, and vice versa. In some applications, this may create a problem. To solve this problem, a  $V_{DD}$ -independent clock source should be created. A quick-and-easy way to create such a clock source is shown in Figure 5-17.

MOSFET  $M_1$  is generally a low  $V_t$  transistor. When a constant reference voltage,  $V_{ref}$ , is applied at the gate of NMOS  $M_1$ , the source voltage at node  $C - V_{DD}$  will always be at the level  $V_{ref} + V_t$ , regardless of  $V_{DD}$  variation, provided  $V_{DD}$  is higher than  $V_{ref} + V_t$ .  $V_{ref}$  voltage is usually derived from a bandgap reference voltage generator. It must be noted that regulating supply voltages through an NMOS-type transistor has some drawbacks. Referring to Figure 2-20 in Chapter 2, consider the

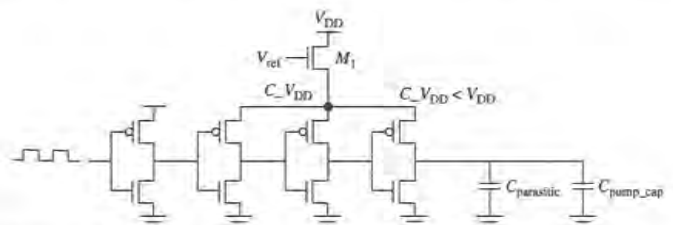


Figure 5-17 A clock signal driving buffer with constant amplitude.

normal operating point of the present NMOS  $M_1$  corresponding to  $V_{gs4}$ . When there is a huge current demand from the  $C - V_{DD}$  node, the voltage at node  $C - V_{DD}$  will invariably dip down significantly. This happens because as  $M_1$ 's  $I_{ds}$  increases,  $V_{gs}$  has to increase. However, because the gate voltage is tied to a constant  $V_{ref}$ ,  $C - V_{DD}$  has to dip down to accommodate the high current demand. In this current scenario, to get around this problem, ample decoupling capacitors should be used on the node  $C - V_{DD}$  to filter out voltage dips.

### 5.7 Parasitic R and C

The parasitic capacitance in the individual Dickson charge pump stages consumes charge during the transferring. It is one of the major factors that can increase the internal impedance of charge pump. Recall the following from a previous equation:

$$V'_\phi = \left( \frac{C}{C + C_s} \right) V_\phi \quad (5-21)$$

Here,  $V'_\phi$  is the voltage induced at each stage, which is a function of the parasitic capacitance  $C_s$ . The larger the parasitic  $C_s$ , the lower the voltage that could be coupled during each clock cycle.

Figures 5-18 and 5-19 show the primary location of the parasitic capacitance  $C_s$ . As capacitor and MOSFET sizes get bigger, they need to be

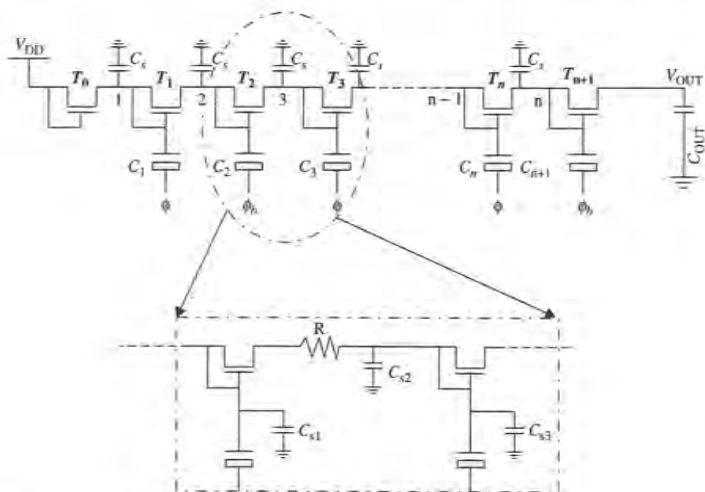


Figure 5-18 Interconnect parasitics.

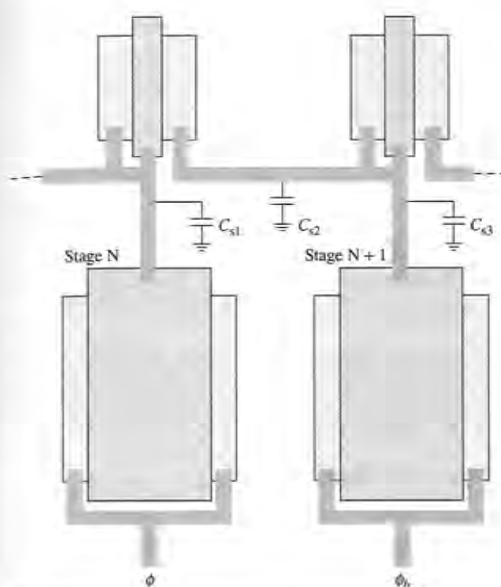


Figure 5-19 Interconnect parasitics, layout view.

split apart into smaller parallel sections. Partitioning large devices into smaller components will increase the total area and the parasitic capacitance associated with each internal node. Proper effort must be made to create an optimum layout floor plan and interconnect metal lengths and widths to make sure the parasitics are kept to a minimum.<sup>16-18</sup>

As shown in Figure 5-20, because of the inherently large size of the capacitors, the total layout area tends to be large. If global clock drivers are used to drive the clocks near X, the clocks will have a large rise/fall times at the end, Y. This will significantly reduce the maximum frequency of operation. In addition, the clock drivers have to be sized up to achieve tolerable clock slopes. This trend tends to consume more power.

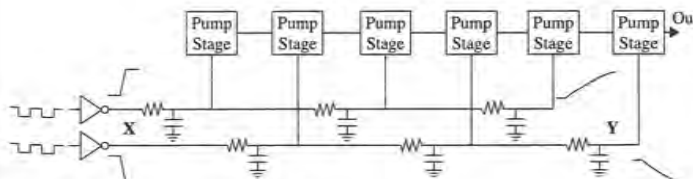


Figure 5-20 Parasitics in clock signal line.

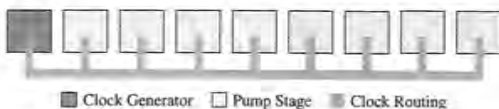


Figure 5-21 Linear layout floor plan.

The clock distribution scheme is also important in reducing the clock line parasitics, which allows faster operation and efficient driver sizing. A linear layout is shown in Figure 5-21. This scheme, though simple, has a lot of clock line parasitics associated with it. Further, the clock rise/fall time at the last block will invariably be longer than those at the initial blocks.

A better way is to use the floral layout pattern shown in Figure 5-22 to reduce clock-signal line-parasitic capacitance. The main idea is to design a floor plan in such a way so as to reduce the parasitic on each signal line.

### 5.8 Power Bus and Bower Bus Capacitance

The charge pump's inherent inefficiency, coupled with the large amount of required switching current for the drivers, is itself the power hog on the chip. On many occasions the chip exhibits poor performance because of the huge current demand from the charge pump on the internal supply bus  $V_{DD}$ . Large  $IR$  drop on  $V_{DD}$  bus could make the actual power supply near the pump deviate significantly from worst case scenario.

In general, it is a better idea to place the charge pump closer to the supply pin pad, or to dedicate a special power supply pad just for the charge pump. The width of power bus from pad to the pump blocks has to be wide enough to address the worst case  $IR$  drop. Appropriate layout area must be allocated for supply decoupling capacitors in order to smooth out the peak current from the clock buffer switching.

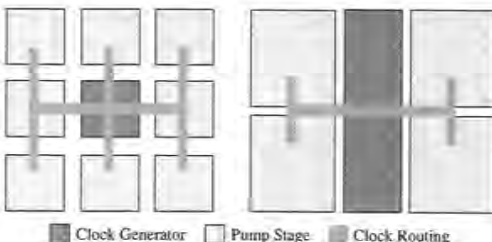


Figure 5-22 Patterned layout floor plan.

### 5.9 Conclusion

This chapter introduced the various charge pump specifications that circuit designers need to determine as an initial step before designing the charge pump. It introduced nine main design specifications and analyzed each of the major parameters in detail and discussed how to create trade-offs between each one of them. Through these steps readers can determine the initial values of the number of pump stages, the pump capacitor size, the pump clock frequency and various other parameters. This chapter also showed how to design the various supporting charge pump circuit blocks. Starting with a balanced non-overlapping clock source generator, the chapter introduced clock amplitude doubler and tripler circuits. Next, the concept of Logical Effort for clock buffer sizing was introduced, along with a detailed quantitative analysis, as well as the impact of parasitic resistors and capacitances on circuit performance. This chapter concluded with a discussion on layout implementation, floor planning and clock distribution—all of which aid in the design of an efficient charge pump.

The following chapter will take your understanding of the basic operation of the 2-phase charge pump to the next level and illustrate how to design a better charge pump.

### References

1. Witters, J.S., G. Groeseneken, H.E. Maes. "Analysis and modeling of on-chip high-voltage generator circuits for use in EEPROM circuits." *IEEE Journal of Solid-State Circuits*, Vol. 24, No. 5, pp. 1372-1380, October 1989.
2. Pelliconi, R., et al. "Power Efficient Charge Pump in Deep Submicron Standard CMOS Technology." *Solid-State Circuits Conference*, 2001, ESSCIRC 2001. Proceedings of the 27th European.
3. Favrat, P., P.H. Deval, M. Declercq. "A New High-Efficiency CMOS Voltage Doubler." *IEEE Custom Integrated Circuits Conference (CICC'97)*, pp. 259-262, May 5-8.
4. Choi, Ki-Hwan et al. "Floating-well Charge Pump Circuits For Sub-2.0 V Single Power Supply Flash Memories." *Digest of Technical Papers. Symposium on VLSI Circuits*, pp. 61-62, June 12-14, 1997.
5. Naso, et al. "Negative-voltage charge pump with feedback control." U.S. Patent 5,168,174.
6. Papaix, Caroline and Jean-Michel Daga. "High Voltage Generation for Low Power Large VDD Range Non Volatile Memories." *PATMOS 2001*. <http://patmos2001.eivd.ch>.
7. Young et al. "Power supply insensitive substrate bias voltage detector circuit." U.S. patent 6,172,554.
8. Sung, Ha Min. "Substrate voltage detection control circuit." U.S. patent 6,281,742.
9. Lee et al. "Circuit for sensing back-bias level in a semiconductor memory device." U.S. patent 5,262,989.
10. Brugler, J.S. "Theoretical Performance of Voltage Multiplier Circuits." *IEEE Journal of Solid-State Circuits*, pp. 132-135, June 1971.
11. Favrat, Pierre, Philippe Deval, and Michel J. Declercq. "A High-Efficiency CMOS Voltage Doubler." *IEEE Journal of Solid-State Circuits*, Vol. 33, No. 3, pp. 410-416, March 1998.

12. Tanzawa, T. and A. Shigeru. "Optimization of Word-Line Booster Circuits for Low-Voltage Flash Memories." *IEEE Journal of Solid-State Circuits*, Vol. 34, No. 8, pp. 1091–1095, AUGUST 1999
13. Sutherland, I., F.R. Sproull, and D. Harris. *Logical Effort: Designing Fast CMOS Circuits*. Morgan Kaufmann, 1999.
14. Bateman, Bruce. "High-Speed SRAM Design." ISSCC'98 Tutorial.
15. Horowitz, M. "VLSI Scaling for Architects." Presentation slides. Computer Systems Laboratory, Stanford University.
16. Chern, J. et al. "Multilevel metal capacitance models for CAD Design Synthesis Systems." *IEEE Electron Device Letters*, Vol. 13, No. 1, pp. 32–34, January 1992.
17. Barke, Erich. "Line-to-Ground Capacitance Calculation for VLSI: A Comparison." *IEEE Transactions on Computer-Aided Design*, Vol. 7, No. 2, pp. 295–298. February 1988.
18. Wong, Shyh-Chyi, Gwo-Yann Lee, and Dye-Jyun Ma. "Modeling of Interconnect Capacitance, Delay, and Crosstalk in VLSI." *IEEE Transactions on Semiconductor Manufacturing*, Vol. 13, No. 1, pp. 108–111, February 2000.

# The First-Ever Guide to Designing and Implementing Charge Pumps for Today's Low-Cost, High-Performance Mobile Devices

A groundbreaking tool for circuit design engineers, **Charge Pump Circuit Design** is the first book to focus solely on the design and implementation of charge pumps used in EEPROMs, Flash memory, White LED drivers, and a myriad of other circuits finding mass applications in PDAs, digital cameras, MP3 players, video recorders, cell phones, USB drives, and more.

Written by two of today's leading circuit designers, **Charge Pump Circuit Design** explores the basic operations, design criteria, and newest approaches for designing state-of-the-art charge pumps. The authors explain the different architectures and requirements, providing comprehensive information for each stage in the design process. Filled with 100 detailed illustrations, this time-saving reference also presents a wealth of practical design tips and potential pitfalls to avoid.

## **Charge Pump Circuit Design** features:

- The latest design techniques for creating highly efficient charge pumps for any type of application requirement
- Step-by-step guidelines for completing a charge pump design—from initial concept to implementation of actual layout
- Thorough mathematical derivations and analyses of operations that are applicable to all charge pump requirements, regardless of the system being designed

## **Inside This Landmark Design Reference:**

- History of High-Voltage Charge Pumps
- Basic Operations of Charge Pumps
- Criteria of a Generic Charge Pump
- How to Design a Basic Charge Pump
- How to Design a Better Charge Pump
- Charge Pump Architectures
- Future Design Reference

The **McGraw-Hill** Companies

[books.mcgraw-hill.com](http://books.mcgraw-hill.com)

Electronic Engineering

ISBN 978-0-07-147045-2  
ISBN 0-07-147045-X



9 780071 470452