High-Voltage Generation and Drive in Low-Voltage CMOS Technology

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy in Electrical Engineering

By

Yousr Ismail

2015
ABSTRACT OF THE DISSERTATION

High-Voltage Generation and Drive in Low-Voltage CMOS Technology

by

Yousr Ismail

Doctor of Philosophy in Electrical Engineering

University of California, Los Angeles, 2015

Professor Chih-Kong Ken Yang, Co-Chair

Professor Mau-Chung Frank Chang, Co-Chair

High-voltage dc and switching waveforms are needed in many of today’s electronic systems. Various MEMS applications require output voltage signals that are several 10's of volts. Advanced CMOS technology nodes allow for smaller, lower-cost electronics, but are not engineered to handle such high voltages directly. High-voltage systems are often implemented in older, voltage-tolerant technology nodes or other specialized processes; driving the overall system size and cost up.
This dissertation introduces technology and circuit methods that extend the voltage range of a standard, fine-linewidth CMOS process beyond its conventional breakdown limit. Examples of high-voltage generation and drive circuits introduced in this dissertation include: 1) voltage charge pumps and 2) output voltage drivers. The introduced circuits are fully compatible with standard low-voltage CMOS process and maintain long-term device reliability.

For high-voltage generation, we introduce a new Hybrid Charge Pump architecture. The hybrid architecture extends the voltage tolerance of a nanometer scale CMOS substrate by ~8x while enabling improved power efficiency. We provide an analytical power model for the Hybrid Charge Pump, and outline a systematic method to optimize its power efficiency. For high-voltage drive, we introduce a Charge Pump-Based output stage suitable for driving high-impedance loads. The output driver enables seamless stacking of 10's of devices with little power and area overhead, enabling output waveforms with extended voltage ranges in a low-voltage CMOS process.

Practical results presented in the dissertation include the measurement results of a 36V 49% efficiency Hybrid Charge Pump in 65nm bulk CMOS technology, and a bipolar 44V Charge Pump-Based driver with a 21KΩ output resistance in 45nm SOI CMOS technology.
The dissertation of Yousr Ismail is approved.

Chang-Jin Kim

Sudhakar Pamarti

Mau-Chung Frank Chang, Committee Co-Chair

Chih-Kong Ken Yang, Committee Co-Chair

University of California, Los Angeles

2015
To my dad who, in my mind, is always standing tall.

To my mom who, in my heart, is always sitting comfortably.
# Table of Contents

ABSTRACT .......................................................................................................................... ii  

Table of Contents ............................................................................................................... vi  

Acknowledgments ............................................................................................................. x  

VITA .................................................................................................................................. xii  

CHAPTER 1  Introduction .................................................................................................. 1  

1.1. Examples of High-Voltage Low-Current Applications .................................................. 2  

1.2. Motivation .................................................................................................................... 3  

1.3. Thesis Organization ..................................................................................................... 6  

CHAPTER 2  Background .................................................................................................. 9  

2.1. Introduction .................................................................................................................. 9  

2.2. Charge Pump DC-DC Converters ................................................................................. 9  

2.3. High-Voltage Design Constraints .............................................................................. 13  

   2.3.1. Switch Limitations ............................................................................................... 13  

   2.3.2. Capacitor Limitations .......................................................................................... 17  

   2.3.3. Substrate Voltage Tolerance ............................................................................... 20  

2.4. Review of High-Voltage Methods .............................................................................. 28  

   2.4.1. High-Voltage Generation .................................................................................... 28  

   2.4.2. High-Voltage Drive .............................................................................................. 29  

2.5. Summary .................................................................................................................... 30
<table>
<thead>
<tr>
<th>CHAPTER 3</th>
<th>Extended Voltage-Range CMOS Substrates</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1.</td>
<td>Introduction</td>
</tr>
<tr>
<td>3.2.</td>
<td>Double-Diode Substrate Isolation</td>
</tr>
<tr>
<td>3.3.</td>
<td>Field-Oxide Substrate Isolation</td>
</tr>
<tr>
<td>3.3.1.</td>
<td>Polysilicon Diode Design</td>
</tr>
<tr>
<td>3.3.2.</td>
<td>Polysilicon Diode-Based Dickson Charge Pumps</td>
</tr>
<tr>
<td>3.4.</td>
<td>Substrate Stacking</td>
</tr>
<tr>
<td>3.5.</td>
<td>Summary</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 4</th>
<th>High-Efficiency Voltage Pumping Cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1.</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2.</td>
<td>Complementary-Type Switch Charge Pumps</td>
</tr>
<tr>
<td>4.2.1.</td>
<td>A CMOS Four-Phase Pump Cell</td>
</tr>
<tr>
<td>4.2.2.</td>
<td>A CMOS Six-Phase Pump Cell</td>
</tr>
<tr>
<td>4.3.</td>
<td>Same-Type Switch Charge Pumps</td>
</tr>
<tr>
<td>4.3.1.</td>
<td>An All-NMOS Four-Phase Pump Cell</td>
</tr>
<tr>
<td>4.3.2.</td>
<td>An All-NMOS Six-Phase Pump Cell</td>
</tr>
<tr>
<td>4.3.3.</td>
<td>An All-NMOS Eight-Phase Pump Cell</td>
</tr>
<tr>
<td>4.4.</td>
<td>Improved-Drive Dickson Charge Pump</td>
</tr>
<tr>
<td>4.5.</td>
<td>Clock Generation Circuit</td>
</tr>
<tr>
<td>4.6.</td>
<td>Summary</td>
</tr>
</tbody>
</table>
CHAPTER 5  Hybrid Charge Pumps............................................................................. 79

5.1. Introduction........................................................................................................... 79

5.2. Hybrid Charge Pump Architecture ...................................................................... 79

5.3. Power Analysis Model.......................................................................................... 81
  5.3.1. Idealized Power Model....................................................................................... 83
  5.3.2. Practical Power Model....................................................................................... 87

5.4. Noise Analysis Model............................................................................................ 119

5.5. Design Examples .................................................................................................. 127
  5.5.1. A +36V Hybrid Charge Pump............................................................................ 127
  5.5.2. A -29V Hybrid Charge Pump............................................................................ 129

5.6. Summary ............................................................................................................... 130

CHAPTER 6  High-Voltage Drive.................................................................................... 132

6.1. Introduction........................................................................................................... 132

6.2. Stacked-Device Drivers ....................................................................................... 132
  6.2.1. A Doubly-Stacked Output Driver........................................................................ 135
  6.2.2. A Triply-Stacked Output Driver......................................................................... 140

6.3. Charge Pump-Based Drivers................................................................................ 143
  6.3.1. Unipolar Output Drive....................................................................................... 146
  6.3.2. Bipolar Output Drive......................................................................................... 151

6.4. Summary ............................................................................................................... 154
CHAPTER 7  Experimental Results ................................................................. 155

7.1. Introduction ........................................................................................................ 155

7.2. Hybrid Charge Pumps ....................................................................................... 155

7.2.1. A +36V Hybrid Charge Pump ...................................................................... 156

7.2.2. A -29V Hybrid Charge Pump ...................................................................... 162

7.3. Charge Pump-Based Drivers ............................................................................... 166

7.3.1. A Unipolar 12V Output Driver ..................................................................... 167

7.3.2. A Bipolar 44V Output Driver ...................................................................... 169

7.4. Summary ............................................................................................................ 173

CHAPTER 8  Conclusion .......................................................................................... 176

References .............................................................................................................. 179
Acknowledgments

I am particularly fortunate to have Professor Ken Yang as my advisor. He fosters a personal relationship with his students that is based on mutual respect, first and foremost. I am grateful to him for providing the environment necessary to grow, develop, and pursue my passion. I appreciate his advice, support, and guidance over the years. His insightful remarks and suggestions proved very helpful at critical moments. Always providing the bigger picture, I have learned a lot from him.

I would like to extend my appreciation to the Department Chair Prof. Frank Chang for his continuous support. He has generously provided silicon real-estate for us to test our ideas, and for that I am grateful.

Special thanks are due to my doctoral committee members: Prof. Sudhakar Pamarti, for his helpful ideas, comments, and manuscript reviews, and Prof. C.-J. Kim for providing the opportunity to explore the high-voltage domain through the ITMARS project. This work has been supported in part by DARPA and Broadcom Corporation.

I am much obliged to the Department's Office of Graduate Student Affairs whose staff members are always there to answer questions, give advice and process paper work promptly. They play a key, behind-the-seen, role in advancing research at UCLA. Deeona Columbia, Mandy Smith, and Kyle Jung, thank you.

Throughout my PhD, I had the chance to do two industry-related internships that provided me with a lot of valuable experience. The first was with the Central Engineering Audio group at
Broadcom, Irvine, for which, I would like to thank Sherif Galal, and Xicheng Jiang generously. The second internship was with the ASIC Design group at SiTime in Sunnyvale. The project that I have been assigned to there was the initial motivation for what latter became an important part of my dissertation. For this, I would like to thank Haechang Lee, and Jim Salvia.

A special thanks goes to a large group of friends here in Los Angeles, whom without, the journey of graduate school would have been much more difficult. Thanks are due to my dear friends, Tamer Ali, Ramy Yousry, Amr Amin, Ahmed Hegab, Ayman Mahfouz, Mohamed Selim, Mahmoud Elgammal, Barzan Mozafari, Maher Elkady, Yavuz Nuri, Eyad Alnaslah, Yasser Shoukry, Omar Alsedeeq, Jean-Michel Madague, Hussam Qasem, Bandar Almangour, Ahmed Hareedy, Amr Alanwar, Mohammed Karmoose, Moustafa Alzantoot, Yahya Ezzeldin, and Ahmed Alaa.

I am heavily indebted to all my family members: my dad, a man like no other, my mom, an epitome of unconditional love, my younger siblings, Sameh, Ragy, and Moataz, and my inspiration Nora.
# VITA

<table>
<thead>
<tr>
<th>Year(s)</th>
<th>Position/Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>B.Sc. in Electrical Engineering, Ain Shams University, Egypt.</td>
</tr>
<tr>
<td>2005-2009</td>
<td>Teaching Assistant, Department of Electronics and Communications Engineering, Ain Shams University, Egypt.</td>
</tr>
<tr>
<td>2006-2009</td>
<td>IC Design Engineer, Si-Ware Systems, Cairo, Egypt.</td>
</tr>
<tr>
<td>2009</td>
<td>M.Sc. in Electrical Engineering, Ain Shams University, Egypt.</td>
</tr>
<tr>
<td>2010-2015</td>
<td>Research and Teaching Assistant, Department of Electrical Engineering, University of California, Los Angeles, California, USA.</td>
</tr>
<tr>
<td>2010</td>
<td>IC Design Intern, Broadcom, Irvine, California, USA.</td>
</tr>
<tr>
<td>2012</td>
<td>IC Design Intern, SiTime, Sunnyvale, California, USA.</td>
</tr>
</tbody>
</table>

# PUBLICATIONS


AWARDS

Broadcom Foundation Fellowship Award (2011-2012).

CHAPTER 1

Introduction

This dissertation introduces technology and circuit methods that enable high-voltage integration into nanometer-scale low-voltage CMOS technology.

High voltages are almost ubiquitous in many of today’s electronic systems. In IC technology, the term high voltage often refers to a wide range of signal amplitudes, and typically spans circuits with voltages that are slightly higher than the technology's nominal supply voltage, all the way to those handling 100's of volts. High-voltage signals are used in many applications including MEMS, automobiles, telecommunications, and biomedical applications. High-voltage implementations differ greatly depending on the current drive requirements. This dissertation focuses on circuit implementations targeting applications with low-current drive such as, MEMS capacitive sensors, piezoelectric transducers and Flash memory programmers. Many such applications require signal amplitudes that are several 10's of volts. Applications requiring large load currents such as motors, audio amplifiers, and LED drivers are not addressed as part of this dissertation. Moreover, the required high-voltage signals can either be dc or switching waveforms. In this dissertation, we discuss circuits for both high-voltage generation and drive.

In this chapter, section 1.1 describes in greater detail the MEMS applications that led to this work. Section 1.2 describes the motivation for using nanometer-scale CMOS technology and the challenges addressed by the work in this dissertation. This chapter concludes with the organization of the dissertation.
1.1. Examples of High-Voltage Low-Current Applications

Generating high dc voltages is critical for a range of applications. Shown in Fig. 1 is a simplified MEMS-based resonator system [1]. The MEMS device sets the system's oscillation frequency and requires a dc bias voltage for operation. Higher MEMS bias voltages result in improved charge sensitivity and lower MEMS electro-motional impedance. One possible way to improve the resonator's signal-to-noise ratio (SNR) is to apply higher bias voltages to the MEMS device [2]. Improved phase noise performance enables MEMS-based timing solutions that better meet the tough jitter requirements of many high-end timing applications, e.g. cellular, GPS and high-speed serial links. Voltages >30V are not uncommon, and a dc-dc converter is typically required to step up the system's low-voltage supply to the higher voltages needed. A similar bias voltage argument holds for many other MEMS device types that need to meet high dynamic range specifications, such as acceleration sensing platforms, and gyroscopes [3]-[4]. Additionally, for MEMS gyroscopes, a programmable bias voltage is used to fine tune the device's resonant frequency. A large voltage range spanning a few volts up to 60V is typical. Other applications requiring a programmable high-voltage bias include tuning antennas and filters that employ ferroelectric BST (Barium-Strontium-Titanate) capacitors.
High-voltage switching waveforms are also needed in other applications. Shown in Fig. 1.2 is an example of an ultrasonic transceiver system. An ultrasonic transducer is made of a piezoelectric material and can be fabricated using MEMS technology. The transducer is used to transmit and receive ultrasonic waves, and can be used for a range of applications including gesture recognition, range finding, and calculating body fat composition [5]-[8]. The transmit pulse amplitude affects the transceiver performance. Higher transmit voltages result in higher SNR at the receiver, enabling operation at longer ranges. Voltages >30V are used in [8]. Another example of devices requiring high-voltage waveforms are electrostatically actuated transducers. For example, integrated MEMS micromotors require large electric fields to overcome the drag forces and operate the rotor [9]-[10]. Voltages as large as 80V may be required.

### 1.2. Motivation

Nanometer-scale CMOS technology nodes allow for a high degree of integration of highly-efficient processing blocks, leading to low-cost, small form-factor, low-power, and high-performance electronics. In the MEMS applications described previously, the integration enables efficient signal processing, feedback control, and finer physical features for the MEMS structures. Typical devices and substrates in advanced technologies, however, suffer from low
voltage tolerances, making it difficult to integrate the high-voltage circuits with the rest of the system. Consequently, in order to integrate low-voltage and high-voltage circuits on the same die, such systems are commonly designed in specialized high-voltage technologies or using a high-voltage option in a standard advanced process. A high-voltage option offers only a limited voltage extension to few devices limiting its use to fewer applications.

A specialized high-voltage (HV) IC technology makes available transistors and capacitors with sufficient voltage tolerances to handle the required signal levels. Examples of high-voltage processes include Smart Power, BCD (Bipolar-CMOS-DMOS), and HV CMOS technologies [11]-[13]. Smart Power and BCD are dedicated technologies with specialized process steps and substrate layers. These steps also accommodate low-voltage devices for higher performance electronics. A HV CMOS technology is one in which a standard CMOS process is tweaked to accommodate high-voltage devices. High-voltage MOS transistors are enabled by engineering the drain region of the device. High-voltage transistors come in many flavors, e.g. DMOS (Double Diffused MOS), LDMOS (Laterally Diffused MOS), and DDDMOS (Double Diffused Drain MOS). A good survey of the different HV device types is presented in [14]-[17]. These devices have larger minimum features compared to their low-voltage counterparts. HV technologies are supported by many foundries including DALSA, XFAB, AMS, and TSMC. An advantage of adopting HV technologies is that it considerably simplifies the circuit complexity, as circuit structures similar to those used in low-voltage designs can be used readily without voltage stress issues. However, typically, HV technologies are not aggressively scaled, and impose large design rules. To truly take advantage of the many benefits of deeply scaled devices, this dissertation introduces technology and circuit techniques that extend the voltage-tolerance of standard, nanometer-scale CMOS technology beyond its conventional voltage limitations.
Maintaining high-voltage tolerance with technology scaling is challenging [18]. As modern CMOS technologies continue to scale, we are faced with two limitations. First, transistor voltage ratings scale down with the technology's nominal supply voltage, as shown in Fig. 1.3(a). Therefore, for the same voltage levels, taller device stacks are required, exacting a toll on power and area efficiency. Second, as shown in Fig. 1.3(b), the substrate voltage-tolerance also scales down as higher doping doses are used to suppress short channel effects in more advanced technology nodes. A CMOS substrate voltage tolerance is set by the reverse breakdown voltage of its well diodes. The substrate voltage tolerance in a 22nm node is <10V.

Creating high-voltage switching waveforms in a low-voltage technology is even more challenging. Driving high-voltage outputs often requires both a high-voltage generation circuit, and high blocking-voltage switches. Thus, high-voltage drive inherits all the challenges of high-voltage generation in addition to the need for a high-voltage tolerant switch implementation.

Moreover, for low-power applications (e.g. portables, wearable devices and medical implants), a high-voltage may need to be generated from a low voltage supply of around 1V. The efficiency of the dc-dc power conversion is critical, imposing additional design constraints.
1.3. Thesis Organization

This dissertation explains methods and circuits that enable the integration of 10's of volt signals onto a standard low-voltage CMOS substrate at improved efficiencies.

In Chapter 2, the dissertation begins with background information on high-voltage designs in IC technology. For high-voltage applications operating off a low-voltage supply, a step-up voltage conversion circuit is needed. A dc-dc converter can be transformed-based, inductor-based, or capacitor-based. A capacitor based dc-dc converter is called a voltage charge pump. For low output currents, charge pumps are attractive in highly-integrated solutions due to the difficulty of integrating high-quality inductors and transformers on chip. As advanced CMOS technologies make available faster switches and denser capacitor implementations, switched-capacitor circuits become more desirable. In this chapter, we first provide an overview of the different types of charge-pump circuits. Then, we discuss the reliability constraints on switch and capacitor design in high-voltage circuits. Finally, we review a number of high-voltage designs presented in literature.

The main contributions of this dissertation are split into two parts. The first part discusses on-chip high-voltage generation, whereas the second part discusses on-chip high-voltage drive. In Chapter 3, we introduce three technology methods to extend the voltage range of CMOS substrates. We explain how standard technology features in a bulk CMOS process are used to extend the substrate voltage tolerance beyond its conventional limit. The proposed methods increase the voltage range of integrated charge-pumps by more than an order of magnitude.

Chapters 4 and 5 describe high-voltage circuit methods that go hand in hand with the technology methods of Chapter 3. The technology methods introduced in Chapter 3 are shown to
extend the voltage range of charge pump circuits from 12V to 100V in a 65nm technology node. To enable these technology methods, special types of charge pump cells need to be implemented. Chapter 4 explains the circuit implementation details of these different pump cells, namely, complementary-type switch voltage doublers, same-type switch voltage doublers, and improved drive Dickson-type voltage doublers. The tradeoff between the voltage range and power efficiency of the different pump types is discussed. We also explain how multiple clock phases (>2) are used to improve the efficiency of transistor-based pump cells.

In order to attain improved power efficiencies at higher voltage ranges, we introduce a Hybrid Charge Pump architecture in Chapter 5. The hybrid pump consists of smaller sub-pumps, and optimally mixes higher voltage-tolerant, lower-efficiency cells with less voltage-tolerant, higher-efficiency cells. As a result, the hybrid pump achieves higher overall power efficiency than that of its least efficient sub-pump. In order to optimize the design of Hybrid Charge Pumps, we introduce an accurate analytical power analysis model in Chapter 5. The model explains a design procedure to size the capacitors of the individual sub-pumps for best efficiency. Also, a noise analysis model is included to estimate the output noise power of the charge pump circuit. Hybrid Charge Pump circuits for both positive and negative output voltages are explained.

The second part of the dissertation is dedicated to high voltage drive. In Chapter 6, two different drive approaches are introduced. In the first approach, we introduce a new, compact, device stacking method to create high blocking-voltage switches without transient voltage stress during switching transitions. This type of driver is suitable for a limited amount of device stacking (2-3 devices), however, it targets faster switching applications (>250MHz). The second driver type is slower but targets much higher output voltages. Therefore, a Charge Pump-Based
driver is proposed to enable the extensive stacking of 10's of devices reliably and with little overhead. The driver speed depends on the pumping frequency and can reach up to 10 MHz.

Multiple test-chips were built as part of this dissertation to validate the various design approaches. Chapter 7 discusses each of these experiments. First, we present the results of 2 Hybrid Charge Pumps in 65nm CMOS technology for both positive and negative output voltages. Second, we present the results of 2 Charge Pump-Based driver designs, one in 65nm CMOS and the other is in 45nm SOI CMOS technology, for both unipolar and bipolar output drive respectively. Chapter 8 summarizes this dissertation with some conclusions and avenues for future work.
CHAPTER 2

Background

2.1. Introduction

This chapter provides a brief overview of the issues in high-voltage circuit design. First, we review some of the basic charge pump circuit architectures, and compare between them from an IC implementation stand point for high-voltage outputs. Next, we discuss the main reliability constraints on the switch and capacitor design in a low-voltage CMOS technology. We also explain how some of these constraints can be relaxed for our final circuit implementation. Next, we explain the upper boundaries on voltage ranges of conventional charge pump designs in CMOS technology. This discussion is important to understand our extended voltage range methods, later introduced in Chapter 3. Finally, we provide a quick review of the different methods reported in literature to enable on-chip high-voltage integration.

2.2. Charge Pump DC-DC Converters

The transition across voltage domains from low to high voltages is performed by a dc-dc converter such as a charge pump. A charge pump is a switched capacitor circuit that relies on the charge redistribution principle to create voltage levels different from the input voltage. A charge pump output can be higher, lower or of opposite polarity to the input voltage. In our context, we are mainly interested in charge pump architectures that step up the input voltage. Because a charge pump implementation needs only switches and capacitors only, it is highly-integrated.
Fig. 2.1. Charge pumps: (a) Heap (b) Dickson (c) Cockcroft-Walton (d) Fibonacci (e) Exponential Gain.
There are many ways by which we can build a charge pump circuit [19]-[21]. Shown in Fig. 2.1 is a conceptual representation of 5 different types of charge pump circuits. Different charge pump implementations have different circuit properties. The most critical of such properties is the stress voltage that different circuit components are exposed to. Depending on the circuit arrangement, the switches and the capacitors can be subjected to different stress voltages.

One convenient way to look at a large class of charge pump circuits is by analyzing them in terms of two-phase switched capacitor H-bridge cells. Depending on how the cells are cascaded, we can generate many of the already known charge pump architectures. Shown in Fig 2.1 are 5 different charge pump architectures, namely, the Heap, Dickson, Crockcroft-Walton, Fibonacci, and Exponential Gain charge pumps. These are referred to here as pump types I through V respectively. Naturally, these architectures have different conversion gain, output resistance, and components voltage rating. The performance of these different pumps is summarized in Table 2.1 in terms of the pump's number of stages (n), switching frequency (f) and stage capacitance (C). All the pump cells are assumed identical. Also, for simplicity, a slow switching limit (SSL) scenario [22] is assumed, and all the parasitic capacitances ignored. For a type I pump, it can be seen that the capacitor maintains a uniform stress voltage for all stages, while the switches are subjected to a progressively larger stress voltage with the number of stages. For a type II pump, the situation is reversed, i.e. the switches are subjected to a uniform voltage stress whereas the capacitors are subjected to a progressively increasing stress voltage with the number of stages. For a type III pump, a uniform voltage rating is maintained across both the switches and capacitors, and is independent of the number of stages. Finally, for type IV and type V pumps, both the switches and capacitors are subjected to a progressively increasing stress voltage with the number of stages.
Table 2.1. Comparison between the charge pump architectures depicted in Fig. 2.1.

<table>
<thead>
<tr>
<th>Type</th>
<th>Conversion Gain ((V_{out}/V_{in}))</th>
<th>Switch Rating</th>
<th>Capacitor Rating</th>
<th>(R_{out}) (Slow Switching Limit)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>((n+1))</td>
<td>(nV_{in})</td>
<td>(V_{in})</td>
<td>(n/fC)</td>
</tr>
<tr>
<td>II</td>
<td>((n+1))</td>
<td>(V_{in})</td>
<td>(nV_{in})</td>
<td>(n/fC)</td>
</tr>
<tr>
<td>III</td>
<td>((n+1))</td>
<td>(V_{in})</td>
<td>(V_{in}) for (n=1) (2V_{in}) for (n&gt;1)</td>
<td>([n(n+1)(n+2)]/12fC) ([((n+1)(n^2+2n+3))/12fC) for n even odd</td>
</tr>
<tr>
<td>IV*</td>
<td>(F_{n+2})</td>
<td>(F_{n+1}V_{in})</td>
<td>(F_{n+1}V_{in})</td>
<td>(F_{n}F_{n+1}/fC)</td>
</tr>
<tr>
<td>V</td>
<td>(2^n)</td>
<td>(2^{n-1}V_{in})</td>
<td>(2^{n-1}V_{in})</td>
<td>((5\cdot4^{n-1}-2)/3fC)</td>
</tr>
</tbody>
</table>

* \(F_n\) refers to the \(n^{th}\) term of the Fibonacci sequence: \(\{1 1 2 3 5 8 13 \ldots\}\).

In CMOS technology, transistor voltage ratings are typically small and are often constrained by the reliability of a relatively thin gate oxide. Transistor ratings are typically reported at the technology's nominal supply voltage with a ±10% voltage tolerance. Metal finger capacitors, on the other hand, have a much thicker inter-finger dielectric, and can thus have higher breakdown voltages. Moreover, capacitor designs for different voltage ratings are much simpler than transistor designs. Metal capacitors can be designed for higher voltage ratings by increasing their finger spacing, whereas devices require stacking, which mandates special handling of their gate drive. Accordingly, for on-chip final voltages that are multiple times the supply voltage, type IV and V pumps are more practical from a design stand-point, since they stress the capacitors rather than the transistors. However, a type IV pump suffers from a diminishing conversion gain in the latter stages. This results in a lower overall pump efficiency and a larger implementation area for the same output voltage. The diminishing gain is due to the capacitive divider forming between the pumping capacitor and its plate parasitics, resulting in lower clock swings in the later stages. Consequently, a type III implementation, also known as the voltage doubler, is always preferred.
2.3. **High-Voltage Design Constraints**

In high-voltage designs, device reliability is of prime concern. Guaranteeing the long term reliability of circuit components directly reflects on the overall circuit mean time-to-failure.

In modern CMOS technology, there exist 2 main reliability constraints on a circuit's output voltage range. The first limitation is related to the components voltage tolerance, and the second limitation is related to the substrate's voltage tolerance. Device ratings constrain the voltage difference across any 2 device terminals to within its breakdown limits. These ratings shrink with finer technology feature sizes, and hence taller device stacks are needed to attain the same voltage levels. This holds equally true for transistors and capacitors. The substrate voltage tolerance is what ultimately limits the amount of device stacking possible. Since the chip substrate is a shared grounded terminal, the circuit's output voltage appears in full across some physical layer in the process substrate. The breakdown voltage of this substrate structure sets the upper limit on a circuit's voltage range. In this section we examine different reliability constraints on both CMOS components and CMOS substrates.

2.3.1. **Switch Limitations**

From a reliability perspective, a transistor is the most sensitive device in a CMOS process. Different breakdown mechanisms exist for a transistor. Some of these mechanisms are related to a relatively thin gate-oxide while others are due to highly energetic carriers near the transistor drain. Oxide breakdown is a slow degradation process and happens when a certain amount of charge per unit area flows through the oxide in response to an applied electric field. Another form of slow transistor degradation is called Bias Temperature Instability (BTI), and is more pronounced for PMOS type transistors. BTI is the slow drift of the transistor threshold voltage...
due to extended exposures of the gate oxide to bias voltages. To relieve gate oxide breakdown and BTI, a transistor's gate-to-source ($V_{gs}$) and gate-to-drain ($V_{gd}$) voltages are constrained to below certain values. Another slow degradation process is Hot-Carrier Injection (HCI). HCI occurs when a transistor is biased in saturation region and highly-energetic carriers are injected into the gate oxide. To relieve HCI, a transistor's drain-to-source ($V_{ds}$) is constrained to below certain values. Shown in Fig. 2.2 is a depiction of the voltage limitations across different device terminals. As mentioned earlier, the upper limit on a transistor source/drain potential ($V_{BD}$) is set by the breakdown voltage of one of the substrate diodes. A comprehensive review of the mentioned breakdown mechanisms is provided in [23]-[24].

![Fig. 2.2. Transistor voltage constraints.](image)

Depending on the circuit type, high-voltage tolerant transistors may or may not be needed. We have seen that for certain charge pump architectures, the switches are not stressed and low-voltage transistors can be used. For high-voltage output stages, high blocking-voltage switches are required. A CMOS process offers different transistor flavors, all of which have a relatively limited $V_{gs}$ voltage rating. Shown in Fig. 2.3 is a representation of the different transistor types available in standard CMOS technology. Core devices are the fastest and have the lowest voltage rating, usually $<1V$ in advanced processes. Input/Output (I/O) devices typically have thicker gate-oxides, and a longer feature size. Hence, they have a higher voltage tolerance but are slower.
switches. Depending on the technology node and the foundry, I/O devices come in different ratings, e.g. 1.8V, 2.5V, and 3.3V. Both core and I/O devices are readily available for a CMOS process. Another transistor option is LDMOS. LDMOS devices are laterally diffused MOSFETS with an enhanced drain-to-source breakdown voltage. These devices offer a high blocking-voltage switch and are widely used in output stage designs such as RF and audio power amplifiers. The $V_{ds}$ voltage tolerance of LDMOS devices can reach as high as 60V, however, the $V_{gs}$ voltage tolerance is still limited to a relatively thin gate-oxide (<5V). LDMOS devices are compatible with a standard CMOS process flow, but require an extra mask set, increasing the fabrication cost. Moreover, an LDMOS design requires certain care to preserve the device gate-oxide reliability. In order to do so, the gate drive of an N-type LDMOS device must not exceed its gate-oxide voltage rating. A P-type LDMOS device requires a level shifting circuit which references the gate drive signal with respect to the high-voltage supply rail. This is better depicted in the high-voltage output stage shown in Fig. 2.4(a).

![Diagram](image)

Fig. 2.3. CMOS device flavors: (a) Core devices (b) I/O devices (c) LDMOS devices.

One possible way to implement a high blocking-voltage switch is to stack multiple low-voltage devices in series [25]. In principle, stacking devices allows voltages to divide equally across a larger number of devices. Thus, higher blocking-voltages can be tolerated without
stressing any device. However, some circuit complexity is necessary to generate the gate drive signal for all the individual devices in the stack. Each transistor in the stack requires some level shifting to maintain its gate-oxide reliability. Moreover, more components may be necessary to guarantee that the device $V_{ds}$ does not exceed the transistor rating during switching transitions. Shown in Fig. 2.4(b) is a high-voltage output stage that is implemented using device stacking. The predriver circuit performs all the level shifting and maintains the $V_{ds}$ reliability of all the stacked devices. The predriver circuitry represents an additional area and power overhead. Notice that the high-voltage stage can switch the output node between the ground and the supply rail only. Implementing a floating high-voltage switch that uses device stacking is more challenging, since all the gate drive signals need to be level shifted with respect to a dynamic node potential.

![Diagram](image)

**Fig. 2.4.** High-Voltage output stage using: (a) LDMOS devices (b) low-voltage device stacking.

In summary, LDMOS devices incur an extra mask cost and are not used in any of our designs. Because I/O devices have higher voltage tolerances compared to the core devices, we implement all the switches in our design using thick-oxide devices. Device stacking enables extended voltage range switches but ensues extra complexity and does not support floating switch implementations easily. Consequently, device stacking is adopted only in circuits that do
not require floating switches, such as high-voltage output stages. While in circuits that need floating switches, such as charge pumps, a single transistor switch implementation proves to be more cost and power efficient.

2.3.2. Capacitor Limitations

A CMOS process makes available different capacitor implementations. One capacitor type relies on a MOS device gate capacitance. MOS capacitors provide the highest capacitor density but have a limited voltage tolerance due to a thin gate oxide. A MOS capacitor has similar voltage ratings to those of a MOSFET and cannot be used in Heap or Fibonacci type pumps. A different capacitor type is an MIM capacitor. These are more linear, high density metal capacitors that incur special processing and an extra mask cost. MIM capacitors are not high-voltage tolerant because the capacitor plates are very tightly separated. One capacitor type that is useful in high-voltage designs is an MOM or finger capacitor. This capacitor relies on the fringe and parallel plate fields occurring naturally between tightly coupled metal layers. Metal finger capacitors offer the lowest capacitor density but have the highest voltage tolerance.

Finger capacitors can be adjusted to have a variable voltage rating by design. Increasing the spacing between the capacitor fingers reduces the electric field across the inter-finger dielectric, and higher voltage tolerant capacitors are attained. In a certain light, spacing out the capacitor fingers can be seen similar to stacking capacitors in series. For the same unit capacitor, wider finger spacing results in a larger implementation area and higher plate parasitics to ground. Consequently, higher voltage tolerant capacitors result in lower power efficiencies. We highlight this property because it is a recurrent theme throughout this dissertation. It is found that extending a circuit's voltage range is often associated with a reduction in power efficiency.
The reliability of metal-finger capacitors is mainly determined by the time-dependent dielectric breakdown (TDDB) failure rates [26]-[27]. TDDB is a long term failure mechanism due to the breakdown of the inter-layer dielectric. TDDB is a function of the finger spacing and lateral area. Our TDDB model is based on the following Weibull distribution,

\[
CDF = 1 - \exp \left[ -\left( \frac{t}{\tau} \right)^{\beta} \right]
\]

(2.1)

The above cumulative distribution function (CDF) expresses the capacitor failure rate as a function of time. If we wait long enough eventually all the samples fail. The CDF in Equation (2.1) has 2 technology dependent parameters, \( \beta \) and \( \tau \). The time constant \( \tau \) represents the time needed at which 63% of the samples fail. A larger \( \tau \) corresponds to longer capacitor lifetimes. If \( \tau \) is known for a given stress voltage (V), absolute temperature (T), and Area (A), then \( \tau \) can be estimated for any V, T, and A using the field accelerated \( \sqrt{E} \)-model, Arrhenius temperature relation, and Poisson area-scaling model specified in Equations (2.2)-(2.4), respectively,

\[
\tau \propto \exp \left(-\gamma \sqrt{V} \right)
\]

(2.2)

\[
\tau \propto \exp \left( \frac{E_a}{KT} \right)
\]

(2.3)

\[
\tau \propto \left( \frac{1}{A} \right)^{\frac{1}{\sigma}}
\]

(2.4)

The values of the parameters \( \gamma \), \( E_a \) and \( \sigma \) are considered sensitive Foundry information and change from one technology node to another. Typically, a designer targets a specific CDF over a specified circuit lifetime. Based on this a time constant \( \tau \) is calculated from equation (2.1). Knowing the capacitor area and operating temperature, the voltage that a capacitor handles is calculated using equations (2.2)-(2.3). The \( \tau \) dependence on metal spacing is not captured by an
accurate formula and foundry measurements performed at different metal spacing are necessary. In this work, we use the data reported in [27] to infer \( \tau \) dependence on the finger spacing.

Shown in Fig. 2.5 is an estimate of the capacitor voltage tolerance versus the finger spacing normalized to the minimum metal pitch (\( \sim 100\text{nm} \) in 65nm CMOS technology). A unit 4pF metal capacitor with minimum finger spacing is estimated to sustain \( \sim 12\text{V} \) based on a CDF of less than 100 ppm over 10 years at 85\(^\circ\)C. It is noticed that voltage tolerance is a superlinear function in finger spacing, and that by doubling the finger spacing the voltage tolerance more than doubles.

![Fig. 2.5. Voltage tolerance of a 4pF metal capacitor versus its finger spacing in 65nm CMOS technology.](image)

In our designs, we gradually taper the capacitor finger spacing based on the voltage requirements of each stage. This optimizes the circuit area and power efficiency as higher voltage tolerant capacitors cost area and power. Shown in Fig. 2.6 are the layouts for 3 different capacitor designs using metal layers M2 through M4. The first metal layer M1 is not used to cut down the bottom plate parasitic capacitance. The first design uses minimally spaced fingers (1x spaced) corresponding to a capacitance density of 1.4fF/\( \text{um}^2 \). The second and third sub-pumps use 1.5x and 2x finger spacing, and correspond to capacitance density of 0.93fF/\( \text{um}^2 \) and 0.6fF/\( \text{um}^2 \), respectively. It is noted that we have control over the metal lateral spacing only while the metal vertical spacing remains unchanged. This raises a problem for capacitors with finger
spacing larger than the metal vertical spacing. In which case, a capacitor implementation that relies on side-wall capacitance is adopted. An example of a capacitor that alleviates vertical fields completely is shown in Fig. 2.6(c). Charge pump efficiency depends on the plate parasitic to main pumping capacitance ratio ($C_p/C$). Capacitors with wider finger spacing have a higher $C_p/C$ that corresponds to 3.6%, 5.2% and 8.8%, for the 1x, 1.5x, and 2x capacitors respectively.

![Fig. 2.6. High-Voltage capacitor layout with different finger spacing: (a) 100nm (b) 150nm (c) 200nm.](image)

**2.3.3. Substrate Voltage Tolerance**

So far we have shown that transistors and capacitors can handle higher voltages through stacking. In principle, higher voltage ranges are attainable as more devices are placed in the stack. However, there exists an upper limit on the number of devices that can be stacked on-chip. This upper limit stems from the breakdown of some physical structure appearing across the process substrate. This structure depends on the process substrate and in a bulk technology, for example, is different from that in an SOI technology. In order to determine the upper limit on on-chip voltages for a given process, the substrate process layers must be analyzed.
Shown in Fig. 2.7 is the process cross-section of 3 different device types commonly available in a p-type substrate, bulk CMOS technology. In older twin-tub processes, all the NMOS devices share a common bulk connection that is the same as the chip substrate. A triple-well process adds an extra deep n-well layer to separate the NMOS devices inside electrically isolated wells. A deep nwell NMOS device can have its source and bulk terminals shorted to eliminate back-gate bias effects. Moreover, a deep n-well layer helps mitigate deleterious substrate coupling effects in a mixed-signal design. In today’s CMOS technology, a deep nwell is almost a standard option.

Fig. 2.7. CMOS devices cross-sections: (a) PMOS device (b) NMOS device (c) Deep nwell NMOS device.

Here we use charge pumps circuits as a vehicle to quantify the voltage tolerance of a CMOS substrate. Charge pump designs are traditionally limited to single-diode substrate isolation. This diode could be that of a source/drain implant, an nwell, or a deep nwell. Depending on the type of devices used and the polarity of the pump output, different diodes can clamp the final voltage.

Shown in Fig. 2.8 are simplified representations of a CMOS charge pump cell using different types of NMOS devices. The pumping cell has 2 inputs $V_{in}$ and $V_{dd}$, and 2 charge transfer switches, one of type NMOS and another of type PMOS. The switches operate on 2 complementary clock phases. During one clock phase, the bottom plate of the pumping capacitor ($C$) is connected to ground and $C$ is charged to $V_{in}$. During the second clock phase, the capacitor bottom plate is connected to $V_{dd}$ and the output voltage is pumped to $V_{in}+V_{dd}$, at steady state, to
a first order. It is assumed that the switches are implemented as efficient charge-transfer switches using the appropriate level shifting circuits. Also, it is assumed that the input voltage $V_{in}$ is generated using similar voltage pumping cells preceding this one, and that $C$ has a voltage tolerance well above $V_{in}$. In this case, the pump's voltage range is only limited to the breakdown voltage of the substrate. Depending on the polarity of the charge pump output and the type of transistors used, different diodes clamp the charge pump's maximum output voltage. Table 2.2 summarizes the different types of diodes and their approximate breakdown voltages in a triple-well 65nm CMOS process.

![Diagrams](a) and (b)

Fig. 2.8. A simplified positive CMOS charge pump cell using (a) NMOS devices (b) deep nwell NMOS devices.

Table 2.2. Approximate breakdown voltages for the different substrate diodes in 65nm CMOS technology.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Breakdown Voltage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{NP}$</td>
<td>8V</td>
<td>n+/pwell diode</td>
</tr>
<tr>
<td>$V_{PN}$</td>
<td>10V</td>
<td>p+/nwell diode</td>
</tr>
<tr>
<td>$V_{NW}$</td>
<td>12V</td>
<td>nwell/psub diode</td>
</tr>
<tr>
<td>$V_{PW}$</td>
<td>10V</td>
<td>pwell/deep nwell diode</td>
</tr>
<tr>
<td>$V_{DNW}$</td>
<td>12V</td>
<td>deep nwell/psub diode</td>
</tr>
</tbody>
</table>
Fig. 2.9. Voltage range limitations on positive output voltage charge pumps: (a) NMOS (b) PMOS (c) deep nwell NMOS (d) deep nwell PMOS.

First, let’s consider the case for positive output voltage charge pumps. Because a p-type substrate is always grounded, a bulk NMOS device cannot have its source/drain potentials exceed the reverse breakdown voltage of the n+/psub diode denoted here by $V_{NP}$. For a PMOS device, the source and bulk terminals are always shorted to prevent forward-bias junctions, in which case the maximum source/drain potential is limited to the reverse breakdown voltage of the nwell/psub diode denoted by $V_{NW}$. For a deep nwell NMOS device, the deep nwell terminal is grounded to prevent forward-bias junctions, in which case the maximum source/drain potential is limited by the minimum of 2 voltages $V_{NW}$ and $V_{DNW}$, where $V_{DNW}$ is the reverse breakdown voltage of the deep nwell/psub diode. Note that a PMOS device placed in a deep nwell also has its source/drain potentials limited by the minimum of $V_{DNW}$ and $V_{NW}$. Shown in Fig. 2.9 are the process cross-sections summarizing these different scenarios. Based on this, it can be shown that
the pumping cell using a bulk NMOS device in Fig. 2.8(a) has a maximum output voltage that is equal to \( \min(V_{NP}, V_{NW}) \). Where \( \min(a, b) \) is the minimum of 2 values a and b. Also, the voltage pumping cell using a deep nwell NMOS device in Fig. 2.8(b) has a maximum output voltage that is equal to \( \min(V_{NW}, V_{DNW}+V_{dd}) \). From an efficiency stand point, the cell in Fig. 2.8(b) is preferred because it does not suffer from threshold voltage deterioration due to back-gate effects.

Next, we consider the case of negative output voltage charge pumps. Shown in Fig. 2.10 are simplified implementations of 2 negative output voltage pumps. The one on the left relies on a CMOS implementation, whereas the on the right relies on an all-NMOS implementation. Only PMOS and deep nwell NMOS devices can be used to generate negative voltages because a bulk NMOS device suffers from a forward biased source/drain junction. For a PMOS device, the bulk terminal must be grounded and thus the device suffers from a deteriorating threshold voltage \( (V_{th}) \) due to back-gate effects. The highest negative potential of a PMOS source/drain is limited by the reverse breakdown voltage of the p+/nwell diode denoted by \( V_{PN} \). For a deep nwell NMOS transistor, the deep nwell terminal is grounded and the source and bulk terminals are shorted together to prevent forward biased junctions. The largest negative potential of an NMOS source/drain is then limited by the reverse breakdown voltage of the pwell/deep nwell diode denoted by \( V_{PW} \). Deep nwell NMOS devices do not suffer from back-gate bias effects. A PMOS device placed in a deep nwell is no different than a regular PMOS device. Shown in Fig. 2.11 are process cross-sections summarizing these different scenarios. Based on the previous discussion, it can be shown that the CMOS pumping cell in Fig. 2.10(a) has a maximum negative output voltage that is equal to \( -\min(V_{PN}, V_{PW}) \). Whereas, the all-NMOS cell in Fig. 2.10(b) has a maximum output voltage that is equal to \( -V_{PW} \). From an efficiency stand point, the all-NMOS implementation is preferred because the switches do not suffer from back-gate bias effects.
Fig. 2.10. A simplified negative CMOS charge pump cell using (a) CMOS devices (b) all- NMOS devices.

Fig. 2.11. Voltage range limitations on negative output voltage charge pumps: (a) NMOS (b) PMOS (c) deep nwell NMOS (d) deep nwell PMOS.

Shown in Fig. 2.12 are the measurement results of charge pump circuits implemented in a 65nm CMOS technology. The measured output voltage of a CMOS charge pump similar to the design in Fig 2.8(b) is plotted versus its input voltage ($V_{in}$) as shown in Fig. 2.12(a). We observe that the pump's output voltage clips near 12.5V as expected. Similarly, the measured output
voltage of an all-NMOS charge pump similar to the design in Fig. 2.10(b) is plotted versus \( V_{in} \) as shown in Fig. 2.12(b). Interestingly, the foundry model predicts \( V_{PW}=10\text{V} \), while measurements show that the pump's output voltage clips near \(-12.5\text{V}\).

![Graph](image)

Fig. 2.12. Measured charge pump output voltage versus \( V_{in} \) for: (a) positive output voltage pump (b) negative output voltage pump.

![Diagram](image)

Fig. 2.13. A channel-implant block mask used to extend the substrate voltage tolerance for: (a) an nwell device (b) a deep nwell device.
One possible way to improve the breakdown voltage of a CMOS substrate is proposed in [11]. It is based on the property that junctions with lower doping concentrations have higher breakdown voltages. In many CMOS processes, a pwell implant is created by default in regions where nwell or deep nwell implants are not present. Thus, the implant block mask used for native devices is employed to create a low doped p-type buffer region surrounding the nwell as shown in Fig. 2.13. This technique reduces the doping concentration on one side of the junction and may improve the substrate breakdown by a few volts.

SOI CMOS technology is an almost standard CMOS process with relatively advanced technology nodes. In SOI CMOS, the active chip devices are isolated from the silicon substrate via a buried oxide (BOX) layer as shown in Fig. 2.14. The substrate voltage tolerance in SOI CMOS is determined by the breakdown voltage of its BOX layer. BOX breakdown voltages are typically several times higher than that of the substrate diodes in nanometer scale bulk processes, and thus SOI CMOS is an attractive choice for many high-voltage applications. The breakdown voltage of a 145nm thick BOX layer used in ultra thin oxide SOI CMOS technology is estimated to be ~60V. Furthermore, because SOI CMOS technology does not rely on reverse biased junctions for isolation, it handles both positive and negative output voltages equally well.

---

![Fig. 2.14. SOI CMOS pump cell: (a) simplified circuit schematic (b) process cross-section.](image-url)
2.4. **Review of High-Voltage Methods**

In this section, we provide a brief review of the different high-voltage generation and drive circuits reported in literature. Some of these methods are technology specific, i.e. rely on a HV CMOS process and devices, whereas other solutions are compatible with standard CMOS process. We first discuss high-voltage charge pumps then high-voltage output stages.

### 2.4.1. High-Voltage Generation

Generating high output voltages on chip requires a high-voltage tolerant substrate. A number of high-voltage charge pumps are reported in literature using different substrate isolation methods [28]-[36]. Older CMOS technology nodes provide higher voltage tolerances than those of more advanced nodes. Due to the lower implant doses used in older technology nodes, the substrate diodes can handle higher reverse breakdown voltages. The measurement results in [28] and [29] report 50V and 15V output charge pumps in 0.6μm and 0.18μm technology nodes, respectively. A HV CMOS technology readily provides more voltage tolerant substrates by design. The work in [30] leverages the high breakdown voltage of HV wells in a DALSA process to build a 50V output charge pump. The work in [31] reports a 40V charge pump implemented in an i2T100 HV technology from AMI Semiconductors. Because BOX provides higher voltage isolation than reverse-biased diodes, 20V and 27V output voltage charge pumps are reported in [33] and [34], respectively in 0.35μm SOI CMOS technology. Similar to the role of BOX in SOI process, the field-oxide (FOX) layer is used to provide substrate isolation in a bulk process as explained in [35] and [36]. Because very primitive active devices can be built on top of FOX, simple polysilicon diode structures are used to implement the switches of a Dickson-type charge pump. The diodes are fully compatible with the standard CMOS process flow, and a 43V output voltage Dickson charge pump is reported in 0.18μm CMOS technology.
Finally, the work in [37] reports a high-voltage charge pump design that avoids the breakdown voltage limitations of CMOS substrates by using MEMS resonant switches. The design method speculates output voltages >50V are possible using these MEMS devices.

Based on this short survey, we find that the only methods enabling high-voltage outputs in a standard, fine line-width CMOS process is by leveraging the BOX isolation in an SOI CMOS technology, or the FOX isolation in a bulk CMOS technology.

2.4.2. High-Voltage Drive

In addition to high-voltage tolerant substrate isolation, output drive requires a high blocking-voltage switch. This makes the problem of high-voltage drive more challenging than that of high-voltage generation. Common ways to implement a high voltage switch are reported in [37]-[47]. In [37], because mechanical switches are used instead of transistors, the switches can in principle enable much higher output voltage waveforms than possible using standard CMOS transistors. The work in [38]-[41] supports HV switching devices in a standard CMOS process. HV LDMOS transistors can support drain-to-source voltages that are much higher than a technology’s nominal supply voltages as earlier explained. An LDMOS device can either be designed manually using the standard process mask set by the designer, or provided automatically by the foundry as a high-voltage module option. The work in [38] reports 60V output waveforms in a 2μm CMOS process, whereas the work in [40] reports an 18V output waveform in 0.35μm SOI CMOS process. More recently, another approach that relies on stacking low-voltage devices in series is gaining popularity [42]-[47]. Stacking devices is performed in such a way that none of the devices $V_{gs}$ or $V_{ds}$ are stressed at any point of time. This requires an intricate predriver circuit design that limits the number of stacked devices to 4-5
transistors in both the pull-up and pull-down networks. The highest output voltage switching waveform reported using device stacking is 10V in a 0.18μm CMOS process.

Based on this survey, we find that in order to implement high-output voltage waveforms in standard, fine line-width CMOS process, we need to adopt extended device stacking techniques to preserves the reliability of individual devices. Shown in Fig. 2.15 is a summary of the different high-voltage generation and drive methods covered so far.

Fig. 2.15. A literature survey of high-voltage methods for voltage generation and voltage drive circuits.

2.5. **Summary**

In this chapter, we have shown that the choice of charge pump architecture is critical in high voltage designs. In a low-voltage CMOS process, either standard devices or capacitors can be stacked in series to attain higher voltage tolerances. However, stacking capacitors is much less complicated than stacking transistors. Thus, picking a charge pump architecture that stresses the capacitors rather than the switches simplifies the charge pump design significantly. We have also shown that the maximum voltage range of a conventional charge pump circuits in bulk CMOS
technology is limited to the reverse breakdown voltage of one substrate diode. This breakdown voltage is fairly limited and shrinks as technologies scale down. In a 65nm technology node, the maximum output voltage of a bulk charge pump is ~12V. In order to achieve output voltages that are several 10's of volts, new substrate isolation methods are required.

In the next chapter, we introduce a number of technology methods that enable us to expand the voltage range of bulk CMOS substrates by multiple times. In order for these techniques to be effective, special pump cell types are required. These pump cells are introduced in Chapter 4.
CHAPTER 3

Extended Voltage-Range CMOS Substrates

3.1. Introduction

In this chapter we introduce technology methods to extend the voltage range of a standard bulk CMOS substrate by multiple times [48]. Mainly, we explain 3 techniques leveraging standard technology features in a triple-well CMOS technology to enable improved substrate voltage isolation. The first technique provides charge pump circuits with double-diode substrate isolation as opposed to conventional designs leading to single-diode substrate isolation. The second technique improves on the FOX isolation method in [35] by providing higher voltage tolerance or improved device reliability using deep nwell. To leverage FOX isolation, only diode-based charge pumps can be used. We explain how polysilicon PIN diodes are implemented and discuss their measured I-V results. Also, measurement results from various polysilicon diode-based pumps are explained to highlight their poor efficiency performance. This serves as a motivation to the improved-drive Dickson-type pumps introduced in Chapter 4. In the third technique, we show how multiple substrates can be stacked to further extend the voltage range of charge pumps to target the several 100's of volt outputs in a nanometer-scale technology.

3.2. Double-Diode Substrate Isolation

In the previous chapter, we have shown that traditionally the voltage range of charge pump circuits in a bulk technology is limited to the breakdown voltage of one of the substrate diodes, i.e. single-diode isolation. In this section, we introduce a well-biasing scheme that enables charge
pump designs to attain double-diode substrate isolation. The scheme relies on the availability of a deep nwell option which is a standard feature in today's nanometer scale technology nodes.

The introduced well-biasing scheme attempts to leverage the extra substrate isolation layer provided by a deep nwell. Because wells in a bulk CMOS process rely on reverse biased junctions for isolation, a deep nwell comes with a separate substrate diode. It can be shown that for the case of positive output voltage pumps, 2 reverse substrate diodes can be stacked in series to attain higher voltage outputs. To realize this substrate diode stacking, we choose *not* to short the source and bulk terminals of a deep nwell NMOS device as commonly done. Although this defeats the original purpose of a deep nwell, it extends the process voltage range. Instead, only the bulk and deep nwell terminals are shorted together and their potential is allowed to rise slowly to an intermediate potential $V_{\text{mid}}$, ideally, equal to the breakdown voltage of the deep nwell/psubstrate diode ($V_{\text{DNW}}$), as shown in Fig. 3.1. This bias voltage can be tapped off a previous charge pump stage. Now, the transistor’s source/drain potentials can rise to voltages higher than $V_{\text{mid}}$ by an additional diode breakdown voltage, namely, $V_{\text{NP}}$. In a 65nm node, this well-biasing scheme extends the voltage range of bulk charge pumps by 66%, from 12 to 20V.

![Fig. 3.1. Double-Diode substrate isolation: (a) simplified pump schematic (b) process cross-section.](image)

33
A charge pump similar to the design in Fig 3.1(a) is implemented in 65nm CMOS technology. The measured pump output voltage is shown in Fig. 3.2 versus the pumping voltage \( V_{dd} \). As expected, the pump's output voltage clips near \( V_{DNW} + V_{NP} \) (~20V).

![Graph showing measured output voltage of the extended voltage range All-NMOS charge pump versus \( V_{dd} \).](image)

This voltage extension method, however, has a few limitations. First, the gate potential of the NMOS devices is also limited to \( V_{DNW} + V_{NP} \). Therefore, to provide pass devices with enough gate overdrive to pass its input voltage, we need to back off from the maximum attainable output voltage by ~0.5V. Second, because the transistor source and bulk terminals are not shorted, back-gate bias results in a larger NMOS \( V_{th} \). Because \( V_{th} \) is a function of the bulk to source voltage difference, this effect is more pronounced in the latter pump stages as we move closer to the output node. Larger switch \( V_{th} \) results in poorer pump efficiency compared to conventional designs. Last, this technique is not applicable to negative output voltage charge pumps. In the case of negative voltages, the NMOS source and bulk terminals must be shorted, also the deep nwell terminal must be grounded to prevent all forward-biased junctions. This leaves only a single diode available for substrate isolation, namely, the pwell/deep nwell diode.
Furthermore, the proposed well biasing scheme cannot be directly extended to PMOS devices even if placed inside a deep nwell. This is because the deep nwell automatically shorts to the nwell containing the PMOS device, and only single-diode isolation is feasible. As a result, only deep nwell NMOS devices can be used in the circuit implementation and an all-NMOS charge pump cell is necessary to enable double-diode substrate isolation. The implementation details of this cell type are explained in the next chapter.

3.3. Field-Oxide Substrate Isolation

For output voltage ranges >20V, we need a fundamentally different voltage isolation method. In this section, we explain how standard technology layers, namely, field oxide and polysilicon can be used to generate on-chip DC voltages as high as 100V.

In CMOS technology, shallow trench isolation (STI), also referred to as field oxide, is primarily used to provide electrical isolation between adjacent CMOS devices. In principle, circuits that are built on top of FOX are well isolated from the substrate by a few hundred nanometers of thick oxide, as is the case with BOX in SOI substrates. Unlike SOI however, only polysilicon structures can be realized on top of STI. Fortunately, simple diode structures can be implemented in polysilicon using the n+ and p+ implant masks used to form a transistor's source and drain regions. Even though diodes are inferior switches compared to transistors due to their knee voltage drop, they still can be used in voltage pumping circuits as shown in Fig. 3.3(a). Consequently, only diode-based pumps can enable >20V output voltages in this technology node. The maximum output voltage is now limited to the breakdown of STI as clear from the cross-section in Fig. 3.3(b), and is measured to be 88V in 65nm CMOS technology. A 43V Dickson charge pump using polysilicon PIN diodes on top of FOX is reported in [35] using
$0.25\mu m$ CMOS technology. The reported Dickson pump in suffers from poor power efficiency due to the inferior diode-based switches, and once again we come across a clear power efficiency, voltage range trade-off.

In this dissertation, we introduce an improved polysilicon diode implementation [48]. This implementation can be used to either extend the voltage range of FOX isolation or improve its long-term reliability. In the improved implementation, we combine the deep nwell biasing scheme from the previous section with FOX isolation to accrue even higher voltage isolation. As shown in Fig. 3.4, the polysilicon diode is placed on top of a deep nwell guard ring for better substrate isolation. In the case of positive output voltage pumps, the diode's pwell and deep nwell terminals are shorted together and allowed to rise slowly to an intermediate potential $V_{\text{mid}}$, ideally equals to $V_{DNW}$. In which case, the substrate's voltage tolerance can now extend by an additional diode breakdown voltage, namely $V_{DNW}$. This bias voltage can be tapped off a previous charge pump stage. Accordingly, the maximum output voltage of the charge pump circuit in Fig 3.5(a) is equal to $V_{FOX}+V_{DNW}$ ($\sim$100V in 65nm technology). Similarly, in the case of negative output voltage pumps, the diode's deep nwell terminal is grounded while the pwell
terminal is allowed to drop slowly to an intermediate negative potential $V_{\text{mid}}$, ideally equals to $-V_{PW}$. In which case, the substrate's absolute voltage tolerance is increased by an additional diode voltage tolerance, namely $V_{PW}$. And the maximum negative voltage of the charge pump circuit in Fig 3.5(b) is equal to $-V_{FOX}-V_{PW} (~ -96V)$.

![Diode well-biasing arrangement](image)

**Fig. 3.4.** Polysilicon diode well-biasing arrangement: (a) positive output voltage pumps (b) negative output voltage pumps.

![Simplified Diode-based charge pump cells](image)

**Fig. 3.5.** Simplified Diode-based charge pump cells with: (a) positive output voltage (b) negative output voltage.

A diode-based pump similar to the design in Fig 3.5(a) is implemented in a 65nm CMOS node. The pump's measured output voltage ($V_{out}$) is plotted in Fig. 3.6(a) versus its input voltage ($V_{in}$). $V_{out}$ increases with $V_{in}$ until the FOX breaks down at $V_{out} \approx 100V$. Shown in Fig. 3.6(b) is the pump's micrograph after breakdown highlighting the failure location near the pump's output.
3.3.1. Polysilicon Diode Design

Polysilicon diodes are attractive devices for building high voltage-tolerant charge pumps. The device process cross-section is depicted in Fig. 3.7(a). It is shown that the device is compatible with the standard CMOS process flow and can be integrated on-chip using readily available process masks at no extra cost. The device layout design is indicated in Fig. 3.7(b). The diode's P and N type regions are created using the same P+ and N+ masks used to create the transistors source and drain implants. We choose to implement the diode as a PIN structure. A PIN diode has an intrinsic polysilicon region sandwiched by a P-type and an N-type region. The undoped intrinsic region enhances the diode's reverse breakdown voltage. That breakdown voltage can be adjusted by tuning the intrinsic region length ($L_i$). For many CMOS processes, the polysilicon is automatically silicided unless otherwise specified. A silicide block mask is extended over the edges of the diode intrinsic region to prevent a short circuit shunting the diode.
The diode is isolated from the substrate through a thick FOX layer (~300nm) with $V_{\text{FOX}} \approx 88\text{V}$ in 65nm technology. Generating high voltages on polysilicon layers may result in undesired inversion channels underneath the FOX. For improved isolation, the diode is guarded by a deep nwell. This is a modification to the original design introduced in [35]. Compared to our previous design in [48], we enhance this diode design in 2 ways. First, when used in a charge pump, the deep nwell is biased near its reverse break down voltage ($V_{\text{DNW}}$). This bias voltage is tapped off a previous stage in the charge pump circuit. Consequently, the FOX region underneath the diodes is subjected to lower stress voltages, improving long term reliability. Second, a lightly doped drain (LDD) block mask is used to protect the polysilicon region from double implants. Not using this mask would result in large variability of device behavior, sensitivity to device orientation, and in some cases device failure. Typically, design rules mandate that all P and N implants extend fully over the polysilicon region. Because this requirement conflicts with the diode structure, this design rule is waived as it does not impact the process yield.

Fig. 3.7. Polysilicon PIN diode: (a) process cross-section (b) layout design.
The diode design includes 3 important parameters: the diode width (W), the intrinsic region length (L_i), and the unsilicided region length (L_{sb}). The diode W is chosen based on the pump load current, and for the same diode forward drop, the current scales proportionally with W. The silicide block mask must extend beyond both intrinsic region edges by at least the diode's depletion width, i.e. L_{sb}>L_i. Also, technology design rules may impose a minimum length requirement on the silicide block mask. In 65nm technology this rule is such that L_{sb} ≥ 0.4μm. By combining these 2 design requirements, we define a feasible diode design space depicted in Fig. 3.8(a). Diodes with excessive L_i suffer from poor diode characteristics. Consequently, for our implemented diode sample space, we impose an additional constraint on L_i such that L_i ≤ 0.75μm. The actual diode instances implemented for characterization are enclosed by the blue outline and have a granularity of 50nm in the x direction and 100nm in the y direction as indicated by Fig. 3.8(b).

![Diode Design Parameters](image)

---

**Fig. 3.8. Polysilicon diode design space: (a) permissible design space highlighted in blue (b) implemented diode sample space enclosed by the blue outline.**
The diode's \( L_i \) value is a critical design parameter. Shown in Fig. 3.9 are the measured I-V characteristics of a number of polysilicon PIN diodes with different \( L_i \). These diodes are implemented with \( W=20\mu m \) and a varying \( L_{sb} \) such that \( L_{sb}=L_i+0.85\mu m \). We notice that diodes with larger \( L_i \) have a larger reverse breakdown voltage (\( V_{BD} \)), and a larger knee voltage (\( V_D \)).

![Graph showing measured polysilicon PIN diode characteristics for different \( L_i \) with \( W=20\mu m \) and \( L_{sp}=L_i+0.85\mu m \): (a) reverse bias (b) forward bias.]

Shown in Fig. 3.10(a) is the measured diode \( V_{BD} \) for different \( L_i \) values. \( V_{BD} \) is defined as the reverse voltage at which the diode current equals 0.1\( \mu A \). Shown in Fig. 3.10(b) is the measured diode forward voltage at a 20\( \mu A \) current for different \( L_i \) values. Notice that the diode has a large forward drop and poor I-V characteristics for large \( L_i \). In a Dickson-type pump with a pumping clock supply \( V_{dd} \), the diodes are subjected to a \( 2V_{dd} \) reverse voltage stress. Therefore, the diode \( L_i \) must be chosen large enough such that \( V_{BD}>2V_{dd} \) by a good margin, and small enough such that \( V_{dd}>V_D \) at the target load current by a good margin. Based on the measurement results, we find that a good range for \( L_i \) is such that \( 0.35\mu m \leq L_i \leq 0.6\mu m \).
Fig. 3.10. Measured polysilicon diode characteristics versus $L_i$ with $W=20\mu m$ and $L_{sp}=L_i+0.85\mu m$: (a) reverse breakdown voltage at $I_r=0.1\mu A$ (b) forward voltage drop at $I_f=20\mu A$.

To fully characterize the diode, we do measurements for a $20\mu m$ wide diode across 3 directions in the design space. The first direction is diagonal with $L_{ab}-L_i=0.85\mu m$ as shown in Fig. 3.10. The other 2 directions are orthogonal, one horizontal and another vertical. For the horizontal direction, $L_i$ varies for a constant $L_{ab}=1.25\mu m$. The measurement results for this diode set are shown in Fig. 3.11. Comparing with the results from the diagonal sample set, we observe that the differences in $V_{BD}$ and $V_D$ are insignificant and are mostly dominated by the process variation. To isolate this design impact of $L_{ab}$, we characterize the diode in the vertical direction, such that $L_{ab}$ changes for a constant $L_i=0.4\mu m$. The measurement results for this set are shown in Fig. 3.12. We notice that the impact of $L_{ab}$ on the diode characteristics is insignificant and is rather swamped by the random process variations. Also, we observe that the process variations in $V_{BD}$ are much larger than the variations in $V_D$. $V_{BD}$ varies between 9.4V and 10.5V in this sample space corresponding to 10.6% peak-to-peak variation, whereas $V_D$ varies between 0.847V and 0.864V and corresponds to 1.75% peak-to-peak variation. Finally, we point out that there exists a strong correlation between the diode's $V_{BD}$ and $V_D$. 
Fig. 3.11. Measured polysilicon diode characteristics versus $L_i$ with $W=20\mu m$ and $L_{sp}=1.25\mu m$: (a) reverse breakdown voltage at $I_r=0.1\mu A$ (b) forward voltage drop at $I_f=20\mu A$.

Fig. 3.12. Measured polysilicon diode characteristics versus $L_{sb}$ with $W=20\mu m$ and $L_i=0.4\mu m$: (a) reverse breakdown voltage at $I_r=0.1\mu A$ (b) forward voltage drop at $I_f=20\mu A$. 

43
Diode matching is characterized for devices across the same die and across different dies. The diodes used have \( W=20\mu m \), \( L_i=0.4\mu m \), and \( L_{sp}=1.25\mu m \). Shown in Fig. 3.13 are the measured I-V characteristics of identical diodes across 7 different chips. We notice that the measured diodes exhibit a fairly good reproducibility. To further quantify the diode variations, we show in Fig. 3.14 the measured diode's \( V_D \) and \( V_{BD} \) versus the chip number. As usual, \( V_{BD} \) exhibits a larger peak-to-peak variation of 3.5%, whereas \( V_D \) has a smaller peak-to-peak variation of 0.9%. Furthermore, diode matching is characterized for 3 devices placed on the same die. Shown in 3.15 are the measured diode's \( V_D \) and \( V_{BD} \) versus the diode number. We notice that \( V_{BD} \) exhibits a peak-to-peak variation of 1.75%, whereas \( V_D \) exhibits peak-to-peak variation of 0.18% for identical diodes placed on the same chip.

Fig. 3.13. Measurements of identical diodes on different dies with \( W=20\mu m \), \( L_i=0.4 \mu m \) and \( L_{sp}=1.25\mu m \): (a) reverse bias (b) forward bias.
Fig. 3.14. Measurements of identical diodes on different dies with \( W=20\mu m \), \( L_i=0.4 \mu m \) and \( L_{sp}=1.25\mu m \): (a) reverse breakdown voltage at \( I_r=0.1\mu A \) (b) forward voltage drop at \( I_f=20\mu A \).}

We also characterize the diode performance for different layout orientations. The default diode orientation for all the previously reported measurements is horizontal with respect to the chip. Measurements of identical diodes having a horizontal and a vertical orientation are shown.
in Fig. 3.16. A systematic effect of orientation is observed across diodes having different \( L_i \). It is noticed that diodes with a vertical orientation have a slightly improved forward characteristics and a lower breakdown voltage. This effect is better depicted by the \( V_{BD} \) and \( V_D \) measurements for different diode orientations shown in Fig. 3.17.

![Graph showing the effect of orientation on diode characteristics](image)

**Fig. 3.16.** Measurements of diodes having different orientations with \( W=20\mu m, \text{ and } L_{sp} = L_i +0.85\mu m \): (a) reverse bias (b) forward bias.

![Graph showing reverse breakdown voltage and forward voltage drop](image)

**Fig. 3.17.** Measurements of diodes having different orientation with \( W=20\mu m, \text{ and } L_{sp} = L_i +0.85\mu m \): (a) reverse breakdown voltage at \( I_r=0.1\mu A \) (b) forward voltage drop at \( I_f=20\mu A \).
Interestingly, we observe that if the diodes are stressed by subjecting them to a large forward current flow (~10-40 mA), their forward I-V characteristics are shifted to the right as shown in Fig. 3.18(a). Shown in Fig. 3.18(b) is the measured diode's $V_D$ before and after stress for diodes with different $L_i$. We notice that diodes with a smaller $L_i$ can handle higher forward current flow before their I-V characteristics shift compared to diodes with a larger $L_i$. The stress currents for diodes with $L_i = 0.4 \mu m$, $0.5 \mu m$ and $0.6 \mu m$ are found to be 40mA, 20mA and 10mA, respectively.

![Fig. 3.18. Measurements of diodes before and after stress with $W=20 \mu m$, and $L_{sb}=L_i+0.85\mu m$: (a) forward bias (b) forward voltage drop at $I_f=20\mu A$.](image)

### 3.3.2. Polysilicon Diode-Based Dickson Charge Pumps

In this section, we report on the measurement results of a number of Dickson charge pumps using polysilicon diodes with different $L_i$ values. All pumps consist of 8 stages and adopt a dual path scheme as shown in Fig. 3.19(a). The diodes used in the pump design have $W=20 \mu m$, and $L_{sb}=L_i+0.85\mu m$. All capacitors are identical with 2x finger spacing similar to the implementation in Fig. 2.6(c) with $C_p/C=8\%$. More efficient capacitors with 1x finger spacing can be used; however, we choose the 2x capacitor implementation to maintain design reliability for when we
cascade multiple charge pumps to attain higher output voltages. Shown in Fig. 3.19(b) are the measured I-V characteristics of a Dickson-type pump using diodes with $L_i=0.4\mu m$, at pumping frequency $f_{\text{pump}}=4\text{MHz}$ and different pumping voltages ($V_{\text{dd}}$). The diode has a maximum output voltage of 17V at $V_{\text{dd}}=2.75V$. The same measurements are repeated at different $f_{\text{pump}}$ values as shown in Fig. 3.20(a). The pump output resistance decreases for higher $f_{\text{pump}}$ as expected of switched capacitor circuits, and is shown in Fig. 3.20(a). In the slow switching limit (SSL) [22], the output resistance exhibits $1/fC$ dependence, whereas, in the fast switching limit (FSL), the output resistance is dominated by the switches series resistance and saturates at 66KΩ.

In general, diode-based pumps exhibit poor power efficiency due to the switch knee voltage ($V_D$). A diode maintains a relatively constant voltage drop ($>V_D$) across its terminals to maintain forward conduction resulting in power dissipation. Shown in Fig. 3.21(a) is the measured pump efficiency versus load current at $f_{\text{pump}}=4\text{MHz}$. We notice that the pump peak efficiency increases with $V_{\text{dd}}$. This is because at higher $V_{\text{dd}}$, a smaller percentage of the pumping voltage is allocated to the switch and the remaining portion is used to build the final output voltage. Shown in Fig. 3.21(b) is the measured peak efficiency versus $V_{\text{dd}}$. A peak efficiency of 46% is measured at $V_{\text{dd}}=2.75V$. Three points are worth noticing here: (1) more efficient pumps are possible using 1x spaced finger capacitors, (2) the improvement in peak efficiency diminishes as $V_{\text{dd}}$ increases and eventually disappears for $V_{\text{dd}}>>V_D$, and (3) for higher $V_{\text{dd}}$, the current at which peak efficiency occurs increases which corresponds to a larger diode drop. Because our devices have a poor diode behavior, this increased drop limits the maximum peak efficiency to even lower values.

In order to improve the efficiency of diode-based pumps, we introduce in the next chapter, an improved-drive Dickson pump that uses a boosted amplitude clock to drive the pump capacitors.
In order to boost the clock amplitude an additional charge pump stage that incurs additional losses is required. We demonstrate in the next chapter that our design still attains overall higher pump efficiency after accounting for the additional power losses.

**Fig. 3.19.** Polysilicon diode-based Dickson charge pump: (a) simplified circuit schematic (b) measured I-V characteristics for diodes with \( L_i = 0.4\mu m \), and \( f_{\text{pump}} = 4\text{MHz} \).

**Fig. 3.20.** Measurement results of a polysilicon diode-based pump with \( L_i = 0.4\mu m \): (a) I-V characteristics at \( V_{dd} = 2.5V \) and different \( f_{\text{pump}} \) (b) output resistance versus \( f_{\text{pump}} \).
Fig. 3.21. Measurement results of a polysilicon diode-based pump with \( L_i = 0.4 \mu m \): (a) power efficiency at \( f_{\text{pump}} = 4 \text{MHz} \) and different \( V_{dd} \) (b) peak efficiency versus \( V_{dd} \) at \( f_{\text{pump}} = 4 \text{MHz} \).

We next show the measurement results of 3 identical pumps using diodes with \( L_i = 0.4 \mu m, 0.5 \mu m, \) and \( 0.6 \mu m \) respectively. All diodes can handle a reverse voltage of at least 10V based on the measurements in Fig. 3.10(a). Shown in Fig. 3.22(a) are the measured I-V characteristics at \( f_{\text{pump}} = 4 \text{MHz} \) and \( V_{dd} = 2.5 \text{V} \) for the 3 pumps. Clearly, diodes with larger \( L_i \) result in lower voltages. The pumps output resistance is plotted in Fig. 3.22(b). In the SSL, all pumps exhibit an identical resistance since they use the same capacitor value. However, in the FSL, pumps with larger \( L_i \) have a higher switch resistance and exhibit an overall higher pump output resistance.

The measured peak efficiencies of all 3 pumps are shown in Fig. 3.23(a). As expected, the pump with \( L_i = 0.4 \mu m \) is the most efficient. The peak efficiencies at \( V_{dd} = 2.75 \text{V} \) are 46%, 42%, and 33% respectively. The pumps maximum output voltages are measured at no load current and \( V_{dd} = 2.5 \text{V} \), and are shown in Fig. 3.23(b). These voltages are 15.6 V, 14.8 V, and 12.7 V respectively. The pumps minimum output resistances are measured in the FSL and are also shown in Fig. 3.23(b). These resistances are 66 K\( \Omega \), 73.2 K\( \Omega \), and 98.2 K\( \Omega \), respectively.
3.4. Substrate Stacking

In case there is a need for >100V outputs in a nanometer scale technology, we can in principle extend our stacking approach to include multiple dies as shown in Fig. 3.24. This
solution is specifically attractive in the case of chip stacked multi-chip modules or 3D-IC packaging technologies. The first die in the stack can be the main system chip with its substrate grounded. This die contains all the low-voltage circuitry and a charge pump circuit capable of a 100V output. The second and third dies have floated substrates and are stacked on top of the first die. These dies should contain high-voltage charge pumps capable of 100V outputs referenced with respect to their substrate potential. The substrate of the second die is biased from the high-voltage output of the first die, and the substrate of the third die is biased by the high-voltage output of the second die in a daisy-chain fashion. All the necessary pumping clock phases for the charge pumps on the second and third dies come off the first die and can be routed via bond wires. In principle, more dies can be stacked in a similar fashion as long as the packaging thickness constraints are met.

![Diagram of substrate bias and pumping clocks](image)

**Fig. 3.24. Stacking CMOS substrates to attain 300V outputs.**

To reiterate, a common theme for enabling high output voltages in a low-voltage technology is stacking. Conceptually, stacking is cumulating the voltage tolerance of multiple structures in
series such that higher voltage ranges are achieved. Stacking in general, applies to individual
devices, different technology layers, and separate dies. First, devices are stacked until we are
limited by the breakdown voltage of a substrate layer. Then, the different substrate layers are
stacked until no more layers are available. And finally, individual dies are stacked until the final
target voltages are met.

3.5. Summary

In this chapter, we demonstrated a deep nwell biasing scheme to enable double-diode
substrate isolation in voltage charge pumps instead of the conventional single-diode isolation
method. As a result, the introduced well-biasing scheme extends the voltage range of bulk charge
pumps in a 65nm CMOS technology from 12V to 20V. In order to enable this method an all-
NMOS charge pump cell implementation is required. A new power efficient all-NMOS charge
pump implementation is introduced in Chapter 4. We have also demonstrated that a polysilicon
diode implemented on top of a deep nwell is either used to increase the voltage tolerance of FOX
or improve its long-term reliability. Measurement results show that using the new polysilicon
diode design, Dickson pumps attain output voltages as high as 100V in a 65nm technology node.
Polysilicon diode-based Dickson pumps suffer from low power efficiency due to the diode's poor
switch characteristics. In Chapter 4, we explain how to improve the efficiency of a Dickson-type
pump by boosting its pumping clock amplitude. Also, in Chapter 5 we explain how a hybrid-type
charge pump enables improved power efficiency at extended voltage ranges compared to a non-
hybrid Dickson pump design. In order to attain even higher voltage ranges, multiple dies can, in
principle, be stacked to achieve voltage ranges that are >100V.
CHAPTER 4

High-Efficiency Voltage Pumping Cells

4.1. Introduction

In Chapter 3, we have introduced a number of technology methods that extend the voltage tolerance of bulk CMOS substrates. Some of these methods require a special pump type implementation, e.g. all-NMOS pump cells and Dickson-type pumps. We have also indicated that a voltage range, power efficiency trade-off exists, i.e. pumps with higher voltage ranges suffer from poorer efficiency, mainly due to the restrictions imposed on the architecture choice and the capacitor layout. In this chapter, we introduce a number of voltage pump cells that go hand-in-hand with the extended-voltage range methods explained in the previous chapter. First, we introduce efficient pump cell designs using complementary-type switches. We show how multiple clock phases (>2) are used to design such pump cells for voltage reliability and power efficiency. Next, we extend our scope to include efficient pumps cells using same-type switches. Next, we introduce an improved-efficiency, diode-based pump design that uses boosted clock swings to improve power and area efficiency. Finally, we explain a reliable and robust CMOS clock generation circuit to provide all the clock phases necessary for our pump designs.

4.2. Complementary-Type Switch Charge Pumps

The first practical integrated charge pump circuit was introduced by Dickson in 1976 [49]. The Dickson pump, however, suffered from 2 drawbacks: (1) a diode drop switch loss of >V_{th}, and (2) a deteriorating switch V_{th} due to back-gate effects. A number of methods since were introduced to alleviate these 2 problems and improve the pump efficiency and voltage range
In effect, to eliminate the switch diode drop, transistors must be operated in a charge-transfer switch (CTS) mode instead of a traditional diode-connected switch mode [51]. To implement reliable charge transfer switches, the transistor gate drive needs to be clock boosted and bootstrapped to the source potential within one $V_{dd}$ voltage difference. To eliminate back-gate effect in modern triple well technologies, it is as simple as tying the transistors bulk terminals to their corresponding source terminal.

![Fig. 4.1. A conventional 2PVD pump cell.](image)

Shown in Fig. 4.1 is a standard two-phase voltage doubler (2PVD) [55]. If $V_{in}=V_{dd}$, then ideally, the maximum steady-state output voltage $V_{out}=2V_{dd}$, and thus the title doubler. The circuit consists of 2 cross-coupled CMOS inverters operating on opposite clock phases, and provides a compact and elegant way to generate the clock boosted signals necessary to drive the transistors into a charge-transfer switch mode. As is the case with logic design, the availability of complementary switches is of convenience and enables design simplicity. NMOS and PMOS devices enable the transmission of low and high voltage levels respectively across pass devices.
with equal efficacy; given that control signals are bounded to within one $V_{dd}$ swing. Also, conveniently, the same gate control signal that turns off NMOS devices can be used to turn on PMOS devices and vice versa, reducing the overall number of gate drive signals.

However, one drawback of the circuit in Fig. 4.1 is reversion losses. It can be shown that regardless of the phase relationship between clocks $\phi_1$ and $\phi_2$, reversion currents always exist [58]. If clocks with overlapping phases are used, devices $M_1$ and $M_2$ are simultaneously on. If the non-overlapping clock phases are used, devices $M_3$ and $M_4$ are simultaneously on. If clock phases with 50% duty cycle are used, devices $M_1$ and $M_3$ ($M_2$ and $M_4$) are simultaneously on, experiencing a shoot-through current loss. The exact impact of reversion currents on the pump efficiency is variable and depends on the overlap (non-overlap) time interval or the clock edges rise and fall times. A number of methods were introduced to resolve reversion power losses in voltage doubler circuits [57]-[59]. Most of these methods, however, insert an extra resistance or a blocking switch in series with some of the pass devices and/or require controlling the slope of the gate control signals. These methods increase the switch effective series resistance and circuit complexity to generate the required clock waveforms.

In the next section, we introduce an efficient and more compact method to alleviate reversion losses in pump cells through the use multiple clock phases. Multiple clock phases are used to guarantee a break-before-make operation for all switch pairs forming a reversion current path.

4.2.1. A CMOS Four-Phase Pump Cell

To enable a break-before-make switch operation, cross-coupled NMOS devices require non-overlapping clock phases, whereas, PMOS devices require overlapping clock phases. Therefore, to alleviate reversion losses from the circuit in Fig. 4.1, we need to decouple the NMOS devices
gate control from the PMOS devices gate control. By separating the PMOS and NMOS gate drive signals, a proper clock phase relationship can be provided for each device pair separately to eliminate its reversion currents. Moreover, to eliminate shoot-through currents through devices $M_1$ and $M_3$ ($M_2$ and $M_4$), the relationship between the overlap period ($t_{ov}$) and the non-overlap period ($t_{nov}$) is chosen such that these devices never conduct simultaneously.

Shown in Fig. 4.2(a) is a four-phase voltage doubler (4PVD) that eliminates reversion losses using phases $\phi_1$-$\phi_4$. Phases $\phi_1$ and $\phi_2$ drive the NMOS gates with non-overlap period $t_{nov}$, whereas phases $\phi_3$ and $\phi_4$ drive the PMOS gates with overlap period $t_{ov}$, such that $t_{nov} < t_{ov}$ as shown in the timing diagram. Even though 2 extra capacitors $C_3$ and $C_4$ are needed, they are chosen much smaller than $C_1$ and $C_2$ since they do not contribute to the output load current.

To explain the circuit's operation, we assume the cell's input voltage is equal to $V_{dd}$. At steady state, capacitors $C_1$ and $C_2$ are charged with $V_{dd}$ across their terminals. Focusing first on the input pair $M_1$ and $M_2$, as $\phi_2$ transitions from low to high, $C_2$ top plate potential transitions from $V_{dd}$ to $2V_{dd}$ and device $M_1$ has a sufficient gate overdrive to charge $C_1$ fully to $V_{dd}$ as presumed and like-wise for capacitor $C_2$. Moreover, $\phi_2$ transitions back low to turn $M_1$ off before $\phi_1$ transitions high and turns $M_2$ on, preventing reversion current losses through $M_1$. Now focusing on the output pair $M_3$ and $M_4$, since at steady state the doubler's output voltage is $2V_{dd}$, capacitors $C_3$ and $C_4$ top plate potentials transition between $V_{dd}$ and $2V_{dd}$. First, $\phi_1$ transitions high, then $\phi_3$ transitions low. Hence, $C_1$ top plate potential is at $2V_{dd}$ while $C_3$ top plate potential is at $V_{dd}$, and device $M_3$ has a sufficient gate overdrive to charge the output capacitor fully to $2V_{dd}$ as presumed. Again, $\phi_3$ transitions back high to turn $M_3$ off before $\phi_1$ transitions low, preventing reversion currents through $M_3$. 

57
Fig. 4.2. Proposed 4PVD pump cells: (a) type I (b) type II.

By flipping the input and output ports and replacing all the NMOS devices with PMOS devices and vice versa, the dual 4PVD implementation is obtained as shown in Fig. 4.2(b). In the dual implementation, phases $\phi_1$ & $\phi_2$ are overlapping while phases $\phi_3$ & $\phi_4$ are non-overlapping.
such that \( t_{ov} < t_{nov} \). Since capacitors \( C_3 \) and \( C_4 \) in the dual implementation are charged directly from \( V_{in} \), an improved NMOS gate drive is possible regardless of the \( C_1 \) and \( C_2 \) charge level. This leads to lower voltage droops at high load currents compared to the design in Fig. 4.2(a).

### 4.2.2. A CMOS Six-Phase Pump Cell

In principle, we can further decouple the gate drive for all the switches from the pumping capacitors drive. This requires an extra pair of clock phases and capacitors as shown in the six-phase voltage doubler (6PVD) cell in Fig. 4.3. Phases \( \phi_3 \) and \( \phi_4 \) have a non-overlap period \( t_{nov} \), whereas phases \( \phi_5 \) & \( \phi_6 \) have an overlap period \( t_{ov} \). Phases \( \phi_1 \) and \( \phi_2 \) can either be overlapping or non-overlapping as long as the rising edge of \( \phi_1 \) falls between two successive falling edges of \( \phi_3 \) and \( \phi_5 \); and vice versa for the rising edge. For low load current scenarios, a 4PVD could be more power efficient than the 6PVD because it uses fewer capacitors, resulting in a smaller power overhead to generate the auxiliary clock phases.

![Proposed 6PVD pump cell](image)

Fig. 4.3. Proposed 6PVD pump cell.
4.3. **Same-Type Switch Charge Pumps**

In Chapter 3, we have shown that an all-NMOS charge pump cell is needed to extend the voltage range of CMOS substrates. More generally, same-type switch pump cells, i.e. all-NMOS or all-PMOS, find a wide range of applications.

One useful application of same-type switch pump designs arises when using older twin-well technologies to generate positive output voltages. In a twin-well technology, all NMOS devices share a common grounded bulk terminal and suffer from back-gate effects. A gradually increasing switch $V_{th}$ with increasing number of stages, not only reduces the pump efficiency, but also limits the maximum voltage range as $V_{th}$ exceeds $V_{dd}$. Consequently, an all-PMOS voltage doubler solution is more power efficient than CMOS doublers in twin-well technologies. The exact mirror scenario arises if PMOS devices are used to generate negative output voltages. In order to prevent forward biased junctions, the bulk terminal of PMOS devices needs to be grounded to the CMOS substrate. This again results in back-gate effects for the PMOS devices, limiting the pump's voltage range and power efficiency. Consequently, an all-NMOS voltage doubler solution is more power efficient than a CMOS doubler in a triple-well technology. A deep nwell option is necessary because bulk NMOS devices can not be used to generate negative output voltages. A third application is extending the substrate voltage tolerance to double-diode isolation, as previously explained.

4.3.1. **An All-NMOS Four-Phase Pump Cell**

In [48], we have introduced one possible implementation of an all-NMOS voltage doubler, as shown in Fig. 4.4(a). This implementation is derived from the CMOS case by substituting the cross-coupled PMOS device pair with a diode-connected NMOS device pair. The same approach
is applicable to all-PMOS cells as well, and requires no extra clock phases or design complexity. However, diode-connected devices have poor switch characteristics due to their $V_{th}$ voltage drop and result in lower efficiencies. This $V_{th}$ problem is even more aggravated when the NMOS bulk and source terminals are not shorted together as is the case with double-diode substrate isolation.

![Fig. 4.4. Proposed same-type switch 4PVD pump cells: (a) all-NMOS (b) all-PMOS.](image-url)
In order to attain efficiencies comparable to those of complementary-type switch cells, we need to replace the diode-connected devices with charge-transfer devices. Assuming $V_{in} = V_{dd}$, in order for $M_3$ and $M_4$ to turn on and fully charge the output capacitor to $2V_{dd}$, their gate potentials need to rise to $3V_{dd}$. Whereas, to turn $M_3$ and $M_4$ off, their gate potentials need to fall all the way to $V_{dd}$ as $C_1$ and $C_2$ top plate potentials fall to $V_{dd}$. Therefore, in order for $M_3$ and $M_4$ to operate in a charge-transfer mode, their gate drive needs to swing between $V_{dd}$ and $3V_{dd}$. However, the direct application of a $2V_{dd}$ signal swing to the transistor gates causes gate-oxide reliability issues. In the next section, we propose a circuit that maintains device reliability by performing the transition from $V_{dd}$ to $3V_{dd}$ on 2 separate transitions, each with one $V_{dd}$ step.

4.3.2. An All-NMOS Six-Phase Pump Cell

The all-NMOS 6PVD shown in Fig. 4.5 is proposed to eliminate the diode drop in a same-type switch 4PVD cell while maintaining its device reliability. As shown in figure, 4 additional transistors ($M_5$-$M_8$), 2 additional capacitors ($C_5$ and $C_6$), and 2 additional clock phases ($\phi_5$ and $\phi_6$) are needed to drive the gates of $M_3$ and $M_4$ reliably. Capacitors $C_3$-$C_6$ can are chosen much smaller in size than the pumping capacitors $C_1$ and $C_2$ since they do not carry any load current. A careful timing relationship between all 6 phases needs to be maintained to eliminate all reversion losses and guarantee that none of the devices are stressed. Phases $\phi_1$ and $\phi_2$, and phases $\phi_3$ and $\phi_4$ have non-overlap periods $t_{nov1}$, and $t_{nov2}$, respectively, such that $t_{nov1} < t_{nov2}$. Also, phases $\phi_5$ and $\phi_6$ are narrow non-overlapping pulses such that the $\phi_5$ pulse falls precisely in between 2 successive rising edges of $\phi_1$ and $\phi_3$. Similarly, the $\phi_6$ pulse falls precisely in between 2 successive rising edges of $\phi_2$ and $\phi_4$. The $\phi_5$ and $\phi_6$ pulses are not required at the rising edges. However, for clock generation purposes, it is simpler to generate the $\phi_5$ and $\phi_6$ pulses for both the rising and falling edges of $\phi_1$ and $\phi_3$, and $\phi_2$ and $\phi_4$, respectively.
To explain the circuit's operation, we assume again that $V_{in}=V_{dd}$. At steady state, capacitors $C_1$ and $C_2$ are charged with $V_{dd}$ across their terminals, and the top plate potential transitions between $V_{dd}$ and $2V_{dd}$. Also, since the doubler's steady state output voltage is $2V_{dd}$, capacitors $C_5$ and $C_6$ top plate potentials transition between $2V_{dd}$ and $3V_{dd}$. As previously mentioned, the $M_3$ gate potential makes the transition to $3V_{dd}$ on 2 controlled steps: (1) a transition to $2V_{dd}$ on the $\phi_5$ rising edge, (2) a transition to $3V_{dd}$ on the $\phi_3$ rising edge. Similarly, the $M_3$ gate potential makes the transition from $3V_{dd}$ on 2 controlled steps: (1) a transition to $2V_{dd}$ on the $\phi_3$ falling edge, (2) a transition to $V_{dd}$ on the $\phi_1$ falling edge. To explain these transitions in more detail, we follow this switching sequence. First, $\phi_1$ transitions high and the $C_1$ top plate potential transitions to $2V_{dd}$, charging capacitor $C_3$. However, in order to fully charge $C_3$ to $2V_{dd}$, the $M_5$ gate potential needs to be one $V_{dd}$ higher. Therefore, $\phi_5$ next transitions high such that the $C_5$ top plate potential is at $3V_{dd}$. Now $M_5$ has enough overdrive voltage to charge $C_3$ to $2V_{dd}$. Next, $\phi_5$ transitions back low.
to prevent reverse current conduction from C₃ into C₁ through M₅ as φ₃ transitions high. Next, φ₃ transitions high, and C₃ top plate potential transitions to 3V₉d to fully charge the output capacitor to 2V₉d through M₃ as presumed. Next, φ₃ transitions back low and C₃ top plate potential discharges to 2V₉d to maintain M₃ reliability as φ₁ transitions low. Finally, φ₁ transitions low and the C₁ top plate potential falls to V₉d. As a result C₃ discharges back into C₁ through M₅ and the C₃ top plate potential falls to V₉d turning M₃ off. Because of the controlled voltage transitions, this switching order guarantees that none of the devices terminals are stressed.

Since at every clock cycle a fixed amount of charge C₃V₉d is transferred from the supply onto C₃ and discharged back to ground, the pump efficiency improves for smaller C₃. Also, a smaller C₃ value allows the M₃ gate potential follows more closely its source potential as φ₁ transitions from high to low. This guarantees that M₃ remains turned off during this entire duration and no reversion current flows from the output node onto C₁. All the device bulk terminals are carefully tied to their corresponding nodes with lower potential to provide full device isolation.

Shown in Fig. 4.6 are various flavors of same-type switch 6PVD designs targeting different applications. The all-PMOS cell in Fig. 4.6(a) is for positive voltage generation in technologies without a deep nwell option. All the PMOS devices have their bulks tied to their source terminals, and suffer from no back-gate effects. The all-NMOS cell in Fig. 4.6(b) is for negative voltage generation in triple-well technologies. All the NMOS devices have their bulks tied to their source terminals, and suffer from no back-gate effects. Finally, shown in Fig. 4.6(c) is an extended range all-NMOS positive voltage pump cell. Note that all the devices have their bulks connected to a common bias V₉mid, and suffer from back-gate effects. Shown in Fig. 4.7 is the simulated V₉th of an NMOS device with W/L=1μm/0.25μm versus its bulk-source potential (V₉bs).
Fig. 4.6. Proposed same-type switch 6PVD pump cells: (a) positive output voltage all-PMOS (b) negative output voltage all-NMOS (c) extended-range positive output voltage all-NMOS.

Fig. 4.7. Simulation results of an NMOS device $V_{th}$ with $W/L=1.4\mu m/0.5\mu m$ versus $V_{sb}$.

One critical point that should not be dismissed when designing all-NMOS pumps with extended voltage range, is the maximum gate potential of the NMOS devices. This too is limited
to \(V_{DNW} + V_{NP}\). Therefore, to provide enough overdrive voltage to turn the pass devices on, we need, by design, to back off the maximum attainable output voltage by some margin (~0.5V). If either \(V_{in}\) or \(V_{dd}\) is increased until source potential of the final NMOS device connected to the output node, is equal to the gate potential, the device is turned off. When the device turns off, no current flows to the load, and the output node discharges at a faster than the device's gate and drain potentials, subjecting the transistor to stress. Measurement results show degradation in the all-NMOS pump output voltage, whenever it is increased to \(V_{DNW} + V_{NP}\).

4.3.3. An All-NMOS Eight-Phase Pump Cell

Lastly, for the sake of completeness, we may choose to fully decouple the gate drive of all the switches from the pumping capacitors for an improved voltage droop at high load currents. Shown in Fig. 4.8(a) is an all-NMOS eight-phase voltage doubler (8PVD) requiring 2 additional transistors and 2 additional clock phases. Phases \(\phi_3\) and \(\phi_4\), and phases \(\phi_5\) and \(\phi_6\) have non-overlap periods \(t_{nov1}\) and \(t_{nov2}\) respectively. Phases \(\phi_1\) and \(\phi_2\) can either be overlapping or non-overlapping as long as the rising edge of \(\phi_1\) falls between 2 successive falling edges of \(\phi_3\) and \(\phi_5\), and the falling edge of \(\phi_1\) falls between 2 successive rising edges of \(\phi_3\) and \(\phi_5\). Moreover, phases \(\phi_7\) and \(\phi_8\) are narrow non-overlapping pulses such that \(\phi_7\) pulses fall precisely between 2 successive rising (falling) edges of \(\phi_1\) and \(\phi_5\). Similarly, \(\phi_8\) pulses fall precisely between 2 successive rising (falling) edges of \(\phi_2\) and \(\phi_6\). The corresponding all-PMOS 8PVD cell is shown in Fig. 4.8(b), where all the NMOS devices are replaced with PMOS devices and all the non-overlapping clock phases are replaced with overlapping ones. For low load current scenarios, the 6PVD in Fig. 4.6(a) could be more power efficient than the 8PVD in Fig. 4.8(b) because it uses fewer capacitors, resulting in a smaller power overhead to generate the auxiliary clock phases.
Fig. 4.8. Proposed same-type switch 8PVD pump cells: (a) all-NMOS (b) all-PMOS.
4.4. Improved-Drive Dickson Charge Pump

In Chapter 3, we have established the need for diode-based Dickson-type pumps to leverage the voltage isolation of FOX. We have also indicated, from measurement results, that the efficiency of diode-based pumps improves with higher pumping voltages. In this section, we introduce an improved-drive Dickson charge pump with $2V_{dd}$ clock swings. One possible way to implement a reliable clock buffer with $2V_{dd}$ output swing is based on [60] and shown in Fig. 4.9. A voltage doubler circuit is needed to provide a boosted supply voltage for higher clock swings. Cascode devices are used to limit the voltage difference across the output devices to within $V_{dd}$. To minimize the system complexity, the same output drive clocks are used in the voltage doubler as well. However, capacitors $C_1$ and $C_2$ must be chosen large enough to provide the necessary currents for all the Dickson pump stages. Capacitors $C_3$ and $C_4$, however, are used for level shifting to maintain the reliability of $M_7$ and $M_8$, and can be much smaller in size.

![Fig. 4.9. Cascode clock buffers with 2$V_{dd}$ output swing based on [60].](image-url)
We notice that for the circuit in Fig. 4.9, the potentials of nodes n_1 and n_3 (n_2 and n_4) are identical. Therefore, we can use C_1 and C_2 to perform both the voltage pumping and the level-shifting, in which case devices M_3-M_6 can be eliminated. Moreover, we notice that the voltage swings at the source of M_9 and M_10 are identical to the swings at nodes n_3 and n_4. Similarly, the voltage swings at the source of M_{13} and M_{14} are identical to the swings at nodes n_5 and n_6. Therefore, devices M_7-M_8 and M_{13}-M_{14} can also be eliminated.

We introduce in Fig. 4.10(a) a more compact and power efficient implementation of a doubly-stacked clock driver with 2V_{dd} swing. The clock driver maintains the reliability of all devices. Devices M_1 and M_2 together with capacitors C_1 and C_2 level-shift the input clocks \( \phi_{1_{nov}} \) & \( \phi_{2_{nov}} \) by V_{dd}. Devices M_3-M_6 share a fixed gate potential (V_{dd}) and operate as source switched output buffers. As a result, the output nodes \( \phi_1 \) and \( \phi_2 \) swing at 2V_{dd} without stressing any of the devices. Capacitors C_1 and C_2 are much larger than the Dickson pump capacitors to provide sufficient charge. Since capacitors C_1 and C_2 are subjected to a voltage difference of V_{dd} only during operation, they are implemented using area efficient MOS capacitors.

Now, that the pump operates off higher clock swings, a fewer number of stages can be used to attain the same voltage levels of a conventional Dickson pump. Each stage consists of a pair of diodes and capacitors operating on opposite clock phases. The sub-pump is laid out in a fashion identical to the schematic depicted in Fig. 4.10(b). This avoids routing both clock phases to the top and bottom sides of the pump and hence reduces the total parasitic capacitance associated with the clock net by \( \sim 50\% \). Because of the higher clock swings, the diodes are exposed to a reverse voltage stress <4V_{dd}. The diodes used in this design have \( L_i=0.5\mu m \) and can thus handle supply voltage as high as 2.5V easily, based on the measurement results in Fig. 3.10(a).
4.5. Clock Generation Circuit

So far, we have introduced a number of efficient voltage pump cells that use multiple clock phases, e.g. 4, 6 and 8 phases. In order to alleviate reversion power losses and maintain device reliability, these phases need to preserve a fixed timing relationship between one another. In this section, we introduce a CMOS clock generation circuit that provides the various clock phases needed by the proposed pump cells. The circuit is designed such that the relationship between these clock phases stays relatively insensitive to process, voltage and temperature variations. Moreover, the delay elements of the clock generation circuit are voltage controlled so that the phases overlap (non-overlap) periods can be adjusted as necessary.

Before explaining the proposed clock generation circuit, we make the following observations on voltage doubler circuits: 1) cross-coupled PMOS devices require overlapping clock phases for their gate drive, 2) cross-coupled NMOS devices require non-overlapping clock phases for their
gate drive, and 3) clock phases not driving transistor gates, e.g. φ₁ and φ₂ in the cell shown in Fig. 4.3 can either be overlapping or non-overlapping as long as they satisfy other timing criteria. Next, we make the following observations regarding clock generation circuits: (1) Overlapping clock phases can be turned into non-overlapping phases with t_{nov} = t_{ov} using a simple logic inversion and vice versa, assuming that the inverters’ rise and fall delays are symmetric. (2) If the inverter delays are not symmetric, yet they are much smaller than t_{ov} or t_{nov}, it can still be assumed that t_{nov} ≈ t_{ov}, and moreover, the input and output waveforms are considered to have the mid-point of their overlap/non-overlap intervals coincide in time.

![Clock Generation Circuits Diagram](image)

Fig. 4.11. Conventional CMOS clock generation circuits using: (a) NAND gates (b) NOR gates.
One common and reliable method for generating overlapping/non-overlapping clock phases is using cross-coupled NAND/NOR gates with delay elements placed in the feedback loop as shown in Fig. 4.11. Depending on the type of logic gates used and the amount of delay in the feedback loop, overlapping (non-overlapping) clock phases can be generated with a specified $t_{ov}$ ($t_{nov}$). Assuming the delay of one delay element is $t_d$ and the logic gates delay is negligible ($<<t_d$), for a total number $2n$ delay elements inserted in the feedback loop, the final output $t_{ov}$ ($t_{nov}$) = $nt_d$.

For cases in which clock phase pairs with different $t_{ov}$ ($t_{nov}$) are required, multiple clock generation circuits having different feedback delays are needed. However, the time interval delay measured between the circuit's input and output clock edges depends on the loop delay. For example, it can be shown from the timing diagram in Fig. 4.11 that the time delay between an input clock edge and the midpoint of the corresponding overlap/non-overlap time interval at the output is $1.5nt_d$. Thus, clock phase pairs derived from the same input clock edge and having different $t_{ov}$ ($t_{nov}$) are time skewed with respect to each other. To equalize this difference in circuit delay, extra delay elements are inserted before the clock generation circuit using fewer delay elements in the feedback loop. Assuming that 2 circuits with $2n_1$ and $2n_2$ delay elements in the feedback loop exist, such that $n_1>n_2$, the number of extra delay elements needed before the circuit with fewer delays = $1.5(n_1-n_2)$. Shown in Fig. 4.12 is an example of a clock generation circuit producing 8 clock phases with 2 different $t_{ov}$ ($t_{nov}$) all centered at the same time instance. The clock phases are such that pairs $\varphi_{ov1}-\varphi_{ov2}$ ($\varphi_{nov1}-\varphi_{nov2}$) have $t_{ov}$ ($t_{nov}$) = $2t_d$, whereas pairs $\varphi_{ov3}-\varphi_{ov4}$ ($\varphi_{nov3}-\varphi_{nov4}$) have $t_{ov}$ ($t_{nov}$) = $4t_d$. Three extra delay elements are inserted before the lower clock generation circuit to align all phases. This circuit provides all the necessary phases needed by the 4PVD circuits previously explained. In order to generate the extra clock phases with narrow pulse width needed by the 6PVD and 8PVD circuits, more logic blocks are needed.
Shown in Fig. 4.13 is the full implementation of a clock generation circuit providing all the clock phases needed by the voltage doubler circuits proposed in this chapter. The clock phase pair $\varphi_1$-$\varphi_2$ is generated with $t_{ov} (t_{nov}) = t_d$, whereas the clock phase pair $\varphi_3$-$\varphi_4$ is generated with $t_{ov} (t_{nov}) = 7t_d$. In order to align the $\varphi_1$-$\varphi_2$ with the $\varphi_3$-$\varphi_4$ phases in time; a total delay of $9t_d$ is inserted before the $\varphi_1$-$\varphi_2$ clock generation circuit. To generate the narrow pulse width $\varphi_5$-$\varphi_6$ phases, first, we generate the intermediate clock signals $\varphi_a$-$\varphi_b$ having a pulse width $= t_d$. The intermediate $\varphi_a$-$\varphi_b$ phases are generated via an AND gate having 2 overlapping phases with $t_{ov}=t_d$ applied to its inputs. The $\varphi_a$ pulses need to occur before the $\varphi_1$-$\varphi_2$ overlap/non-overlap intervals, whereas the $\varphi_b$ pulses need to occur after the $\varphi_1$-$\varphi_2$ overlap/non-overlap intervals; and proper delay elements are inserted accordingly to space out the $\varphi_a$ and $\varphi_b$ pulses by $4t_d$. Finally, phase $\varphi_5$ is generated by selecting the $\varphi_a$ and $\varphi_b$ pulses only that occur when $\varphi_1$ is high. Similarly, phase $\varphi_6$
is generated by selecting the $\phi_a$ and $\phi_b$ pulses only that occur when $\phi_2$ is high. The corresponding timing diagram and phase relationship is depicted in Fig. 4.14.

Fig. 4.13. Proposed PVT insensitive CMOS clock generation circuit.

To maintain careful clock phase alignment and guarantee that delayed clock inputs preserve a 50% duty cycles, delay elements with symmetric rise and fall delays must be used. In order to do this, a non-inverting delay element consisting of two inverting delay stages is used. By doing so, each clock transition experiences one rise time delay ($t_r$) plus one fall time ($t_f$) delay irrespective of the transition type. The total cell delay now is independent of the input clock transition and is always equal to $t_r + t_f$. The circuit implementation of the delay element is shown in Fig. 4.15. The delay of the cell is voltage controlled by starving the bias current of its inverters.
Fig. 4.14. Timing diagram of the different clock phases in the clock generation circuit: (a) non-overlapping clock phases (b) overlapping clock phases.

Fig. 4.15. Voltage programmable delay element with symmetric rise and fall time delays.
All delay elements must be matched and placed within close proximity to guarantee that the clock phases maintain a fixed timing relationship insensitive to process and external variations. Moreover, all delay elements are loaded with dummy inverters (not shown here) and logic gates are sized such that all delay elements see an equal load. For an 8PVD with $t_{ov1}=t_{ov2}$ ($t_{nov1}=t_{nov2}$), the circuit in Fig. 4.13 suffices. However, if $t_{ov1}$ and $t_{ov2}$ are not chosen equal, an additional cross-coupled NAND gate loop is needed and additional path delay equalization delay elements.

Since phases $\phi_1$ and $\phi_2$ always drive much larger pumping capacitors $C_1$ and $C_2$ with large plate parasitic capacitance, we choose to drive these phases through tri-state buffers and a charge equalization switch. This allows for the recycling of half the charges stored on one parasitic capacitor to be reused in charging the second capacitor in a ping-pong fashion [61]. Consequently, this cuts down the power dissipated in charging these parasitic capacitors by a factor of $\sim 2$, not accounting for the extra power dissipation in the tri-state logic gates.

![Figure 4.16. Charge recycling tri-state clock driver based on [61].](image)
4.6. Summary

In this chapter, we have introduced a number of power-efficient charge pump cells targeting different output voltage ranges. Power efficient cell designs include: (1) complementary-type switch cells that alleviate reversion current losses using 4/6 clock phases, (2) same-type switch cells that alleviate diode drops using 6/8 clock phases, and (3) improved-drive Dickson-type cells using doubly-stacked clock drivers. Moreover, we have demonstrated that same-type switch designs, in addition to enabling double-diode substrate isolation, can be used to enhance the efficiency of pump cells implemented in older twin-well technologies, or those targeting negative output voltages. Also, we have introduced a robust CMOS clock generation circuit to provide the clock phases required by the different pump designs. The clock generation circuit preserves the timing relationship between the different phases across PVT variations.

So far, we have shown that for our pump cell designs, an inherent power-efficiency, voltage-range trade-off exists. In other words, extended voltage range pump cells mandate circuit architectures that are inherently less power efficient. In order to relax this trade-off and achieve the combined goals of extended voltage range and improved power efficiency, we introduce in the next chapter the Hybrid Charge Pump architecture.
CHAPTER 5

Hybrid Charge Pumps

5.1. Introduction

In this chapter, we introduce a new class of charge pump circuits, namely, the Hybrid Charge Pump. Hybrid Charge Pumps employ higher voltage-tolerant cells, only incrementally, to extend the output voltage range of more efficient pump cells suffering from limited ranges. First, we start by explaining the architecture concept and structure. Next, we propose an accurate power analysis model to explore the design space and estimate the efficiency of Hybrid Charge Pumps. Based on this model, the optimal numbers of stages and pumping capacitor values of the individual pump cells can be chosen for best efficiency. Also, we propose a noise analysis model to define the contribution of different noise sources on the final output noise power. Finally, we explain 2 Hybrid Charge Pump design examples, one targeting positive output voltages and the other targeting negative output voltages.

5.2. Hybrid Charge Pump Architecture

We have shown in Chapter 3 that depending on the substrate isolation method, charge pumps with different voltage ranges can be conceived. We have also shown that higher voltage-tolerant pumps are less power efficient, mainly due to the circuit architecture imposed by the choice of isolation method. For applications with extended voltage range, rather than using a low-efficiency, diode-based Dickson pump, we propose here a Hybrid Charge Pump architecture. A Hybrid Charge Pump achieves extended voltage ranges at improved efficiency by optimally mixing high efficiency, low voltage-tolerant cells with lower efficiency, high voltage-tolerant
one. Through combining cells having different properties, the final design captures the benefits of its constituent cells.

The proposed charge pump architecture is shown in Fig. 5.1. In the most general case, it is composed of a cascade of 3 smaller sub-pumps. The first sub-pump is a limited voltage range, high efficiency CMOS charge pump. It consists of m stages and uses pumping capacitor value \( C_1 \). The second sub-pump is a medium voltage range, medium efficiency all-NMOS in deep nwell charge pump. It consists of n stages and uses pumping capacitor value \( C_2 \). The third sub-pump is a extended voltage range, low efficiency polysilicon diode Dickson charge pump. It consists of p stages and uses pumping capacitor value \( C_3 \). Various possible implementations of such charge pump types have been proposed in Chapter 4. In principle, this architecture is valid for both positive and negative output pumps. For positive-type pumps, we prefer to tie the pump's input voltage to \( V_{dd} \) for best efficiency. For negative-type pumps with positive \( V_{dd} \), we ground the pump's input. Since, the pump's output voltage is maximum at zero load current, upper bounds on the number of sub-pump stages, assuming positive-type pumps, are given in terms of the different process parameters as shown in Table 5.1.

\[
\begin{align*}
\text{Sub-pump I} & \quad (m \text{ stages, } C_1) \\
\text{Range: } V_{\text{max}1} & \quad \text{Efficiency: } \eta_1 \\
\text{Sub-pump II} & \quad (n \text{ stages, } C_2) \\
\text{Range: } V_{\text{max}2} & \quad \text{Efficiency: } \eta_2 \\
\text{Sub-pump III} & \quad (p \text{ stages, } C_3) \\
\text{Range: } V_{\text{max}3} & \quad \text{Efficiency: } \eta_3
\end{align*}
\]

\[V_{\text{max}1} < V_{\text{max}} < V_{\text{max}1} \]
\[\eta_3 < \eta_2 < \eta_1\]

Fig. 5.1. Charge pump block diagram representation with (a) arbitrary input (\( V_{in} \)) (b) grounded input.
Table 5.1. Expressions for the maximum number of stages for each sub-pump at no-load condition

<table>
<thead>
<tr>
<th>Sub-pump Type</th>
<th>Number of stages</th>
<th>Parameters Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk CMOS Voltage Doubler</td>
<td>( m \leq \left( \frac{V_{DNW}}{V_{dd}} - 1 \right) \left( 1 + \frac{C_p}{C} \right) )</td>
<td>( V_{out} = \text{Charge Pump Output Voltage} ) ( V_d = \text{Diode Knee Voltage} )</td>
</tr>
<tr>
<td>DNW All-NMOS Voltage Doubler</td>
<td>( m \leq \left( \frac{V_{PW}}{V_{dd}} \right) \left( 1 + \frac{C_p}{C} \right) )</td>
<td>( V_{dd} = \text{Supply Voltage} ) ( V_{DNW} = \text{DNW/PSUB Breakdown Voltage} )</td>
</tr>
<tr>
<td>Polysilicon Diode Dickson pump</td>
<td>( p \leq \left( \frac{V_{out} - V_{DNW} - V_{PW} + V_D}{C} \right) \left( \frac{C}{C + C_p} \right) V_{dd} - V_D )</td>
<td>( V_{PW} = \text{N+/PWELL Breakdown Voltage} ) ( V_{FOX} = \text{FOX Breakdown Voltage} ) ( C_p = \text{Parasitic Plate Capacitance} )</td>
</tr>
</tbody>
</table>

More generally, a charge pump design is required to simultaneously meet certain output voltage (\( V_{out} \)) and load current (\( I_{out} \)) requirements. For a single sub-pump design, there exists an infinite number of ways by which we can pick the number of stages (\( n \)) and pumping capacitor value (\( C \)) to meet the same (\( V_{out}, I_{out} \)) constraint. In other words, a large \( n \) value can be coupled with a small \( C \) value and vice-versa. Moreover, for a Hybrid Charge Pump design, the design space is even larger as we need to pick the number of stages and pumping capacitor values for each of the individual sub-pumps to meet the final (\( V_{out}, I_{out} \)) requirement. To better understand the impact of the different design parameters on the pump efficiency and explore its design space, we propose the following accurate power analysis model.

### 5.3. Power Analysis Model

In many battery-powered applications such as wearable devices, biomedical implants, and hand-held electronics, power efficiency is crucial. To better understand and optimize our hybrid pump efficiency, we propose a mathematical model to calculate the hybrid pump efficiency in
terms of its constituent pump efficiencies. First, we assume an idealized pump model, to grasp the weighted average essence of our model. Then, we propose a more realistic model that incorporates the main losses mechanisms for different pump cell types.

![Charge Pump Block Diagram](image)

**Fig. 5.2.** Charge pump block diagram representation with: (a) grounded input (b) arbitrary input $V_{\text{in}}$.

Shown in Fig. 5.2 is an abstract representation of an $n$ stage charge pump circuit having 2 inputs $V_{\text{dd}}$ and $V_{\text{in}}$; where $V_{\text{dd}}$ is the pumping voltage and $V_{\text{in}}$ is an arbitrary input voltage which may or may not be equal to $V_{\text{dd}}$. The dc currents flowing through the $V_{\text{in}}$ and $V_{\text{dd}}$ ports are $I_{\text{in}}$ and $I_{\text{dd}}$ respectively. The pump's output voltage is $V_{\text{out}}$ and corresponding load current is $I_{\text{out}}$. Typically, for positive-type charge pumps, $V_{\text{in}}=V_{\text{dd}}$ and hence, for efficiency calculations, pumps are regarded as having a single input, $V_{\text{dd}}$. This single input representation is lacking here, since for hybrid charge pumps, the second and third sub-pumps are connected to arbitrary input voltages. Moreover, for negative-type pumps, $V_{\text{in}}=0$.

Generally speaking, efficiency is defined as the ratio of the load power to the supply power. Effectively, a charge pump with $V_{\text{in}}=V_{\text{dd}}$ has a higher efficiency than a pump with $V_{\text{in}}=0$, since the pump's input voltage provides an additional voltage boost to the final output voltage at no power loss, when compared to adding more pumping cells. Assuming that $V_{\text{in}}$ is supplied by an ideal voltage source, the higher the voltage applied to the pump input, the higher the efficiency.
In order for power efficiency to be a useful metric, we need to decouple this \( V_{in} \) dependency such that efficiency is an intrinsic pump property that reflects the circuit implementation and technology parameters. Thus, we define \( \eta_0 \), the pump efficiency at \( V_{in}=0 \) as follows,

\[
\eta_0 = \frac{P_{out}}{P_{in} \mid V_{in}=0}
\]

In our next analysis, we express all pump efficiencies in terms of this intrinsic efficiency \( \eta_0 \).

5.3.1. Idealized Power Model

For simplicity, we start first with an ideal charge pump scenario. In an ideal pump, all switches and capacitors are perfect, and charge redistribution losses are the only mechanism contributing to efficiency losses. To calculate the efficiency of an \( n \) stage pump with arbitrary \( V_{in} \), we notice that, at steady state and assuming voltage doublers or Dickson-type pumps.

\[
I_{in} = I_{out}
\]

(5.2)

\[
I_{dd} = nI_{out}
\]

(5.3)

This is valid for both cases in Fig. 5.2(a) and 5.2(b). And, \( \eta_0 \) can be expressed as follows,

\[
\eta_0 = \frac{V_{out}I_{out}}{V_{dd}I_{dd} \mid V_{in}=0} = \frac{I_{out}V_{out} \mid V_{in}=0}{nI_{out}V_{dd} \mid V_{in}=0} = \frac{V_{out} \mid V_{in}=0}{nV_{dd}}
\]

(5.4)

Since losses due to charge redistribution between capacitors can be modeled as an equivalent series resistance, the output voltage \( V_{out} \leq nV_{dd} \) and \( \eta < 1 \), for non-zero load currents. Moreover, \( V_{out} \) for any arbitrary \( V_{in} \) can be related to \( V_{out} \) when \( V_{in}=0 \) as follows,

\[
V_{out} = V_{in} + V_{out} \mid V_{in}=0
\]

(5.5)

If we define \( \eta \) to be the efficiency for an arbitrary \( V_{in} \), then we find that
\[ \eta = \frac{V_{\text{out}} I_{\text{out}}}{V_{\text{in}} I_{\text{out}} + V_{dd} I_{dd}} = \frac{V_{\text{in}} + V_{\text{out}} I_{\text{out}}}{V_{\text{in}} + n V_{dd}} \]  

(5.6)

By substituting from Equation (5.4) into Equations (5.6), \( \eta \) can be expressed in terms of the intrinsic efficiency \( \eta_o \) as follows,

\[ \eta = \frac{V_{\text{in}}/V_{dd} + n \eta_o}{V_{\text{in}}/V_{dd} + n} \]  

(5.7)

Considering special cases we find the following: for \( V_{\text{in}}=0 \), \( \eta = \eta_o \) as expected, for \( V_{\text{in}} \gg V_{dd} \), \( \eta \approx 1 \) for a fixed number of stages, and finally for \( V_{\text{in}} = V_{dd} \), we find that

\[ \eta = \frac{1 + n \eta_o}{1 + n} \]  

(5.8)

Equation (5.8) shows that as \( n \to \infty \), \( \eta \to \eta_o \). Shown in Fig. 5.3 is the plot of Equation (5.8) assuming \( \eta_o = 0.6 \) versus the number of stages. Once again, we see the charge pump efficiency dropping at higher output voltages. This is consistent with a recurrent theme that manifested itself previously in different forms.

Fig. 5.3. Ideal charge pump efficiency versus number of stages for \( V_{\text{in}} = V_{dd} \).
To sum up, so far, we found that higher output voltage pumps are less power efficient due to 3 main factors: (1) from a voltage reliability stand point, higher voltage pump cells require wider spaced capacitor fingers resulting in higher plate parasitics, (2) from an architecture stand point, voltage-tolerant substrate isolation methods mandate circuit implementations that are inherently less efficient, and (3) from a power analysis stand point, higher voltages require more pump stages. As the number of stages increases, their relative contribution to the final output voltage compared to $V_{\text{in}}$ increases, incurring more losses compared to a perfect input voltage source.

Now, we consider the case for hybrid pumps. We assume a Hybrid Charge Pump is composed of $m$ sub-pumps and has $V_{\text{in}}=0$, as shown in Fig. 5.4. We also assume that sub-pump $i$ consists of $n_i$ stages and has efficiency $\eta_{oi}$ when $V_{\text{in}}=0$.

![Fig. 5.4. Hybrid charge pump block diagram representation with $V_{\text{in}}=0$.](image)

It automatically follows that,

$$I_{\text{in}} = I_{\text{out}} = I_{\text{out}}$$  \hspace{1cm} (5.9)

$$I_{\text{ddi}} = n_i I_{\text{out}}$$  \hspace{1cm} (5.10)

Accordingly, $\eta_o$ for a Hybrid Charge Pump is given by,

$$\eta_o = \frac{I_{\text{out}} V_{\text{out}}|_{V_{\text{in}}=0}}{(n_1 I_{\text{out}} + n_2 I_{\text{out}} + \ldots + n_m I_{\text{out}}) V_{\text{dd}}} = \frac{V_{\text{out}}|_{V_{\text{in}}=0}}{(n_1 + n_2 + \ldots + n_m) V_{\text{dd}}}$$  \hspace{1cm} (5.11)

Moreover, the final output voltage is related to the individual sub-pump outputs as follows,
\[ V_{out} |_{V_{in}=0} = V_{out1} |_{V_{in}=0} + V_{out2} |_{V_{in}=0} + ... + V_{outm} |_{V_{in}=0} \] (5.12)

By substituting from Equations (5.4) and (5.12) into Equation (5.11), the intrinsic efficiency \( \eta_o \) of a Hybrid Charge Pump can be expressed in terms of its sub-pumps \( \eta_{oi} \) as follows,

\[
\eta_o = \frac{\sum_{i=1}^{m} n_i \eta_{oi}}{\sum_{i=1}^{m} n_i} \tag{5.13}
\]

Equation (5.13) expresses the overall pump efficiency at \( V_{in}=0 \) as a weighted average of the constituent sub-pump efficiencies, such that a sub-pump efficiency is weighted by its number of stages. To find the efficiency \( \eta \) for arbitrary inputs, we refer to Equation (5.7) such that,

\[
\eta = \frac{V_{in}/V_{dd} + \sum_{i=1}^{m} n_i \eta_{oi}}{V_{in}/V_{dd} + \sum_{i=1}^{m} n_i} \tag{5.14}
\]

Interestingly, if the input voltage source \( V_{in} \) is thought of as an ideal charge pump with \( \eta_o = 1 \) and number of stages \( n = \frac{V_{in}}{V_{dd}} \), we can read Equation (5.14) as the weighted average form of Equation (5.13). Sometimes it is useful to express \( \eta \) of a Hybrid Charge Pump in terms of its intrinsic efficiency as follows,

\[
\eta = \frac{V_{in}/V_{dd} + \eta_o \sum_{i=1}^{m} n_i}{V_{in}/V_{dd} + \sum_{i=1}^{m} n_i} \tag{5.15}
\]

More generally, if the sub-pumps are operated from different supply voltages \( k_i V_{dd} \) as shown in Fig. 5.5, then Equation (5.15) needs to reflect that change. Assuming \( k_i \) is the supply scaling factor for sub-pump \( i \), the new efficiency for arbitrary \( V_{in} \) is given by,
Since we have established the weighted average nature of the Hybrid Charge Pump efficiency, we can now move to a more realistic power analysis model.

5.3.2. Practical Power Model

Power losses in a charge pump circuits are broadly characterized into 4 types [58], namely, redistribution loss, switching loss, conduction loss, and reversion loss. Redistribution loss is the most fundamental of all 4 types and sets an upper bound on the peak pump efficiency. Because there is always energy dissipation associated with charge redistribution between capacitors, even in the most ideal of settings, redistribution loss is a characteristic of all charge pumps. Switching loss is due to the dynamic switching of capacitors between 2 voltage levels periodically, and is more or less dependent on technology parameters, e.g. the ratio of the parasitic plate capacitance to the main pumping capacitance, and the transistor gate and drain capacitances per unit width. Although some circuit techniques such as charge recycling [62] are used to cut down switching loss, the biggest improvements on that front are enabled by improved technology capacitors [63]-[64]. Conduction losses need to be accounted for mostly when high pumping frequencies are used, i.e. FSL, and can be ignored for lower frequencies. This frequency divide can be drawn by
calculating the $f_{pump}$ value at which the switch on-resistance is comparable to the equivalent $1/fC$ switched capacitor resistance. Finally, we have shown that Reversion loss can be eliminated by careful design and the use of multiple clock phases to perform break-before-make switching.

\[ R = \frac{1}{f(C + C_p)} \coth\left( \frac{1}{4r_{on}f(C + C_p)} \right) \]  
\[ (5.17) \]

Where $C$ is the main capacitor and $C_p$ is the top plate parasitic capacitance, and $r_{on}$ is the switch on resistance [64]. In the SSL, where $f << r_{on}(C+C_p)$, redistribution loss dominates and $R \approx 1/(C+C_p)$. In the FSL, conduction loss dominates and $R \approx 4r_{on}$. In the FSL, we notice that $C_p$ plays an interesting role in voltage doubler circuits. Although $C_p$ results in an output voltage loss due to the capacitive divider between $C$ and $C_p$, it also reduces the doubler's output resistance.

Based on the observations made so far, we present in Fig. 5.7 an equivalent circuit model of a voltage pumping cell having $n$ stages and with $V_{in}=0$. This model incorporates redistribution and switching losses only. The main pumping capacitor is $C$, while the bottom and top plate
parasitics are $\alpha_1 C$ and $\alpha_2 C$ respectively, where $\alpha_1$ and $\alpha_2$ are technology dependent parameters with $0<\alpha_1, \alpha_2<1$. In principle, $\alpha_1$ would account for all the clock driver drain parasitics, while $\alpha_2$ would account for the switches drain parasitics, and other gate capacitance if present.

The voltage conversion nature of the circuit is modeled using an ideal lossless transformer with a turns ratio $1:n$. The power dissipated per stage due to charge redistribution is modeled as an equivalent resistance $R=1/fC$. A nearly fixed switching loss exists due to the periodic charging and discharging of $C_{p1}$ and $C_{p2}$ regardless of the load current. This power loss, per stage, is modeled using 2 resistors $R/\alpha_1$ and $R/\alpha_2$ shunting the primary and secondary terms respectively. Notice that the resistive divider arrangement in the secondary turns accurately reflects both the output voltage division and the reduced output resistance actions of $C_{p2}$.

In the SSL, we can, to a first order, treat the effects of $f$ and $C$ similarly, i.e. identical $R$ values are possible for identical $fC$ products. Hence, for the coming derivations, we express pump efficiencies in terms of $R$, where

$$R = \frac{1}{fC} \quad (5.18)$$
This approximation assumes that the same pump efficiency can be attained for any fixed fC product. This is not entirely true since at higher f and lower C, parameters $\alpha_1$ and $\alpha_2$ grow larger due to the fixed switch drain and gate parasitic component, and the pump efficiency decreases.

First, we derive the intrinsic efficiency $\eta_o$ for the circuit in Fig. 5.7. By treating $V_{dd}$ and $I_{out}$ as independent variables, we use the superposition principle to find expressions for $V_{out}$ and $I_{dd}$ in terms of $V_{dd}$ and $I_{out}$ as follows,

$$V_{out} = \frac{n}{1+\alpha_2} (V_{dd} - I_{out}R)$$

$$I_{dd} = \frac{nV_{dd}}{R} \left( \frac{\alpha_1 + \frac{\alpha_2}{1+\alpha_2}}{1+\alpha_2} \right) + \frac{nI_{out}}{1+\alpha_2}$$

Accordingly, expressions for $P_{out}$ and $P_{in}$ are given as follows,

$$P_{out} = \frac{nV_{dd}}{1+\alpha_2} \left( 1 - \frac{I_{out}R}{V_{dd}} \right) I_{out}$$

$$P_{in} = \frac{nV_{dd}}{1+\alpha_2} \left[ \frac{(\alpha_1 + \alpha_2 + \alpha_1\alpha_2) V_{dd}}{R} + I_{out} \right]$$

Finally, the intrinsic power efficiency $\eta_o$ is defined as,

$$\eta_o = \left( \frac{I_{out}}{I_{out} + (\alpha_1 + \alpha_2 + \alpha_1\alpha_2) V_{dd}/R} \right) \left( 1 - \frac{I_{out}R}{V_{dd}} \right)$$

For a symmetric finger capacitor design, and ignoring other parasitic contributions, we can assume that $\alpha_1=\alpha_2=\alpha$ and $\alpha<<1$, and the efficiency is expressed as follows,

$$\eta_o \approx \left( \frac{I_{out}}{I_{out} + 2\alpha V_{dd}/R} \right) \left( 1 - \frac{I_{out}R}{V_{dd}} \right)$$
Equation (5.24) expresses the pump efficiency in terms of its load current and has a form that is characteristic of most charge pump circuits. Equation (5.24) predicts 2 nulls in power efficiency, one at $I_{out}=0$ and another at $I_{out}=V_{dd}/R$. Efficiency is a function of the load current. The first null occurs because for $I_{out}=0$, there is no power is delivered to the load, yet switching losses exist. The other null occurs because for $I_{out}=V_{dd}/R$, $V_{out}=0$ and no power is delivered to the load, yet switching and redistribution losses exist. In between these 2 nulls, $\eta_o$ is nonzero and a maximum exists. Because Equation (5.24) manifests itself many times in this analysis, we choose to rewrite it on the following generic form,

$$\eta_o = \left( \frac{I_{out}}{I_{out} + \alpha_o I_o} \right) \left( 1 - \frac{I_{out}}{I_o} \right)$$  \hspace{1cm} (5.25)

$$\alpha_o = \alpha_1 + \alpha_2 + \alpha_1 \alpha_2$$ \hspace{1cm} (5.26)

$$I_o = \frac{V_{dd}}{R}$$ \hspace{1cm} (5.27)

where $\alpha_o$ represents an effective parasitic capacitance coefficient that is technology dependent, while $I_o$ represents the null load current at which $V_{out}=0$.

To further analyze Equation (5.25), we break it down into 2 different functions defined by the brackets. The first function, $f(I_{out})$, is defined by the equation in the second bracket, whereas the second function, $g(I_{out})$, is defined by the equation in the first bracket. Interestingly, $f(I_{out})$ represents the efficiency of an ideal charge pump with no parasitic loss, i.e. $\alpha_1=\alpha_2=0$. In an ideal charge pump, redistribution loss is inevitable and $f(I_{out})$ drops linearly with $V_{out}$ for higher $I_{out}$, till $\eta_o=0$ at $I_{out}=I_o$. The second function $g(I_{out})$ captures switching loss and inserts another null at $I_{out}=0$. Overall, $g(I_{out})$ shapes the ideal $f(I_{out})$ efficiency to predict a more practical efficiency dependence on $I_{out}$ as shown in Fig. 5.8, with a peak efficiency $\eta_{max}$ occurring at $I_{out}=I_{max}$.
Interestingly, we notice that larger $\alpha_o$ has 2 effects on $\eta_o$, first it leads to smaller $\eta_{\text{max}}$, and second, it shifts $I_{\text{max}}$ to higher load currents. To calculate $\eta_{\text{max}}$ we differentiate Equation (5.25) with respect to $I_{\text{out}}$ and equate to zero. By solving for $I_{\text{out}}$, we find that,

$$I_{\text{out}}^2 + 2\alpha_o I_o I_{\text{out}} - \alpha_o I_o^2 = 0$$ \hspace{1cm} (5.28)

$$I_{\text{max}} = (\sqrt{\alpha_o (1 + \alpha_o)} - \alpha_o)I_o$$ \hspace{1cm} (5.29)
\[
\approx \sqrt{\alpha_o} (1 + \alpha_o/2 - \sqrt{\alpha_o}) I_o
\]  

(5.30)

By substituting from Equation (5.29) into Equation (5.25), \( \eta_{\text{max}} \) is found to be,

\[
\eta_{\text{max}} = (\sqrt{1+\alpha_o} - \sqrt{\alpha_o})^2
\]  

(5.31)

\[
\approx (1 + \alpha_o/2 - \sqrt{\alpha_o})^2
\]  

(5.32)

Notice that \( \eta_{\text{max}} \) is only a function of \( \alpha_o \) and does not depend on \( f, C, n \) or \( V_{\text{dd}} \). For the special case, \( \alpha_1=\alpha_2=\alpha \), we find that,

\[
I_{\text{max}} \approx \sqrt{2\alpha} (1 + \alpha - \sqrt{2\alpha}) I_o
\]  

(5.33)

\[
\eta_{\text{max}} \approx (1 + \alpha - \sqrt{2\alpha})^2
\]  

(5.34)

Shown in Fig. 5.9 is a plot of \( \eta_{\text{max}} \) versus \( \alpha \) as per Equation (5.34). For \( 0.01 < \alpha < 0.05 \), the corresponding efficiency range is \( 0.53 < \eta_{\text{max}} < 0.75 \). We notice that in order to maximize \( \eta_{\text{max}} \) we need to minimize \( \alpha_o \). This property is used later to optimize Hybrid Charge Pump designs.

![Fig. 5.9. Calculated peak \( \eta_o \) versus the parasitic capacitance to pumping capacitor ratio.](image)

To verify our proposed model, we plot the calculated and measured efficiencies of a CMOS 4PVD in Fig. 5.10 for \( V_{\text{in}}=0, n=4, C=8\text{pF}, f=4\text{MHz}, V_{\text{dd}}=2.5\text{V}, \) and \( \alpha=0.025 \). Equation (5.25)
predicts the pump peak efficiency accurately, and deviates slightly for larger $I_{\text{out}}$. This deviation is attributed to the poor process control over the absolute value of the implemented capacitors.

Typically charge pumps are operated with $V_{\text{in}}=V_{\text{dd}}$, thus we repeat the same derivations for the circuit model assuming arbitrary $V_{\text{in}}$ as shown in Fig. 5.11.

Assuming an additional independent variable $V_{\text{in}}$ and using the superposition principle, we can rewrite the expressions for $V_{\text{out}}$ and $I_{\text{dd}}$ in terms of the previously derived values in Equations (5.19) and (5.20) as follows,
\[ V_{out} = V_{in} + V_{out}\bigg|_{\tilde{V}_a=0} \]  
(5.35)

\[ I_{dd} = I_{dd}\bigg|_{\tilde{V}_a=0} \]  
(5.36)

Moreover,

\[ I_{in} = I_{out} \]  
(5.37)

As a result the efficiency \( \eta \) for arbitrary \( V_{in} \) can be calculated as follows,

\[ \eta = \frac{V_{out} I_{out}}{V_{in} I_{in} + V_{dd} I_{dd}} = \frac{(V_{in} + V_{out}\bigg|_{\tilde{V}_a=0}) I_{out}}{V_{in} I_{out} + V_{dd} I_{dd}} \]  
(5.38)

But we know by definition that

\[ V_{out}\bigg|_{\tilde{V}_a=0} I_{out} = \eta_o V_{dd} I_{dd} \]  
(5.39)

Therefore \( \eta \) is expressed in terms of \( \eta_o \) as follows,

\[ \eta = \frac{I_{out} V_{in}/V_{dd} + I_{dd}\eta_o}{I_{out} V_{in}/V_{dd} + I_{dd}} \]  
(5.40)

For the special case when \( V_{in} = V_{dd} \), \( \eta \) is given by

\[ \eta = \frac{I_{out} + I_{dd}\eta_o}{I_{out} + I_{dd}} \]  
(5.41)

\[ \eta = \eta_o + \frac{I_{out}(1-\eta_o)}{I_{out} + I_{dd}} \]  
(5.42)

Since \( 0<\eta_o<1 \), Equation (5.42) guarantees that \( \eta>\eta_o \), as we should anticipate. To verify our model, we substitute into Equation (5.41) with the measurement results of the CMOS 4PVD in Fig. 5.10 for \( V_{in} = 0 \). Shown in Fig. 5.11 are the calculated and measured efficiencies of the CMOS 4PVD circuit at \( V_{in} = V_{dd} \). It can be seen that the measurements and calculations are in good agreement.
Interestingly, the pump efficiency $\eta$ at arbitrary $V_{\text{in}}$ can be expressed in a more common form as shown below, by substituting for $\eta_0$ from Equation (5.23) into Equation (5.40).

$$\eta = \left( \frac{I_{\text{out}}}{I_{\text{out}} + \alpha_o I_o} \right) \left( 1 - \frac{I_{\text{out}}}{I_o} \right)$$

(5.43)

$$\alpha_o = \frac{1}{\lambda^2} (\alpha_1 + \alpha_2 + \alpha_1^2 \alpha_2)$$

(5.44)

$$I_o = \lambda \frac{V_{\text{dd}}}{R}$$

(5.45)

$$\lambda = 1 + \frac{(1 + \alpha_2)V_{\text{in}}}{nV_{\text{dd}}}$$

(5.46)

Notice the new parameter $\lambda$ captures the impact of $V_{\text{in}}$ and $n$ on the pump efficiency. For $V_{\text{in}}=0$, we find that $\lambda=1$, and $\eta$ is identical to that predicted by Equation (5.23). For $V_{\text{in}}>0$, we find that $\lambda>1$. This effectively scales down the $\alpha_o$ parasitic coefficient by $1/\lambda^2$ leading to larger $\eta_{\text{max}}$. Also, $I_o$ scales up by $\lambda$ resulting in higher $I_{\text{max}}$ values. Moreover, we notice that as $n \to \infty$, $\eta \to \eta_0$ as predicted earlier by Equation (5.8). Because $\eta$ maintains the same functional form as $\eta_0$, Equations (5.29) and (5.31) are still used to predict $I_{\text{max}}$ and $\eta_{\text{max}}$, respectively.
More generally, the pump switches could incur a diode voltage drop as in Dickson-type pumps. As a result, the equivalent circuit model needs to be modified as shown in Fig. 5.13. An n-stage Dickson pump contains n+1 diodes. For simplicity, we assume an ideal diode model with a knee voltage $V_D$. For $V_{in}=0$, $V_{out}$ and $I_{out}$ are calculated using superposition as follows,

$$V_{out} = \frac{n}{1+\alpha_2} (V_{dd} - I_{out} R) - (n+1)V_D$$

$$I_{dd} = \frac{nV_{dd}}{R} \left( \frac{\alpha_1 + \frac{\alpha_2}{1+\alpha_2}}{1+\alpha_2} \right) + \frac{nI_{out}}{1+\alpha_2}$$

Consequently, $P_{out}$ and $P_{in}$ are given by,

$$P_{out} = \left( \frac{nV_{dd} - (1+\alpha_2)(n+1)V_D}{1+\alpha_2} \right) \left( 1 - \frac{nI_{out}R}{nV_{dd} - (1+\alpha_2)(n+1)V_D} \right) I_{out}$$

$$P_{in} = \frac{nV_{dd}}{1+\alpha_2} \left( \frac{(\alpha_1 + \alpha_2 + \alpha_1\alpha_2)V_{dd}}{R} + I_{out} \right)$$

Thus, the efficiency $\eta_o$ is given by,

$$\eta_o = \left( 1 - \frac{(1+\alpha_2)(n+1)V_D}{nV_{dd}} \right) \left( \frac{I_{out}}{I_{out} + (\alpha_1 + \alpha_2 + \alpha_1\alpha_2)V_{dd}/R} \right) \left( 1 - \frac{nI_{out}R}{nV_{dd} - (1+\alpha_2)(n+1)V_D} \right)$$

To accommodate this more general case, we express $\eta_o$ on the following form,
\[
\eta_o = \beta \left( \frac{I_{out}}{I_{out} + \alpha_o I_o} \right) \left( 1 - \frac{I_{out}}{I_o} \right)
\]  \hspace{1cm} (5.52)

\[
\alpha_o = \frac{1}{\beta} (\alpha_1 + \alpha_2 + \alpha_1 \alpha_2)
\]  \hspace{1cm} (5.53)

\[
I_o = \beta \frac{V_{dd}}{R}
\]  \hspace{1cm} (5.54)

\[
\beta = 1 - \frac{(1 + \alpha_2)(n + 1)V_D}{nV_{dd}}
\]  \hspace{1cm} (5.55)

A new parameter $\beta$ is introduced to capture the impact of $V_D$ on the pump efficiency. For $V_D=0$, we find that $\beta=1$, and $\eta_o$ is identical to that predicted by Equation (5.23). For $V_D>0$, we find that $\beta<1$. This effectively scales up $\alpha_o$ by a factor $\beta$ leading to lower $\eta_{\text{max}}$. Also, $I_o$ scales down by the same factor resulting in lower $I_{\text{max}}$ values. Since $\eta_o$ maintains the same functional form except for the scaling factor $\beta$, $I_{\text{max}}$ and $\eta_{\text{max}}$ are given by,

\[
I_{\text{max}} = (\sqrt{\alpha_o (1+\alpha_o)} - \alpha_o) I_o
\]  \hspace{1cm} (5.56)

\[
\eta_{\text{max}} = \beta (\sqrt{1+\alpha_o} - \sqrt{\alpha_o})^2
\]  \hspace{1cm} (5.57)

Notice that for diode-based pumps $\eta_{\text{max}}<1$, even with zero switching loss. This is expected because of the diode switch voltage drop, i.e. $\eta_{\text{max}}=\beta$ for $\alpha_o=0$. More importantly, we notice that increasing $V_{dd}$ increases $\beta$ and decreases $\alpha_o$. This results in higher peak efficiency, e.g. for $V_{dd}>>V_D$ and $\alpha_o=0$, we find that $\eta_{\text{max}}\to 1$. Shown in Fig. 5.14 is a plot of the peak efficiency of a diode-based pump versus $V_{dd}$. The pump has the following parameters: $V_{in}=0$, $V_D=0.7V$, $n=8$ and $\alpha=0.025$. As expected the pump efficiency increases with $V_{dd}$ and approaches 60% for very large $V_{dd}$. Shown in Fig. 5.15, is a plot of the measured and calculated efficiencies of the diode-based pump versus $I_{out}$ based on Equation (5.51). It can be seen that the calculated efficiency is slightly off due to the poor process control over the absolute value of implemented capacitors.
For arbitrary $V_{in}$, $\eta$ is calculated using Equation (5.40). After some manipulations, $\eta$ can be expressed on a form identical to Equation (5.52), but having the following parameters,

$$\alpha_o = \frac{1}{\beta \lambda^2} (\alpha_1 + \alpha_2 + \alpha_i \alpha_2)$$  \hfil (5.58)

$$I_o = \beta \lambda \frac{V_{dd}}{R}$$  \hfil (5.59)

$$\beta = 1 - \frac{(1 + \alpha_2)(n + 1)V_D}{nV_{dd} + (1 + \alpha_2)V_{in}}$$  \hfil (5.60)

$$\lambda = 1 + \frac{(1 + \alpha_2)V_{in}}{nV_{dd}}$$  \hfil (5.61)
More generally, the diode forward drop is a logarithmic function of current given by,

\[ V_D(I_{\text{out}}) = mV_T \ln \left( 1 + \frac{I_{\text{out}}}{I_s} \right) \]  

(5.62)

where \( m \) is the diode ideality factor, \( V_T \) is the thermal voltage, and \( I_s \) is the diode reverse saturation current. For more accurate estimates of efficiency, Equation (5.62) is used to substitute for \( V_D \) in Equation (5.55).

Based on our analysis, we now highlight a design methodology to target optimum power efficiency. Typically, a charge pump is specified to target an output voltage \( (V_{\text{out}}) \) for a given load current \( (I_{\text{out}}) \). In general, there are 2 types of parameters: (1) technology parameters: \( \alpha_1, \alpha_2, V_D, \) and \( V_{\text{dd}} \), and (2) design parameters: \( n, f, \) and \( C \). First, we start the design by picking the pump architecture. Whenever possible, we pick a pump cell design that uses capacitors with the smallest \( \alpha_1, \alpha_2 \), and devices that can handle the highest \( V_{\text{dd}} \) with, preferably, no \( V_D \) drops. Also, we choose \( V_{\text{in}}=V_{\text{dd}} \) in positive-type pumps, for higher efficiency. Next, we need to pick \( n, f, \) and \( C \) to meet the \( (I_{\text{out}}, V_{\text{out}}) \) requirement and achieve optimum efficiency. The pump's \( V_{\text{out}} \) and \( I_{\text{out}} \) are related through the following equation,

\[ V_{\text{out}} = V_{\text{in}} - (n + 1)V_D + \frac{n}{1 + \alpha_2} \left( V_{\text{dd}} - \frac{I_{\text{out}}}{fC} \right) \]  

(5.63)

For a given \( (V_{\text{out}}, I_{\text{out}}) \) specification, \( fC \) is related to the number of stages \( n \) as follows,

\[ fC = \frac{I_{\text{out}}}{(V_{dd} - (1 + \alpha_2)V_D) - \frac{(1 + \alpha_2)}{n} \left( V_{\text{out}} - V_{\text{in}} + V_D \right)} \]  

(5.64)

Shown in Fig. 5.16 are the required \( fC \) values for different \( n \) to maintain a fixed \( V_{\text{out}}=40V \) at \( I_{\text{out}}=20\mu A \). The following parameters are assumed: \( V_{\text{dd}}=2.5V, V_{\text{in}}=2.5V, V_D=0 \), and \( \alpha=0.025 \).
As expected a larger fC product is needed to maintain $V_{out}$ at smaller n. The calculated (n, fC) values are substituted into Equations (5.58)-(5.61) and the pump efficiency at $I_{out} = 20\mu A$ is calculated using Equation (5.52). Shown in Fig. 5.17 is the calculated efficiency versus the number of stages. For very small n, the required fC product is too high and the peak efficiency current $I_{max} > 20\mu A$, whereas for very large n, the required fC product is too small and $I_{max} < 20\mu A$. The optimum number of stages that matches $I_{max}$ to 20uA is found to be 19. The corresponding optimum efficiency $\eta_{opt} = 65.4\%$. 

Fig. 5.16. Calculated fC values required to produce (40V, 20μA) outputs for different n.

Fig. 5.17. Calculated pump efficiency versus n for (40V, 20μA) outputs.
To summarize our design procedure, we outline the following steps,

1. First, we start with the minimum number of stages meeting the $V_{out}$ specification for a given $V_{dd}$ and $\alpha_2$. $n_{\text{min}}$ is calculated using,

$$n_{\text{min}} = \frac{(1 + \alpha_2)(V_{out} - V_{in} + V_D)}{V_{dd} - (1 + \alpha_2)V_D} \quad (5.65)$$

In general $n_{\text{min}}$ is a non-integer, this $n_{\text{min}}$ assumes $fC=\infty$ and is impractical, and we round it off to the next integer.

2. Given the calculated $n$ in the previous step, we use Equation (5.64) to calculate the required $fC$ product to meet the $(V_{out}, I_{out})$ specification.

3. Given the $(n, fC)$ data point calculated in step 2, $\alpha_0$ and $I_o$ are calculated using Equations (5.58) and (5.59).

4. Knowing $\alpha_0$ and $I_o$, we calculate $I_{\text{max}}$ using Equation (5.56). By comparing $I_{\text{max}}$ and $I_{\text{out}}$, we decide whether to move to the next iteration or stop here. If $I_{\text{max}}>I_{\text{out}}$, we increase $n$ to $n+1$ and repeat steps 2 through 4. If $I_{\text{max}}\leq I_{\text{out}}$ we stop at the current iteration.

5. For a given pumping frequency $f$, usually specified by the system, we find the pumping capacitor value $C$. Generally, we can trade off $f$ and $C$ while maintaining the same $fC$ product. Operating at lower $f$ results in higher pump efficiencies given a fixed switch size, however, using a higher $f$ saves the capacitor area.

6. The switches are sized such that their on resistance ($r_{\text{on}}$) meets the SSL operation requirement and alleviate conduction losses, i.e. $4r_{\text{on}}fC<<1$.

This design methodology is summarized in the flow chart in Fig. 5.18.
Now, we are ready to extend our power model analysis to the more general case of Hybrid Charge Pumps. Shown in Fig. 5.19 is the equivalent circuit model of a Hybrid Charge Pump consisting of \( m \) sub-pumps, where each sub-pump \( i \) has parameters: \( n_i, R_i, \alpha_{1i} \) and \( \alpha_{2i} \).

For the proposed circuit model, the following relations hold,

\[
I_{in} = I_{out} = I_{out} \quad (5.66)
\]

\[
\eta_o = \frac{V_{out}}{V_{dd}(I_{dd1} + I_{dd2} + \ldots + I_{ddm})} \quad (5.67)
\]
Also,

\[ V_{\text{out}}|_{V_{\text{in}}=0} = V_{\text{out}}|_{V_{\text{in}}=1} + V_{\text{out}}|_{V_{\text{in}}=2} + \ldots + V_{\text{out}}|_{V_{\text{in}}=m} \]

(5.68)

\[ V_{\text{out}}|_{V_{\text{in}}=0} \cdot I_{\text{out}} = V_{\text{dd}} I_{\text{dd}} \eta_o \]

(5.69)

By substituting from Equations (5.68) and (5.69) into Equation (5.67), the intrinsic efficiency \( \eta_o \) can be rewritten on the following form,

\[ \eta_o = \frac{\sum_{i=1}^{m} I_{\text{dd}} \eta_{oi}}{\sum_{i=1}^{m} I_{\text{dd}i}} \]

(5.70)

Interestingly, the Hybrid Charge Pump efficiency \( \eta_o \) manifests itself on a weighted average form, where each sub-pump \( \eta_{oi} \) is weighted by its supply current \( I_{\text{dd}i} \). Moreover, for arbitrary \( V_{\text{in}} \), the hybrid pump efficiency \( \eta \) is expressed as,

\[ \eta = \frac{V_{\text{in}} I_{\text{out}}}{V_{\text{dd}}} + \sum_{i=1}^{m} I_{\text{dd}} \eta_{oi} \]

(5.71)

Equation (5.71) is no different than Equation (5.70) in the sense that it too represents a weighted average form. If \( V_{\text{in}} \) is treated as an ideal charge pump with \( \eta_o=1 \), and accordingly has a supply current \( I_{\text{dd}}=V_{\text{in}} I_{\text{out}}/V_{\text{dd}} \), we find that Equations (5.70) and (5.71) are identical.

Shown in Fig. 5.20 are the measured efficiencies of a Hybrid Charge Pump composed of 3 smaller sub-pumps. The individual sub-pump efficiencies are plotted for \( V_{\text{in}}=0 \), while the overall Hybrid Charge Pump efficiency is calculated at \( V_{\text{in}}=V_{\text{dd}} \) using Equation (5.71). It is clear that the measured and calculated hybrid efficiencies are in close agreement.
Now, we are interested in calculating $\eta_{\text{max}}$ and $I_{\text{max}}$ for the Hybrid Charge Pump. To do so, we assume that each sub-pump $i$ has $\eta_{oi}$ that assumes the generic form given by Equation (5.52) such that,

$$\eta_{oi} = \beta_i \left( \frac{I_{\text{out}}}{I_{\text{out}} + \alpha_{oi} I_{oi}} \right) \left( 1 - \frac{I_{\text{out}}}{I_{oi}} \right)$$

(5.72)

$$\alpha_{oi} = \frac{1}{\beta_i} \left( \alpha_{ii} + \alpha_{zi} + \alpha_{ii} \alpha_{zi} \right)$$

(5.73)

$$I_{oi} = \beta_i \frac{V_{dd}}{R_i}$$

(5.74)

$$\beta_i = 1 - \frac{(1 + \alpha_{zi})(n_i + 1)V_{Di}}{n_i V_{dd}}$$

(5.75)

Moreover, the supply current of each sub-pump $I_{ddi}$ is independent of $V_{in}$ and is expressed by Equation (5.48). For convenience, we rewrite Equation (5.48) on the following form,

$$I_{ddi} = \frac{n_i}{1 + \alpha_{zi}} \left( I_{\text{out}} + \alpha_{oi} I_{oi} \right)$$

(5.76)

After substituting from Equations (5.72) and (5.76) into Equation (5.70), and carrying out some mathematical steps, the hybrid pump efficiency can be written on the following form.
\[ \eta_o = \left( \sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}} \right) \frac{I_{out}}{I_{out} + \sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}} \alpha_{oi} I_{oi}} \right) \left( \sum_{i=1}^{m} \frac{\beta_i n_i}{1 + \alpha_{2i}} \frac{1}{I_{oi}} \right) \] (5.77)

Interestingly enough, \( \eta_o \) of the Hybrid Charge Pump holds the same functional form of the single sub-pump \( \eta_o \) given by Equation (5.52), and has the following parameters,

\[ \alpha_{oi} I_{oi} = \frac{\sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}} \alpha_{oi} I_{oi}}{\sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}}} \] (5.78)

\[ \frac{1}{I_o} = \frac{\sum_{i=1}^{m} \frac{\beta_i n_i}{1 + \alpha_{2i}}}{\sum_{i=1}^{m} \frac{\beta_i n_i}{1 + \alpha_{2i}}} \] (5.79)

\[ \beta = \frac{\sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}} \beta_i}{\sum_{i=1}^{m} \frac{n_i}{1 + \alpha_{2i}}} \] (5.80)

Equations (5.78)-(5.80) inform us that if all the sub-pumps have identical \( n \) and \( \alpha \), the sub-pump with the largest \( \beta \) dominates the hybrid pump's \( \beta \), whereas the pump with the smallest \( I_o \) dominates the hybrid pump's \( I_o \). Moreover, if all the sub-pumps have identical \( I_o \), the sub-pump with the largest \( \alpha \) dominates the hybrid pump's \( \alpha \). An expression for the hybrid pump \( \alpha \) can be obtained by multiplying Equations (5.78) and (5.79).

\[ \alpha_o = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\beta_i n_i n_j}{(1 + \alpha_{2i})(1 + \alpha_{2j})} \frac{\alpha_{oi} I_{oj}}{I_{oi}}}{\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\beta_i n_i n_j}{(1 + \alpha_{2i})(1 + \alpha_{2j})}} \] (5.81)
Equation (5.81) is important because we have established earlier that \( \eta_{\text{max}} \) is a function of \( \alpha_o \) as shown in Fig. 5.9. For a maximum \( \eta_{\text{max}} \), we need to minimize \( \alpha_o \). By differentiating Equation (5.81) with respect to \( I_{ok} \) for all \( 1 \leq k \leq m \) and equating the partial derivatives to zero we get,

\[
\frac{\partial \alpha_o}{\partial I_{ok}} = 0 \quad \text{for} \quad k = 1, 2, \ldots, m \quad (5.82)
\]

After some manipulations, the set of Equations (5.82) can be expressed as follows,

\[
\frac{\alpha_{ok} I_{ok}^2}{\beta_k} = \sum_{i=1}^{m} \frac{n_i \alpha_{oi} I_{oi}}{(1 + \alpha_{2i}) I_{oi}} \quad \text{for} \quad k = 1, 2, \ldots, m \quad (5.83)
\]

This system of equations has rank \( m-1 \) and is equivalent to adjusting the peak efficiency currents of all sub-pumps such that they align at the same current value. It can be shown that this set of equations has a non-trivial solution if the following relation holds for all sub-pumps \( i \) and \( j \),

\[
\frac{I_{oi}}{I_{oj}} = \frac{\alpha_{oj} \beta_j}{\alpha_{oi} \beta_i} \quad (5.84)
\]

By normalizing all the sub-pumps \( I_{oi} \) with respect to the first sub-pump, Equation (5.84) is rewritten as follows,

\[
I_{oi} = \sqrt{\frac{\alpha_{oi} \beta_i}{\alpha_{oj} \beta_j}} I_{o1} \quad \text{for} \quad i = 2, \ldots, m \quad (5.85)
\]

Because \( I_o \) and \( C \) are related as in Equation (5.74), by substituting back into Equation (5.85), the different sub-pump capacitors can be sized relative to the first sub-pump capacitor as follows,

\[
C_i = \sqrt{\frac{\alpha_{1i} + \alpha_{2i} + \alpha_{1i} \alpha_{2i} C_1}{\alpha_{1i} + \alpha_{2i} + \alpha_{1i} \alpha_{2i} C_1}} \quad \text{for} \quad i = 2, \ldots, m \quad (5.86)
\]
Interestingly, Equation (5.86) indicates that relative capacitor sizing is only a function of the capacitor's parasitic coefficients $\alpha_1$ and $\alpha_2$. Furthermore, sub-pumps having a larger parasitic capacitance require a smaller $C$ value. This is consistent with our understanding of the impact of $\alpha_1$ and $\alpha_2$ on peak efficiency current values. Since larger $\alpha_1$ and $\alpha_2$ values shift the peak efficiency current towards higher values, using a smaller $C$ offsets that effect and shifts peak efficiency currents towards smaller values. Moreover, Equation (5.86) indicates that using identical $C$ values for all cells with identical cell design is not only convenient but also more power efficient.

By substituting from Equation (5.84) back into Equation (5.81), the optimum $\alpha_o$ for the hybrid pump is given by,

$$\alpha_{o\min} = \left( \sum_{i=1}^{m} \frac{n_i}{(1 + \alpha_{2i})} \right)^2 \frac{\alpha_{1i} + \alpha_{2i} + \alpha_{1i}\alpha_{2i}}{\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{\beta_i n_i n_j}{(1 + \alpha_{2i})(1 + \alpha_{2j})}}$$  \hspace{1cm} (5.87)

Based on this $\alpha_{o\min}$ value, the pump's optimum current $I_{opt}$ and optimum peak efficiency $\eta_{opt}$ are calculated using Equations (5.56) and (5.57) and are given by,

$$I_{opt} = (\sqrt{\alpha_{o\min}(1 + \alpha_{o\min})} - \alpha_{o\min})I_o$$  \hspace{1cm} (5.88)

$$\eta_{opt} = \beta(\sqrt{1 + \alpha_{o\min}} - \sqrt{\alpha_{o\min}})^2$$  \hspace{1cm} (5.89)

where $I_o$ and $\beta$ and are calculated using Equations (5.79) and (5.80) respectively. To get a better understanding of the design space, we assume a Hybrid Charge Pump design consisting of 3 cascaded sub-pumps having the following parameters: $V_{dd}=2.5V$, $V_{in}=0$, $f=4MHz$, $n_1=4$, $\alpha_{11}=\alpha_{21}=0.025$, $V_{D1}=0$, $n_2=4$, $\alpha_{12}=\alpha_{22}=0.0375$, $V_{D1}=0$, $n_3=4$, $\alpha_{13}=\alpha_{23}=0.05$, $V_{D3}=0.7V$, and $C_1=8pF$.

Shown in Fig. 5.21 is the calculated $\eta_{max}$ of the Hybrid Charge Pump versus the second and third
sub-pumps normalized capacitor values, $C_2/C_1$ and $C_3/C_1$. This $\eta_{\text{max}}$ is calculated based on the $\beta$ and $\alpha_o$ values given by Equations (5.80) and (5.81) respectively.

![Diagram](image)

**Fig. 5.21.** Hybrid Charge Pump peak efficiency plotted over the design space for $V_{\text{in}}=0$.

**Table 5.2.** Hybrid Charge Pump peak efficiency versus different $C_2/C_1$ and $C_3/C_1$ values for $V_{\text{in}}=0$.

<table>
<thead>
<tr>
<th></th>
<th>$C_2/C_1$</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_3/C_1$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td>47.68%</td>
<td>47.72%</td>
<td>47.14%</td>
<td>46.34%</td>
<td>45.47%</td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td>48.37%</td>
<td>49.14%</td>
<td>49.08%</td>
<td>48.68%</td>
<td>48.12%</td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td>47.9%</td>
<td>49.16%</td>
<td><strong>49.44%</strong></td>
<td>49.31%</td>
<td>48.97%</td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td>47.06%</td>
<td>48.68%</td>
<td>49.22%</td>
<td>49.29%</td>
<td>49.12%</td>
</tr>
<tr>
<td>1.1</td>
<td></td>
<td>46.11%</td>
<td>48.01%</td>
<td>48.75%</td>
<td>48.75%</td>
<td>48.94%</td>
</tr>
</tbody>
</table>

Each point in the design space has $\eta(I_{\text{out}})$ with a different peak efficiency $\eta_{\text{max}}$. The largest $\eta_{\text{max}}$ value over the design space is $\eta_{\text{opt}}=49.4\%$ and occurs for $C_2/C_1=0.814$ and $C_3/C_1=0.703$. 109
These optimum capacitor values are found to be in close agreement with the values estimated using Equation (5.86). Moreover, we find that the design space is relatively flat around this optimum efficiency point. We summarize in Table 5.2 the pump's $\eta_{\text{max}}$ values for different $C_2/C_1$ and $C_3/C_1$ values. It is noticed that even when $C_2$ and $C_3$ are off from their optimum values by as much as 50%, $\eta_{\text{max}}$ drops by no more than 4%. Interestingly, for this design if we choose $C_1=C_2=C_3$, we find $\eta_{\text{max}}=49.16\%$ and is near optimum. This is due to the fact that if we skew both $C_2$ and $C_3$ in the same direction, e.g. $C_2/C_1=C_3/C_1$, the decrease in $\eta_{\text{max}}$ is much smaller than the decrease we get when we skew $C_2$ and $C_3$ in opposite directions, e.g. $C_3/C_1=1-C_2/C_1$. This works out in the design's favor since for closely matched capacitors, the process skew in capacitor values is likely to occur in the same direction.

Now, we consider a more general case in which the Hybrid Charge Pump has an arbitrary $V_{\text{in}}$, and each sub-pump is powered from a different supply $k_i V_{\text{dd}}$ as shown in Fig. 5.22.

![Fig. 5.22. Hybrid Charge Pump equivalent circuit model with arbitrary $V_{\text{in}}$ and different supplies.](image)

In this case, each sub-pump parameters need to be modified to reflect the effect of the supply scaling factor $k_i$ as follows,

$$
\alpha_{ai} = \frac{1}{\beta_i} \left( \alpha_{i1} + \alpha_{2i} + \alpha_{i1} \alpha_{2i} \right)
$$

(5.90)
\[ I_{oi} = k_i \beta_i \frac{V_{dd}}{R_i} \]  
(5.91)

\[ \beta_i = 1 - (1 + \alpha_{2i})(n_i + 1)\frac{V_{Di}}{k_n V_{dd}} \]  
(5.92)

Moreover, the weighted average forms of Equations (5.70) and (5.71) can be modified to accommodate \( k_i \) as follows,

\[ \eta_o = \frac{\sum_{i=1}^{m} k_i I_{dd} \eta_{oi}}{\sum_{i=1}^{m} k_i I_{ddi}} \]  
(5.93)

\[ \eta = \frac{V_{in} I_{out}}{V_{dd}} \frac{\sum_{i=1}^{m} k_i I_{dd} \eta_{oi}}{\sum_{i=1}^{m} k_i I_{ddi}} \]  
(5.94)

By using the sub-pump parameters given in Equations (5.90)-(5.92) and the weighted average form of Equation (5.94), the efficiency of a Hybrid Charge Pump with arbitrary \( V_{in} \) can be expressed on the same canonical form for efficiency such that,

\[ \eta = \beta \left( \frac{I_{out}}{I_{out} + \alpha_o I_o} \right) \left( 1 - \frac{I_{out}}{I_o} \right) \]  
(5.95)

\[ \alpha_o I_o = \frac{\sum_{i=1}^{m} k_i n_i \alpha_{oi} I_{oi}}{V_{in} + \sum_{i=1}^{m} k_i n_i} \]  
(5.96)

\[ \frac{1}{I_o} = \frac{\sum_{i=1}^{m} k_i \beta_i n_i}{V_{in} + \sum_{i=1}^{m} k_i \beta_i n_i} \]  
(5.97)
\[
\beta = \frac{V_{in} + \sum_{i=1}^{m} k_i n_i}{V_{dd} + \sum_{i=1}^{m} k_i n_i} \frac{1}{1 + \alpha_{2i}}
\]

By multiplying Equations (5.95) and (5.96) we get an expression for \(\alpha_o\) as follows,

\[
\alpha_o = A \cdot B \cdot \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} k_i k_j \beta_n n_{ij} \frac{\alpha_o I_{oj}}{I_{oi}}}{\sum_{i=1}^{m} \sum_{j=1}^{m} \frac{k_i k_j \beta_n n_{ij}}{(1 + \alpha_{2i})(1 + \alpha_{2j})}}
\]

\[
A = \frac{\sum_{i=1}^{m} k_i n_i}{V_{in} + \sum_{i=1}^{m} k_i n_i}
\]

\[
B = \frac{\sum_{i=1}^{m} \frac{k_i \beta_n n_i}{1 + \alpha_{2i}}}{V_{in} + \sum_{i=1}^{m} \frac{k_i \beta_n n_i}{1 + \alpha_{2i}}}
\]

Notice that the A and B terms given by Equations (5.100) and (5.101) respectively are independent of the sub-pump \(I_o\) used to size its pumping capacitor value. Accordingly, to minimize \(\alpha_o\) and maximize \(\eta\), we need to minimize only the rightmost term only Equation (5.99), which is identical in form to the expression derived for \(\alpha_o\) previously in Equation (5.81) at \(V_{in}=0\). Conveniently, it can be shown that \(\alpha_o\) in Equation (5.99) is minimum when the same conditions of Equation (5.84) are upheld. Similarly, the different sub-pump capacitors can be sized relative to the first sub-pump capacitor based on the relation in Equation (5.86). However, for \(V_{in}>0\), we expect higher pump efficiencies than that for \(V_{in}=0\), because smaller \(\alpha_o\) values are attainable as per Equation (5.98). Table 5.3 summarizes the \(\eta_{max}\) values of a Hybrid Charge Pump with identical parameter values to the one previously explained, except for \(V_{in}=2.5V\). The largest \(\eta_{max}\)
value over the design space is $\eta_{opt}=52.4\%$ and occurs for $C_2/C_1=0.814$ and $C_3/C_1=0.703$. As expected, these are the same capacitor values that optimize the design efficiency for $V_{in}=0$. Moreover, the dependence of $\eta_{max}$ on the different capacitor values still maintains its same general shape of Fig. 5.21, except for a positive offset in the z direction.

Table 5.3. Hybrid Charge Pump peak efficiency versus different $C_2/C_1$ and $C_3/C_1$ values for $V_{in}=V_{dd}$.

<table>
<thead>
<tr>
<th>$C_2/C_1$</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
<th>1.2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>50.69%</td>
<td>50.73%</td>
<td>50.16%</td>
<td>49.38%</td>
<td>48.52%</td>
</tr>
<tr>
<td>0.5</td>
<td>51.37%</td>
<td>52.13%</td>
<td>52.06%</td>
<td>51.67%</td>
<td>51.13%</td>
</tr>
<tr>
<td><strong>0.7</strong></td>
<td>50.9%</td>
<td>52.14%</td>
<td><strong>52.41%</strong></td>
<td>52.28%</td>
<td>51.95%</td>
</tr>
<tr>
<td>0.9</td>
<td>50.09%</td>
<td>51.67%</td>
<td>52.2%</td>
<td>52.27%</td>
<td>52.1%</td>
</tr>
<tr>
<td>1.1</td>
<td>49.15%</td>
<td>51.01%</td>
<td>51.74%</td>
<td>51.97%</td>
<td>51.93%</td>
</tr>
</tbody>
</table>

Now, we explain a design methodology to optimize hybrid-type pumps. Our basic premise is that there exist $m$ charge pump types each with a voltage tolerance $V_{max}$ and peak efficiency $\eta_{max}$. Furthermore, we assume that voltage tolerance and peak efficiency are conflicting attributes, i.e. $V_{max_1} < V_{max_2} < ... < V_{max_m}$ and $\eta_{max_1} > \eta_{max_2} > ... > \eta_{max_m}$. We also assume that all sub-pumps are operate at the same frequency. The design procedure requires picking the pumping frequency ($f$), the number of sub-pumps ($m$), and the number of stages ($n$) and all sub-pumps pumping capacitor value ($C$). Once again, we start with a ($V_{out}$, $I_{out}$) specification. Ideally, we would like to use pump cells with maximum efficiency only. Thus, the first sub-pump used should be the sub-pump with $\eta_{max_1}$, followed by the sub-pump with $\eta_{max_2}$, and so on. First, we design the first sub-
pump for an output specification \((V_{\text{max},i}, I_{\text{out}})\), and choose the pumping capacitor value such that its peak efficiency current aligns with \(I_{\text{out}}\). Next, we size the pumping capacitors of the remaining pumps to align their peak efficiency currents with the first sub-pump, and choose their number of stages to meet the required \(V_{\text{out}}\) specification. This design procedure is outlined as follows,

1. First, we start by comparing \(V_{\text{out}}\) and \(V_{\text{max},i}\). If \(V_{\text{out}} < V_{\text{max},i}\), then \(m=1\) and a single charge pump is designed using the methodology previously outlined in Fig. 5.18 with a \((V_{\text{out}}, I_{\text{out}})\) specification. If \(V_{\text{out}} > V_{\text{max},i}\), then \(m>1\) and the first sub-pump is designed as outlined in Fig. 5.18 with a \((V_{\text{max},i}, I_{\text{out}})\) specification. Thus, \(n_1, f\) and \(C_1\) are calculated.

2. Now, we consider the next sub-pump \(i\), where \(2 \leq i \leq m\). First, we compare \(V_{\text{out}}\) and \(V_{\text{max},i}\). If \(V_{\text{out}} > V_{\text{max},i}\), the sub-pump \(i\) output voltage is chosen such that \(V_{\text{out}_i} = V_{\text{max},i}\). If \(V_{\text{out}} < V_{\text{max},i}\), this is the final sub-pump with \(V_{\text{out}_i} = V_{\text{out}}\), and the number of sub-pumps \(m=i\).

3. For \(2 \leq i \leq m\), the number of stages \(n_i\) is calculated by finding the number of stages needed by each sub-pump to meet \(V_{\text{out}_i}\) and then rounding it off to the nearest smaller integer.

\[
 n_i = \text{floor} \left( \frac{(1 + \alpha_{2i})(V_{\text{out}_i} - V_{\text{max},i} + V_{Dh})}{k_i V_{dd} - (1 + \alpha_{2i})V_{Dh} - I_{\text{out}_i}/fC_i} \right) 
\]  

(5.102)

4. The sub-pump capacitor \(C_i\) is sized based on Equation (5.86). Due to the rounding off in step 4, the final \(C_i\) sizes may need to be slightly scaled up to meet the final \(V_{\text{out}}\) specification.

5. The switches for the different sub-pumps are sized such that their on resistance \((r_{\text{on}})\) meets the SSL operation requirement and alleviate conduction losses, i.e. \(4r_{\text{on}}fC<<1\).

The above design methodology is summarized in the flow chart in Fig. 5.23.
So far, we have ignored the impact of pumping frequency on the charge pump peak efficiency. In effect, as $f_{\text{pump}}$ increases, the switch series on resistance ($r_{\text{on}}$) becomes more significant compared to $1/fC$. In which case, the pump's equivalent series resistance does not decrease at the same rate by which dynamic power loss increases, and the pump efficiency drops with $f_{\text{pump}}$. A better circuit model that incorporates conduction losses of a charge pump is shown in Fig. 5.24. Based on this model, the pump's output resistance is given by,

$$R_{\text{out}} = \frac{n}{(1 + \alpha_2) fC} + 4nr_{\text{on}}$$  \hspace{1cm} (5.103)

This is a much better approximation of the accurate resistance expression provided by Equation (5.17) than Equation (5.18) is.
Fig. 5.24. A switch-based pumping cell equivalent circuit including the switch on resistance.

The model assumes an ideal clock buffer and a 50% duty cycle clock. It can be seen that for pumping frequencies $f > > 1/4r_{on}C$, the pump output resistance is $\sim 4r_{on}$ as predicted from Equation (5.17). For the new circuit model, it can be shown that the circuit power efficiency attains the same functional form of Equation (5.43) but having the following parameters,

$$
\alpha_o = \frac{1}{\lambda^2} (\alpha_1 + \alpha_2 + \alpha_1\alpha_2)(1 + 4(1 + \alpha_2)fCr_{on})
$$

(5.104)

$$
I_o = \frac{CVdd}{1 + 4(1 + \alpha_2)fCr_{on}}
$$

(5.105)

$$
\lambda = 1 + \frac{(1 + \alpha_2)V_{in}}{nV_{dd}}
$$

(5.106)

As expected, the dependence of $\alpha_o$ on $f$ is now captured by Equation (104). As $f$ increases, $\alpha_o$ increases and lower peak efficiencies are attained. Furthermore, by substituting with $r_{on} = 0$ into Equations (5.104)-(5.106), we get Equations (5.44)-(5.46) as should be the case. It is noticed that $\alpha_o$ has a corner frequency inversely proportional to $C$ and $r_{on}$ such that $f_{3-db} = 1/4(1 + \alpha_2)r_{on}C$. This is verified by the plot in Fig. 5.25 where the calculated peak efficiency based on Equation (5.104) is plotted versus $f_{pump}$, assuming the following model parameters: $\alpha = 0.025$, $C = 8pF$, $r_{on} = 1K\Omega$, $n = 4$ and $V_{in} = V_{dd} = 2.5V$. 

116
Moreover, if the pumping frequency is further increased, there comes a point when the clock buffer can no longer switch fast enough to fully charge the lower plate parasitic capacitance, i.e. $f \gg 1/4\alpha_2 r_{on} C$, and the circuit looses its voltage pumping ability. For a more inclusive model, the clock buffer on resistance must be accounted for. If we again assume a 50% duty cycle clock and a buffer on resistance equal to $r_b$ per stage, then for $n$ identical cascaded stages, the buffer's conduction losses can be modeled using a single resistance $4r_b/n$ as shown in Fig. 5.26. For a more efficiency calculations versus $f$, this is used.

**Fig. 5.25.** Calculated charge pump peak efficiency versus the pumping frequency.

**Fig. 5.26.** A switch-based pumping cell equivalent circuit including the switch and buffer on resistance.
It can be shown that for this model and $V_{in}=0$, $\alpha_o$ is given by,

$$\alpha_o = \left( \alpha_1 + \alpha_2 + \alpha_1 \alpha_2 \right) \left[ 1 + 4(1 + \alpha_2) f C r_{on} \right] \left[ 1 + 4(1 + \alpha_1) f C r_b \right] - 16 f^2 C^2 r_b r_{on} \right)$$

(5.107)

Shown in Fig. 5.27 is the measured peak efficiency of a Hybrid Charge Pump consisting of 3 sub-pumps versus the pumping frequency. Overlaid on top of the measured data is the calculated peak efficiency based on the model Equation (5.107). The model parameters are fitted to the measurement data. The model parameters used are: $\alpha=0.09$, $C=8pF$, $r_{on}=1.5K\Omega$, $r_b=1.5K\Omega$, $n=4$ and $V_{in}=V_{dd}=2.5V$. It is noticed that the model and measurements are in good agreement. The buffer resistance $r_b$ also contributes to the roll-off of $\alpha_o$ with frequency and effectively lowers the peak efficiency corner frequency.

In the low frequency range, Equation (5.107) can be approximated as,

$$\alpha_o \approx \left[ 1 + 4 \left( (1 + \alpha_1) r_b + (1 + \alpha_2) r_{on} \right) f C \right] \left( \alpha_1 + \alpha_2 + \alpha_1 \alpha_2 \right)$$

(5.108)

Thus, the $\alpha_o$ corner frequency can be more accurately approximated with $f_{3-dB} \approx 1/(4(r_{on}+r_b)C)$.

Fig. 5.27. Measured and calculated Hybrid Charge Pump peak efficiency versus the pumping frequency.
5.4. Noise Analysis Model

Low output noise power is crucial for high-end applications requiring a large dynamic range, e.g. reference clocks and inertial sensing. Understanding the impact of different noise sources on the final pump output noise is important. As depicted in Fig. 5.28, there are 3 main sources of noise that need to be addressed in a charge pump design: 1) switch noise, 2) voltage supply noise, and 3) clock jitter. Here, we quantitatively identify the contributions of the first and second noise sources only, and qualitatively discuss the third noise source.

![Fig. 5.28. A switch-based pumping cell noise circuit model.](image)

First, we start by the charge pump switch noise. Transistor switches have both thermal and flicker noise contributions. We show in the following analysis that the noise power at the charge pump output evolves in time until it reaches a steady state value, the same way the pump output voltage does. Moreover, we show that this steady state thermal noise power depends only on the final load capacitor and does not depend on the number of pump stage, the pumping capacitor, or the switch on resistance. For simplicity we ignore all parasitic capacitances in our analysis.
Here, we derive an expression for the output noise power of a single pump cell considering only the thermal noise contribution of the switches. In principle, our same analysis can be applied to charge pumps with more stages. In the most general case, we assume different switch on resistance values and different pumping and load capacitor values. The switches thermal noise is assumed uncorrelated. Shown in Fig. 5.29 is the equivalent noise model of a simple 2 phase, voltage pump cell. For now, we ignore the load resistance. Initially, all capacitors are assumed to have no charge stored on them. By the end of the first clock phase, the mean square noise voltage sampled on $C$ is due to $r_{on1}$ and is given by,

$$\overline{v_{n1}^2} = \frac{KT}{C} \quad (5.109)$$

By the end of the second clock phase, the mean square noise voltage sampled on $C_L$ is due to $r_{on2}$ and charge sharing from $C$, and is given by,

$$\overline{v_{n2}^2} = \left( \frac{KT}{C} + \frac{KT}{C/L_{eq}} \right) \left( \frac{C}{C + C_L} \right)^2 \quad (5.110)$$

Where $C/L_{eq}$ is the equivalent capacitance of the series combination $C$ and $C_L$. For the following clock cycle, the noise voltage sampled on $C$ by the end of the first phase is again given by Equation (5.109). However, the noise voltage sampled on $C_L$, by the end of the second phase,
is given by Equation (5.110) plus an additional term to account for charge sharing from the previous noise voltage stored on $C_L$. Thus, the updated mean square noise voltage is given by,

$$\overline{v_{no}^2} = \left(\frac{KT}{C} + \frac{KT}{C/C_L//C_L}\right)\left(\frac{C}{C/C_L+1}\right)^2 \left[1 + \left(\frac{C_L}{C+C_L}\right)^2\right]$$  \hspace{1cm} (5.111)

Similarly, at the end of each new clock cycle, the stored mean square noise voltage on $C_L$ is updated to a larger value. At steady state, the final sampled noise voltage on $C_L$ is given by,

$$\overline{v_{no}^2} = \left(\frac{KT}{C} + \frac{KT}{C/C_L//C_L}\right)\left(\frac{C}{C/C_L+1}\right)^2 \left[1 + \left(\frac{C_L}{C+C_L}\right)^2 + \left(\frac{C_L}{C+C_L}\right)^4 + \ldots \right]$$ \hspace{1cm} (5.112)

Equation (5.112) can be rewritten on the following form,

$$\overline{v_{no}^2} = \left(\frac{KT}{C} + \frac{KT}{C/C_L//C_L}\right)\left(\frac{C}{C/C_L+1}\right)^2 \frac{1}{1 - \left(\frac{C_L}{C+C_L}\right)^2}$$ \hspace{1cm} (5.113)

Equation (5.113) can be read as the product of 3 terms. The first term accounts for the noise contributions of switches $r_{on1}$ and $r_{on2}$, the second term accounts for the charge sharing action between capacitors $C$ and $C_L$, and the third term accounts for the integration effect of the final pump capacitor which stores and accumulates noise samples over many clock cycles. Interestingly, after a few algebraic reductions, Equation (5.113) can be rewritten as follows,

$$\overline{v_{no}^2} = \frac{KT}{C_L}$$ \hspace{1cm} (5.114)

Equation (5.114) shows that the mean square noise of a single pump stage depends only on the final load capacitor ($C_L$) as is the case with a simple RC circuit. Moreover, it can be shown that this result still holds for charge pumps with multiple stages, where each has a different
pumping capacitor value, and the final noise power depends only on the value of $C_L$. This is better explained by referring to our charge pump circuit model in Fig. 5.11, consisting of an ideal transformer and a series resistance proportional to $1/\omega C$. If $f$ or $C$ varies, the pump's series resistance varies such that the noise power spectral density-bandwidth product remains the same. If the charge pump drives a resistive load, our noise analysis is not affected as the thermal output noise depends only on the value of $C_L$ and not on the value of the Thevinin equivalent resistance.

Because a charge pump circuit multiplies dc signals, the switches flicker noise contributes to the charge pump output noise as well. The pump's output flicker noise evolves in time in a manner similar to that shown for thermal noise; and the steady state value depends on the flicker noise of the individual switches. The flicker noise of a switch is inversely proportional to its area and a larger switch size results in lower output noise, at the cost of higher switching power loss. However, because switching resets the transistor noise accumulation, a lower flicker noise power spectral density is expected from switched transistors [65]. The flicker noise contribution of the switch can be analyzed in a similar manner to our thermal noise analysis, except that the final noise output depends on the switches sizing as well. The more stages used, the larger the output flicker noise power, however, the overall SNR improves because the signal power increases with the square of the number of stages, whereas noise power increases with the number of stages.

Next, we address the second noise source in charge pump circuits, namely, supply noise. We show here that the input supply noise is effectively low-pass filtered by the transfer function from the supply input to the pump output. The bandwidth of this low-pass filter depends on the pumping capacitor to load capacitor ratio ($C/C_L$), and the pumping clock frequency ($f_{\text{pump}}$). A larger $C_L$ and a smaller $f_{\text{pump}}$ result in an overall lower noise bandwidth.
In the most general case, a charge pump circuit should be treated as a linear time-variant system with the supply voltage \( (V_{dd}) \) as its input. To simplify our analysis, we focus on the sampled nature of the charge pump circuit, and hence we analyze the circuit as a discrete-time LTI system, with a sampling rate equal to \( 2f_{pump} \). In order for the discrete-time analysis to hold, we assume that the bandwidth of the input signal is limited to half the sampling rate as specified by the Nyquist sampling theorem. If the supply noise spectrum extends beyond its Nyquist bandwidth, noise folding occurs and a time-variant system model must be adopted to accurately predict these noise folding effects.

![Fig. 5.30 Equivalent discrete-time LTI model of a single pump cell.](image)

Shown in Fig. 5.30 is a simplified depiction of the proposed discrete-time LTI model, where \( V_{in} \) is the supply voltage, and \( V_{out} \) is the pump's output voltage. The switch resistance is assumed small enough so that the sampled input voltage settles to its final voltage within one sampling period. By applying the charge conservation principle, the following difference equation relating \( V_{out} \) and \( V_{in} \) is obtained,

\[
CV_{in}[n-1] + C_LV_{out}[n-2] = C(V_{out}[n] - V_{in}[n]) + C_LV_{out}[n]
\]

\[ (5.115) \]
By taking the Z transform of Equation (5.115), we find that the pump's z-domain transfer function can be expressed as follows,

\[
H(z) = \mu \frac{(1 + z^{-1})}{1 - (1 - \mu)z^{-2}}
\]

(5.116)

\[
\mu = \frac{1}{1 + C_L/C}
\]

(5.117)

Equation (5.116) represents the transfer function of a low-pass filter. The bandwidth of the filter depends on the ratio C/C_L. Shown in Fig. 5.31 is the magnitude squared of the frequency response plotted for different C/C_L values. Typically C_L is larger than C by orders of magnitude. First, we notice that the dc gain of the pump cell is independent of C/C_L and is equal to 2, as expected of voltage doublers. It is also noticed that the pump bandwidth decreases with larger C_L or smaller C. This is consistent with our understanding that the pump bandwidth is inversely proportional to its output resistance and load capacitance. Moreover, the charge pump bandwidth is proportional to f_{pump} and can be reduced by using a smaller f_{pump}.

![Fig. 5.31 The supply noise transfer function of a single stage charge pump with different C/C_L values.](image-url)
Because we are only interested in amplifying the signal at dc, we can choose to reduce the pump bandwidth arbitrarily. A smaller pump bandwidth effectively filters off the higher frequency supply noise, and results in lower output noise power. This however, negatively impacts the pump's start-up time, and longer periods are required for the pump to reach its steady state voltage. This may be a critical consideration for systems with a fast wake-up time requirement. To alleviate this trade-off, a higher $f_{\text{pump}}$ can be used at start-up for faster settling. And after reaching steady state, a lower $f_{\text{pump}}$ can be used for improved supply noise filtering.

Similarly, the supply noise transfer function for pumps with arbitrary number of stages can be derived. Shown in Fig. 5.32 is the magnitude squared of the frequency response for a pump having 1, 2, and 3 stages, and $C/C_L=10$. Naturally, the dc gain increases with the number of stages, since higher pump gains are expected with more stages. Also, the pump bandwidth decreases with a higher number of stages. This is because the pump's output resistance increases linearly with the number of stages. This means that, by increasing the number of pump stages, the supply high-frequency noise is better filtered, whereas the low-frequency noise is amplified as its dc input signal. Consequently, the supply flicker noise does not obtain much filtering from the pump, and is amplified by almost the same gain as the supply voltage. Therefore, the pump output maintains the same SNR as its input when considering only very low frequency noise contributions. Thus, for improved flicker noise at the output, any necessary noise filtering has to be performed before applying the supply voltage to the pump input. Typically, the pump's supply voltage is generated through a bandgap circuit and an output buffer. Flicker noise reduction techniques such as chopper stabilization and/or correlated double sampling can be used in these circuits. Moreover, a large noise filtering capacitor can be used at the output of the voltage supply buffer to limit its noise bandwidth to below $f_{\text{pump}}$ to prevent noise folding.
Interestingly, it is found that, for pumps with an odd number of stages, the noise transfer function has a null at $f_{\text{pump}}$ and multiples thereof. This useful property can be used to reject all supply spurious tones occurring at $f_{\text{pump}}$ and its harmonics.

Finally, we qualitatively explain the impact of the third noise source on charge pumps, namely, clock jitter. It can be shown that, when accounting for the switch on resistance, the pump output voltage depends on the clock duty cycle ($D$) [66]. Assuming the switch resistance is $r_{\text{on}}$, the top plate parasitic capacitance is $\alpha_2 C$, and an ideal clock buffer is used, the output voltage of a single pump cell is related to the load current as follows,

$$V_{\text{out}} = V_{\text{in}} + \frac{V_{\text{dd}}}{1 + \alpha_2} \left( \frac{1}{(1 + \alpha_2) f C} + \frac{r_{\text{on}}}{D(1 - D)} \right) I_{\text{out}}$$  \hspace{1cm} (5.118)

Equation (5.118) indicates that uncertainties in the clock frequency and/or duty cycle modulate the pump's output resistance; this in turn modulates the pump's output voltage. Consequently, the pumping clock phase noise translates into voltage noise at the pump's output. More accurate clock references with faster rise and fall transitions improve the pump's noise.
performance. Equation (5.1.18) can be used to derive an upper bound on the clock jitter requirements to meet a given output noise specification.

### 5.5. Design Examples

In this section, we explain 2 hybrid charge pump designs targeted for both positive and negative output voltages in 65nm bulk CMOS technology. Because of the reverse breakdown of the well diodes, conventional charge pump designs in this technology are limited to 12V for positive output voltage pumps and to -10V for negative output voltage pumps.

#### 5.5.1. A +36V Hybrid Charge Pump

This Hybrid Charge Pump design is composed of a cascade of 3 smaller sub-pumps with different voltage ranges as shown in Fig. 5.33. Each sub-pump consists of 4 identical cascaded stages and is capable of boosting positive output voltages. For the first sub-pump, the unit cell is a CMOS 4PVD identical to the one depicted in Fig. 4.2(a). The output voltage of this sub-pump is limited to 12V. For the second sub-pump, the unit cell is an all-NMOS 6PVD identical to the one depicted in Fig. 4.6(c). All the NMOS devices in the second sub-pump share the same bulk connection that is shorted to the deep nwell tie. The deep nwell is biased from the output voltage of the first sub-pump. The output voltage of this sub-pump is limited to 20V. The third sub-pump is a polysilicon diode-based Dickson pump. To improve the efficiency per stage and reduce the number of stages, the sub-pump is driven by reliable doubly-stacked clock drivers with \(2V_{dd}\) swing as depicted in Fig. 4.10. The polysilicon diodes deep nwell is biased from the first sub-pump’s output voltage. This bias improves the FOX time-to-failure by \(~5000\)x based on the square-root \(E\) TDDDB model equation explained in Chapter 2. The output voltage of this sub-pump is limited to 100V. To optimize the design power efficiency, we want to align peak
efficiency load currents for all sub-pumps at the same 20uA value. Power analysis indicates that choosing identical pumping capacitors for all sub-pumps results in near-optimum performance. Two unit pumping capacitors of 4pF are used per cell.

![Diagram of Sub-pumps](image)

**Fig. 5.33. A +36V Hybrid Charge Pump design.**

To further optimize the design power and area efficiency, metal capacitors with gradually tapered finger spacing are used based on the data in Fig. 2.5. In this technology node, the minimum metal pitch is ~100nm. Metal finger capacitors with 1x finger spacing are used in this first sub-pump. Metal finger capacitors with 1.5x and 2x finger spacing are used in this second and third sub-pumps respectively. Since the $\phi_1$ and $\phi_2$ phases drive the main pumping capacitors in each sub-pump, they suffer from the largest plate parasitics power losses. Consequently, the $\phi_1$ and $\phi_2$ clock drivers of the first and second sub-pumps perform charge recycling via tri-state buffers and a charge-equalization switch as explained in Fig. 4.16. This charge recycling is not applied for the third sub-pump because the $2V_{dd}$ swing of its clock drivers stresses the charge-equalization switch. Maintaining a fixed timing relationship between all clock phases avoids
power loss due to reverse current and maintains device reliability. A PVT robust clock generation circuit is implemented as proposed in Fig. 4.13. All delay elements are matched and placed in close proximity. Dummy gates are used to equalize all the loads.

Only thick-oxide devices with 2.5V±10% voltage rating are used. For the voltage doublers, the transistors gate-source and drain-source potentials, by design, can not exceed the pumping clock amplitude. To maintain the reliability of all devices, the supply voltage is constrained to $V_{dd}<2.75V$. For the Dickson pump, the diodes are exposed to roughly twice the clock amplitude when switched off. Because doubly-stacked clock drivers are used in the Dickson sub-pump, the diodes must be able to sustain a reverse voltage as high as $4V_{dd}$. Accordingly, the polysilicon diodes are designed with $L_i=0.5\mu m$ based on the data in Fig. 3.10(a).

### 5.5.2. A -29V Hybrid Charge Pump

As explained in Chapter 3, double-diode substrate isolation can not be applied to negative-type pumps. Accordingly, this Hybrid Charge Pump design is composed of a cascade of only 2 sub-pumps as shown in Fig. 5.34. Each sub-pump consists of 4 identical cascaded stages and is capable of boosting negative output voltages. For the first sub-pump, the unit cell is an all-NMOS 6PVD identical to the one depicted in Fig. 4.6(b). All NMOS device in the first sub-pump have a separate bulk connection and not shorted to the deep nwell tie. The deep nwell tie is grounded instead. The output voltage of this sub-pump is limited to 10V. The second sub-pump is an improved-drive polysilicon diode Dickson pump as depicted. To turn the positive output voltage design in Fig. 4.10 to a negative output voltage pump, we simply swap the pump's input and output terminals. In this sub-pump, the polysilicon diodes deep nwell is grounded while the pwell is biased from the first sub-pump’s output to improve its FOX time-to-failure. To optimize
the design power efficiency, we want to align peak efficiency load currents for both sub-pumps at the same 20uA value. Identical pumping capacitors for both sub-pumps are chosen. Two unit pumping capacitors of 4pF are used per cell.

![Diagram of Hybrid Charge Pump design]

Fig. 5.34. A -29V Hybrid Charge Pump design.

Also, metal capacitors with gradually tapered finger spacing are used. Metal finger capacitors with 1x, and 2x finger spacing are used in the first and second sub-pumps respectively. The φ₁ and φ₂ clock drivers, in the first sub-pump only, use tri-state buffers and a charge-equalization switch to perform charge recycling. The same clock generation circuit explained in Fig. 4.13 is used to provide the different sub-pump clock phases.

5.6. Summary

In this chapter, we have introduced an improved efficiency, extended voltage-range Hybrid Charge Pump architecture. A Hybrid Charge Pump design optimally mixes cells having extended voltage range with cells having improved power efficiency to reap both benefits. To help explore
the design space and optimize the efficiency of hybrid pumps, we proposed an accurate power analysis model. The model expresses the hybrid pump efficiency as a weighted average of the individual sub-pump efficiencies. Based on the model, we can estimate the optimum number of stages and pumping capacitor value for each sub-pump. We have also proposed a noise analysis model to estimate the final charge pump output noise power. To minimize the final output noise power, we need to increase either/both the output load capacitance, and the pump output resistance. This effectively limits the pump's equivalent noise bandwidth but increases the pump's start-up time. Finally, we have walked through 2 Hybrid Charge Pump designs targeting different voltage polarities. The measurement results of these pumps are discussed in Chapter 7.

This chapter concludes the first part of the dissertation addressing high-voltage generation methods. In the next chapter, we switch to the topic of high-voltage drive. For high-voltage drive, we leverage some of the concepts and techniques explained so far, and introduce some new methods to enable high-voltage switching in nanometer-scale CMOS technology nodes.
CHAPTER 6

High-Voltage Drive

6.1. Introduction

As motivated in Chapter 1, high-voltage waveform signals are needed in many applications. In this chapter, we introduce 2 types of high-voltage output stages that are compatible with low-voltage CMOS technology while targeting different requirements. The first driver is a stacked-device type driver suitable for fast switching applications. Device stacking allows extended voltage range outputs beyond the voltage tolerance of a single device; but requires an intricate predriver circuit to maintain the stacked devices reliability. The predriver circuit is often non-scalable and uses large passive components. In this chapter, we introduce compact doubly- and triply-stacked drivers in 65nm CMOS technology capable of 5V and 7.5V swings respectively, and running at >200MHz. Because the complexity of the predriver circuit often limits the number of stacked devices to few devices only, we introduce a second type of output drivers that is based on charge pumps. The Charge Pump-Based architecture is suitable for slow switching applications (<10MHz), but enables seamless and reliable stacking of 10's of devices. Bipolar >40V output swings in 45nm SOI CMOS technology are possible using this driver.

6.2. Stacked-Device Drivers

One common technique to extend a circuit's high-voltage capability is to stack multiple devices in series [67]-[68]. Device stacking allows for higher voltage drops to divide equally across more devices, so that no single device is stressed. In principle, the achievable drive level
is proportional to the number of stacked devices and is eventually bounded by the substrate voltage tolerance. In practice, stacking is limited by the rapidly growing complexity of a reliable driver circuit. One simple implementation of a stacked-device driver is presented in [60]. However, such an implementation maintains device reliability at the steady state drive levels only and does not account for the transient voltage stress during switching transitions. Transient voltage stress may degrade device characteristics and reduces the circuit’s mean time to failure. Other stacking architectures that address reliability during switching transitions are proposed in [46]-[47]. The stacked-device driver in [46] invests in an intricate predriver circuit to provide the required controls for a reliable device switching. However, the predriver requires capacitors and resistors whose values need to be tuned within the desired tolerance to guarantee reliability. Also, the design is not scalable, and a new predriver design is needed for different numbers of stacked devices. While the driver design in [47] is scalable, it uses diodes and resistors to maintain device reliability. And, larger resistors are needed to reduce the off-current leakage.

Fig. 6.1. (a) A doubly-stacked driver (b) high-level drive (c) low-level drive.
In principle, stacking 2 devices as shown in Fig. 6.1(a), enables output swings up to 2x the transistor nominal voltage rating ($V_{dd}$). Simple stacking reduces the drain-source voltage of the devices, whereas gate-oxides are still subject to stress. Hence, a level-shifted gate drive is necessary to maintain the reliability of the stacked devices. Consequently, the gate potential of $M_{n1}$ is constrained to swing between ground and $V_{dd}$, the gate potential of $M_{p1}$ is constrained to swing between $V_{dd}$ and $2V_{dd}$ and the gates of $M_{n2}$ and $M_{p2}$ are always connected to $V_{dd}$. All bulk terminals are not shown and assumed connected to their respective source terminal. Simple device stacking as such, poses reliability issues during switching transitions. For instance, during the low-to-high transition, as $M_{n2}$ turns off, its source potential initially rises to one threshold voltage below the gate potential (i.e. $V_{dd}-V_T$), then sub-threshold conduction slowly charges this node close to $V_{dd}$. Hence, during the low-to-high transition, the drain-source potential of $M_{n2}$ is momentarily stressed to $V_{dd}+V_T$ and then slowly recovers to $V_{dd}$. Furthermore, $M_{p2}$ is also stressed during this transition. As the gate potential of $M_{p1}$ falls to $V_{dd}$, its drain potential starts charging to $2V_{dd}$ instantly, however, the drain potential of $M_{p2}$ does not start charging until its source potential is at least one $|V_T|$ above $V_{dd}$. Again, this results in overstressing the drain-source voltage of $M_{p2}$ by $\sim V_T$. Similar stress events can be demonstrated for $M_{n2}$ and $M_{p2}$ during the high-to-low transitions. These scenarios are better depicted by the schematics in Fig. 6.1(b) and 6.1(c). The devices in black are on devices while the ones in grey are off devices.

To better quantify the significance of these transient stress events, we observe that transistors are often rated by foundries to handle voltages exceeding the nominal supply voltage by 10%. Threshold voltages in modern CMOS technologies typically constitute 25-35% of the nominal supply voltage. The impact of such transient stress on device lifetimes depends on the driver activity factor and switching frequencies. Higher switching activity and frequencies exacerbate
the problem. Using a lower supply voltage can possibly alleviate such stress risks. Supply-voltage reduction, however, requires a larger device stack to attain the same drive levels, resulting in slower and less efficient drive.

In the next 2 sections, we present a compact stacked-device driver design in 65nm CMOS technology. Doubly- and triply-stacked drivers capable of 2x and 3x voltage swings while maintaining device reliability at switching transitions are introduced. No passive components are needed; instead, *assisted-charging* transistors enable an area-efficient implementation.

6.2.1. A Doubly-Stacked Output Driver

In this section, we propose the doubly-stacked output driver shown in Fig. 6.2. One key feature of the proposed circuit is to avoid the slow charging and discharging intermediate nodes $n_1$ and $p_1$. These nodes are slow due to sub-threshold conduction of stacked devices $M_{p2}$ and $M_{n2}$. Towards that end; we add four extra transistors to the device stack. We call these devices assisted-charging devices because they provide an enhanced conduction path to the intermediate slow nodes in the stack. Two assisted-charging devices of type PMOS are used in the pull-down network. These are shown in blue and are necessary to maintain the reliability of device $M_{n2}$. Similarly, two assisted-charging devices of type NMOS are used in the pull-up network. These devices are shown in red and are necessary to maintain the reliability of device $M_{p2}$.

For the low-to-high transition, first, the gate potential of $M_{n1}$ drops to ground turning $M_{n1}$ off and $M_{p3}$ enters saturation. As a result, node $n_1$ now charges almost instantly to its steady state value of $V_{dd}$ through $M_{p3}$ as opposed to the slow sub-threshold conduction of $M_{n2}$. Next, the gate potential of $M_{p4}$ rises to $2V_{dd}$ turning $M_{p4}$ off and the output node charges to $2V_{dd}$ through the pull up stack. For the high-to-low transition, first the gate potential of $M_{p4}$ drops to $V_{dd}$ and $M_{p4}$
enters saturation. As a result, the drain of M_{n2} is rapidly discharged to V_{dd}+V_T through M_{p4}. Next, the gate potential of M_{n1} rises to V_{dd} turning M_{n1} on, M_{p3} off and discharges nodes n_1 and the output to ground. And even though M_{n2} does not discharge its drain until node n_1 potential is at least one V_T below V_{dd}, M_{n2} is not stressed because its drain potential has already been reduced from 2V_{dd} to V_{dd}+V_T. Thus, the drain-source potential of M_{n2} does not exceed 2V_T during this transition. Similarly, devices M_{n3} and M_{p4} play an identical role to maintain the reliability of device M_{p2} during switching transitions. The necessary level shifting is implemented using clock boosting circuits as shown in Fig. 6.2. The clock phases timing diagram is also shown in figure. Note that two clocks with identical frequencies and different duty cycles are needed to maintain the described switching order. Fig. 6.3(a) annotates the steady-state node potentials for a high-level drive scenario. The stacked NMOS devices are off, stacked PMOS devices are on and V_{out} = 2V_{dd}. Device M_{p3} is the only assisted-charging device that is on and connects node n_1 to V_{dd}. Fig. 6.3(b) depicts the situation for a low-level drive scenario. The device sizing of the doubly-stacked driver is summarized in Table 6.1, and is chosen such that the drive on-resistance is 12\Omega.

![Fig. 6.2. The proposed doubly-stacked driver and clock phases timing diagram.](image)
Fig. 6.3. The proposed doubly-stacked driver operation: (a) high-level drive (b) low-level drive.

Table 6.1. Doubly-stacked driver device sizes.

<table>
<thead>
<tr>
<th>Device</th>
<th>Size (µm/µm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{n1}$-$M_{n2}$</td>
<td>160/0.28</td>
</tr>
<tr>
<td>$M_{p1}$-$M_{p2}$</td>
<td>400/0.28</td>
</tr>
<tr>
<td>$M_{n3}$</td>
<td>80/0.28</td>
</tr>
<tr>
<td>$M_{n4}$</td>
<td>16/0.28</td>
</tr>
<tr>
<td>$M_{p3}$</td>
<td>200/0.28</td>
</tr>
<tr>
<td>$M_{p4}$</td>
<td>40/0.28</td>
</tr>
</tbody>
</table>

Simulations focus on verifying that no voltage difference across any 2 device terminals exceeds the maximum rating for a given load and switching frequency. Only thick oxide I/O devices with voltage tolerance of 2.5V+10% are used. Shown in Fig. 6.4 are the simulated transient voltage waveforms for the doubly-stacked driver in response to a 200MHz, 2.5V input square wave. The driver is loaded with a 1pF capacitance. It is shown that for high-to-low
transitions, because of the assisted-charging devices, nodes n₁ and p₁ settle quickly to ground and \( V_{dd} \) potentials, respectively. Similarly, for low-to-high transitions these nodes settle quickly to \( V_{dd} \) and \( 2V_{dd} \) potentials. By plotting the \( V_{ds} \) and \( V_{gs} \) transient waveforms of a device versus one another, a signature curve is produced as shown in Fig. 6.5. This curve is useful to visualize transient stress across the device terminals. The curve in Fig. 6.5 is for device Mₙ₂ in a doubly-stacked driver. Two cases are depicted, one in which no assisted-charging devices are used (red) and another in which assisted-charging devices are used (blue). It is shown that for the non-assisted-charging case, device Mₙ₂ is subjected to hot-carrier degradation, while for the assisted-charging case, the gate-source and drain-source potentials do not exceed 2.75\( V \) for all times. Device stress is further reduced in the case of a resistive load due to the potential divider arising between the load and the switch resistance. Though the \( V_{ds}-V_{gs} \) locus plot is shown for device Mₙ₂ only (topmost stack device), all the individual driver devices are tested for reliability in a similar manner.

![Fig. 6.4. Transient waveforms of the doubly-stacked driver.](image-url)
Fig. 6.5. $V_{ds}$-$V_{gs}$ locus plot for device $M_{a2}$ in the doubly-stacked driver.

Fig. 6.6. (a) A triply-stacked driver (b) high-level drive (c) low-level drive.
6.2.2. A Triply-Stacked Output Driver

To enable higher voltages, a triply-stacked output driver is shown in Fig. 6.3. Similar to the doubly-stacked case, this driver suffers from drain-source stresses upon switching transitions. As shown in Fig. 6.6, the gate potential of $M_{n1}$ is constrained to swing between ground and $V_{dd}$, the gate potential of $M_{p1}$ is constrained to swing between $2V_{dd}$ and $3V_{dd}$ and the gate potentials of $M_{n3}$ and $M_{p3}$ are constrained to swing between $V_{dd}$ and $2V_{dd}$. Also, the gates of $M_{n2}$ and $M_{p2}$ are always connected to $V_{dd}$ and $2V_{dd}$, respectively. Notice that devices $M_{n3}$ and $M_{p3}$ are momentarily stressed during switching transitions as depicted by the schematics in Fig. 6.6(b) and 6.6(c).

A reliable triply-stacked driver is shown in Fig. 6.7. Assisted-charging devices are used to enhance the settling speed of the slow node potentials. Four assisted-charging PMOS devices (in blue) maintain the reliability of $M_{n2}$ and $M_{n3}$, whereas 4 assisted-charging NMOS devices (in
red) maintain the reliability of $M_{p2}$ and $M_{p3}$. For the low-to-high transition, first, the gate potential of $M_{n1}$ drops to ground turning $M_{n1}$ off and $M_{p4}$ enters saturation charging node $n_1$ instantly to its steady state value of $V_{dd}$. Next, the gate potential of $M_{p5}$ and source potential of $M_{p6}$ rise to $2V_{dd}$ turning $M_{p5}$ off and $M_{p6}$ on, and node $n_2$ charges instantly to $2V_{dd}$. Finally, the gate potential of $M_{p7}$ rises to $3V_{dd}$ turning $M_{p7}$ off and the output node charges to $3V_{dd}$ through the pull up stack. For the high-to-low transition, first the gate potential of $M_{p7}$ drops to $V_{dd}$ and $M_{p7}$ enters saturation discharging node $n_2$ to $2V_{dd}$ through $M_{p7}$. Next, the gate potential of $M_{p5}$ and source potential of $M_{p6}$ drop to $V_{dd}$ turning $M_{p5}$ on, $M_{p6}$ off and discharges nodes $n_2$ and the output to $V_{dd}+V_T$. Finally, the gate potential of $M_{n1}$ rises to $V_{dd}$ turning $M_{n1}$ on, $M_{p4}$ off, and discharges nodes $n_1$, $n_2$ and the output to ground. Devices $M_{n4}$-$M_{n7}$ play an identical role in to maintain the reliability of the stacked devices $M_{p2}$ and $M_{p3}$. Gate drive for devices $M_{p7}$ and $M_{n7}$ requires $2V_{dd}$ swings and thus doubly-stacked drivers are used. The necessary level shifting and predrive circuitry are also shown in Fig. 6.7. The timing diagram for the different clock phases is depicted by Fig. 6.8. Note that 3 clocks with identical frequencies and different duty cycles are needed. The unassigned generated clock phases can be used to drive a similar triply-stacked stage for fully differential drive. The device sizing of the driver is summarized in Table 6.2, and is chosen such that the drive on-resistance is $12\Omega$.

![Fig. 6.8. Clock timing diagram for the triply-stacked driver.](image)
Table 6.2. Triply-stacked driver device sizes.

<table>
<thead>
<tr>
<th>Device</th>
<th>Size (μm/μm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M_{n1}-M_{n3})</td>
<td>240/0.28</td>
</tr>
<tr>
<td>(M_{p1}-M_{p3})</td>
<td>600/0.28</td>
</tr>
<tr>
<td>(M_{n4}, M_{n6})</td>
<td>120/0.28</td>
</tr>
<tr>
<td>(M_{n5}, M_{n7})</td>
<td>24/0.28</td>
</tr>
<tr>
<td>(M_{p4}, M_{p6})</td>
<td>300/0.28</td>
</tr>
<tr>
<td>(M_{p5}, M_{p7})</td>
<td>60/0.28</td>
</tr>
</tbody>
</table>

Shown in Fig. 6.9 are the simulated transient voltage waveforms for the triply-stacked driver in response to a 200MHz, 2.5V input square wave. The driver is loaded with a 1pF capacitance. Again all intermediate nodes settle rapidly to their designated stress-free levels. The \(V_{ds}-V_{gs}\) locus plot for device \(M_{n3}\) is shown in Fig. 6.10. It can be seen that with no assisted charging \(M_{n3}\) is subjected to a severe hot carrier degradation that a simple supply lowering can not relieve. The proposed assisted-charging devices, however, completely alleviate this issue.

![Fig. 6.9. Transient waveforms of the triply-stacked driver.](image-url)
The proposed output driver using assisted-charging devices can in principle be applied to even taller stacks by introducing more drive levels and timing constraints. The complexity of the circuit, however, grows very rapidly, with the number of stacked devices. This makes theses drivers impractical when targeting voltage ranges >20V. In the next section, we introduce a voltage scalable output stage. The output stage is charge pump based, so it is inherently slow, yet it enables modular and reliable stacking of 10's of devices easily.

6.3. Charge Pump-Based Drivers

In this section, we present a high-voltage output stage producing signals well beyond the voltage ratings of standard devices in a nanometer-scale CMOS technology. A novel, bidirectional, switched capacitor output stage combining both voltage conversion and pulse drive is introduced. The design is highly modular and enables extensive stacking of standard devices seamlessly and with little overhead. Consequently, output voltage ranges are only limited to the substrate voltage tolerance rather than the availability of HV devices, or predriver complexity.
One common realization of an on-chip high-voltage driver is shown in Fig. 6.11. A charge pump circuit steps up the chip supply voltage to the desired output drive level (HV<sub>dd</sub>), and an output stage performs the final load switching. Since the full HV<sub>dd</sub> voltage difference appears across the output devices, high-voltage-tolerant switches are required in the output stage. Switches with a high blocking-voltage are implemented using specialized LDMOS devices, [38] incurring an extra mask cost, or by stacking standard devices [42]-[46], and [69]. Stacked-device output stages are capable of driving large current loads (8-50Ω) but require a carefully designed predrive circuitry. The stacked devices require different gate drive levels, voltage swings, and must maintain a switching order. The predrive circuit is designed to guarantee device reliability by monitoring the gate-source, drain-source, and gate-drain transient waveforms, and validating that the devices maintain their reliability limits at different process corners, temperatures and supply voltages. As more constraints are imposed on taller device stacks, the design complexity increases rapidly with the number of devices. As a result, stacked-device drivers are typically limited to short stacks consisting of 3-5 devices only.

![Fig. 6.11. High-voltage drive: (a) conventional (b) proposed architecture.](image)
Fig. 6.12. High-voltage drive: (a) conventional (b) proposed architecture.

To simplify the drive problem, a voltage-scalable output stage is introduced [70]. Shown in Fig. 6.12 is a high-voltage driver architecture composed of a cascade of identical stages. Each stage is a switched capacitor circuit capable of bilateral charge transfer, to and from the load. Because capacitors are available within a stage, both voltage pumping and voltage switching functionalities can be integrated into the same block. The number of stages depends on the final output drive level. In the pump-up mode, an up/down control signal is high and the driver is configured as a forward pump, gradually building towards its final output voltage. Conversely, in pump-down mode, the up/down signal is low and the driver is configured as the reverse pump, gradually discharging the load in a similar fashion. For both pump up and pump down modes, charge transfer is quantized and well-controlled which guarantees device reliability. As with charge pumps, the timing and amplitude of all voltage transitions are well controlled and depend on the pumping clock swings. Since voltage swings for all transistor terminals cannot exceed $V_{dd}$, by carefully timing a stage's clock phases, all devices are guaranteed reliability for $V_{dd}$ within the transistor voltage rating. Moreover, because this is a switched capacitor circuit, the output drive resistance is tunable through the pumping frequency $f_{pump}$. 
Due to design modularity, maintaining the reliability of a single stage is sufficient to guarantee the reliability of the whole driver, independent of the number of stages. A direct consequence of this feature is the ability to stack 10's of devices reliably in a scalable fashion. Because the predriver design complexity is alleviated, the stack length becomes limited by the technology's voltage-handling capability. When using an advanced technology with a voltage-tolerant substrate like SOI CMOS, voltages >40V are feasible [71]. Due to the charge-pump nature of the circuit, the output current drive is limited by the stage capacitor size and pumping frequency. The stage output resistance is also proportional to the number of stages. To enable large output currents (>10mA), large off-chip capacitors must be used. Since for high-voltage outputs, multiple stages are cascaded, the number of off-chip capacitors can be prohibitive. Moreover, if we double the number the stages and want to maintain the same drive resistance, the total capacitor sizing quadruples. Thus, this drive architecture is only useful for applications requiring extended device stacking but with low current drive requirements.

In the next 2 subsections, we discuss circuit implementation details of the driver stage for 2 types of output drive, namely, unipolar and bipolar outputs.

6.3.1. Unipolar Output Drive

A general depiction of a driver single stage is shown in Fig. 6.13. Focusing only on the top half-circuit, we have an NMOS charge-transfer switch connected to the stage input (lower voltage side), a PMOS charge-transfer switch connected to the stage output (higher voltage side), and a pumping capacitor C connected in between. The switches operate on 2 complementary phases. During one clock phase, the NMOS switch turns on and the capacitor stores the input voltage (V_{in}). During the second clock phase, the PMOS switch turns on and the capacitor boosts V_{in} by one clock supply (V_{dd}) developing the stage output voltage such that V_{out} = V_{in} + V_{dd}. This
is the pump up mode. For the pump-down mode, this switching order is reversed for. Because the capacitor is needed to step up the voltage in the charging mode only, its bottom plate potential is switched between $V_{dd}$ and ground in this mode, and is tied to ground in the discharge mode. A phase assortment circuit is assumed to control the relationship between the different clock phases for the pump-up and pump-down modes. For complete charging, the gate potential of the NMOS switch needs to exceed $V_{in}$ by $V_{dd}$ to provide an adequate switch overdrive. Similarly, for a complete discharge, the gate potential of the PMOS device needs to be lower than $V_{out}$ by $V_{dd}$. To optimize the use of the complementary clock phases from the clock level-shifting circuits, an identical switch-capacitor arrangement operates in a ping-pong fashion with the top half-circuit.

Since different gate drive levels are required, the switches and pumping capacitor must be driven independently. This decoupling of the clock phases controlling the switches and the pumping capacitor enables improved power efficiency as explained shortly. Clock phases $\Phi_a$ and $\Phi_b$ drive the NMOS gates, phases $\Phi_c$ and $\Phi_d$ drive the PMOS gates, and phases $\Phi_1$ and $\Phi_2$ drive the pumping capacitors bottom plates. One possible implementation of a phase assortment circuit that provides the required level-shifted gate drive signals is also shown in Fig. 6.13. The gate drive circuit uses cross-coupled level-shifting circuits with capacitors. Two non-overlapping phases boosted by $V_{dd}$ are generated using a cross-coupled NMOS pair tied at the source to $V_{in}$ and a pair of capacitors ($C_s$). Similarly, 2 overlapping phases suppressed by $V_{dd}$ are generated using a cross-coupled PMOS pair tied at the source to $V_{out}$ and a pair of capacitors ($C_s$). Signals $\Phi_a$ and $\Phi_c$ are both in phase, but are out of phase with $\Phi_1$. All six signals are derived from one master clock, and all driver mode control is built on the low-voltage side. The final unit stage implementation is shown in Fig. 6.14.
The overall driver block diagram is shown in Fig. 6.15. The driver consists of 5 stages to produce 12V outputs. Higher power efficiency is achievable by connecting the first stage input to
the up/dn control signal. A clock generation circuit derives all 6 phases from a master clock and routes them to all stages. Only phases \( \phi_1 \) and \( \phi_2 \) are gated by the up/dn control before being routed to the driver stages.

![Clock and Phase Generator Diagram](image)

Fig. 6.15. Overall driver block diagram.

Decoupling the switches gate drive enables a break-before-make switching operation, reducing reverse conduction losses and improves efficiency. The phases timing relationship within a single stage is shown in Fig. 6.16(a). Three different duty cycles are needed. Phase \( \phi_a \) has the smallest duty cycle while phase \( \phi_c \) has the largest. The \( \phi_1 \) rising edge occurs between the \( \phi_a \) and \( \phi_c \) falling edges and vice versa. The diagram in Fig. 6.16(b) illustrates the timing relationship between the up/dn control, \( \phi_1 \) and \( \phi_2 \) phases, and the driver output. In pump-up mode, phases \( \phi_1 \) and \( \phi_2 \) are active, and the driver charges the output load at a time constant \( \tau \) depending on the driver resistance and the load capacitance. This time constant sets the maximum output waveform switching frequency \( (f_{out}) \). When configured for pump-down, \( \phi_1 \) and \( \phi_2 \) are inactive and the driver gradually discharges the load at the same time constant \( \tau \).

Only thick-oxide devices with 2.5V±10% voltage rating and metal finger capacitors are used. All transistors have their source and bulk terminals shorted. The current drive capability of a
stage depends on values of $f_{\text{pump}}$ and $C$. For a given $f_{\text{pump}}$, larger $C$ results in higher output currents. The gate drive capacitor ($C_s$) does not contribute to the load current and can be made much smaller than $C$. In our design, $C$ and $C_s$ are chosen to be 3.8pF and 380fF, respectively. As $f_{\text{pump}}$ increases the driver resistance drops and becomes eventually limited by the transistors series on resistance. $M_1$-$M_4$ transistor sizing sets a lower bound on the stage drive resistance. However, oversized switches lead to reduced power efficiency as their diffusion capacitance becomes comparable to the capacitors plate parasitics. Again, $M_5$-$M_8$ device sizing can be made considerably smaller than $M_1$-$M_4$ since they do not accommodate any load currents. In our design, device sizing is chosen as follows: 16µm/0.28µm for $M_1$-$M_2$, 32µm/0.28µm for $M_3$-$M_4$, 2µm/0.28µm for $M_5$-$M_6$, and 4µm/ 0.28µm for $M_7$-$M_8$.

Finger capacitors with minimum spacing (~100 nm) are used. Based on our TDDB model, the capacitors are estimated to sustain ~12V for a cumulative failure rate of less than 0.01% over 10 years at 85°C.

![Diagram](image_url)

Fig. 6.16. (a) Clock phases relationship (b) Driver clock timing diagram.
6.3.2. Bipolar Output Drive

To enable bipolar output drive a different type of driver stage is needed. Shown in Fig. 6.17 is a unit driver stage that is capable of bipolar output signals. The stage consists of 2 identical, yet, oppositely-oriented TPVD's used for bilateral charge transfer. Because of the 2 separate charge-transfer paths and the complementary nature of the drive levels, only one path is active per stage at a time. A control signal gates the clocks in each charge-transfer path to ensure a mutually exclusive up/down operation. However, because one path only is active at a time, while the capacitors in the active path are continually refreshed, those of the idle path are slowly discharging. This results in voltage stresses across the transistors of the idle path. To alleviate this issue, both charge-transfer paths in each stage are bootstrapped to one another.

Figure 6.18(a) depicts the circuit schematic of two successive driver stages. For the stages closest to the output, and as the capacitors in the idle path start discharging, the devices connected to the output node experience a large voltage stress. To remedy this issue, nodes n_1 and n_3 within a stage are connected to nodes n_2 and n_4, respectively. This connectivity ensures that the node voltages in the idle path continuously follow those of the active path. As a result, the idle-path capacitors are automatically refreshed from the active-path capacitors. This behavior is better depicted by observing the simulation waveforms of a representative four-stage driver shown in Fig. 6.18(b). During the output positive half-cycle, only the clocks of the charge-path are active, nonetheless, the voltages of the discharge-path capacitors are shown to track the voltages of their charge-path counterparts. Similarly, during the output negative half-cycle, only the clocks of the discharge-path are active, nonetheless, the voltages of the charge-path capacitors are shown to track the voltages of their discharge-path counterparts.
Fig. 6.17. Bipolar driver unit stage implementation.

Fig. 6.18. (a) 2 successive driver stages (b) intermediate node voltage waveforms during pump-down and pump-up modes of a 4 stage bipolar driver.
In the next chapter, we discuss the measurement results of a bipolar output driver implemented in 45nm SOI CMOS technology using the proposed unit driver stage. The higher voltage-handling capability of SOI enables us to demonstrate extensive device stacking. The implemented driver is composed of a cascade of 48 stages and is capable of >40V outputs. Another advantage of SOI technology is that it does not rely on reverse biased junctions for device isolation as is the case with bulk technology. This convenient feature makes generating negative voltages similar to generating positive ones.

Thick gate-oxide devices with 1.5V±10% voltage tolerance are used. The minimum horizontal spacing between wires is 70 nm. Metal capacitors with minimum finger spacing are estimated to sustain ~15 V based on a TDDB cumulative failure rate of less than 0.1% over 10 years at 85°C. The BOX thickness is ~150 nm and is estimated to sustain ~51 V based on the same criteria. Note that the only areas of the BOX that are exposed to voltage stress are the transistors’ active areas. No data for the BOX reliability is available, so its TDDB estimates are based on the low-k dielectric parameters. To minimize area, different stages use capacitors with variable finger spacing depending on the voltage requirement of the stage. For the first 16 stages, minimally-spaced finger capacitors from the technology library with a capacitance density of 2.3fF/um² are used. Capacitors with 2x (1.3 fF/um²) and 3x (0.75 fF/um²) finger spacing are used for the second and third 16 stages, respectively. For the 2x and 3x capacitors, the vertical metal spacing is smaller than the finger spacing, and capacitors with side-wall fields only are devised.

The driver is laid out while carefully considering the necessary lateral and vertical clearances between low- and high-voltage signals. Supply rails and control wires are separated from the high-voltage interconnects by a sufficient number of metal layers whenever crossing.
6.4. **Summary**

In this chapter, we have introduced 2 types of high-voltage output stages in standard CMOS technology. The first driver type is a compact, stacked-device driver using assisted-charging devices. Assisted-charging devices facilitate faster settling of the slow intermediate nodes in the stack and maintain the reliability of the devices during switching transitions. Because, the complexity of stacked-device drivers grows rapidly with the number of stacked device, a stacked-device driver is a better alternative for applications requiring limited device stacking (2-3) but higher switching speeds (>200MHz). The second driver type is a bidirectional, charge pump-based output stage. This driver architecture inherits the modularity, voltage reliability, and power efficiency of voltage charge pumps and has a frequency-tunable output resistance. The driver maximum voltage range is limited by the substrate breakdown rather than the predriver design complexity. Because of its gate drive modularity, the design enables extended device stacking (48 devices), and enables >40V output drive in 45nm technology as demonstrated in Chapter 7. A Charge Pump-Based driver facilitates the integration of 10's of volts switching in low-voltage CMOS processes, leading to cheaper and more compact SoC solutions.

This chapter concludes the second part of the dissertation addressing high-voltage drive methods. In the next chapter, we discuss measurement results from various test chips demonstrating the high-voltage techniques introduced so far in the dissertation.
CHAPTER 7

Experimental Results

7.1. Introduction

In this chapter, we discuss the measurement results of the different high-voltage circuits explained in the previous 2 chapters. First, we start by the measurement results of the Hybrid Charge Pump designs explained in Chapter 5. Then, we discuss the measurement results of the Charge Pump-Based driver designs explained in Chapter 6.

7.2. Hybrid Charge Pumps

In this section, we present the measurement results of 2 Hybrid Charge Pumps in 65nm bulk CMOS technology. The first is a +36V charge pump, whereas the second is a -29V charge pump. The die micrographs of both pumps are shown in Fig. 7.1(a) and 7.1(b), respectively.

Fig. 7.1. Die micrographs in 65nm bulk CMOS technology: (a) +36V hybrid charge pump (b) -29V hybrid charge pumps.
7.2.1. A +36V Hybrid Charge Pump

This Hybrid Charge Pump is composed of 3 smaller sub-pumps capable of positive output voltage, as explained in Chapter 5. Each sub-pump is characterized separately with its input connected to the supply voltage, i.e. $V_{\text{in}}=V_{\text{dd}}$. First, we present the measurement results of the individual sub-pumps then present the measurement results of the overall Hybrid Charge Pump.

The first sub-pump is a positive output voltage CMOS 4PVD. Its dc I-V characteristics are measured versus different pumping voltages and frequencies as shown in Fig. 7.2(a) and 7.2(b) respectively. As expected, the pump's maximum output voltage is clipped at 12.5V, and its current drive capability improves with higher $f_{\text{pump}}$. The pump's efficiency and output resistance measurements are shown in Fig. 7.3(a) and 7.3(b) respectively. The pump's peak efficiency is independent of $V_{\text{dd}}$ and is 70.4% at 20$\mu$A load current. The pump's output resistance exhibits a $1/fC$ dependence as expected of switched capacitor circuits and is limited by the switches series on resistance. In the FSL, the pump's output resistance is 29$\Omega$.

![Fig. 7.2. Measured dc I-V characteristics of the first sub-pump at: (a) $f_{\text{pump}}=4$MHz (b) $V_{\text{dd}}=2.5$V.](image-url)
The second sub-pump is a positive output voltage all-NMOS 6PVD. Its dc I-V characteristics are measured versus different pumping voltages and frequencies as shown in Fig. 7.4(a) and 7.4(b) respectively. For this pump, output voltage clipping can be shown for $V_{dd}>2.75\text{V}$. The pump's efficiency and output resistance measurements are shown in Fig. 7.5(a) and 7.5(b) respectively. The pump's peak efficiency improves slightly with $V_{dd}$ and is 55.6% at 20μA load current for $V_{dd}=2.75\text{V}$. In the FSL, the pump's output resistance is 27kΩ.

Fig. 7.3. Measurement results of the first sub-pump: (a) efficiency at $f_{pump}=4\text{MHz}$ (b) output resistance.

Fig. 7.4. Measured dc I-V characteristics of the second sub-pump for: (a) $f_{pump}=4\text{MHz}$ (b) $V_{dd}=2.5\text{V}$. 
The third sub-pump is a positive output voltage improved-drive polysilicon diode Dickson pump. Its dc I-V characteristics are measured versus different pumping voltages and frequencies as shown in Fig. 7.6(a) and 7.6(b) respectively. Our improved-drive Dickson design is compared with a regular Dickson design using more stages (9 stages) and targets the same output voltages as shown in Fig. 7.7(a). Measurements of the pump's output resistance are shown in Fig. 7.7(b).
In the FSL, the improved design output resistance is 39KΩ and equals roughly half of the regular design's output resistance since it uses fewer stages. The pump's peak efficiency improves significantly with $V_{dd}$ as shown in Fig. 7.8(a), and is equal to 43% at $V_{dd}=2.75V$. A comparison between the peak efficiencies of the regular and improved designs is shown in Fig. 7.8(b). We observe that for higher $V_{dd}$, the improvement in peak efficiency diminishes.

![Graphs showing output voltage and output resistance](image)

Fig. 7.7. Measurement results of a regular and improved-drive Dickson pumps with $L_i=0.5\mu$m: (a) I-V characteristics at $V_{dd}=2.5V$ (b) output resistance.

![Graphs showing efficiency vs load current and pumping voltage](image)

Fig. 7.8. (a) Measured efficiency of an improved-drive Dickson pump with $L_i=0.5\mu$m at $f_{pump}=4MHz$ (b) measured peak efficiencies of a regular and improved-drive Dickson pumps.
Next, we discuss the measurement results of the overall Hybrid Charge Pump. The measured output voltages of all 3 cascaded sub-pumps are plotted versus $V_{dd}$ at no load current as shown in Fig. 7.9(a). As shown, the first and second sub-pump outputs clip at 12.5V and 20V, respectively, while the third sub-pump output voltage continues increasing with $V_{dd}$. The pump’s maximum output voltage is 36V at $V_{dd}$=2.75V. The pump’s dc I-V characteristics are measured at $f_{pump}$=4MHz and different $V_{dd}$ as shown in Fig. 7.9(b). The pump maintains output voltages >34V for load currents <10µA at $V_{dd}$=2.75V. Because of the hybrid nature of the pump, the overall pump efficiency is a weighted average of its constituent sub-pump efficiencies as explained earlier in Chapter 5. Consequently, the overall pump efficiency falls somewhere between the first sub-pump and the third sub-pump efficiencies. The overall pump efficiency is measured at $f_{pump}$=4MHz and different $V_{dd}$ as shown in Fig. 7.10(a). The pump’s peak efficiency improves with $V_{dd}$ and is equal to 49% at 20uA load current and $V_{dd}$=2.75V. The pump's output resistance is equal to the sum of the individual sub-pumps' output resistance. In the FSL, the hybrid pump's output resistance is equal to 100KΩ as shown in Fig. 7.10(b).

![Graphs showing measurement results of the overall Hybrid Charge Pump](image-url)
Fig. 7.10. Hybrid Charge Pump measurement results: (a) efficiency at $f_{\text{pump}}=4$ MHz (b) output resistance.

Because the third sub-pump is the only pump capable of 36V outputs, it is useful to compare its efficiency with the hybrid pump efficiency as shown in Fig. 7.11(a). Using a hybrid pump design improves the power efficiency from 43% to 49%. Finally, we characterize the pump peak efficiency versus $f_{\text{pump}}$ as shown in Fig. 7.12(b). As expected, the pump efficiency starts to drop rapidly for $f_{\text{pump}}>10$ MHz, as the switch conduction losses become comparable to the charge redistribution losses. The pump occupies 0.1625 mm$^2$ in area.

Fig. 7.11. Hybrid Charge Pump measurements: (a) efficiency comparison (b) peak efficiency versus $f_{\text{pump}}$. 
7.2.2. A -29V Hybrid Charge Pump

This Hybrid Charge Pump is composed of 2 smaller sub-pumps capable of negative output voltage, as explained in Chapter 5. Each sub-pump is characterized separately with its input connected to ground, i.e. \( V_{in} = 0 \), since only positive supply voltages are available to power the chip. First, we present the measurement results of the individual sub-pumps then present the measurement results of the overall Hybrid Charge Pump.

The first sub-pump is a negative output voltage all-NMOS 6PVD. Its dc I-V characteristics are measured versus different pumping voltages and frequencies as shown in Fig. 7.12(a) and 7.12(b) respectively. As shown, the pump's maximum output voltage is clipped at -12.5V, and its current drive capability improves with higher \( f_{pump} \). The pump's efficiency and output resistance measurements are shown in Fig. 7.13(a) and 7.13(b) respectively. The pump's peak efficiency is relatively independent of \( V_{dd} \) and is 63.5% at 20\( \mu \)A load current. In the FSL, the pump's output resistance is 46K\( \Omega \).

![Fig. 7.12. Measured dc I-V characteristics of the first sub-pump at: (a) \( f_{pump} = 4\text{MHz} \) (b) \( V_{dd} = 2.5\text{V} \).](image)
The second sub-pump is a negative output voltage all-NMOS 6PVD. Its dc I-V characteristics are measured versus different pumping voltages and frequencies as shown in Fig. 7.14(a) and 7.14(b) respectively. Because of FOX isolation, no voltage clipping is observed. The pump's efficiency and output resistance measurements are shown in Fig. 7.15(a) and 7.15(b) respectively. The pump's peak efficiency improves with $V_{dd}$ and is 38.5% at 20μA load current for $V_{dd}$=2.75V. In the FSL, the pump's output resistance is 39KΩ.
Fig. 7.15. Measurement results of the second sub-pump: (a) efficiency at $f_{pump}=4$MHz (b) output resistance.

Next, we discuss the measurement results of the Hybrid Charge Pump. The measured output voltages of the individual cascaded sub-pumps are plotted versus $V_{dd}$ at no load current as shown in Fig. 7.16(a). As shown, the first sub-pump output clips at -12.5V while the second sub-pump output voltage continues increasing with $V_{dd}$. The pump’s maximum output voltage is 29V at $V_{dd}=2.75V$. The pump’s dc I-V characteristics are measured at $f_{pump}=4$MHz and different $V_{dd}$ as shown in Fig. 7.16(b). The pump maintains output voltages < -27V for load currents < 10µA at $V_{dd}=2.75V$. Again, the overall pump efficiency is a weighted average of its constituent sub-pump efficiencies. Consequently, the overall pump efficiency falls somewhere between the first sub-pump and the second sub-pump efficiencies. The overall pump efficiency is measured at $f_{pump}=4$MHz and different $V_{dd}$ as shown in Fig. 7.17(a). The pump’s peak efficiency improves with $V_{dd}$ and is equal to 47% at 20µA load current and $V_{dd}=2.75V$. The pump's output resistance is equal to the sum of the individual sub-pumps' output resistance. In the FSL, the hybrid pump's output resistance is equal to 91KΩ as shown in Fig. 7.17(b).
Fig. 7.16. Hybrid Charge Pump: (a) unloaded sub-pumps outputs (b) dc I-V characteristics at $f_{\text{pump}}$=4MHz.

Fig. 7.17. Hybrid Charge Pump measurement results: (a) efficiency at $f_{\text{pump}}$=4MHz (b) output resistance.

Because the second sub-pump is the only pump capable of -29V outputs, it is useful to compare its efficiency with the hybrid pump efficiency as shown in Fig. 7.18(a). Using a hybrid pump design improves the power efficiency from 43% to 38%. Finally, we characterize the pump peak efficiency versus $f_{\text{pump}}$ as shown in Fig. 7.18(b). As expected, the pump efficiency starts to drop rapidly for $f_{\text{pump}}>10$MHz, as the switch conduction losses become comparable to the charge redistribution losses. The pump occupies 0.115 mm$^2$ in area.
Fig. 7.18. Hybrid Charge Pump measurement results: (a) efficiency comparison (b) peak efficiency vs. \(f_{\text{pump}}\)

7.3. Charge Pump-Based Drivers

In this section, we present the measurement results of 2 types of Charge Pump-Based output drivers as explained in Chapter 6. The first driver is a 12V unipolar output driver implemented in 65nm bulk CMOS technology. The second driver is a >40V bipolar output driver implemented in 45nm SOI CMOS technology. The die micrographs for both drivers are shown in Fig. 7.19(a) and 7.19(b), respectively.

Fig. 7.19. Die micrographs: (a) 12V unipolar output driver (b) 40V bipolar output driver.
7.3.1. A Unipolar 12V Output Driver

This driver is characterized over supply voltages ranging from 2.25V to 2.75V, and pumping frequencies ranging from 1 MHz to 0.5 GHz. The circuit is tested for both square-wave outputs, and DC outputs driving different load currents.

Shown in Fig. 7.20 is a measured 12V, 250KHz square-wave output driving a 67pF capacitive load. The driver operates at $V_{dd}=2.5V$ and $f_{pump}=250MHz$. Arbitrary duty cycles can also be attained via the low-voltage side control signal.

![Image of measured driver output waveform](image)

**Fig. 7.20.** Measured 12V, 250KHz driver output waveform.

The driver dc I-V curves are shown in Fig. 7.21(a) at $f_{pump}=50MHz$ and different $V_{dd}$ values. The output voltage increases linearly with $V_{dd}$. It is seen that for $V_{dd}=2.75V$ and low output currents, the driver voltage is clipped at 12.5V, as expected. The driver maintains >11V output voltages for load currents <80µA. Similar measurements are shown for $V_{dd}=2.5V$ and different $f_{pump}$ in Fig. 7.21(b). The driver power efficiency is measured at $V_{in}=0V$. Shown in Fig. 7.22(a) is the measured driver efficiency at $f_{pump}=50MHz$ and different $V_{dd}$. The driver peak efficiency
slightly degrades for higher $V_{dd}$. Shown in Fig. 7.22(b) is the measured driver efficiency at different $V_{dd}=2.5V$ and different $f_{pump}$. The driver peak efficiency degrades with higher $f_{pump}$.

The design has power efficiencies comparable to those of charge pumps implemented in the same technology. The driver peak efficiency versus $f_{pump}$ is shown in Fig. 7.23(a) for different inputs. Connecting the up/dn signal to the driver input results in higher output voltages for the same number of stages, improving the driver efficiency. The peak efficiency is 63.5% at $f_{pump}=1$MHz and $V_{in}=V_{dd}$. Peak efficiency drops rapidly for $f_{pump} > 150$MHz. This corresponds to
pumping clock periods longer than the RC time constant formed by the switch on resistance and the pumping capacitor. The output drive resistance is plotted versus $f_{\text{pump}}$ as shown in Fig. 7.23(b). For $f_{\text{pump}} > 150$ MHz, the drive resistance is limited by the switch on-resistance and is limited to 3.7KΩ. The driver power dissipation at no load current is overlaid on top of Fig. 7.23(b). The dissipation is proportional to $f_{\text{pump}}$ as is the case with CMOS logic and is equal to 4.5mW at $f_{\text{pump}} = 150$ MHz. The driver occupies a total area of 0.056mm².

![Fig. 7.23. Driver measurement results versus $f_{\text{pump}}$: (a) peak efficiency (b) output resistance.](image)

### 7.3.2. A Bipolar 44V Output Driver

This driver is characterized over supply voltages ranging from 0.9 V to 1.65 V, and pumping frequencies ranging from 100 MHz to 1 GHz. The circuit is tested for both square-wave outputs, and DC outputs at different load currents.

Shown in Fig. 7.24(a) is a measured 20KHz, 44V, bipolar square-wave output drive (appears slightly lower due to the oscilloscope loading). The driver operates from a 1.65V supply and a 450MHz pumping frequency while driving a 35pF capacitive load. Interestingly, when the driver is operated at very low output drive frequencies, the positive output level settles to the lower
steady state value of 32 V as shown by the 400 mHz waveform in Fig. 7.24(b). This is an artifact of the slow settling time constant of back-gate effect in SOI CMOS [72], and limits the output drive frequencies to >10 Hz. Because transistor bodies that are closer to the driver output are floating at a much higher potential than that of the chip substrate, the substrate and the BOX form a back-gate and an inversion layer is formed above the BOX. The formed back-channel causes reverse conduction losses and leads to a lower absolute output voltage. Note that the effect is only observed for positive output voltages.

Fig. 7.24. Measured ±44 V square waveforms at: (a) 20 KHz (b) 400 mHz.
Because of the back-gate effect, the dc response is characterized for the negative drive level. The driver dc I-V characteristics are shown in Fig. 7.25(a) for $f_{\text{pump}} = 450\,\text{MHz}$ and different $V_{dd}$. The driver is capable of providing >40 V outputs for <100 $\mu\text{A}$ load currents. The driver dc I-V characteristics are shown in Fig. 7.25(b) for $V_{dd} = 1.5\,\text{V}$ and different $f_{\text{pump}}$. The driver output resistance is plotted versus $f_{\text{pump}}$ as shown in Fig. 7.26. The driver current consumption drawn from a 1.5V supply at no load current is overlaid over the same plot. The output resistance drops with higher $f_{\text{pump}}$ and is eventually limited to 21.5 K$\Omega$ in the FSL. Due to the output resistance,
power dissipation trade-off, we choose to report the driver performance at the \((36\,\text{K}\Omega, 28\,\text{mA})\) data point. The driver's maximum output voltage and minimum output resistance are plotted versus the number of driver stages as shown in Fig. 7.27(a) and 7.27(b), respectively. The ripple voltage is measured to be 1.2 mV at \(f_{\text{pump}} = 450\,\text{MHz}\) and a 35 pF load capacitance. The circuit occupies \(0.21\,\text{mm}^2\) in area and achieves the highest voltage drive reported to date in a CMOS technology while adopting an all-low-voltage device implementation.

![Graph](image_url)

**Fig. 7.26.** Measured driver output resistance and current consumption at \(V_{\text{dd}}=1.5\,\text{V}\).

![Graph](image_url)

**Fig. 7.27.** (a) Maximum output voltage versus the driver number of stages (b) Minimum output resistance versus the driver number of stages.
7.4. Summary

In this chapter, we have discussed the measurement results of various high-voltage circuits implemented in nanometer-scale CMOS technology. A 36V 49% efficiency Hybrid Charge Pump circuit has been demonstrated in 65nm bulk CMOS technology. This design represents a three-fold increase in the output voltage range compared to conventional designs implemented in the same technology [48]. Also, a 44V bipolar output driver has been implemented in 45nm SOI CMOS technology, and demonstrates the modular stacking of 48 thick-oxide devices. This design achieves the highest voltage drive reported to date in a CMOS technology while adopting an all-low-voltage device implementation [71]. The design performance of this work is summarized and compared to previous published work for voltage generation and drive in Tables 7.1 and 7.2, respectively. By noticing the implementation technology and the voltage range for various designs, both tables show our ability to integrate several 10's of volt signals in fairly advanced, nanometer scale CMOS technologies, in a manner unattempted previously.
Table 7.1. Performance Summary of HV Charge Pumps

<table>
<thead>
<tr>
<th>Reference</th>
<th>Technology</th>
<th>Maximum Voltage</th>
<th>Load Current</th>
<th>Supply Voltage</th>
<th>Number of Stages</th>
<th>Pumping Capacitor</th>
<th>Pumping Frequency</th>
<th>Active Devices</th>
<th>Area</th>
<th>Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6µm CMOS</td>
<td>51 V</td>
<td>4 µA</td>
<td>6 V</td>
<td>18</td>
<td>50 pF</td>
<td>0.5 MHz</td>
<td>PMOS</td>
<td>2.42 mm²</td>
<td>&lt;30%</td>
</tr>
<tr>
<td></td>
<td>0.18µm CMOS</td>
<td>14.8 V</td>
<td>0.7 µA</td>
<td>1.8 V</td>
<td>10</td>
<td>10 pF</td>
<td>50 MHz</td>
<td>PMOS</td>
<td>0.129 mm²</td>
<td>Not Available</td>
</tr>
<tr>
<td></td>
<td>0.8µm HV CMOS</td>
<td>50 V</td>
<td>50 µA</td>
<td>5 V</td>
<td>16</td>
<td>0.5 pF</td>
<td>10 MHz</td>
<td>PMOS</td>
<td>0.33 mm²</td>
<td>Not Available</td>
</tr>
<tr>
<td></td>
<td>0.35µm SOI CMOS</td>
<td>19.6 V</td>
<td>20 uA</td>
<td>3.3 V</td>
<td>6</td>
<td>16 pF</td>
<td>4 MHz</td>
<td>MOS Diodes</td>
<td>Not Available</td>
<td>Not Available</td>
</tr>
<tr>
<td></td>
<td>0.25µm CMOS</td>
<td>28 V</td>
<td>2 uA</td>
<td>2.5 V</td>
<td>12</td>
<td>Not Available</td>
<td>1 MHz</td>
<td>Polysilicon Diodes</td>
<td>Not Available</td>
<td>Not Available</td>
</tr>
<tr>
<td></td>
<td>65nm CMOS</td>
<td>36 V (up to 100V)</td>
<td>20 µA</td>
<td>2.5 V</td>
<td>12</td>
<td>4 pF</td>
<td>4 MHz</td>
<td>PMOS/ NMOS Polysilicon Diodes</td>
<td>0.1625 mm²</td>
<td>49%</td>
</tr>
</tbody>
</table>
Table 7.2. Performance Summary of HV Drivers

<table>
<thead>
<tr>
<th>Reference</th>
<th>[39]</th>
<th>[41]</th>
<th>[42]</th>
<th>[47]</th>
<th>This Work</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
<td>2μm CMOS</td>
<td>0.35μm HV CMOS</td>
<td>0.18μm CMOS</td>
<td>0.35μm SOI CMOS</td>
<td>45nm SOI CMOS</td>
</tr>
<tr>
<td><strong>Maximum Voltage</strong></td>
<td>50 V</td>
<td>10 V</td>
<td>10 V</td>
<td>10 V</td>
<td>44 V</td>
</tr>
<tr>
<td><strong>Delay</strong></td>
<td>80 ns</td>
<td>2.4 ns</td>
<td>20 ns</td>
<td>5.5 ns</td>
<td>12 us</td>
</tr>
<tr>
<td><strong>Drive Resistance</strong></td>
<td>Not Available</td>
<td>Not Available</td>
<td>2.26 Ω</td>
<td>Not Available</td>
<td>21.5 KΩ</td>
</tr>
<tr>
<td><strong>Active Devices</strong></td>
<td>80V LDMOS</td>
<td>10V LDMOS</td>
<td>3.3V CMOS</td>
<td>5V CMOS</td>
<td>1.5V CMOS</td>
</tr>
<tr>
<td><strong>Output Frequency</strong></td>
<td>100 KHz</td>
<td>25 MHz</td>
<td>20 KHz</td>
<td>20 MHz</td>
<td>20 KHz</td>
</tr>
<tr>
<td><strong>Load Capacitance</strong></td>
<td>30 pF</td>
<td>Not Available</td>
<td>Not Available</td>
<td>Not Available</td>
<td>35 pF</td>
</tr>
</tbody>
</table>
CHAPTER 8

Conclusion

In this dissertation, we introduced technology and circuit methods that increase achievable voltage ranges in a standard fine line-width CMOS technology by orders of magnitude. The dissertation was split into 2 parts dealing with different classes of high-voltage circuits, mainly, voltage charge pumps, and output voltage drivers.

In the first part of the dissertation, we introduced 3 techniques to enable extended voltage-range charge pumps in bulk CMOS technology, namely, double-diode substrate isolation, FOX over deep nwell isolation, and substrate stacking. To enable these technology methods, improved efficiency, special type pump cells were devised. We proposed a new, all-NMOS 6PVD that alleviates diode drops in same-type switch pump designs, and an improved-drive polysilicon diode-based Dickson pump in deep nwell. We also highlighted a fundamental, voltage range-power efficiency trade off that arises with our extended-voltage range technology methods.

To enable extended voltage range pumps at improved power efficiency, we introduced a novel Hybrid Charge Pump architecture. Hybrid Charge Pumps optimally mix sub-pump trading off voltage range and power efficiency. We have provided an accurate mathematical model to calculate the power efficiency and output noise power of hybrid-type pumps. We have shown that the hybrid pump efficiency is a weighted average of its individual sub-pump intrinsic efficiencies. We have also derived expressions for the normalized pumping capacitor values needed in each sub-pump to optimize efficiency. Based on our model equations, a systematic
approach for the design of Hybrid Charge pumps has been outlined. A 36V charge pump with 49% efficiency has been demonstrated in 65nm CMOS technology [48]. This design exhibits a 3x increase in voltage range compared to conventional designs using the same technology.

In the second part of the dissertation, we addressed the problem of high-voltage drive in standard nanometer-scale technology. Hence, we introduced 2 types of output stages. The first is a compact stacked-device driver that uses assisted-charging devices to maintain the reliability of stacked devices during switching transitions. This driver type is limited to only few stacked devices (~2-3), but is capable of fast switching, (>200MHz in 65nm CMOS technology) [69]. The second driver type is a Charge Pump-Based driver design [70]. This design inherits the modularity, reliability, and power efficiency of voltage charge pumps and has a frequency-tunable output resistance. The modularity of switch gate drive enables extended device stacking, and voltage ranges are limited by the substrate breakdown rather than by design complexity. In [71], we have demonstrated the stacking of 48 devices enabling >40V outputs in 45nm SOI CMOS technology.

Proposals for future work may include accurate characterization and modeling of the polysilicon diodes. In low-power applications, knowledge of the diode's reverse saturation current as a function of the intrinsic region length ($L_i$) could be necessary. In high dynamic range applications, low output noise power is necessary. Deriving a relationship between the pumping clock jitter and the final charge pump output noise, can help save power by choosing the right clock reference jitter specification. Also, as is the case with semiconductor junctions, the PIN diode's forward bias is a function of temperature. This results in an output voltage variation with temperature. Such temperature variations can be cancelled, to a first order, by using a pumping
voltage that exhibits an identical temperature dependence to drive the diode-based pumps. This can be realized by using the PIN diodes to generate a CTAT supply voltage to power the diode-based charge pumps. More generally, the pump can be placed in a negative feedback to provide accurate voltages and track peak efficiencies at different load currents. Finally, multiple CMOS dies can be stacked in the manner proposed in Chapter 3 to demonstrate >100V dc outputs.

Extending substrate voltage tolerances and enabling the integration of high-voltage switching waveforms into a nanometer-scale technology allow us to harvest CMOS scaling benefits, and enable better integrated SoC solutions.
References


