INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®
HIGH PERFORMANCE GHZ RF CMOS IC's for INTEGRATED PHASE-LOCKED LOOPS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Shenggao Li, B.S., M.S.

* * * * *

The Ohio State University

2000

Dissertation Committee:

Professor Mohammed Ismail, Adviser
Professor Roberto Rojas
Professor Furrukh Khan

Approved by

Adviser
Department of Electrical Engineering
ABSTRACT

The implementation of high-performance integrated phase-locked loops (PLLs) represents one of the primary design challenges in integrating a complete wireless RF transceiver into a single chip.

In this work, we explore the design trade-offs to achieve high-performance for fully integrated RF CMOS PLLs. Based on investigation at system-level, critical design parameters are chosen to achieve low phase-noise, fast settling behavior, and low power consumption. New circuit blocks and design techniques are presented. A novel dynamic phase-frequency detector is proposed which demonstrates no visible dead-zone, extended phase-detection range, and high frequency operation. Programmable charge-pump techniques are proposed to improve system performance in terms of loop dynamics and spur reduction. Optimization techniques for fractional-X control are also discussed. With the adoption of a state-of-the-art 0.18 \( \mu m \) Cu CMOS technology, high-speed dynamic-logic frequency dividers, and high-Q on-chip spiral inductors are obtained. The design techniques and high-performance circuit blocks are then applied to the design of a fully integrated 5.8\( GHz \) PLL for WLAN applications.
To my parents and my wife
ACKNOWLEDGMENTS

I would like to express my whole-hearted thanks to my advisor Professor Mohammed Ismail for his endless help in supporting my Ph.D. research. He guided me into the exploration of the wireless RF IC design frontier, and several times he provided me with so valuable advices when I fell into a low spirit in research and deviated from my academic endeavor. Without his understanding and encouragement, it is hardly possible that I would have fulfilled the mission to pursue a doctoral degree. Without his insightful advice, I wouldn't have started my career and research enthusiastically in phase-locked loops.

Grateful thanks should be extended to Professor Patrick Roblin for his teaching and suggestions in RF and microwave circuits design. Same grateful thanks are to be expressed to Professor Roberto Rojas. Professor Rojas gave me a lot of very valuable suggestions during my dissertation defense, and helped me to correct many of the mistakes in my dissertation. During this research, he and one of his student spent their precious time in participating our discussions on PLL design issues. I would also like to thanks Professor Furrukh Khan for his time in reading my dissertation, and for his humor regarding bluetooth radio.

It's been a pleasant experience for me to work with people in the Analog VLSI Lab. Much of my knowledge in analog and RF ICs was obtained from close discussion with students and visiting scholars in the Analog VLSI Lab. I would like to mention
Hassan Elwan, Chi-huang Lin, and Mona Hella, who were always open to a discussion without any reservation. Chunlei Shi, Yue Wu, with whom I worked together to gain important design experience at the early stage of our research. I would also like to thank Tuna Tarim for her friendship and help during my transition to Ph.D. study. Thanks to all the people who worked and are working hard to make a better Analog VLSI Lab environment.

Thanks to Jerasimos Zohios for his partnership in working with me together in the Copper Design Contest. Cheers to the numerous days and nights, let it be Christmas Eve, New year's Eve, or Chinese Lunar-year's Eve, that we spent together in the lab in struggling with technical troubles and up-coming report deadlines. Some credits in the discussion of inductors and VCOs should also be attributed to Zohios for his earlier work in the related areas within the Analog VLSI Lab.

Thanks to Texas Instruments and Intel Corporation for providing financial sponsorship for my Ph.D. research, and thanks to UMC and Novellus for providing the opportunity of access to the state-of-the-art CMOS technology.

I owe so much to my wife - since she joined me after our new marriage, I have been heavily involved in research projects and stayed in the lab all the time. When I was writing my dissertation in the lab, she accompanied me late in the evening in the lab even though she herself had heavy duties in study.

I am deeply indebted to my parents for their support and understanding through all the years. I have kept them waiting for over 10 years when each time they felt that I were ready to take a break from the endless study.

Shenggao Li

Columbus, Nov. 17, 2000
VITA

Nov. 24, 1972 ...........................................Born - Hunan Province, China

1994 .....................................................B.S. Dept. of Automatic Control.
Northwestern Polytechnical University.
Xi'an

1997 .....................................................M.S. Dept. of Automation.
Tsinghua University, Beijing

1997-2000 ..........................................Graduate Fellow, Research Associate
The Ohio-state University, Columbus

Since March 2000 .................................Senior Analog/RF Design Engineer
Intel Corporation

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Electrical Engineering

Studies in Analog and RF Integrated Circuits: Dr. Mohammed Ismail
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>x</td>
</tr>
</tbody>
</table>

Chapters:

1. Introduction | 1 |
   1.1 The Information Era | 1 |
   1.1.1 Broadband communication | 2 |
   1.1.2 Wireless communication | 3 |
   1.2 System On A Chip – Design Challenges | 5 |
   1.3 Phase-Locked Loops | 7 |
   1.3.1 Frequency synthesizer | 7 |
   1.3.2 Clock recovery | 8 |
   1.4 Current Work | 10 |

2. Design of High-performance Phase-locked Loops | 12 |
   2.1 Introduction | 12 |
   2.2 Phase-noise Fundamentals | 15 |
   2.3 Discussion on PLL Dynamic Behaviors | 19 |
   2.3.1 PLL linear model | 19 |
   2.3.2 Analysis of charge-pump PLLs | 25 |
2.3.3 Noise transfer characteristics ............................................................ 32
2.4 Topology Discussion .............................................................................. 40
2.4.1 Fractional-N PLL ............................................................................... 41
2.4.2 Dual-loop PLLs .................................................................................. 43

3. Phase-Frequency Detector ........................................................................ 47
3.1 Phase Detectors Review ......................................................................... 47
3.1.1 Multiplier ............................................................................................ 48
3.1.2 Exclusive-OR gate .............................................................................. 50
3.1.3 JK-flipflop .......................................................................................... 51
3.1.4 Tri-state phase detector ..................................................................... 51
3.2 Design Issues in Phase-frequency Detectors ......................................... 56
3.2.1 Dead-zone .......................................................................................... 56
3.2.2 Blind-zone ......................................................................................... 58
3.3 Dynamic Logic Phase-frequency Detectors ........................................... 61
3.4 A Novel Dynamic-logic Phase-frequency Detector ............................... 66
3.4.1 Circuit operation ............................................................................... 66
3.4.2 Performance evaluation .................................................................... 69
3.5 Conclusion .............................................................................................. 74

4. Charge Pump ........................................................................................... 77
4.1 Considerations in Charge Pump Design ............................................... 77
4.1.1 Current mismatch ............................................................................ 78
4.1.2 Mismatch reduction ......................................................................... 80
4.2 System Performance Optimization ......................................................... 82
4.2.1 Division ratio variation ..................................................................... 82
4.2.2 Fast settling ...................................................................................... 83
4.3 A Programmable Charge Pump ............................................................. 85
4.3.1 Circuit description ............................................................................ 85
4.3.2 Reference spurs reduction ................................................................. 89
4.4 Programmable Charge-pump for Multi-standard Applications ............. 92

5. High-Speed Frequency Divider ................................................................. 94
5.1 Prescaler .................................................................................................. 95
5.1.1 Dual-modulus prescaler .................................................................... 96
5.1.2 TSPC D-Flipflop ............................................................................. 98
5.1.3 Trapezoidal scaling and optimization .............................................. 100
5.1.4 Charge-sharing problem .................................................................. 101
5.1.5 Performance ..................................................................................... 102
5.2 Programmable Divider ................................................................. 104

6. On-chip Integrated Inductor and VCO ........................................... 108
   6.1 Introduction .................................................................................. 108
   6.2 Design of High-Q Inductors ......................................................... 111
      6.2.1 On-chip spiral inductors in Copper CMOS process .......... 111
      6.2.2 Parasitics and layout consideration ................................. 113
      6.2.3 Inductor de-embedding and measurement results .......... 115
   6.3 LC Oscillator .............................................................................. 121
      6.3.1 Brief on oscillator phase noise ........................................... 121
      6.3.2 Circuit description and design consideration .................. 122
      6.3.3 Layout and simulation results ............................................ 125

7. Implementation of a 5.8 GHz PLL .................................................... 128
   7.1 Some Practical Design Issues .................................................... 128
      7.1.1 Frequency-hopping .............................................................. 128
      7.1.2 Loop parameters ................................................................. 134
      7.1.3 Settling time ....................................................................... 135
      7.1.4 Reference spurs ................................................................. 136
      7.1.5 Process variation related issues ....................................... 138

8. Conclusions and Future Work ....................................................... 140

Bibliography ..................................................................................... 143
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 A heterodyne RF receiver</td>
<td>5</td>
</tr>
<tr>
<td>1.2 A homodyne RF receiver</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Phase noise and spurs</td>
<td>8</td>
</tr>
<tr>
<td>1.4 The effect of phase noise and spurious tones in a RF receiver</td>
<td>9</td>
</tr>
<tr>
<td>1.5 Phase-jitter in time domain</td>
<td>10</td>
</tr>
<tr>
<td>2.1 Block diagram of a Phase-locked loop</td>
<td>12</td>
</tr>
<tr>
<td>2.2 The convergence of a dynamic system</td>
<td>14</td>
</tr>
<tr>
<td>2.3 Phase noise and spurs</td>
<td>18</td>
</tr>
<tr>
<td>2.4 Linear model of a phase-locked loop</td>
<td>20</td>
</tr>
<tr>
<td>2.5 Transient response with regard to different $\xi$</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Settling time with regard to different $\xi$</td>
<td>24</td>
</tr>
<tr>
<td>2.7 Diagram of a Charge-pump PLL</td>
<td>25</td>
</tr>
<tr>
<td>2.8 Root locus of a 3rd-order charge-pump PLL</td>
<td>28</td>
</tr>
<tr>
<td>2.9 Open-loop frequency-domain response</td>
<td>29</td>
</tr>
<tr>
<td>2.10 Transient step response</td>
<td>30</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>2.11 Relationship between open-loop frequency specifications and transient behavior</td>
<td>31</td>
</tr>
<tr>
<td>2.12 Transient step response of the 3rd-order PLL with different phase margin</td>
<td>32</td>
</tr>
<tr>
<td>2.13 Settling time with regard to different phase margin</td>
<td>33</td>
</tr>
<tr>
<td>2.14 Linear model of a phase-locked loop</td>
<td>34</td>
</tr>
<tr>
<td>2.15 Noise transfer function ( N=1 )</td>
<td>35</td>
</tr>
<tr>
<td>2.16 Noise transfer function ( N \neq 1 )</td>
<td>36</td>
</tr>
<tr>
<td>2.17 Loop filter noise transfer function ( N \neq 1 )</td>
<td>38</td>
</tr>
<tr>
<td>2.18 Noise transfer response for different open-loop phase margin</td>
<td>39</td>
</tr>
<tr>
<td>2.19 Output phase noise for different open-loop phase margin</td>
<td>40</td>
</tr>
<tr>
<td>2.20 Fractional-N frequency synthesizer</td>
<td>43</td>
</tr>
<tr>
<td>2.21 A Dual-loop PLL using phase-interpolation technique</td>
<td>44</td>
</tr>
<tr>
<td>2.22 Linearity of phase interpolation</td>
<td>46</td>
</tr>
<tr>
<td>3.1 Multiplier used as phase detector</td>
<td>49</td>
</tr>
<tr>
<td>3.2 Phase characteristics of the multiplier</td>
<td>50</td>
</tr>
<tr>
<td>3.3 A Exclusive-OR PD and its phase-characteristics</td>
<td>52</td>
</tr>
<tr>
<td>3.4 A JK-flipflop PD and its phase-characteristics</td>
<td>53</td>
</tr>
<tr>
<td>3.5 A tri-state PFD with charge-pump and its ideal phase characteristics</td>
<td>54</td>
</tr>
<tr>
<td>3.6 Conventional PFD</td>
<td>57</td>
</tr>
<tr>
<td>3.7 Dead-zone and blind-zone</td>
<td>58</td>
</tr>
<tr>
<td>3.8 Maximum frequency estimation</td>
<td>59</td>
</tr>
</tbody>
</table>
4.8 Current mismatch in a charge-pump .............................................................. 90
4.9 Mismatch compensation using programmable current sources ................ 91
5.1 Block diagram of a dual-modulus prescaler .................................................. 97
5.2 Schematic of the TSPC DFF ........................................................................ 99
5.3 Transient analysis of the TSPC toggle flip-flop with 10 GHz input frequency and 1.5 V supply voltage ................................................................. 101
5.4 Waveforms of the TSPC DFF ........................................................................ 103
5.5 Waveforms of the prescaler with $f_{clk}$ at maximum frequency ............... 104
5.6 Diagram of a programmable divider ............................................................. 105
5.7 End-of-count detector .................................................................................. 106
5.8 Reload circuit ............................................................................................... 107
6.1 A simple lumped inductor model ................................................................. 109
6.2 Layout of a spiral inductor ........................................................................... 112
6.3 Cross section of spiral inductor and inter-cap ............................................. 114
6.4 Layout consideration .................................................................................... 115
6.5 Physical representation of the lumped parasitics ....................................... 116
6.6 Measured S-parameter (dotted-line: raw data; solid-line: de-embedded data) ........................................................................................................... 119
6.7 Inductance versus frequency ......................................................................... 120
6.8 Q-factor versus frequency ........................................................................... 121
6.9 Typical phase noise region ........................................................................... 122
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.10</td>
<td>Oscillator schematic</td>
<td>123</td>
</tr>
<tr>
<td>6.11</td>
<td>VCO layout</td>
<td>125</td>
</tr>
<tr>
<td>6.12</td>
<td>Oscillation start-up</td>
<td>126</td>
</tr>
<tr>
<td>6.13</td>
<td>Phase noise</td>
<td>127</td>
</tr>
<tr>
<td>7.1</td>
<td>Testing chip</td>
<td>129</td>
</tr>
<tr>
<td>7.2</td>
<td>Bounding curves</td>
<td>131</td>
</tr>
<tr>
<td>7.3</td>
<td>N=128</td>
<td>132</td>
</tr>
<tr>
<td>7.4</td>
<td>N=64</td>
<td>133</td>
</tr>
<tr>
<td>7.5</td>
<td>Fractional-N PLL diagram</td>
<td>136</td>
</tr>
<tr>
<td>7.6</td>
<td>Frequency hopping analysis</td>
<td>137</td>
</tr>
<tr>
<td>7.7</td>
<td>Fractional-N control</td>
<td>137</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 The Information Era

During the past three decades, we have seen the booming development of personal computers, internet, mobile telecommunication, and multimedia. Never before, people are able to obtain information so timely and so conveniently. Yet in a great sense, one can state that this is not fast and convenient enough — technologies are still advancing, and there will be certainly some more technology revolutions to come.

The power of high-speed computing, worldwide internet access, mobile telecommunication, and multimedia, shall be greatly attributed to the development of integrated circuits. It should be noted that, compared with other technologies such as GaAs, bipolar, and BiCMOS, it is the standard CMOS technology that contributes most significantly to the fast expanding consumer electronics market. The ever-decreasing feature-size of CMOS technologies has gone from around 10 \( \mu m \) in the 1970’s, around 1 \( \mu m \) in the late 1980’s, to 0.12 \( \mu m \) in the present market. Feature size of 0.1 \( \mu m \) and smaller have been in development. While there exists a limit for CMOS downsizing, some newly evolved CMOS processes, such as triple-well CMOS, Cu CMOS.
Silicon-on-Insulator (SOI), Silicon-on-Sapphire (SOS), and etc. appear to provide better performance and possibly higher integration.

With the down-scaling of CMOS feature sizes, it is possible to implement more and more complex computing and signal-processing functions in a single chip. System-on-a-chip has become a popular terminology in recent years. This trend is driven by several factors from the consumer market: power consumption reduction, volume and weight reduction, and product cost reduction.

In light of the advancement in CMOS technologies and the motivation for full-system integration, it is our interest to look into the two most active communication areas in the IC industry.

1.1.1 Broadband communication

This mainly involves the high-speed transmission of data, voice, and video over existing communication infrastructure over twisted pairs, optical fiber and etc. Two representative applications, DSL and Gigabit Ethernet, are addressed below. Fully integrated CMOS transceivers for these two applications involve quite complicated on-chip digital filtering, encoding, decoding, echo-cancellation, and equalization techniques, presenting quite a lot challenges. Optical fiber communication is also a fast-going business area in broadband communication. Due to the author's lack of knowledge in this field, it will not be discussed here. Yet, since our focus in this work is phase-locked loops, it is worthwhile to point out that the phase-jitter reduction is one of the primary design concerns in this type of application.

- DSL (Digital Subscribe Line): DSL technologies are currently in a stage of fast growth in the world. DSL is intended for expanding the bandwidth of existing
residential and business telephone lines (twisted copper pairs), in pace with the vast demand for high-speed internet and intranet access from home and office. In most countries, twisted-pair is still the most prevalent infrastructure for telecommunication. This tells why DSL presents a huge and profitable market.

- **Ethernet**: The focus of Ethernet is LAN (local area network). Fully integrated Gigabit Ethernet transceivers have been in the market for a short while and will soon obsolete the existing 10M/100M-bit Ethernet transceivers. At this time, people are paying attention to 10-Gigabit Ethernet transceivers, which will very likely become the mainstream in a few years. The development of 10-Gigabit Ethernet transceivers will further increase communication capacity, and might provide a cheaper method for WAX (wide area network) access.

### 1.1.2 Wireless communication

The wireless and radio frequency (RF) market has undergone tremendous expansion over the last 10 years.

The most aggressive wireless market is cellular phone. First-generation mobile communication, analog cellular, appeared as an entirely new form of communications, and required a system-wide deployment of infrastructure for a market that didn’t yet exist. Second-generation mobile communication, digital cellular or PCS, evolved from the first generation mobile phone quickly, and offered more advanced digital services and better quality. Just a few years after the digital cellular, the wireless communication is entering the third generation (3G). In 3G wireless, mobile telephony and internet access have been combined under the same ceiling. A wireless user will
have global access to voice, data and video services easily from anywhere using any terminal.

Another important trend in the wireless business is the development of wireless products in the unlicensed ISM\textsuperscript{1} bands: 900 MHz (902-928 MHz), 2.4 GHz (2400-2483.5 MHz) and 5.8 GHz (5725-5850 MHz). To facilitate multiple users sharing the same spectrum in an unlicensed environment, spread spectrum techniques are employed. Two types of spread spectrum techniques are available: the Direct Sequence Spread Spectrum (DSSS) and Frequency Hopped Spread Spectrum (FHSS). Several efforts are currently undergoing fast growth in parallel, which include Wireless Local Area Network (WLAN), Bluetooth (BT), and HomeRF. Although WLAN, BT, and HomeRF share the same frequency spectrum, each of them targets different level of users, and operation range. WLAN aims to replace wired LAN in some occasions by providing fast network construction and low cost. Bluetooth is a wireless technology that enables the instant wireless connection between communication devices in a 'piconet', e.g., home or office environment, as well as instant wireless access to internet. HomeRF targets uniquely home environment. In fact, there is overlapping between these three different application domains. Mutual interference and interoperability are in the list of system design concerns.

The evolution of wireless communication is accompanied by people’s enthusiasm to put as many as RF components in integrated circuits, or more desirably, to implement a full RF transceiver and baseband processing circuits in a single-chip using standard digital CMOS technologies. Compared with the bulky and power hungry cellular phones several years ago, today’s cellular phones are already quite small (palm size).

\textsuperscript{1}ISM stands for Industry, Science, and Medicine.
with a battery life of around one week. Yet, the fight for the reduction of volume, weight and power consumption in mobile handsets and other RF ICs is going to continue.

1.2 System On A Chip – Design Challenges

Implementing a full system in a single-chip does not come without pain.

Firstly, standard CMOS technologies are generally characterized and optimized for digital baseband IC design. Extra characterization of devices at RF frequencies, such as device and bonding wire parasitics, is necessary to ensure the simulation accuracy of analog and RF circuits.

Secondly, analog and particularly RF circuits often involve large amount of passive components, such as filters, inductors, varactors, and capacitors. Implementing these passive components on-chip may not be practical because they are area consuming. Demonstrated in Figure 1.1 is the RF front-end of a heterodyne receiver architecture (double down-conversion). In order to achieve satisfactory band-select, channel-select, and image-rejection, the band-pass filters in this receiver need to have high Q-factors and linearity. Active on-chip filters can not meet these requirements. instead, external passive SAW (surface acoustic wave) filters or crystal filters are generally adopted.

Figure 1.1: A heterodyne RF receiver
External passive components increase system volume and bonding wiring cost, and therefore it is very important to reduce or eliminate off-chip passive components. This requires a careful trade-off in system-level design. A homodyne receiver architecture (direct down-conversion) can be used to reduce off-chip components, which is shown in Figure 1.2. In this architecture, image-rejection filters are not necessary since $\omega_{IF} = 0$ and therefore no image signal exists. Because of the direct-downconversion, the baseband low-pass filters are working at lower frequencies and can be relaxed to allow monolithic integration. On the other hand, the direct down-conversion architecture has a severe drawback — the local oscillator may leak to the LNA and the antenna, and then is reflected back to the mixer. The reflected signal and the LO signal are at the same frequency, so the self-mixing will generate a DC-offset which can be large enough to saturate subsequent stages [1].

Thirdly, people have sought for a long time to implement on-chip high-Q inductors, transformers and etc. In most CMOS processes, it is very difficult to obtain high-Q on-chip inductors. Reported Q-factor for on-chip inductors without post-processing
or extra mask handling is around 3-6, which makes the design of low phase-noise oscillator very challenging, as will be discussed later.

Finally, analog and RF circuits are more sensitive to noise, and their performance can be seriously degraded by noise through substrate coupling or cross-talk. Since digital circuits are generally more noisy due to their switching activities, analog/RF blocks in a chip should be very well isolated from digital circuits. In addition to interference from neighboring noisy blocks, inherent electronic noise of a device can also become a major design barrier. For instance, in the previous mentioned wireless transceivers, people are often struggling hard to get the noise figure of the low-noise amplifier down to around 2 dB.

1.3 Phase-Locked Loops

Phase-locked loops are widely used as frequency synthesizers and clock recovery circuits. This is described below.

1.3.1 Frequency synthesizer

In wireless communications, phase-locked loops are used as frequency synthesizers to generate LO signals (refer to Figure 1.1 and 1.2). Ideally, LO should be a single-tone frequency signal. In a practical oscillator, due to electronic noise and interference, the LO frequency will fluctuate randomly around its center frequency. As a result, the LO spectrum will have skirts around its center frequency. This is illustrated in Figure 1.3, where the skirts are due to random phase-noise. In addition to random phase-noise, some spurious tones (spurs) also appear in the LO spectrum. Spurious tones originate from unexpected deterministic signals that may appear in the control input of an oscillator. This will be investigated later. Both phase-noise and spurs will
degrade communication quality. To understand this, we look at the down-conversion process of a wireless receiver, as depicted in Figure 1.4.

In this figure, the desired RF signal is weaker than the two blockers appearing in adjacent channels. The desired RF signal is down-converted to IF frequency \( f_{IF} = |f_{LO} - f_{RF}| \) by the LO frequency. Also, the phase-noise at \( f_{IF} \) apart from the first blocker, and the spurious tone at \( f_{IF} \) apart from the second blocker, down-convert the blockers to the same IF frequency. The desire signal therefore suffers from serious SNR degradation due to the blockers. To alleviate this problem, phase-noise and spurs in a LO must be suppressed to a sufficiently low level.

1.3.2 Clock recovery

Another important application of phase-locked loops is clock recovery. In many digital communication systems, data is transmitted without any additional timing
reference. In the transmission path to the destination, the waveform of the transmitted data stream then becomes garbled and distorted due to noise, interference, transmission delay, and etc. To reconstruct the data, an clock signal embedded in the transmitted data has to be firstly recovered at the receiver side. This recovered clock is then used as the timing basis to sample the received data. For accurate timing, the recovered clock has to be very stable. Similar to the case in frequency synthesis, the

Figure 1.4: The effect of phase noise and spurious tones in a RF receiver
recovered clock signal (PLL output) suffers from noise disturbance. This is reflected by random fluctuation of the clock transition time, as illustrated in Figure 1.3. The cycle-to-cycle variation of the clock is called timing jitter (or phase-jitter), which is simply the representation of phase-noise in the time domain.

![Jitter (Transition uncertainty)](image)

Figure 1.3: Phase-jitter in time domain

In order to sample the incoming data accurately, timing-jitter in the recovered clock should be as small as possible. In high-speed data communication, because the clock cycle is reduced, the time budget for the timing-jitter becomes considerably tight.

1.4 Current Work

Based on the above discussion, in this work we will deal with the problems associated with the full integration of phase-locked loops using standard CMOS technologies. In particular, a 0.18 µm 1-poly 6-metal 1.5/3.3 dual-voltage twin-well Cu CMOS technology is used for the design in this work [2] [3] [4] [5].

In Chapter 2, we will focus our discussion on the design trade-offs of PLLs. This is followed by more specific design strategies for the building blocks of a PLL in subsequent chapters: Chapter 3 discusses phase-frequency detectors. A novel dynamic-logic
phase-frequency detector is presented and its merits are compared with previous designs. The phase-frequency detector demonstrates no visible dead-zone and extended linear phase-detection range. In Chapter 4, we investigate the behavior of charge-pump. A programmable charge-pump is proposed which is capable of speeding up settling time, compensating loop gain variation, and reducing reference spurs. In Chapter 5, frequency dividers are investigated. Using the state-of-the-art Cu CMOS technology, a dual-modulus divider using dynamic logic is capable of operating normally for the 5.8 GHz WLAN application. In chapter 6 we present the design and experimental results of on-chip high-Q inductors ($Q > 10$) and their application to the design of LC-VCOs. In Chapter 7, we return to the topic of designing fully integrated Phase-locked loops. A 5.8 GHz frequency synthesizer is designed for wireless LAN applications. Finally, Chapter 8 draws some general conclusions for the presented work.
CHAPTER 2

DESIGN OF HIGH-PERFORMANCE PHASE-LOCKED LOOPS

2.1 Introduction

Phase-locked loops (PLLs) are widely used as frequency synthesizers and data recovery circuits in telecommunication systems. The block diagram in Figure 2.1 shows the general form of a PLL, consisting of a phase detector (PD), a low-pass filter (LPF), a voltage-controlled oscillator (VCO) and a frequency divider.

![Block diagram of a Phase-locked loop](image)

Figure 2.1: Block diagram of a Phase-locked loop

In contrast to most feedback systems in which signal amplitudes and their rate of change are of interest, a PLL is a negative feedback system that operates on the
frequency and phase of periodic signals. Phase-locking is the key to implement signal tracking or frequency multiplication in this unique system.

More specifically, in a PLL, the phase detector serves as an “error indicator” between the input reference signal ($U_{ref}$) and the feedback signal ($U_{dir}$). The phase detector constantly monitors the phase difference of the two signals. If a phase error builds up, the phase detector will generate a correcting signal proportional to the phase error. This signal passes through the low-pass filter with high-frequency AC components suppressed. The remaining DC component adjusts the VCO frequency in such a way that the phase error between $U_{ref}$ and $U_{dir}$ is reduced.

Through the feedback mechanism, the phase error between $U_{ref}$ and $U_{dir}$ is minimized. At balance, the phase error is zero or remains constant, the phase of the feedback signal ($U_{dir}$) is said to be “locked” to the phase of the reference signal ($U_{ref}$). As such the system is referred as a phase-locked loop.

It is important to note that phase-lock alone is not sufficient to keep the loop staying at balance. If the frequency of the reference signal and the frequency of the feedback signal are not equal, phase error will be built up over time. The ultimate acquisition of both frequency-lock and phase-lock may not be immediately perceptible based on the above discussion. This is explained further below: In a transient controlling and adjusting process, the loop might have already reached frequency locking at a certain point of time, but there is still phase error that continues to build up the control voltage, forcing the frequencies unequal again. The acquisition process seems to be reset, but of course, the process does not restart from scratch, since the system is not coming to a new initial condition. If the loop is designed to be stable, the system is now supposed to be closer to its final stable state. This behavior can
be more easily understood by drawing the trace of phase-error and frequency-error in a Cartesian-coordinate plane (phase-plane). We assume the system in discussion is stable, and let the X-axis the phase-error and Y-axis the frequency-error. If (x,y) is originally at a point further away from the origin, after a certain time, it will move to a point that is closer to the origin the next time. The trace represented by (x,y) will be expected to converge to the origin finally (Figure 2.2).

Figure 2.2: The convergence of a dynamic system

It is desirable that the time to converge from an unbalanced state to the origin is as small as possible. Stability and settling time are two important design issues to be explored further in this chapter.
We briefly point out that there exists another type of frequency synthesis circuit—the Direct Digital Frequency Synthesizer (DDFS). Unlike a PLL which is a dynamic feedback system and thus has slow settling time in frequency transition, a DDFS allows its synthesized frequencies to be changed instantly. The drawback of a DDFS is that it is power hungry, because its digital clock of a DDFS is required to run much faster than the synthesized output frequencies. For applications at gigahertz frequency operation, it is possible to up-convert the output of a low frequency DDFS by a PLL of fixed frequency. In general, DDFS is considered not suitable for silicon integration because it is bulky and power-hungry.

To understand PLL design issues, we will discuss phase-noise in the next section. In subsequent sections, we investigate the design trade-offs at system level, and discuss different ways to reduce phase-noise and settling time, and etc.

2.2 Phase-noise Fundamentals

PLLs operate on the phase of signals. Subject to internal device noise and external interference, the loop will be disturbed from time to time, and deviate from its “phase-locked” state. The loop monitors the deviation, reacts to return the system to a balanced state, and is disturbed again ... The output frequency of the PLL will fluctuate around its center frequency.

It is easy to understand the effect of phase-noise in time domain: under the influence of phase-noise, the zero-crossing time of a sine-wave signal (or the transition time of a square-wave signal) will fluctuate in the vicinity of its nominal instant. Consequently, the period of each cycle will be different from cycle to cycle as already illustrated in Figure 1.5. To quantify the random error, we assume that the error...
follows a normal distribution - \( N(0, \sigma^2) \). where \( \sigma \) is RMS phase-jitter. It is known that in normal distribution, the probability of the error falling into one \( \sigma \) is around 2/3. Assume that there is a 1.0 GHz clock signal, and its RMS phase-jitter is around 30 ps. It follows that the RMS phase-jitter takes up 10\(^\circ\) out of a 360\(^\circ\) clock cycle. This will give rise to significant inter-symbol interference error.

Frequency domain representation of phase-noise is more frequently used for wireless communication systems. To gain some insight of how frequency-domain phase-jitter is related to time-domain phase-jitter, we take a look at the output of a VCO, which is given by

\[
U_{\text{out}}(t) = U_0 \sin(\omega_0 t + K_{\text{vco}} \int_0^t v_{\text{out}}(t) dt) + \theta(t)
\]

where \( U_0(t) \) is the amplitude of the output, \( v_{\text{out}}(t) \) is the VCO control voltage, and \( \theta(t) \) is the initial excess phase. All variables are prone to random noise and interference, so we represent each of them as a time-variant function.

We consider the simplest case when each of the three variables subjects to a small sine-wave interference.

With \( U_0(t) = U_0 + a \sin(\omega_n t) \), one can obtain:

\[
U_0(t) = U_0 \sin(\omega_0 t + K_{\text{vco}} V_0 t + \theta_0) + a \sin(\omega_n t) \sin(\omega_0 t + K_{\text{vco}} V_0 t + \theta_0)
\]

\[
= U_0 \sin(\omega_t + \theta_0) - \frac{a}{2} \left[ \sin(\omega_c t + \omega_n t - \theta_0) + \sin(\omega_c t - \omega_n t + \theta_0) \right]
\]

where \( \omega_c = \omega_0 + K_{\text{vco}} V_0 \) is the output carrier frequency.

Obviously, the sinusoidal variation of the VCO amplitude induces two sidebands at \( \omega_c \pm \omega_n \).
In the case of $V_{cmt}(t)$ fluctuation, i.e., $V_{cmt}(t) = V_0 + a \sin(\omega_n t)$, the VCO output is in effect frequency modulated, and

$$ U_a(t) = U_0 \sin[\omega_0 t + K_{cvo} \dot{V}_0 t + a K_{cvo} \int \sin(\omega_n t) dt + \theta_0] $$

$$ = U_0 \sin[\omega_0 t - \frac{a K_{cvo}}{\omega_n} \cos(\omega_n t) + \theta_0] $$

$$ = U_0 \sin(\omega_0 t + \theta_0) \cos[\frac{a K_{cvo}}{\omega_n} \cos(\omega_n t)] $$

$$ - U_0 \cos(\omega_0 t + \theta_0) \sin[\frac{a K_{cvo}}{\omega_n} \cos(\omega_n t)] \tag{2.4} $$

In general, the condition $a K_{cvo} / \omega_n \ll 1$ can be satisfied, then

$$ U_a(t) \approx U_0 \sin(\omega_0 t + \theta_0) - \frac{a K_{cvo} \dot{U}_0}{2 \omega_n} \cos(\omega_0 t + \theta_0) \cos(\omega_n t) $$

$$ = U_0 \sin(\omega_0 t + \theta_0) $$

$$ - \frac{a K_{cvo} \dot{U}_0}{2 \omega_n} \cos(\omega_0 t + \omega_n t + \theta_0) + \cos(\omega_0 t - \omega_n t + \theta_0) \tag{2.5} $$

This result is immediately recognized to be similar to that of equation 2.3.

Let's now consider the excess phase fluctuation, with $\theta(t) = \theta_0 + a \sin(\omega_n t)$, it follows that.

$$ U_a(t) = U_0 \sin[\omega_0 t + K_{cvo} \dot{V}_0 t + \theta_0 + a \sin(\omega_n t)] $$

$$ = U_0 \sin(\omega_0 t + \theta_0) \cos[a \sin(\omega_n t)] $$

$$ + U_0 \cos(\omega_0 t + \theta_0) \sin[a \sin(\omega_n t)] \tag{2.6} $$

Again, assume small phase variation, i.e., $a \sin(\omega_n t) \ll 1$, one gets.

$$ U_a(t) \approx U_0 \sin(\omega_0 t + \theta_0) + a U_0 \cos(\omega_0 t + \theta_0) \sin(\omega_n t) $$

$$ = U_0 \sin(\omega_0 t + \theta_0) $$

$$ + \frac{a U_0}{2} [\sin(\omega_0 t + \omega_n t + \theta_0) - \sin(\omega_0 t - \omega_n t + \theta_0)] \tag{2.7} $$
In the above simple analysis, although the fluctuation originates from different sources, equation 2.3, 2.5 and 2.7 all tell that a deterministic interference will contribute sidebands (spurs) to the output spectrum (refer to Figure 2.3). The strength of spurs is measured in $dB_c$, which is the power level of a spur in reference to the carrier power level.

![Figure 2.3: Phase noise and spurs](image)

If the fluctuation in the above parameters is random, we will obtain the skirts in the output spectrum instead. To distinguish between the skirts and spurs, in a narrow sense, the skirts are referred as "phase-noise". Because of the random nature of "phase-noise", strictly the spectrum skirts in an oscillator is "phase-noise density". This is why phase-noise is measured in dBc/Hz. By definition, phase-noise is given
by the following expression.

\[ \mathcal{L}(\Delta \omega) = \frac{\text{total noise power in a 1 Hz bandwidth at a frequency offset } \Delta \omega}{\text{carrier power}} \]  \hspace{1cm} (2.8)

where \( \Delta \omega \) is in reference to the center carrier frequency.

It might be unrealistic to measure the noise power in a 1-Hz bandwidth in practice. Instead, "resolution bandwidth" is used to quantify the phase-noise. Equation 2.8 is then changed slightly, as follows.

\[ \mathcal{L}(\Delta \omega) = \frac{\text{total noise power in the resolution-bandwidth at offset } \Delta \omega}{(\text{carrier power}) \times (\text{resolution-bandwidth in Hz})} \]  \hspace{1cm} (2.9)

### 2.3 Discussion on PLL Dynamic Behaviors

#### 2.3.1 PLL linear model

The dynamic behavior of PLLs is now analyzed. In the discussion, we all focus on the noise transfer characteristics.

The analysis is based on the linear model as given by Figure 2.-1, in which the phase-detector is represented by \( K_{PD} \), the loop filter is represented by \( F_{LPF}(S) \), the VCO is represented by \( K_{VCO}/S \) and the frequency divider is represented by \( 1/N \).

It is well-known that in a phase-locked loop the primary noise sources are the VCO noise and reference noise. These two noise sources are represented by \( \theta_{ref} \) and \( \theta_{reo} \) respectively.

The open-loop transfer function of the loop is obtained as.

\[ G(S) = K_{PD}F_{LPF}(S)\frac{K_{VCO}}{S} \cdot \frac{1}{N} \]  \hspace{1cm} (2.10)

The close-loop transfer function between the input phase \( \phi_{in} \) and the output phase \( \phi_{out} \) is given as follows.

\[ H(S) = \phi_{out}(S)/\phi_{in}(S) \]
Figure 2.4: Linear model of a phase-locked loop

\[ G(S) = \frac{N \cdot G(S)}{1 + G(S)} \]

\[ = \frac{K_{PD} F_{LPF}(S) K_{VCO}}{1 + K_{PD} F_{LPF}(S) K_{VCO}} \cdot \frac{1}{N} \]

\[ = \frac{N K_{PD} F_{LPF}(S) K_{VCO}}{N S + K_{PD} F_{LPF}(S) K_{VCO}} \]  

(2.11)

(2.12)

For this moment, we assume that the low-pass filter is a lag-lead low-pass filter with a single pole and a single zero, as given in the following transfer function.

\[ F_{LPF}(S) = K_{LPF} \frac{1 + \frac{S}{\omega_p}}{1 + \frac{S}{\omega_z}} \]  

(2.13)

where \( \omega_p < \omega_z \).

We define \( K = K_{PD} K_{LPF} K_{VCO}/N \) as the loop gain. The open-loop transfer function is then rewritten as.

\[ G(S) = K \cdot \frac{1 + \frac{S}{\omega_p}}{S(1 + \frac{S}{\omega_p})} \]  

(2.14)

The close-loop transfer function in equation 2.12 is obtained.

\[ H(S) = \frac{N \cdot K_{VCO}(1 + \frac{S}{\omega_p})}{S^2 + S(\omega_p + \frac{K_{VCO}}{\omega_z}) + K_{VCO}} \]  

(2.15)
This equation is further reduced to the following form in terms of the loop damping factor, $\xi$, and natural frequency, $\omega_n$.

$$H(S) = \frac{Y \cdot \omega_n^2 [2\xi - (\omega_n/K)]S + \omega_n^2}{S^2 + 2\xi \omega_n S + \omega_n^2}$$

(2.16)

where

$$\omega_n = \sqrt{K \omega_p} \text{ rad/sec}$$

(2.17)

and

$$\xi = \frac{1}{2} (1 + K/\omega) \sqrt{\frac{\omega_p}{K}}$$

(2.18)

Equation 2.16 is the transfer function of a typical 2nd-order system. Its transient response is characterized by the denominator of $H(S)$, which is further determined by $\xi$ and $\omega_n$ uniquely. For this reason, we only need to consider the following normalized transfer function.

$$D(S) = \frac{\omega_n^2}{S^2 + 2\xi \omega_n S + \omega_n^2}$$

(2.19)

whose transient step response is given by

$$d(t) = 1 - \frac{e^{-\xi \omega_n t}}{\sqrt{1 - \xi^2}} \sin(\omega_n \sqrt{1 - \xi^2} t + \theta_0)$$

(2.20)

Figure 2.5 shows the transient step response with regard to different values of $\xi$, in which the horizontal time axes is normalized to $\omega_n t$. For a given $\xi$, the settling time $t_s$ of the system is proportional to $1/\omega_n$, that is, the bigger $\omega_n$ is, the shorter the settling time will be.

With regard to $\xi$, when $\xi = 0$, the transient step response of the system is sinusoidal, with the oscillation frequency equal to $\omega_n$. When $0 < \xi < 1$, the transient
Figure 2.5: Transient response with regard to different $\xi$
response has overshoot, but will gradually settle down to its steady state. When \( \xi > 1 \), the transient response will rise to the final steady state monotonically. At \( \xi > 1 \), although there is no overshoot, the response is generally slow.

A precise relationship between \( t_s \) and \( \xi \) is hard to obtain, partly due to the definition of \( t_s \), which is largely dependent on the allowed settling error \( \Delta \). When \( 0 < \xi < 1 \), a rough approximation for \( t_s \) can be obtained by simply considering the time when the envelope of the transient response falls within the allowed error. This is obtained by letting

\[
\frac{e^{-\xi \omega_n t}}{\sqrt{1 - \xi^2}} = \Delta
\]

from which one obtains

\[
\omega_n t_s = -\ln \Delta - \ln \sqrt{1 - \xi^2} / \xi
\]

Equation 2.22 provides an upper bound for \( t_s \). Figure 2.6 illustrates this result graphically.

It turns out that when \( \xi \) is around 0.7, \( \omega_n t_s \) is close to minimum, while the overshoot value \( \sigma_o \) is moderate. In practice, \( \xi \) is chosen to be between 0.5 and 0.8, and \( \omega_n \) is chosen as large as possible.

Now, we return to equation 2.17 and 2.18, and make the following observations.

- One can increase \( K_p \), \( \omega_p \), and \( \omega_z \) simultaneously by the same ratio. By doing so, \( \xi \) remains unchanged, but \( \omega_n \) is increased by the same ratio. According to equation 2.22, the settling time will be decreased by this ratio.

- Increasing \( \omega_n \) will influence the phase-noise performance in two aspects: first, there will be larger reference spurs at the PLL output spectrum; second, random
Figure 2.6: Settling time with regard to different $\xi$

noise from reference input will have more weight on the overall phase noise of the PLL output. These two aspects are to be further explained in the next subsection.

- Increasing $K$ can be done by increasing $K_{PD}$, $K_{LPF}$, $K_{VCO}$ or by decreasing $N$. Increasing $K_{PD}$ generally requires the increase of supply voltage or current. A passive LPF like the previous mentioned lag-lead filter doesn't provide voltage or current gain. However, for the passive low pass filter in a charge-pump PLL (to be discussed later), using smaller capacitors corresponds to the increase in current-to-voltage conversion gain. It is also possible to increase the gain of an active LPF, but using an active LPF isn't favorable because it contributes more
noise. Increasing $K_{VCO}$ isn’t favored either, because that means the VCO will be more sensitive to noise. Finally, decreasing $N$ requires the increase of input reference frequency, which is constrained by output frequency resolution and the operating speed of phase-detectors.

2.3.2 Analysis of charge-pump PLLs

Transfer function

To continue, we pay our attention to the more frequently used charge-pump PLLs. In a charge-pump PLL, a tri-state phase detector (phase-frequency detector) is used together with a charge-pump device. Tri-state phase detector is to be discussed in chapter 3, and charge-pump is to be discussed in chapter 4 respectively. Figure 2.7 illustrates the diagram of a charge-pump PLL. The loop filter is also given in this diagram. In general, $C_1$, $R_2$, and $C_2$ are used to form a 2nd-order low-pass filter. In many applications, however, $R_3$, and $C_3$ or even more RC stages might be used to form a 3rd-order or higher-order filter to suppress reference spurs more effectively. In the following discussion, we will focus on 2nd-order LPF.
The phase-frequency detector in the PLL operates on phase difference between the reference input $U_{ref}$ and the feedback input $U_{dvi}$. and has two output terminals $UP$ and $DN$ to control the current sources $I_{UP}$ and $I_{DN}$ respectively. When $U_{ref}$ leads $U_{dvi}$, the upper switch turns on for a period of $T_{UP}$, thus $I_{UP}$ flows into node $P$ and charges capacitor $C_1$ for the time period of $T_{UP}$, causing node $P$ voltage to increase. Similarly, if $U_{dvi}$ leads $U_{ref}$, capacitor $C_1$ will be discharged by $I_{DN}$ for a period of $T_{DN}$, causing node $P$ voltage to decrease. If $U_{ref}$ and $U_{dvi}$ are in lock, both $I_{UP}$ and $I_{DN}$ will be turned off from node $P$. then, node $P$ is in a high-impedance state and its voltage remains unchanged. In any input cycle $T_{ref}$, there might be both charging and discharging current. The overall average current $I_a$ entering node $P$ is given by

$$I_a = \frac{I_{UP} \cdot T_{UP} - I_{DN} \cdot T_{DN}}{T_{ref}} \quad (2.23)$$

$$= I_{CP} \cdot \frac{T_{UP} - T_{DN}}{T_{ref}} \quad (2.24)$$

in which $I_{UP} = I_{DN} = I_{CP} = const$ is assumed.

The PFD and the charge-pump operate at discrete time, but as before, we will use a linear model to analyze the charge-pump PLL. The combination of the PFD and the charge-pump is in effect a phase-to-current converter. Its equivalent continuous time transfer function can be easily obtained by considering two extreme cases: At zero phase difference, neither $I_{UP}$ nor $I_{DN}$ is connected to node $P$, so $I_a = 0$. In the case that $U_{ref}$ leads $U_{dvi}$ by a phase of $2\pi$, in a whole reference cycle $T_{ref}$, node $P$ is charged by $I_{UP}$, so $I_a = I_{UP} = I_{CP}$. The phase-to-current conversion gain is thus given by.

$$K_{PD} = \frac{I_{CP}}{2\pi} \quad (2.25)$$

26
The transfer function for the 2nd-order low-pass filter is given by.

\[ K_{LPF} = \frac{1 + S R_2 C_2}{C_1 + C_2 S(1 + S R_2 \frac{C_2}{C_1 + C_2})} \] \hspace{1cm} (2.26)

\[ = K_{LPF} \frac{1 + \frac{S}{\omega_z}}{S(1 + \frac{S}{\omega_p})} \] \hspace{1cm} (2.27)

where \( K_{LPF} = 1/(C_1 + C_2) \), \( \omega_z = 1/(R_2 C_2) \), and \( \omega_p = 1/(R_2 \frac{C_2}{C_1 + C_2}) \).

Note that the low-pass filter provides an additional pole at \( S = 0 \). The PLL is now a 3rd-order system, with its open-loop transfer function given by.

\[ G(S) = K \cdot \frac{1 + \frac{S}{\omega_z}}{S^2(1 + \frac{S}{\omega_p})} \] \hspace{1cm} (2.28)

which has two poles at \( S = 0 \).

The close-loop transfer function is given by.

\[ H(S) = \frac{N \cdot K \omega_p(1 + \frac{S}{\omega_z})}{S^3 + S^2 \omega_p + S \frac{K \omega_p}{\omega_z} + K \omega_p} \] \hspace{1cm} (2.29)

3rd-order system analysis

The root locus of the system can be obtained from the open-loop poles and zeros. It is important to note that for the open-loop transfer function given in equation 2.28, because \( \omega_z < \omega_p \), the root locus of the close-loop system will always locate in the left half of the complex plane, as illustrated in Figure 2.8. This means the stability of a 3rd charge-pump PLL is guaranteed. under the assumption that the linear model for the PLL is still valid.

Similar to the 2nd-order PLL discussed earlier, we may also rewrite the denominator of equation 2.29 in terms of \( \xi \) and \( \omega_n \), with an additional pole \(-\omega_r\) on the real axis, as follows.

\[ D(S) = \frac{\omega_r \omega_n^2}{(S + \omega_r)(S^2 + 2\xi \omega_n S + \omega_n^2)} \] \hspace{1cm} (2.30)
where $A_1$, $A_2$, and $A_3$ are constants.

Our previous analysis and conclusions with regard to the transient behavior of a 2nd-order system can still provide some insight here. However, unlike the 2nd-order system, here an analytical form for $\xi$, $\omega_n$, and $\omega_r$ in terms of the design parameters $K$, $\omega_p$, and $\omega_n$ is hard to obtain, due to the extra pole $S = -\omega_r$. If the close-loop pole $S = -\omega_r$ is not far from the two complex poles, it will greatly affect the settling behavior, making the previous transient analysis method invalid.
It will be more convenient to carry out the design of a 3rd-order system in frequency domain based on its open-loop frequency response. In frequency domain, the dynamic behavior of a system is characterized by the unit-gain frequency (or cut-off frequency) $\omega_c$ and phase-margin $\gamma$ of the open-loop frequency response.

Figure 2.9 illustrates the open-loop frequency response of a 3rd-order PLL represented by equation 2.28, in which the phase margin and cut-off frequency are identified. For a $45^\circ$ degree phase margin given by this open-loop frequency response, the corresponding close-loop step response is illustrated in Figure 2.10.

One may wonder what is the relationship between the previous mentioned transient specifications ($\xi, \omega_n$) for transient response and the open-loop specifications ($\omega_c$).
\begin{equation}
\omega_c = \omega_n \sqrt{1 + 4\xi^2 - 2\xi^4} \quad (2.32)
\end{equation}
\begin{equation}
\zeta = \arctan \frac{2\xi}{\sqrt{1 + 4\xi^4 - 2\xi^2}} \quad (2.33)
\end{equation}

In a 2nd order system, for $\xi$ to be between 0.5 and 0.8, $\zeta$ needs to be around $50^\circ \sim 70^\circ$.

In order to relate the open-loops specifications to the transient specifications, people developed some empirical equations [6] in practice, as follows.

\begin{equation}
\sigma_p = 0.16 + 0.4 \left( \frac{1}{\sin \zeta} - 1 \right) \times 100\% \quad (2.34)
\end{equation}

\begin{equation}
t_s(5\%) = \frac{K \pi}{\omega_c} \quad (2.35)
\end{equation}

Figure 2.10: Transient step response
with \( K = 2 + 1.5\left(\frac{1}{\sin \gamma} - 1\right) + 2.5\left(\frac{1}{\sin \gamma} - 1\right)^2 \) and \( 35^\circ < \gamma < 90^\circ \)

The curves given by the above equations are drawn in Figure 2.11.

![Figure 2.11: Relationship between open-loop frequency specifications and transient behavior](image)

It should be noted that the above empirical equations fit well for some systems, but may give rise to large error for other systems. Figure 2.12 shows the transient step response of the 3rd-order PLL. It is found that equation 2.34 give a very good estimation for \( \sigma_p \), while the approximation for the settling time given by equation 2.35 is rather poor. In light of this, we have obtained the relationship between the simulated settling time and phase margin for this 3rd PLL in Figure 2.13.
The accuracy of the empirical equations is not our primary concern. Instead, we are very interested in the order of the settling time. Over a large variation of phase-margin, the setting time approximately falls within $\frac{5}{\omega_c}$ to $\frac{14}{\omega_c}$. This observation helps to draw a general conclusion: in settling time critical systems, the key to reduce $t_s$ is to increase the cut-off frequency.

### 2.3.3 Noise transfer characteristics

We will now discuss the noise performance of the charge-pump PLL. The block diagram of the PLL is displayed here again for convenience (Figure 2.14). The transfer
function with regard to the two major noise sources in a 3rd-order charge-pump PLL are derived below.

The transfer function for the reference noise is given by

\[
\frac{\phi_{out}(S)}{\theta_{ref}(S)} = \frac{N \cdot G(S)}{1 + G(S)} = \frac{N \cdot K \omega_p (1 + \frac{\gamma}{\omega_c})}{S^3 + S^2 \omega_p + S \frac{K}{\omega_c} + K \omega_p}
\]  \hspace{1cm} (2.36)

It will become clear that this function represents a low-pass filter.

The transfer function for the VCO noise is given by

\[
\frac{\phi_{out}(S)}{\theta_{vco}(S)} = \frac{1}{1 + G(S)}
\]
The frequency response of both transfer functions are given in Figure 2.15. In this figure, we illustrate the case when the divider division ratio $N = 1$, so the magnitude of each curve is flat at 0 dB. The frequency at which the two transfer curves intersect is recognized to be exactly the open-loop cut-off frequency. This is because at the cross-over frequency:

$$|\frac{\phi_{out}(j\omega_c)}{\theta_{ref}(j\omega_c)}| = |\frac{\phi_{out}(j\omega_c)}{\theta_{ref}(j\omega_c)}|$$

$$\frac{1}{1 + G(j\omega_c)} = \frac{N \cdot G(j\omega_c)}{1 + G(j\omega_c)}$$

which gives

$$1 = |G(j\omega_c)|. \text{ since } N = 1.$$  (2.40)

Indeed we see that $\omega_c = \omega_e$.

From Figure 2.15, we see that below the cut-off frequency $\omega_c$, reference noise passes through the PLL without being attenuated. In order to reduce reference noise at the
PLL output. $\omega_c$ should be reduced. On the contrary, above the cut-off frequency, VCO noise will appear at the PLL output without being attenuated. $\omega_c$ is also known as “loop bandwidth” in PLLs. Within the loop bandwidth, the output phase noise is primarily contributed by the reference noise, which is referred as in-band noise. While out of the loop bandwidth, the output phase noise is primarily contributed by the VCO noise, which is referred as out-of-band noise. To reduce the total output phase noise, an optimum balance has to be made between the in-band and out-of-band noise. This results in the trade-off to determine a proper “loop bandwidth”.

In the above analysis we purposely chose $N = 1$ to recognize the particular importance of the open-loop cut-off frequency. It is important to notice that when $N > 1$. 
the reference noise is amplified by $N$. This is shown in Figure 2.16, in which the reference noise is $20\log(N)$ dB higher, while the VCO noise remains as before.

![Figure 2.16: Noise transfer function ($N \neq 1$)]

In a frequency synthesizer, the input reference signal is generally from a crystal oscillator of high-quality, which means there is less noise coming out from the reference signal. This may lead one to conclude that the reference noise can be ignored, and thus $\omega_c$ can be increased to favor the reduction of VCO noise. This may not be the case if $N$ is so large that the noise contributed by the reference signal cannot be ignored. This situation is exacerbated when noise from other sources other than the reference input and the VCO are significantly large. This is addressed below.
Other noise sources

All devices in a PLL contribute to output phase noise. We have the noise of each device referred to its output (to be consistent with the reference and VCO noise), and derive the noise transfer function to the PLL output. The final results are expressed in terms of either the reference noise transfer function, or the VCO noise transfer function.

For phase detector (the combination of PFD and charge-pump),

\[
\frac{\phi_{\text{out}}(S)}{\theta_{\text{PFD}}} = \frac{1}{K_{\text{P}}D} \cdot \frac{\phi_{\text{out}}(S)}{\theta_{\text{ref}}} \tag{2.41}
\]

For the divider,

\[
\frac{\phi_{\text{out}}(S)}{\theta_{\text{divider}}} = \frac{\phi_{\text{out}}(S)}{\theta_{\text{ref}}} \tag{2.42}
\]

For the loop filter,

\[
\frac{\phi_{\text{out}}(S)}{\theta_{\text{LPF}}} = \frac{K_{\text{VCO}}}{S} \cdot \frac{\phi_{\text{out}}(S)}{\theta_{\text{VCO}}} \tag{2.43}
\]

It turns out that both the phase detector and divider contribute to in-band noise. In referring to the reference noise, there is a multiplication factor \(1/K_{\text{P}}D\) for the phase detector noise, and a factor of 1 for the divider noise. According to the analysis in reference noise, these two noise sources are also amplified by \(N\). It is therefore obvious that these two noise sources should be taken into serious consideration. With regard to the phase detector noise, given that \(K_{\text{PD}} = I_{\text{LPF}}/2\pi\) in a charge-pump PLL, an increase in the charge-pump current \(I_{\text{LPF}}\) helps to reduce the phase-noise. However, the increase in \(I_{\text{LPF}}\) will generally lead to larger power consumption and larger loop filter capacitors.
The noise transfer function for the loop filter noise is somewhat different. We illustrate the transfer function of the loop filter noise along with the transfer functions of the reference noise and VCO noise in Figure 2.17. The PLL is a band-pass filter to the loop filter noise. Notice that the loop filter contributes significantly both in-band and out-of-band noise. In particular, from equation 2.43, the larger the VCO gain, the larger the noise contribution from the loop-filter. In practice, people use passive loop filter instead of active filter to reduce this noise source. It will be also helpful to reduce the noise influence by using smaller VCO gain.
Figure 2.18: Noise transfer response for different open-loop phase margin

Phase noise peaking

The close-loop transfer functions for the reference noise source and VCO noise source are displayed in Figure 2.18 again for different open-loop phase margin. When phase-margin becomes small, there is significant peaking in the magnitude response near the cut-off frequency. This peaking will be reflected to the total output phase-noise power. Figure 2.19 illustrates the overall noise contributed by the reference input and the VCO together. In this figure, we have assumed that reference noise and VCO noise follow the typical spectrum of an oscillator, with the reference noise power significantly lower than the VCO noise power. For different phase margin, the phase noise density may vary up to 10dB.
2.4 Topology Discussion

So far, we have investigated the dynamic behavior and noise transfer characteristics of a PLL. We made it clear that the performance of a PLL is greatly dependent on the loop-bandwidth and phase margin. We have shown that with a fairly large variation (30° - 70°) in phase margin, there is no critical degradation of performance in terms of settling time, although the total phase-noise performance can be affected. In critical design, the center issue in a PLL is on the design of the loop-bandwidth. Trade-offs have to be made with regard to the following aspects.

- For fast settling time, it is always good to increase the loop bandwidth since the settling time $t_s$ is inversely proportional to $\omega_r$. 

Figure 2.19: Output phase noise for different open-loop phase margin
• To reduce VCO induced noise, larger loop bandwidth should be used, however, increasing loop bandwidth will increase in-band noise induced by the reference input, the phase detector and the divider.

• With large loop bandwidth, the harmonics of the reference clock will leak through the loop filter and modulate the VCO output. This will generate reference spurs at the PLL output, which is as harmful as phase-noise. The problem can be alleviated by using large reference frequency.

The constraints on the loop bandwidth lead to the search of topologies other than the basic type discussed previously. There are also some other practical design issues in PLLs such as output frequency resolution, high-speed frequency dividers, and so on, which need to be solved by using more complex loop topologies.

In the following sections we will discuss fractional-N PLLs and dual-loop PLLs to address some design trade-offs.

2.4.1 Fractional-N PLL

For better noise suppression and good loop dynamics, large loop bandwidth is expected in frequency synthesizers. Nevertheless, the input reference frequency of an integer-N PLL is restricted by the resolution of the synthesizer. To illustrate this, we change the division ratio of the programmable divider from \( M \) to \( M + 1 \), and the change in the output frequency will be

\[
\Delta f = f_{\text{out},M-1} - f_{\text{out},M} = f_{\text{ref}} \cdot (M + 1) - f_{\text{ref}} \cdot M
\]

\[
= f_{\text{ref}}
\]
If the resolution of the synthesizer is $\Delta f_{\text{min}}$, then it is obvious that $f_{\text{ref}}$ needs to be smaller than $\Delta f_{\text{min}}$. In GHz frequency synthesizers, a programmable divider cannot directly handle high frequency, a prescaler of fixed ratio $N$ may precede the programmable divider. In this case, $f_{\text{ref}}$ needs to be smaller than $\Delta f_{\text{min}}/N$.

A method to remove the restriction on the input reference frequency is to use fractional-N PLLs. In a fractional-N PLL, the prescaler is a dual or multi-modulus divider. Let's consider the simpler dual-modulus $N/(N-1)$ divider. The divider is divided by $N$ for a period of time $T_N$, and then divided by $N+1$ for a period of time $T_{N-1}$. The average division ratio is then:

$$
N_{\text{ave}} = \frac{N \cdot T_N + (N + 1) \cdot T_{N-1}}{T_N + T_{N-1}}
$$

Ideally, by changing $T_N$ and $T_{N-1}$, $N_{\text{ave}}$ may adopt any real number between $N$ and $N+1$, and therefore a fractional-N frequency synthesizer can be tuned continuously. This means that the reference frequency can basically take any large value, without being restricted by the synthesizer resolution any more.

The block diagram of a fractional-N synthesizer is shown in Figure 2.20, in which the dual-modulus prescaler is usually controlled by an accumulator. The overflow signal of the accumulator switches the dual-modulus divider between $1/N$ and $1/(N+1)$ operation.

Since the fractional division ratio is achieved dynamically, it presents a serious drawback: When the divider is in the divided-by-$N$ mode, the feedback reference frequency will be instantaneously larger than the reference frequency. When the divider is in the divided-by-$N+1$ mode, the feedback frequency will be instantaneously
smaller than the reference frequency. In either case, the phase detector will drive current pulses to the VCO control input periodically. The consequence is that some significant sidebands will occur in the vicinity of the carrier frequency.

To overcome this problem, a ΔΣ modulator can be utilized as the modulus controller. Through the randomization and noise shaping capability of ΔΣ modulators, fractional spurs can be reduced.

2.4.2 Dual-loop PLLs

Dual-loop PLLs are often used in practice to ease some of the design challenges. Two types of dual-loop PLLs were described in [1]. Here we introduce a third dual-loop PLL that utilizes phase-interpolation techniques. This architecture provides a significant benefit for data recovery applications as clarified below.

Figure 2.21 shows the diagram of the dual-loop PLL. In this figure, the first loop is a regular PLL, which is used to generate an output frequency of very low phase-jitter.
Figure 2.21: A Dual-loop PLL using phase-interpolation technique

Note that the input reference clock to this loop is coming from a crystal oscillator, which contributes much less noise than a regular oscillator. In comparison with the case of a very noisy reference input, here one can choose a larger loop bandwidth to suppress oscillator noise.

The second loop is composed of a phase interpolator, a phase detector and a low-pass filter. The phase interpolator is controlled by the output of the low-pass filter in such a way that the phase of the interpolator output will be adjusted and aligned to the phase of clock signal embedded in the received data stream. Here the bandwidth of the low-pass filter is chosen to be small such that jitter originated from the input data is maximally suppressed.

Notice that by using the dual-loop architecture, the loop bandwidth of the two loops can be designed independently. In the first loop, the loop bandwidth is chosen to be large to suppress VCO noise, while in the second loop, the loop bandwidth is chosen to be small to suppress input noise.

Another merit of this dual-loop architecture is that there is only one VCO. In general, if there are more than one VCO residing in the same die, there will be big
chance that the VCOs will talk to each other, and their output frequencies will be pulled toward each other. By using phase interpolation techniques, one can avoid the frequency pulling hazard. To achieve phase interpolation, firstly, one needs to generate multi-phase clock signals (0, \( \phi_1, \phi_2, \ldots, 2\pi \)) from the VCO. For each two neighboring phases, certain operations can be performed to obtain finer phase steps. Mathematically, given two sine-wave signals of phase \( \phi_a \) and \( \phi_b \), and let

\[
y = (1 - x) \cdot \sin(\omega_t + \phi_a) + x \cdot \sin(\omega_t + \phi_b)
\]

\[\text{(2.48)}\]

where \( x \in [0, 1] \) is the interpolation weight.

\( y \) is still a sine-wave signal, whose amplitude and phase are given by

\[
A_y = \sqrt{[(1 - x) \cos(\phi_a) + x \cos(\phi_b)]^2 + [(1 - x) \sin(\phi_a) + x \sin(\phi_b)]^2} \\
\phi_y = \tan^{-1} \frac{(1 - x) \sin(\phi_a) - x \sin(\phi_b)}{(1 - x) \cos(\phi_a) + x \cos(\phi_b)}
\]

\[\text{(2.49)}\]

\[\text{(2.50)}\]

To look at the linearity of the interpolation, \( \phi_y \) is drawn versus \( x \) in Figure 2.22 (without losing generality, let \( \phi_a = 0 \) and \( \phi_b = 45^\circ, 60^\circ, 90^\circ \) respectively). It turns out that when \( \phi_b - \phi_a \leq 45^\circ \), the interpolation linearity is fairly good. Circuit implementation of phase interpolation is trivial, e.g., a double-balanced Gilbert mixer [7] [5] [8] can be used as a phase interpolator [9]. Finally, it is important to note that some complex digital circuits are needed to implement the weight control in the phase interpolator.
Figure 2.22: Linearity of phase interpolation
A number of different circuits can be used for phase detection. In earlier days, people used multiplier, Exclusive-OR gate and JK-flipflop as phase detectors (PDs). Phase-locked Loops using such PDs require auxiliary circuits to acquire frequency locking. A second class of phase detector, the tri-state phase detector, provides both phase and frequency detection capability. The tri-state phase detector is also called phase-frequency detector, and is favorably (if not exclusively) used in most phase-locked loops.

This chapter will go over the features of different phase detectors, and focus on the design issues of phase-frequency detectors (PFDs). A new dynamic-logic PFD of high-performance is presented to overcome the drawbacks of existing PFDs.

### 3.1 Phase Detectors Review

This section examines the operation and characteristics of different phase detectors. Two features that critically influence the loop dynamics of PLLs are under close inspection: the monotonic detection range and the frequency detection capability.
3.1.1 Multiplier

An analog multiplier (Figure 3.1 a) can be used as a phase detector. A possible implementation of a multiplier is the Gilbert cell (Figure 3.1 b). For proper operation, the reference input signal needs to be a sine wave:

\[ V_{ref} = A_1 \sin(\omega_1 t + \theta_1) \]  

(3.1)

and the other signal is generally a square wave, which can be described by a Fourier series:

\[ V_{dr} = \frac{4A_2}{\pi} \left[ \cos(\omega_2 t + \theta_2) + \frac{1}{3} \cos(3\omega_2 t + \theta_2) + \cdots \right] \]  

(3.2)

Multiplying \( V_{ref} \) and \( V_{dr} \), we get:

\[
V_{out} = V_{ref} \cdot V_{dr} = \frac{4A_1 A_2}{\pi} \sin(\omega_1 t + \theta_1) \cdot \left[ \cos(\omega_2 t + \theta_2) + \frac{1}{3} \cos(3\omega_2 t + \theta_2) + \cdots \right] \]  

(3.3)

When \( \omega_1 = \omega_2 \), \( V_{out} \) is given by

\[
V_{out} = \frac{2A_1 A_2}{\pi} \left| \sin \theta_e + \cdots \right| \]  

(3.4)

where \( \theta_e = \theta_1 - \theta_2 \) is the phase error.

The first term - the "DC" component in the above equation represents the phase error. After low pass filtering, all the other higher frequency AC components will be removed, and the DC term \( \bar{V_d} \)

\[
\bar{V_d} = K_{PD} \sin(\theta_e) 
\]  

(3.5)

where \( K_{PD} = \frac{2A_1 A_2}{\pi} \) is the detection gain. Note that the detection gain is dependent on the amplitude of the inputs \( A_1 \) and \( A_2 \).
Figure 3.1: Multiplier used as phase detector
The transfer characteristics of the analog PD is shown in Figure 3.2. we see that the analog multiplier operates nonlinearly, and the monotonic detection range is \((-\pi/2, \pi/2)\).

![Figure 3.2: Phase characteristics of the multiplier](image)

In the above discussion, we assumed that both \(V_{ref}\) and \(V_{div}\) have no DC component. If DC component exists in both inputs, the PD will have a phase offset.

It is obvious that when the two inputs have different frequencies \((\omega_1 \neq \omega_2)\), the DC term of \(V_{out}\) will be zero, so the analog PD is not frequency sensitive.

### 3.1.2 Exclusive-OR gate

The operation of an Exclusive-OR gate used as a phase detector is shown in Figure 3.3. When the two input square-wave signals are 90° out of phase, the output
will have a 50% duty cycle, giving an average value of zero. When the phase difference deviates from 90°, the output duty cycle is no longer 50%, and the average value of the output is proportional to the phase difference. The monotonic detection range in this case is (0, π). If the waveforms of the input signals are not symmetrical, the detection range may be significantly reduced. This can be illustrated by considering an extreme situation when both two inputs have a small duty cycle, e.g., 1%. In this case, the average output is clipped to a value close to zero and the whole detection range is only \( \frac{2\pi}{100} \). Obviously, this PD is not frequency sensitive.

### 3.1.3 JK-flipflop

An edge triggered JK-flipflop can be used as a PD, as shown in Figure 3.4. In this device, a positive edge at the J input triggers the output to “high” and a positive edge at the K input triggers the output to “low”. Note that the average of the output \( \hat{v}_d \) reaches zero when the two input signals are at opposite phase. The average output reaches maximum when there is a 360° phase difference, and reaches minimum when the two signals are in phase. So the phase detection range of this PD is twice of previous two PDs. In addition, because this PD is edge triggered, the waveforms of the inputs do not need to be symmetrical.

Similar to the Exclusive-OR gate phase detector, the JK-flipflop is not capable of frequency detection.

### 3.1.4 Tri-state phase detector

A tri-state phase detector, or phase-frequency detector (PFD), is shown in Figure 3.5. It is composed of two D-flipflops and an AND-gate, and have two output terminals \( UP \) and \( DV \).
Figure 3.3: A Exclusive-OR PD and its phase-characteristics
Figure 3.4: A JK-flipflop PD and its phase-characteristics
Figure 3.5: A tri-state PFD with charge-pump and its ideal phase characteristics
At any time, the two outputs of the PFD, UP and DN, will be at one of four states: 00, 01, 10, 11, with UP = high, DN = low represented by state 10, and etc. The fourth state is an unstable state (at state 11, the AND-gate will reset the D-flipflops). In general, the tri-state PFD is implemented together with a charge pump, as illustrated in the dashed-line box in Figure 3.5. In this simple illustration, the charge pump is composed of a P current source and a N current source. The two current sources are controlled by the two outputs of the PFD. An active signal at UP (UP = 1) will turn on the P current source, allowing a positive current to flow into the output node (node \( CP_{out} \)) and make the output voltage to rise. Similarly, an active signal at DN (DN = 1) will turn on the N current source, which pulls current from the output node, causing the voltage at the output node to reduce. When both UP and DN are not active (UP=0, DN=0), both current sources are turned off from the output node, then the voltage at node \( CP_{out} \) will remain unchanged.

The operation of the PFD is illustrated in Figure 3.5 as well. A positive transition in the REF signal causes the UP terminal to transit to “high”. Similarly, a positive transition in the DIV signal causes the DN terminal to transit to “high”. When both UP and DN are “high”, the PFD is reset, which brings UP and DN to “low”. The average output of the PFD is given in Figure 3.5. This time, the output put is expressed in current instead of voltage, since our interest is the average current flowing in and out of node \( CP_{out} \).

It turns out that the ideal linear phase detection range of this PFD is \((-2\pi, 2\pi)\), which is twice of the JK-flipflop PD’s detection range. Out of this range, the phase characteristic curve is periodic with a period of \(2\pi\). To look at the extra frequency detection feature provided by this device, we assume the REF frequency is higher.
than the DIV frequency. Notice that now the REF signal will have more positive transitions than the DIV signal, as such the UP terminal will have more chances to stay at “high” than the DN terminal. This means more current will flow into node $CP_{out}$, and causes the voltage to increase.

We conclude that the tri-state phase-detector has superior performance than other phase detectors in terms of phase detection range and frequency sensitivity.

Up to now, our focus is on the ideal behavior of PDs. We will continue to investigate the non-ideal characteristics of PFD.

### 3.2 Design Issues in Phase-frequency Detectors

#### 3.2.1 Dead-zone

Instead of using complex reset D-flipflops, a tri-state PFD is often implemented in the form of Figure 3.6. Note that the reset signal is generated from intermediate signals instead of from the UP and DN outputs directly. If the phase difference of the input signals is small, a positive transition at UP or DN will be closely followed by a reset operation. If the propagation delay at UP and DN is large enough, it is possible that UP or DN may be pulled to low by the reset operation before it completes a positive transition. Consequently, the charge pump output voltage will keep intact at small phase difference. This phenomenon is called “dead-zone”, and is shown in Figure 3.7, with the dead-zone exaggerated. In a PLL, if the input phase error is within the dead-zone, it will have little influence on the VCO control voltage. Consequently, the PLL operates as if the loop is open, and the oscillator noise will appear at the PLL output without being suppressed.
A delay cell is usually inserted in the reset path such that the reset pulse is wide enough to allow the PFD outputs become effective before they are reset (Figure 3.6). By doing so, the UP and DN outputs are given enough time to make a positive transition, and thus dead-zone can be reduced. Nevertheless, the maximum operating frequency of the device will be limited by the total delay of the reset path. Since the maximum operating frequency happens when REF and DIV are in opposite phase [10], an approximate estimation of the operating frequency is given as follows:

Assume equal gate delay $t_g$ for the internal gates and $t_G$ for the driving gates of UP/DN nodes in Fig. 3.6. Also, denote the delay time of the delay element as $t_d$. It follows that the minimum RESET pulse width is $3t_g + t_d$ and the delay time from
the input (REF or DIV) to RESET is $2t_g + t_d$. We state that any transition edge in
REF and DIV will be ignored during the reset operation (RESET = 1), and this is
considered abnormal. To avoid such a situation, a transition in REF or DIV should
happen after the reset operation. The maximum operating frequency is therefore
given by

$$f_{max} = \frac{1}{2(5t_g + 2t_d)}$$  \hspace{1cm} (3.6)

which is illustrated graphically in Figure 3.8.

\subsection{Blind-zone}

The PFD output characteristic curve in Figure 3.7 rises monotonically with phase
error. When the phase error approaches ±$2\pi$, the curve switches to the opposite sign
abruptly. The cause of the abrupt polarity change can be explained through timing
analysis. In Figure 3.9, the occurrence of a REF rising edge coincides with a reset operation \((RESET = 1)\) and therefore does not have any impact on the PFD output. This situation is depicted in Figure 3.9 by the dotted area. Consequently, the leading REF signal is incorrectly indicated by the PFD as lagged \((DN \text{ is wider than UP})\) in a subsequent cycle. As long as a PFD is in reset mode, it will be blind, i.e., any input transition will be unseen by the PFD.

It becomes obvious that the blind-zone of a PFD is directly related to its reset time. Blind-zone is detrimental to the PFD settling behavior, and will slow down the lock time. This can be perceived on the phase-plane of Figure 2.2 - during the settling
Figure 3.9: Timing analysis of blind-zone

procedure, the (phase-error, frequency-error) will follow a trace moving toward the origin. If an input transition of the PFD happens to occur in the blind-zone, the trace will bounce further away from the origin, and therefore it will take a longer time for the trace to converge to the origin.

The situation that an input transition falls in the blind-zone is nevertheless a random event. In the phase-plane of Figure 2.2, given a different initial conditions, a PLL may follow a trace that may or maynot experience a blind-zone event. If it does, the PLL is driven to a follow a new trace. A blind-zone event is hard to capture during a PLL settling process. It also becomes obvious that one can not completely
eliminate the blind-zone through circuit design. In practice, the only effective way to reduce the probability of having a "blind-zone event" is to reduce the blind-zone.

Having a blind-zone is equivalent to a reduced phase detection range. Here we simply point out that the lock range of a PLL is proportional to the phase-detection range \([11]\) of the phase detector used. Based on previous discussion, one can easily estimate the phase detection range \(\phi_{\text{max}}\) according to the reset time of the PFD:

\[
\phi_{\text{max}} = 360\degree - t_{\text{reset}} \times f_{\text{in}} \times 360\degree \tag{3.7}
\]

where \(f_{\text{in}}\) is the frequency of input signals, and maximum input frequency \(f_{\text{in},\text{max}}\) is constrained by \(f_{\text{in},\text{max}} \leq \frac{1}{2} f_{\text{reset}}\).

One can find from equation 3.7 that the phase-detection range is inversely proportional to the input frequency. Therefore, at high-operating frequency, a PFD will experience a blind-zone event more frequently.

### 3.3 Dynamic Logic Phase-frequency Detectors

Dynamic CMOS logic circuits and especially domino logic have been used in high performance designs due to their faster operating speed and potential saving in power. Several dynamic logic phase-frequency detectors were introduced in the last five years \([12]\) \([13]\) \([14]\) \([15]\) \([16]\).

A pioneer dynamic logic PFD is shown in Figure 3.10 \([12]\). In this circuit, the outputs are reset almost immediately following a positive transition at one of the inputs if initially a positive transition has occurred at the other input. As a result, the circuit experiences dead-zone problem, which cannot be effectively solved by adding an extra delay in the reset path similar to Figure 3.6. Furthermore, the detection
range of this circuit is only \((-\pi, \pi)\). To see this problem, consider a situation that a positive transition in \(\text{REF}\) occurs when \(\text{DIV} = 1\): In a normal tri-state PFD, following the rising edge of \(\text{REF}\), \(\text{UP}\) should be pulled to high. In this circuit, \(\text{UP}\) will stay at low instead. This is because node \(U_1\) is discharged and node \(U_2\) is charged immediately after the rising edge of \(\text{REF}\). \(\text{UP}\) gets no time to transit to high.

The above dead-zone problem was partially improved in [13] (U.S. Pat. No. 5,661,419. 1997) (Figure 3.11). This is accompanied by the three inverters and two NMOS transistors at the \(\text{REF}\) (or \(\text{DIV}\)) input, which basically form a pulse generator to delay the discharge of node \(U_1\) (and \(U_2\)) such that \(\text{UP}\) and \(\text{DN}\) can become effective.
for a certain amount of time. A defect in this circuit is that a DC path might be formed from the power supply to the ground through node U1 or U2. For instance, if DIV is transmitting from 0 to 1 while REF=0, a DC path is formed through node U1. Depending on whether the voltage at node U1 is low enough to be below the threshold of the following PMOS transistor, the UP signal may or may not be reset. Because of this, this circuit has only a linear detection range of \((-\pi, \pi)\).

Shown in Figure 3.12 is another dynamic PFD circuit [14]. Unlike the two dynamic PFD discussed above, in this design, the reset action is initiated solely by the PFD outputs. This means that a reset process will not happen until UP and DN become effective. For this reason, dead-zone is not visible in this design. There also exists a defect in this circuit – assume a rising edge of REF has caused a positive transition at
Figure 3.12: A dynamic PFD without dead-zone

UP, and then REF returns to 0. Subsequently, a rising edge of DIV causes a positive transition at DN. Since REF=0, UP=1, and DN=1, a DC path will be formed between the power supply and the ground via node U1. This situation is very similar to that found in Figure 3.11. The linear detection range of this circuit is only \((-\pi, \pi)\) as well.

A common problem in the above three dynamic PFDs is that they are not purely edge sensitive. By using two pulse generators (similar to that used in Figure 3.11), a dynamic logic PFD disclosed in [15] (U.S. Pat. No. 5,963,059. 1999) (Figure 3.13) is made to be purely sensitive. The device then functions in the same way as demonstrated by the original tri-state PFD (Figure 3.5). Ideally, the phase detection range
will approach \((-2\pi, 2\pi)\). Yet, according to our analysis in section 3.2, there exists a blind-zone of approximately three gate-delay. This circuit is relatively very simple, but requires a careful control of the delay of the pulse generators in order to reduce dead-zone. Extra latches are used (the cross connected inverters) in this circuit to ensure proper operation, which also slows down its operating speed.

Figure 3.13: A dynamic PFD with reduced blind-zone
3.4 A Novel Dynamic-logic Phase-frequency Detector

In an effort to design a PFD capable of working at gigahertz frequency with minimum dead-zone, minimum phase offset, and reduced blind-zone, we have proposed a new dynamic PFD using domino logic. Our design goals are accomplished by maintaining circuit integrity, that is, no DC path (short circuit) shall appear in a normal circuit.

The new PFD is shown in Figure 3.14. MU2-3 and MD2-3 form two inverter structures in the precharge stage that basically prevent the device from being short-circuited as happened in Figure 3.11 and Figure 3.12. It should become obvious that the integrity of the circuit and its operation will remain intact with a permutation of MU1 and MU2, MU3 and MU4 (also MD1 and MD2, MD3 and MD4).

The operation and the merits of this new PFD are further analyzed in the following discussion.

3.4.1 Circuit operation

To inspect the behavior of this PFD more closely, we have shown a finite-state diagram of the PFD with all its four possible states included (Figure 3.15). In the diagram, each state transition is accompanied by its corresponding transition condition, which is basically the positive transition of REF or DIV, denoted by $\text{REF}^\uparrow$ and $\text{DIV}^\uparrow$ respectively. Two simultaneously occurring positive transitions are simply denoted by $(\text{REF}^\uparrow\text{DIV}^\uparrow)$.

The operation of the circuit can be explained according to the charging and discharging behavior of typical dynamic circuits. Assuming initially UP and DN are both low (state = 00) and REF and DIV are low, nodes U1 and D1 are precharged to
Figure 3.14: The proposed dynamic PFD
Figure 3.15: Finite-state diagram of PFD operation

high through transistor MU1-2 and MD1-2 respectively. At the rising edge of REF, MU6-7 are turned on, node U2 is pulled to low, which drives UP to high. Similarly, a rising edge of DIV will drive DN to high. If the rising edge of REF comes first, the state will transit from 00 to 10, and then transit from 10 to 11 upon the arrival of a rising edge of DIV. If the rising edge of DIV comes first, the state transition will be from 00 to 01 and then to 11. In the case that the rising edges of REF and DIV come simultaneously, the state will transit from 00 to 11 directly. State 11 is an unstable state, as identified by the dashed circle. At state 11, UP and DN will turn on MU3-4 and MD3-4. Nodes U1 and D1 are discharged accordingly, and nodes U2 and D2 will be charged to high through MU5 and MD5 respectively. This returns the circuit to state 00.

Indeed, all the tri-state PFDs operate in the above manner, i.e., a rising edge of REF drives UP to high, and a rising edge of DIV drives DN to high. When both UP and DN are high, the circuit is reset, returning UP and DN to low. This general statement may disguise the truth. As we have carefully inspected all the previously mentioned dynamic PFDs, when the phase error is beyond \((-\pi, \pi)\), some of the
dynamic PFDs do not follow the statement any more, and in fact, most of them are not capable of properly indicating the phase difference. For this reason, we examine the operation of our PFD cautiously as below.

Assume REF is leading DIV by more than $180^\circ$ degree. Initially, a leading rising edge of REF drives the device to state 10 as before. Now, upon a DIV rising edge, DN will be pulled to low. At this moment, REF is low, but UP is driven high. So, the path formed by MU1-2 is shut off and no charge is injected from $V_{dd}$ to node U1 during the time when U1 is being pulled to low through path MU3-4. UP is therefore reset to low normally. This clarifies that the detector can operate correctly when phase error lies between $\pi$ and $2\pi$.

### 3.4.2 Performance evaluation

High-speed operation is one of our design concerns, so let’s examine the maximum operating frequency of the circuit. As before, we assume DIV lags REF by $180^\circ$. Also we assume that the circuit is initially at state 00 and transits to state 10 upon the arrival of the rising edge of REF. Refer to Figure 3.14. at state 10, an incoming rising edge of DIV causes node D2 to discharge. After $t_1 = t_{D2,f} + t_{D_{N,f}}$, DN reaches high and the circuit transits to state 11. This triggers the reset mode, and after $t_2 = t_{D1,f} + t_{D_{2,f}} + t_{D_{N,f}}$, UP/DN are pulled down to low. Note that REF cannot rise immediately after the falling edge of UP/DN, instead, it has to wait for an additional time $t_3 = t_{l_{1,f}}$ for node U1 to be precharged to high so as to prepare for next cycle’s operation (Figure 3.16). The maximum operating frequency is estimated to be

\[
f_{\text{max}} = \frac{1}{2(t_1 + t_2 + t_3)}
\]

\[
= \frac{1}{2(t_{D1,r} + t_{D1,f} + t_{D2,r} + t_{D2,f} + t_{D_{N,r}} + t_{D_{N,f}})}
\]
In the above representations such as $t_{DN, f}$, the first subscript refers to the corresponding node in charging or discharging, the second subscript refers to a rising or falling transition.

For simplicity, let $t_y = t_{D1, r} = t_{D1, f} = t_{D2, r} = t_{D2, f}$ and $t_G = t_{DX, r} = t_{DN, f}$, the maximum operating frequency is given as

$$f_{\text{max}} = \frac{1}{4(2t_y + t_G)}$$

where $t_G$ depends on external loads at nodes UP and DN.

Note that $t_z$ corresponds to the minimum pulse width of UP and DX, which is one fourth of $1/f_{\text{max}}$. Figure 3.19 demonstrates the waveform when the two inputs are at maximum operating frequency. Indeed, in our implementation the minimum pulse width of UP and DX is around 100 ps, one fourth of the period of the maximum operating frequency.

The phase characteristic of the PFD is obtained as shown by the solid-line in Figure 3.17. This clarifies our previous analysis that the circuit is able to operate up to 360° phase difference. Near the 360° phase difference, the phase curve falls down, presenting a blind-zone. Notice that the blind-zone is considerably smaller than that of the conventional PFD (dashed-line). This is attributed to the smaller reset time (or small minimum pulse width). We shall point out that blind-zone is not totally eliminated, as is common in all other practical implementation. In Figure 3.17, the operating frequency applied is 500 MHz, and the blind-zone is less than 10% of a clock cycle, reasonably small.

The phase curve near small phase error is examined in Figure 3.18. no obvious gain degradation is found even though the phase difference is as small as 1 ps. Also, because the PFD is symmetrical in structure, no phase offset is incurred.
The frequency sensitivity of the proposed PFD is illustrated in the waveform of Figure 3.20. The two input signals are at different frequencies.

Charge-sharing issue does not seem to impact the normal operation of the PFD although charge transfer through transistors MU2, MU3 (MD2, MD3) changes the voltage level at node U1 (D1). This can be attributed to the larger noise threshold.
Figure 3.17: Phase characteristics of the proposed PFD (versus prior arts)

Figure 3.18: Magnified view near zero phase error
It turns out that the cascaded PMOS transistors (MU1-2, MD1-2) are the limiting factor for the PFD to run at higher speed. Nevertheless, compared with other PFDs requiring more complexity to achieve similar performance, speed is not sacrificed in this design. Benefiting from the very simple structure, no particular design efforts and trade-off are required in designing this PFD.

Under different corner and temperature tests, the device appears to be very robust, and still presents a linear detection range greater than 180° even the power supply is as small as 0.7 volt. This is illustrated in Figure 3.21.

At 500 MHz, the power consumption of this device is less than 400 μW at different power supply, which is trivial compared with other blocks in a PLL (Figure 3.22).
3.5 Conclusion

A new domino logic PFD with no visible dead-zone and extended phase-detection range is designed for high-frequency operation. By identifying shortcomings and design defects in previous PFD circuits, we have made effort to maintain the operation integrity of domino logic circuit, and effectively improved the performance of dynamic logic PFDs.
Figure 3.21: Phase characteristics under various power supply
Figure 3.22: Power consumption under various power supply
CHAPTER 4

CHARGE PUMP

4.1 Considerations in Charge Pump Design

In charge-pump PLLs, a charge pump is used together with a 3-state phase-frequency detector as shown in Figure 4.1. The two outputs, UP and DN, of the PFD control the two current sources $I_{UP}$ and $I_{DN}$ respectively. Ideally, $I_{UP}$ and $I_{DN}$ shall be constant and equal. The capacitor at the output is charged by $I_{UP}$ and discharged by $I_{DN}$ depending on which of the two switches are turned on. When a PLL is in lock, the UP signal and the DN signal of the PFD maintain a minimum pulse width, and turn on the two switches in the charge-pump for equal time. Therefore the average charge goes into the capacitor will be zero. In the case the UP pulse is wider than DN pulse, then $I_{UP}$ will drive the output voltage to increase, and vice versa.

A major problem in charge-pump PLL is the mismatch of the $I_{UP}$ and $I_{DN}$. The cause of the mismatch and its effect on PLL performance will be examined in the next two sections.
4.1.1 Current mismatch

When a PLL is in lock, the charge stored in the capacitor should remain constant. If \( I_{UP} \) and \( I_{DN} \) are not equal, \( I_{UP} \) and \( I_{DN} \) need to satisfy the following expression

\[
I_{UP} \times T_{UP} = I_{DN} \times T_{DN}
\]  

(4.1)

for each reference cycle.

Therefore the pulse width of \( I_{UP} \) and \( I_{DN} \) will be different. The output of the charge pump will then have spikes as shown in Figure 4.2. Using Fourier transformation, the voltage signal across the capacitor can be decomposed to a DC component, and many AC components at the harmonic frequencies of the reference signal. The loop filter following the charge-pump attenuate the AC components as long as the AC components fall within the loop bandwidth. The remaining AC components showing up at the VCO control input will modulate the VCO output and give rise to spurs in the output spectrum. According to the analysis in section 2.2, spurs are expected to show up at the harmonics of the reference signal (relative to the carrier frequency).
Component leakage is another very important factor that causes reference spurs. Due to leakage current, charge stored in the capacitor may not be able to remain constant as a PLL stays locked. From time to time, the phase-locked loop needs to temporarily enter an unbalanced state by causing $T_{UP} \neq T_{DN}$, such that the charge-pump will deliver a certain amount of charge to compensate for the charge loss.

We have mentioned the undesirable effect caused by reference spurs in previous chapters. Reducing reference spurs is therefore a major task in designing charge-pumps. This is done by minimizing the $I_{UP}$ and $I_{DN}$ current mismatch through proper circuit design, and by choosing proper loop bandwidth for maximum spur attenuation. We will address these issues further in this chapter.
4.1.2 Mismatch reduction

The current sources of the charge pump are implemented using current mirrors. In CMOS implementation, due to channel length modulation, the drain-source current of a MOS transistor does not maintain constant when its drain-source voltage varies. For the charge-pump, we have the variation of $I_{IP}$ and $I_{DN}$ with regard to the output voltage drawn in the same figure (Figure 4.3). The two output currents are equal only at point $A$. While it is usual that some PLLs may have only a fixed output frequency, and thus a single VCO control voltage at lock state, in general for frequency synthesizer PLLs, the VCO control voltage swings over a wide voltage range. We have to seek methods to reduce the current mismatch.

![Figure 4.3: Charge pump current mismatch due to channel length modulation](image)

Figure 4.3: Charge pump current mismatch due to channel length modulation
A natural way to cope with this problem is to reduce channel length modulation effect in the current mirrors. This can be accomplished by using cascaded current mirror, or by adopting moderately larger channel length without significantly decreasing the turn-on speed of the charge-pump. To further reduce the mismatch due to channel length modulation, one may use a replica amplifier to compensate for the mismatching current. The idea is shown in Figure 4.1, in which the V-I converter delivers to the capacitor a compensation current based on the extra voltage of $V_H - V_A$ as illustrated in Figure 4.3.

Figure 4.4: Charge pump with replica circuit to compensate current mismatch
4.2 System Performance Optimization

For convenience, the open-loop transfer function of a 3rd-order charge pump PLL is rewritten as follows.

\[ G(S) = \frac{I_{CP} \cdot (1 + S/\omega_z) \cdot K_{VCO}}{2\pi \cdot N \cdot (C1 + C2)S^2(1 + S/\omega_p)} \]  

(4.2)

According to our earlier discussion, the loop dynamics and phase noise performance are closely related to the loop bandwidth \( \omega_c \) and phase margin \( \gamma \) (refer to Chapter 2). The expressions for \( \omega_c \) and \( \gamma \) are given as follows.

\[ |G(S)|_{S=j\omega_c} = \left| \frac{I_{CP} \cdot (1 + j\omega_c/\omega_z) \cdot K_{VCO}}{2\pi N(C1 + C2)\omega_z^2(1 + j\omega_c/\omega_p)} \right| = 1 \]  

(4.3)

\[ \gamma = \tan^{-1}\left( \frac{\omega_c}{\omega_z} \right) - \tan^{-1}\left( \frac{\omega_p}{\omega_z} \right) \]  

(4.4)

Equation 4.3 4.4 are the basis of our later discussion in PLL performance optimization with regard to fast settling time and optimum loop dynamics. Reference spurs reduction is also one of the important aspects in system performance optimization, which is discussed in section 4.3.2.

4.2.1 Division ratio variation

An important observation [17] is that in frequency synthesizers, the division ratio \( N \) is programmed to change in a certain range. For example, in 2.4 GHz Wideband CDMA applications, the usable bandwidth is 80 MHz, which gives rise to a change of \( N \) in percentage close to 4%. In some other applications, the change is even bigger. The change in \( N \) will make a PLL to deviate from its designed loop bandwidth and phase-margin. Assume \( \omega_z \ll \omega_c \ll \omega_p \), from equation 4.3, we have the following approximation.

\[ \left| \frac{1 + j\omega_c/\omega_z}{\omega_z^2(1 + j\omega_c/\omega_p)} \right| \approx \frac{\left| j\omega_c/\omega_z \right|}{\omega_z^2 \cdot 1} = \frac{1}{\omega_z \omega_c} \]  

(4.5)
It follows that

\[ |G(S)|_{s=j\omega} = \left| \frac{I_{CP}K_{VCO}}{2\pi f(C1 + C2)\omega_c\omega_r} \right| = 1 \]  \hspace{1cm} (4.6)

So when a change happens to \( N \), in order to satisfy the unit gain condition, \( \omega_c \) needs to change in the opposite direction with approximately the same percentage. A change in \( \omega_c \) causes the phase margin to change as well (refer to equation 4.4).

This means that the design of loop bandwidth and phase margin for optimum loop dynamics and phase noise is valid only at one division ratio. When \( N \) takes other values, the system performance will be degraded.

A remedy to the problem is to adjust other parameters in equation 4.3 in accordance with the variation of \( N \), such that the loop bandwidth and phase margin remain unchanged. In general, it is hard to vary the VCO gain. One may change the capacitance of \( C1 \) and \( C2 \) to compensate for the change in \( N \), but in order to keep the phase margin intact, one needs to simultaneously change the resistance of \( R2 \). Another simple solution is to use a programmable charge-pump. Obviously, when the charge-pump current changes exactly the same percentage as \( N \), \( \omega_c \) and \( \gamma \) remain the same.

### 4.2.2 Fast settling

Settling time can be reduced through the use of large loop bandwidth. However, increasing loop bandwidth may degrade phase-noise performance and cause large reference spurs.

In certain frequency hopping systems, both fast settling behavior and good phase noise performance are important. To achieve both goals, one may design a PLL that provides a wide loop-bandwidth during the transient settling process, while returning
to a moderately lower loop-bandwidth for spurs and phase-noise reduction after the
PLL is in lock [18].

The idea here is similar to the one described in section 4.2.1. Through the manipu-
lation of the loop parameters in equation 4.3 4.4, one can increase the loop bandwidth
instantly during a settling process, while maintaining $\tau$ intact all the time.

Suppose that the loop bandwidth is increased by a factor of $\alpha$, to satisfy the
constant $\tau$ requirement, one can increase $\omega_2$ and $\omega_p$ by the same factor of $\alpha$. This is
further achieved by reducing either $C_1$ and $C_2$, or $R_2$ by the same factor. We favor
the reduction of $C_1$ and $C_2$, as will be clarified shortly.

Now return to equation 4.3, by multiplying $\omega_r$, $\omega_2$, and $\omega_p$ by a factor of $\alpha$, there
will be a term $\alpha^2$ existing in the denominator. This extra term can be canceled out
by properly changing $\frac{I_{cp}K_{1C}}{N(t_1-t_2)}$. If $C_1$ and $C_2$ have been reduced by a factor of $\alpha$, the
$\alpha^2$ term can be canceled out by either increasing $I_{cp}$ or reducing $N$ by a factor of $\alpha$
(Again, it is not easy to change $K_{1C2}^2$). Alternatively, if $R_2$ is reduced by a factor of
$\alpha$, then one needs to increase $I_{cp}$ by a factor of $\alpha^2$, or to reduce $N$ by a factor of $\alpha^2$.

The choice of changing $I_{cp}$ or $N$ is left open, depending on which one is more
conveniently available in a specific design. In either case, it is desirable that the
variation factor in $I_{cp}$ or $N$ is $\alpha$ instead of $\alpha^2$, in order to ease the design and reduce
power consumption.

The change of $I_{cp}$ can be achieved through the use of programmable charge-pump
described in the next section.
4.3 A Programmable Charge Pump

4.3.1 Circuit description

Figure 4.5 is a charge pump capable of low supply voltage operation. Although a 0.18 µm CMOS process is used in this design, the channel length adopted in the circuit is not minimum. This is because with minimum channel length, MOS transistors experience serious channel length modulation problem, which makes the $I_{UP}$ and $I_{DN}$ current sources largely mismatched.

![Charge pump circuit diagram](image)

Figure 4.5: Charge pump circuit

The transistor M1-2 and M3-4 in Figure 4.5 are current mirrors to generate $I_{UP}$ current and $I_{DN}$ current. The two switches Mp1 and Mn1 are controlled by the phase detector outputs $UP$ and $DN$ (where $UPb$ is the inverse of $UP$, and $DNb$ is the inverse of $DN$). When $UP$ becomes active (high), Mp1 is turned off and Mp2 is
turned on, such that the current mirror M1-2 becomes functional, and a current goes down from M2 to the output node \(V_{cp, out}\). When \(UP\) becomes low, Mp1 is turned on, which pulls the gate voltage of M1-2 to the positive power supply, and effectively shuts off M1-2. The transistor Mp2 is used to ensure that M2 gate voltage is pulled high enough to shut off M2. Similar analysis can be applied to the current mirror M3-4.

The two biasing currents in the charge pump are supplied by a 4-bit programmable current source respectively. The programmable current source adopted is illustrated in Figure 4.6. In this circuit, there is also an adjustable analog input, which can be used to further extend the output current range. With this configuration, the charge-pump is able to provide the largest flexibility for several purposes.

- Facilitate the dynamic division ratio compensation.
- Speed up the settling behavior during frequency transition.
- Compensate the current mismatch caused by channel length modulation.

Figure 4.7 illustrates the transient response of the charge pump, with \(UP\) and \(DN\) activated by two signals of same frequency and 90° phase delay. The operation of the charge pump is rather nonlinear, due to the fact that M2 works frequently in triode region whenever the pump output voltage \(V_{cp, out}\) is charged to the positive rail of the power supply. Also, when \(V_{cp, out}\) becomes larger, the average charge delivered to the integrating capacitor will become smaller and smaller, due to the mismatch caused by channel length modulation. This is why the increase of the filter output \(V_{filt, out}\) slows down with time.
Figure 4.6: Programmable current source
Figure 4.7: Charge pump transient response
4.3.2 Reference spurs reduction

We observed that due to the mismatch in $I_{UP}$ and $I_{DN}$, there would be charge being transferred to the integrating capacitor even when the two control inputs are at same frequency and same phase. Figure 4.8 illustrates this situation. In this simulation, we set up $I_{bias1}$ and $I_{bias2}$ equal, with the initial output voltage close to the positive rail of power supply. Notice the output voltage decreases with time because $I_{DN}$ is greater than $I_{UP}$. $V_{fit_{out}}$ will continue to decrease until at a certain point when $I_{UP} = I_{DN}$.

When this is inspected in a PLL, the mismatch will cause the input reference signal and feedback signal to have a phase error. This is discussed earlier. The phase error will further give rise to reference spurs. Assume that in a frequency synthesizer, the currents are perfectly matched for a certain output frequency, it becomes clear that the further a synthesized frequency deviates from this frequency, the stronger reference spurs will be.

To alleviate this problem, we can control the programmable current sources $I_{bias2}$ and $I_{bias1}$ such that $I_{UP}$ and $I_{DN}$ are kept nearly equal when a frequency synthesizer is at different output frequencies. Figure 4.9 demonstrates this idea for the same case shown in Figure 4.8. Now the filter output stays nearly constant at a voltage closer to the positive rail of power supply. The compensation of the mismatch, however, is greatly dependent on the available resolution of the programmable current source.
Figure 4.8: Current mismatch in a charge-pump
Figure 4.9: Mismatch compensation using programmable current sources
4.4 Programmable Charge-pump for Multi-standard Applications

In the mobile communication world, there are currently still many different communication standards coexisting in different countries and different regions in the world. Some of the representative standards are GSM in Europe, IS-54, IS-95, IS-136 in North America, PHS in Japan, and so on. When a mobile phone subscriber enters from one region to another region using a different mobile standard, it is desirable that his mobile handset can support both mobile standards. Ideally, it is desirable that a mobile user will be able to use his mobile handset around the world without being aware of what communication standard he might be using. This is so called global-roaming—the key of this conception is ultimate integration of hardware, in particular, RF transceivers, in a handset to support different communication standards.

Integrating RF transceivers for multi-standard applications isn't an easy task. First of all, each standard operates in different frequency bands, has different channel spacing, adopts different modulation schemes, and etc. Integrating all the functions in a single transceiver is hardly possible. Some designs use several transceivers, but it is still desirable that these transceivers share common components as many as possible in order to reduce area and power consumption.

One of the most challenging tasks here is the design of multi-standard frequency synthesizers. To accommodate different standards operating at different carrier frequencies (900 MHz for GSM, 2.4 GHz for Wideband CDMA, and etc.), and different channel spacing (200 kHz for GSM, 5 MHz for Wideband CDMA, and etc.), a multi-standard frequency synthesizer may need to use multi VCOs, and different frequency
tuning schemes for each standard, let alone that the phase noise specifications are quite different in each standard. The integration of frequency synthesizers for multi-standard applications is very attractive and remains rather challenging.

We make some observations for the design of a multi-standard PLL frequency synthesizer, as follows

- Since the carrier frequency, bandwidth, and channel spacing are different for each standard, multi VCOs or extra frequency dividers are needed, along with the proper configuration of programmable frequency divider and prescaler.

- This means that VCO gain and division ratio in a PLL may vary over a large range.

- Because of different channel spacing, different input reference frequencies might be necessary. Also, phase noise requirements for each standard are different, and it might be necessary to vary the loop bandwidth over a large range.

- Referring to the discussions in chapter 2 and in this chapter, different values for $K_{VCO}$, $K_{LPF}$, and $N$ in different standards need to be compensated in order to maintain loop stability and good dynamic performance. The most convenient way to do so is by changing $K_{PD}$, or in another words, by changing the charge-pump current.

Based on the above observation, we consider that programmable charge-pump is the key for the development of multi-standard phase-locked loops. This conception is yet left to be proved in our further research and design practice.
CHAPTER 5

HIGH-SPEED FREQUENCY DIVIDER

We look into the design of frequency dividers in this chapter. In general, the VCO output signal has the highest frequency in a PLL. This high frequency signal is fed to a frequency divider directly, and therefore maximizing the operating speed is the primary consideration in designing the frequency divider.

In previous chapters, we treated the frequency divider in a PLL as a single block. In practice, it is composed of two parts: a prescaler and a programmable divider. A prescaler has simpler logic, and is capable of operating at higher frequency. On the contrary, a programmable divider is more complex in logic, and is much slower in speed. For this reason, the prescaler is used to convert the VCO signal to lower frequency before it goes into the programmable divider.

The prescaler in an integer-N PLL is generally constructed by modulus-2 dividers. A divided-by-two divider is simply a T-flipflop (a D-flipflop with its \(\bar{Q}\) output connected to input \(D\)). D-flipflop is also the building block for more complicated dividers. Because a single D-flipflop can operate really very fast, in most cases, the prescaler speed in an integer-N PLL isn’t a design problem.

In a fractional-N divider, the prescaler is, in most designs, a dual-modulus divider, or in some situation, a multi-modulus divider. For a dual-modulus \(N/N+1\) divider.
the division ratio is controlled to switch between modulus-\(N\) and modulus-\(N+1\). Due to the additional control logic, its operating frequency is lower than an integer divider. To cater the need of high frequency applications, special care needs to be taken in designing high-speed D-flipflops.

Integer-\(N\) PLLs are still frequently used in many applications. While the speed of an integer prescaler isn't a problem, one would expect the integer prescaler to have small division ratio such that the input reference frequency is not restricted too much by the output frequency resolution of the PLL (Chapter 2). To maximize the reference input frequency, the division ratio of the integer \(N\) divider needs to be as small as possible, but this also means that the programmable divider shall operate at much higher frequency.

In this chapter, we will focus on the above topics. In section 5.1, we discuss the design of prescalers using high-speed dynamic logic. The fast operation of a dual-modulus prescaler implemented in True-Single-Phase-Clock (TSPC) logic [19] is demonstrated using the 0.18 \(\mu m\) Copper CMOS technology. We have also implemented a high-speed programmable divider in section 5.2.

5.1 Prescaler

The implementation of gigahertz prescalers in CMOS process is a challenging task. In the past, quite a lot effort was made to boost the operating speed of dual-modulus dividers [19] [20] [21] [22] [23]. Among them, TSPC logic dual-modulus dividers were investigated in many publications [24] [25] [26] [19] [22] [23]. However, most of the reported dual-modulus dividers were fabricated in above 0.5 \(\mu m\) CMOS processes, and their maximum operating frequency is around 1.5 \(GHz\) to 1.9 \(GHz\).
For the 5.8 GHz WLAN application, we made some attempts to investigate the capability of existing design techniques. Using the 0.18 μm Copper process, we identified that a 64/65 dual-modulus divider designed in traditional static CMOS logic is capable of operating up to 4.0 GHz. As such, we turned our attention to dynamic TSPC logic. Among several different versions of TSPC D-flipflops, we adopted the circuit from [19] based on the consideration of reliability and charge-sharing issue. For a 64/65 dual-modulus divider using this TSPC D-flipflop, we achieved an operating speed of greater than 6.0 GHz by careful design and layout optimization, which is slightly better than what we expect for the 5.8 GHz application.

5.1.1 Dual-modulus prescaler

The general divide-by-\(N/N+1\) dual-modulus prescaler is based on the circuit in Figure 5.1. It consists of a divide-by-4/5 counter (in the dashed-line box), several divided-by-2 T-flipflops and a few logic gates. The \textit{MODE} pin is controlled by an accumulator or a sigma-data modulator to form the so-called fractional-\(N\) prescaler. Take an accumulator for example, given an initial starting number, the accumulator will add up upon the coming of each control clocks. When the accumulator overflows, an overflow signal is generated on the \textit{MODE} pin that switches the division ratio from \(N\) to \(N + 1\) for one clock cycle of the model control signal. By controlling the accumulator input, we basically determine the time that the divider stays in divided-by-\(N\) mode, and the time that the divider stays in divided-by-\(N + 1\) mode. The average of this operation provides a fractional division ratio.

The operation of the dual-modulus divider is described as follows:
When $MODE$ is set to '1', the output of the 5-input NOR-gate is always '0', without being influenced by the output of the TFF chain. This implies that the output of the NAND gate in the divided-by-4/5 stays at '1', and therefore the $Q'$ pin of the succeeding DFF stays at '0'. As a result, the first two DFFs are isolated from the third DFF, and they operate as a divided-by-4 counter. The divide-by-4 counter is further fed to the TFF chain of divided-by-16 divider. The total division ratio is therefore 64.

In the case that the $MODE$ pin is set to '0', as long as one of the TFF outputs is at '1', the output of the 5-input NOR-gate will be '1'. Similar to the above case, the divider operates as a divided-by 64 divider. However, if at a certain time, the outputs of the TFFs reach all '0's, the output of the 5-input NOR-gate will be '1'.
The third DFF in the dashed-line box now comes into operation, and the divided-by-4/5 counter is in the divided-by-5 configuration. This configuration will delay the transition of the Q4 for one input clock cycle (as if an input clock cycle is missing). At the next input clock cycle, Q4 takes a transition, which drives the TFFs to a non-zero state, and the divider is put in divided-by-4 configuration again. Therefore, the whole divider is equivalently in division ratio of 65 when \( MODE = 0 \).

A fractional division ratio between 64 and 65 is achieved by toggling \( MODE \) between '1' and '0' at a certain ratio.

The speed bottleneck of the prescaler is the divided-by-4/5 counter, which is the only part that operates at maximum frequency. The remaining flip-flops operate at frequency four times lower than the input frequency.

### 5.1.2 TSPC D-Flipflop

TSPC logic is used for the implementation of the D-flipflops in the divided-by-4/5. Figure 5.2 is the TSPC DFF schematic [19]. The merit of this circuit is that it allows a systematic way to optimize transistor sizes for high-speed operation. Also, charge-sharing problem, which is typical for this type of dynamic logic, is remedied by adding additional latches at node Y1 and node \( Q' \) respectively. MN5, MP2 and the two inverters serve for such purpose. For the convenience of understanding how the circuit operates, MN5, MP2 and the two inverters are ignored in the following discussion.

The circuit operation is basically divided in two stages: the precharge stage (\( \phi = 0 \)) and the evaluation stage (\( \phi = 1 \)).
Figure 5.2: Schematic of the TSPC DFF

Assume initially $D$ stays at '1'. At the precharge stage, $Y1$ stays low, and $Y2$ is precharged to high. At the rising clock edge, the circuit enters the evaluation stage. Since $Y1$ remains low, $Y3$ will be high, node $Q'$ will be discharged to low through MN3, MN4, and MNS2. The state of node $Q$ is therefore '1', which is the same as that of $D$ input.

On the other hand, if initially $D$ is '0', at the precharge stage, both $Y1$ and $Y2$ are precharged to high. At the evaluation stage, $Y1$ remains high, $Y2$ is discharged to low through MN2 and MNS1, which turns MP2 on, and causing $Q'$ to be charged to high, and $Q$ is forced to low. Note that again, $Q$ is the same as $D$.

In sum, the circuit functions as a DFF as we expect. Be cautious that in the above analysis $D$ is assumed to be fixed during the operation. If $D$ changes during the precharge or evaluation stage, the output of the circuit will change in an unexpected
way, and it will be very different from a normal DFF. Therefore, the TSPC DFF is not a fully functional DFF. By incorporating the TSPC DFF into a divider, however, will ensure that $D$ will remain constant during the transition of the input clock.

5.1.3 Trapezoidal scaling and optimization

In this design, minimum gate length is used for all the transistors. According to [27] [19], to reduce the switching delay, the cascaded $N$ transistors or $P$ transistors in the TSPC DFF should be scaled in a manner that the FET closest to AC ground is the largest, with FET size decreasing monotonically from AC ground to the output node (trapezoidal scaling) [19] [27] [25].

Following the trapezoidal scaling, design trade-offs have to be made in the sizing of each transistor. As each transistor presents load capacitance to its preceding stage, small transistor width is desired. However, each transistor drives the gates of its succeeding stage, for fast operation, large driving current, and therefore large transistor width is desirable. Fortunately, these conflicting requirements are not so overwhelming in this circuit. As analyzed in [19], the operation of the circuit puts some transistors in floating. Transistors in floating present much less load to the driving stage than if we generally consider gate parasitics as the load. Analysis shows that only the two transistors MP2 and MN1 have conflicting requirements, the width of other transistors, by first order estimation, should be either increased or reduced but not both. This greatly simplifies the optimization procedure.

During the optimization simulation, the TSPC DFF is configured as a toggle flip-flop. An inverter of proper size is added to node $Q$ to load the TFF. By careful tuning, the TFF is able to operate normally over $10\,GH\!z$ at $1.5\,V$ supply voltage (Figure 5.3).
5.1.4 Charge-sharing problem

As mentioned previously, MN5, MP3, and the two inverters are used to alleviate the charge sharing problem at node Y1 and node Q'. Refer to fig. 5.2, during the evaluation stage (\( \phi = 1 \)), if \( D \) changes from high to low, normally node Y1 should remain low until the next precharge stage. However, because node A was charged to high before MPS1 is turned off, and remains high. The charges at node A will discharge through MP1 to node Y1 when \( D \) switches from high to low. This may pull the voltage at node Y1 to a sufficiently high level and turn MN2 on. Y2 then is slowly pulled to low. The output at \( Q \) and \( Q' \) changes accordingly and give the wrong logic. The positive feedback loop formed by the inverter and MN5 functions as a latch.
which keeps node $Y_1$ voltage below a relatively low level when charge flows from $A$ to $Y_1$ in the above situation.

Under normal operation, when $\sigma$ and $D$ are both low, node $Y_1$ should be charged to high as if the latch does not exist. Therefore, the size of MN5 and the gain of the inverter should be chosen properly such that the loop only forms a week positive feedback. Otherwise, the circuit may malfunction or even oscillate. The same explanation holds for the latch at node $Q'$. 

It is worthy of mentioning that proper layout structure will decrease the capacitance at node $A$, hence less charge is stored at node $A$. This further alleviates the charge-sharing problem. 

5.1.5 Performance

After speed optimization and charge-sharing reduction, the TSPC DFF is incorporated into the divided-by-4/5 counter as well as the TFF chain in the prescaler. No further relative scaling is done to the three DFFs and the TFFs although it might provide better power consumption and speed performance.

Figure 5.4 illustrates the waveforms of the TSPC DFF operation within a divided-by-4/5 divider. Notice the voltage level variation at node $a_1$, $a_2$, and $Q'$. This is partly due to the charging and discharging behavior, and partly due to charge-sharing. Note that charge-sharing is greatly reduced by the latches.

The waveforms of the final dual-modulus divider are shown in Figure 5.5. When $Mode = 0$, a pulse at node $ModeOut$ will be generated per 65 input clock cycles, and the pulse is 5 clock cycles wide. At mark A in Figure 5.5, an extra input clock cycle is incurred to the output waveform of the divided-by-4/5 divider.
The power consumption obtained in this design is around 3.3 mW at maximum operating frequency and 1.5 V power supply. Considering the down-scaled power supply and increased operating frequency, this value seems reasonable compared with the reported power consumption in other designs [22]. The power consumption in those designs range from 25 to 50 mW when operating at gigahertz frequency and at 3 V or 5 V power supply.
Figure 5.5: Waveforms of the prescaler with $f_{\text{clk}}$ at maximum frequency

5.2 Programmable Divider

The block diagram of a programmable divider is shown in Figure 5.6, where the D-flipflops are configured to be divided-by-2 counters and their outputs are fed to the end-of-count (EOC) detector. In normal operation, the counters are initially loaded with a number (the programmable division ratio $N$). Triggered by the input clock $F_{\text{in}}$, the counter counts down by 1 at each clock cycle. When the outputs of the D-Flipflops all reach zero, the EOC detector is triggered. The counter stage is then
re-initialized (reload) to the state \( V \). The counter will start another cycle of counting. The frequency of the reload signal is then the input clock frequency divided by \( V \).

Figure 3.6: Diagram of a programmable divider

The speed of this programmable divider is restricted by the total delay resulting from the counter, the EOC detector and the reload circuit.

The propagation delay of the counter is determined by the first DFF in this asynchronous counter structure. A regular DFF circuit can be used for this DFF instead of a dynamic type. This is mainly because a simple implementation of a set and reset DFF based on the TSPC DFF is not easy. In addition, the delay incurred by a regular DFF here is not so overwhelming since there is no big capacitance load interacting with it.

The delay incurred by the EOC detector does not seem to be a major problem either. For the down-counting counter, when an all zero condition happens, all the
high bits have already settled down to zero in previous clock cycles, so the EOC delay is purely decided by the lowest bit.

The reload circuit delay needs some special consideration, since the reload signal is driving n-bit data to the counter simultaneously. The high fan-out may incur a big delay for the reloading operation. A buffer stage is needed to reduce this delay. [20] proposed a method that produces the reload signal one clock cycle in advance. The method utilizes dynamic logic circuits to pre-store the reload signal for each bit, and allows the reload signal to be buffered per two bits. By such an arrangement, the circuit achieved a considerably high operating frequency.

Another simpler scheme is proposed in [28] to boost the throughput of the programmable divider. In this scheme, the EOC detector is designed to expand the reloading operation over three input clock cycles. Although the total delay for the counter, EOC detection and reload circuit may be long, they are now distributed over three clock cycles. Therefore, the circuit can operate much faster than the conventional detection method.

We adopted the second scheme. For convenience, Figure 5.7 illustrated the EOC detector that is designed to detect the state 000010 instead of 000000.

![Figure 5.7: End-of-count detector](image-url)
Initially, the *Reload* signal is low. When the counter state changes from 000011 to 000010, the output of the EOC detector ($D_0$) becomes high. A subsequent rising edge of the input clock triggers the output $Q'_0$ to 0, which causes the *reload* signal to reach high. The *Reload* signal resets the counter to $N$. At this clock cycle (000010), $D_0$ becomes low again. So at the next clock rising edge, $Q'_0$ will return to high and $Q_0$ is deactivated. The delay is basically distributed into 3 cycles instead of one.

Finally, Figure 5.8 demonstrates a simple reload circuit for one single digit.

![Reload circuit](image)

**Figure 5.8: Reload circuit**

Without going through a full-custom design process, we implemented the programmable divider using cells from a standard digital cell library. Timing simulation shows that the achievable maximum operating frequency is up to 2.0 GHz. Better performance can be achieved by a full-custom design, with careful design optimization and layout.
CHAPTER 6

ON-CHIP INTEGRATED INDUCTOR AND VCO

6.1 introduction

In the search for fully integrated frequency synthesizers, the design of low phase-noise VCO and on-chip high-Q inductors are the most challenging tasks and have been receiving much attention during the past several years. LC oscillators with on-chip inductors were reported in many recent papers \[29\] \[30\] \[31\]. Some detailed investigations were made on on-chip inductors \[32\] \[33\] \[34\] \[35\] \[36\] \[37\] \[38\]. However, due to process limitations, on-chip spiral inductors formed in conventional CMOS technologies generally have small Q-factors of around 3-6, and hardly achieve a Q-factor of better than 10, which is far lower than that can be attained from off-chip inductors.

A lumped \(\pi\) model for an on-chip inductor is shown in Figure 6.1, in which the series resistor \(R_s\) stands for the resistive loss of the metal conductor, the capacitor \(C_{ox}\) stands for the parasitic capacitance between the conductor layer and the substrate, and the resistance \(R_{sub}\) stands for the substrate loss.

The parasitic resistances, especially the series resistance \(R_s\), need to be minimized whenever possible. This is because they are directly related to the phase noise of LC
oscillators, as derived in [31]:

$$dV_{out}^2 \{\Delta \omega \} = kT \cdot R_{eff} \cdot \left[1 + A \right] \cdot \left( \frac{\omega_0}{\Delta \omega} \right)^2$$  \hspace{1cm} (6.1)

where $dV_{out}^2 \{\Delta \omega \}$ is the oscillator noise density at the frequency $\omega_0 + \Delta \omega$. $R_{eff}$ is the effective resistance which is approximately equal to the total series resistance in a LC tank. $A$ is a coefficient related to the noise from the active devices in an oscillator.

The Q-factor of the inductor is inversely proportional to $R_{eff}$ as given by

$$Q = \frac{\omega_0 L}{R_{eff}}$$  \hspace{1cm} (6.2)

As illustrated by the simple $\pi$ model, in order to design high-Q on-chip inductors, one needs to reduce the series resistance $R_s$ and $R_{sub}$.

The reduction of $R_s$ can be achieved by the use of thicker metal layers (1-3$\mu$m thick in some typical RF CMOS processes), the use of higher conductivity metal
layers (Cu versus Al), or the use of multi-metal layers that increases the effective thickness of inductors.

The reduction of $R_{\text{sub}}$, i.e., the reduction of substrate loss, is similarly important. Many CMOS processes use a highly doped substrate that has low resistivity. Substrate resistivity in some typical CMOS processes ranges from 10 to 20 $\Omega \cdot \text{cm}$ or less. At low frequency, low substrate resistivity does not seem to cause problem. However, when the frequency in a spiral inductor is increased over gigahertz, the magnetic field generated by the oscillating current in the inductor will induce eddy currents in the substrate, as such energy will be dissipated in the substrate. In addition, the magnetic field formed by eddy currents tends to cancel the magnetic field generated formed by the current in the inductor, so the total inductance of the spiral inductor will be reduced.

As a measure of reducing substrate loss, some people used etching techniques to remove the substrate under spiral inductors [39]. Recently, MEMS technology is being utilized to create suspended inductors that can achieve Q-factor of over 40. Nevertheless, the extra etching step required by these methods raises fabrication cost, in addition to possible mechanical failure and yield reduction. As a result, they are not widely accepted by the industry yet. Substrate loss reduction can also be achieved by the use of thicker oxide, by the use of hollow spiral inductors (The inner sections of a spiral inductor contribute less inductance because they span a small area. On the other hand, the skin effect of the inner sections is significant due to the stronger magnetic field at the center of the spiral, which results in larger resistance even though they have smaller length.), or straightforwardly, the use of SOI (silicon-on-insulator) technology.
6.2 Design of High-Q Inductors

6.2.1 On-chip spiral inductors in Copper CMOS process

Although we previously made an argument that the reduction of parasitic resistance can be achieved by the increase of metal thickness, this benefit might be countered by the increase in parasitic capacitance. The availability of Copper CMOS technology allows us to obtain high-Q inductors from another perspective due to the fact that the sheet resistance of Cu process is about 30\% lower than Al process, and via resistance of Cu process is 10x lower than Al process.

In the CMOS technology we are using, there are totally 6 layers of metal for interconnect, all of them are copper. The top two copper layers are 1.0 \( \mu m \) thick, M3 and M4 are 0.45 \( \mu m \) thick, and M1 and M2 are 0.3 \( \mu m \) thick. Inter-metal oxide layers range between 0.5 and 0.8 \( \mu m \), with an oxide dielectric constant of 4.1. The typical value of the substrate resistivity is 20 \( \Omega - cm \). These process parameters basically determine the performance of on-chip inductors.

Figure 6.2 shows the layout of a 2.5-turn octagonal inductor, with a Ground-Signal-Ground (G-S-G) testing pattern. The open testing pattern is used to cancel out the influence of the probing pads (de-embedding). We will discuss de-embedding procedures in more detail later.

Based on the reports from other literature dealing with on-chip inductors, the inductance and Q-factor of a on-chip spiral inductor are also related to the geometry shape, number of turns, conductor width, and conductors spacing of the inductor. In general, inductors with circular geometry provide higher Q-factor than octagonal or rectangular inductors. A thorough optimization of spiral inductors in terms of all the geometry parameters seems to be prohibiting, since the simulation of an inductor
in an electro-magnetic simulator requires long simulation time and large memory. To facilitate the simulation process, we have developed a program to automate the inductor simulation. Firstly, the layout of a spiral inductors is generated by a program that can take arbitrary shape and geometry as input parameters. The layout data is then given to an electro-magnetic solver to obtain the two-port scattering parameters (S parameters). The Q-factor and the inductance can then be extracted by first transforming the S-parameter to Y-parameter (normalized), as given by the following expressions:

\[
Y_{11} = \frac{(1 - S_{11})(1 + S_{22}) + S_{12}S_{21}}{\Delta} \\
Y_{21} = \frac{-2S_{21}}{\Delta} \\
Y_{12} = \frac{-2S_{12}}{\Delta}
\]

(6.3)  (6.4)  (6.5)
\[
Y_{22} = \frac{(1 + S_{11})(1 - S_{22}) + S_{12}S_{21}}{\Delta}
\]  

(6.6)

where \(\Delta = (1 + S_{11})(1 + S_{22}) - S_{12}S_{21}\).

Q-factor is then given by,

\[
Q = \frac{Im(Y_{12})}{Re(Y_{11})}
\]  

(6.7)

In this expression we have assumed that one port of the inductor is connected to ground (this is the case valid for most RF applications). The Q-factor is referred to unloaded-Q. In practice, the inductor is often in parallel or in series with an external resistive load, which will cause a lower effective Q-factor.

\(L\) is given by,

\[
L = \frac{Im(Y_{12})}{2\pi f \cdot [Re^2(Y_{12}) + Im^2(Y_{12})]}
\]  

(6.8)

6.2.2 Parasitics and layout consideration

To minimize series resistance, we have M5 and M6 connected in parallel. Figure 6.3 illustrates the cross section of the M6 and M5 conductors along with the interconnect capacitance. In this figure, \(C_a\) represents the parallel plate area capacitance. \(C_f\) represents the fringe capacitance to the substrate, and \(C_c\) represents the coupling capacitance between adjacent conductors. These parasitic capacitances need to be minimized as well to achieve high resonant frequency.

Both metal layers are 1 \(\mu m\) thick, and they are connected everywhere through M65 vias. The distance between M5 and the substrate is 5.4 \(\mu m\), therefore the parasitic capacitance between the conductor layer and the substrate is relatively small (0.006 \(fF/\mu m^2\)) in comparison with the inter-layer capacitance (0.06 \(fF/\mu m^2\)) of
two neighboring metal layers. Because M6 and M5 are connected by vias, the interlayer capacitance of these two layers is eliminated. However, because the equivalent thickness of the conductor is big, the coupling capacitance ($C_c$) will dominate the parasitic capacitance if the spacing between two adjacent conductors is fairly narrow.

Figure 6.4 demonstrates the layout of a circular inductor, which has M6 and M5 connected in parallel, with the inner port connected to the probing pad by a M4 under-pass. The inductor is surrounded by a p-type guard-ring tied to ground to shield the inductor from other neighboring devices. To avoid any additional doping, the spiral inductor area should be blocked from the formation of P well or N well during fabrication, i.e., underneath the inductor, there is only the raw P-substrate layer.
6.2.3 Inductor de-embedding and measurement results

De-Embedding model representation

The inductor performance can be obtained by measuring the S-parameters of the two-port device shown in Figure 6.4. Note that the large dimension of the probing pads results in large parasitic capacitance, which must be subtracted from the S-parameters. In addition, the contact resistance between the probe and the probing
pads, and interconnect wiring resistance may also need to be removed before the final Q-factor and inductance are calculated.

A physical representation of the lumped parasitics resulting from probing pads and interconnect is shown in Figure 6.5. $G_p$ represents the shunting parasitics resulting from the probing pads and interconnect wires. $Z_r$ represents the contact resistance between probing pads and probes. After a standard substrate calibration the reference plane for S-parameter measurement is extended to the probe tips [40], so ideally, $Z_r$ should be zero. In practice, because the substrate material for calibration and the probing pads are different, a contact resistance of up to several hundred $\mu$Ω exists, which will cause measurement inaccuracy. $Z_w$ represents the wiring parasitics between the probing pads to the actual inductor ports.

**De-embedding procedure**

Assume the raw inductor S-parameter is represented by $S_{ij,\text{raw}}$. To de-embed all the parasitics ($Z_r$, $Z_w$, and $G_p$), some additional measurements need to be performed. The procedures are given as follows.
Do a *short* measurement by shorting $P_1$ and $P_2$ to the ground. The obtained S-parameter is converted to Z-parameter, and this will extract $Z_{c1}, Z_{c2}$ according to.

$$Z_{c1} = Z_{11, \text{short} P}$$

(6.9)

$$Z_{c2} = Z_{22, \text{short} P}$$

(6.10)

By subtracting out $Z_{c1}$ and $Z_{c2}$ we can obtain the de-embedded Z-parameter at the $P_1, P_2$ reference plane.

$$Z_{11, P} = Z_{11, \text{raw}} - Z_{c1}$$

(6.11)

$$Z_{21, P} = Z_{21, \text{raw}}$$

(6.12)

$$Z_{22, P} = Z_{22, \text{raw}} - Z_{c2}$$

(6.13)

$$Z_{12, P} = Z_{12, \text{raw}}$$

(6.14)

Next we do an *open* measurement by leaving $D_1$ and $D_2$ open circuit. $G_{p1}$ and $G_{p2}$ are extracted according to.

$$G_{p1} = Z_{11, \text{open} D} - Z_{c1}$$

(6.15)

$$G_{p2} = Z_{22, \text{open} D} - Z_{c2}$$

(6.16)

$G_{p1}$ and $G_{p2}$ can now be de-embedded from the $P$ plane using Y-parameter.

$$Y_{11, P'} = Y_{11, P} - G_{p1}$$

(6.17)

$$Y_{21, P'} = Y_{21, P}$$

(6.18)

$$Y_{22, P'} = Y_{22, P} - G_{p2}$$

(6.19)

$$Y_{12, P'} = Y_{12, P}$$

(6.20)

Finally, we are going to de-embed $Z_{w1}$ and $Z_{w2}$. We do another *short* measurement by shorting $D_1$ and $D_2$ to ground. $Z_{w1}$ and $Z_{w2}$ is extracted according to the following
expressions.

\[ Z_{u1} = \frac{1}{Z_{11,\text{shortD}} - Z_{c1}} - G_{p1} \]  
(6.21)

\[ Z_{u2} = \frac{1}{Z_{22,\text{shortD}} - Z_{c2}} - G_{p2} \]  
(6.22)

\[ \text{The final results of the de-embedded inductor are given by the following equations.} \]

\[ Z_{11} = Z_{11,p'} - Z_{u1} \]  
(6.24)

\[ Z_{21} = Z_{21,p'} \]  
(6.25)

\[ Z_{22} = Z_{22,p'} - Z_{u2} \]  
(6.26)

\[ Z_{12} = Z_{12,p'} \]  
(6.27)

Finally, the \( Z \) parameters are converted to \( Y \) parameters or \( S \) parameters to obtain the Q-factor and inductance value.

Note that more often only \( G_{p1} \) and \( G_{p2} \) are de-embedded through an open measurement. This is because \( G_{p1} \) and \( G_{p2} \) are more significant due to the large dimension of testing pads, while \( Z_{c1,c2} \) and \( Z_{u1,u2} \) have much less influence on the measurement accuracy.

**Measurement results**

S-parameter measurements were performed using a HP8720 network analyzer. First, the network analyzer was calibrated following a standard calibration procedure. The probing pad parasitics were then de-embedded according to the above procedures (depending on whether the contact resistance and interconnect parasitics are significantly large, the two short de-embedding steps may be skipped).
The measured two-port S parameters for a 1 \( nH \) inductor are shown in Figure 6.6. In this figure we demonstrate both the measurement results before de-embedding and after de-embedding. The differences of these two sets of S-parameters are mainly due to the parasitic capacitance of the probing pads.

![S-parameters](image)

Figure 6.6: Measured S-parameter (dotted-line: raw data; solid-line: de-embedded data)

The final S-parameters are used to extract the Q-factor, inductance as well as other parameters in the lumped inductor model. Figure 6.7 and Figure 6.8 show the plot for the measured inductance and Q-factor. The simulated \( L \) and \( Q \) within the
frequency range of our interest are also plotted in the figures. The results between the simulation and measurement are rather consistent, with the measured Q slightly below the simulated value. In this design, the inductance is 1 nH, and it stays considerably constant up to 7.0 GHz and then ramps up when the frequency approaches the self-resonant frequency of the inductor. The peak Q-factor is greater than 10 at a frequency range of 1.0–6.0 GHz. It can be seen that the Q-factor initially increases with frequency until it reaches the peak value, and then at high frequencies it decreases due to large substrate loss resulting from eddy current.

![Figure 6.7: Inductance versus frequency](image)

Figure 6.7: Inductance versus frequency
6.3 LC Oscillator

6.3.1 Brief on oscillator phase noise

The description of LC oscillator phase noise has been empirically represented by Leeson's equation:

\[ \mathcal{L} (\Delta \omega) = 10 \log \left\{ \frac{2 F k T}{P_{\text{carrier}}} \left[ 1 + \left( \frac{\omega}{2 Q \Delta \omega} \right)^2 \right] \left( 1 + \frac{\Delta \omega}{f_0^2} \right) \right\} \]  \hspace{1cm} (6.28)

In which \( F \) is a fitting factor that needs to be obtained through experiments.

This equation provides a fitting curve for the three regions in the typical phase-noise characteristics of an oscillator, as shown in Figure 6.9. The noise in the \( \Delta \omega^{-2} \) region is the primary concern of oscillator design. Noise in this region results from
thermal noise sources in the oscillator. Thermal noise expands over a large frequency range, it contributes to the out-of-band noise in a PLL significantly, and therefore becomes the design focus, especially when one is choosing the PLL loop bandwidth. The $\Delta \omega^{-3}$ noise appears at very low frequency offset, and is caused by $1/f$ noise. $1/f$ noise generally falls in the loop bandwidth, and can be greatly attenuated. Finally, the flat section at high frequencies is usually seen in measurement, and it can be attributed to the noise floor of the circuit, or measurement instrument, and is not a design concern.

6.3.2 Circuit description and design consideration

The design of low phase-noise integrated VCO depends heavily on the quality factor $Q$ of on-chip spiral inductors as shown by the above equation.
Figure 6.10 illustrates the simplified VCO schematics, which is composed of a LC oscillator core, a double-end to single-end gain stage which serves also as a clock slicer, and finally an inverter buffer stage.

In this design, the oscillator used two 1 nH spiral inductors. For the 5.8 GHz wireless LAN application, the frequency bandwidth extends between 5.725 and 5.850 GHz. Therefore the total capacitance for each branch (ckp, ckn) of the oscillator needs to change from 0.740 pF to 0.773 pF, of which approximately 0.18 pF is from the spiral inductor parasitic capacitance, and other capacitance is allocated to the MOS transistors and varactors. The parasitic capacitance of the active devices is composed of the drain-bulk, gate-source, gate-drain capacitance of the cross-coupled transistors M1 and M2, the gate-drain, gate source capacitance of differential transistor pair M3 and M4, as well as the pn-junction capacitance of the varactors.
The phase noise of the oscillator can be predicted according to [29]:

\[
\mathcal{L}(\Delta \omega) = 10 \log \left[ \frac{8kT \omega_c}{P^3 Q^3 L \Delta \omega^2} \left( \frac{LG + 1.5}{3} \right) \right] \frac{dBc}{Hz}
\]

(6.29)

where \( k \) is Boltzmann's constant \((k = 1.38 \times 10^{-23} \text{ J/K})\), \( T \) is absolute temperature \((290 \text{ K})\), \( I_b \) is the oscillator bias current, \( \omega_c \) is the oscillator center frequency, \( \Delta \omega \) is the frequency offset from \( \omega_c \), and \( LG \) is the loop gain of the oscillator, which is expressed as \( g_m(1 + Q^2)R \).

With \( L = 1nH \), \( Q = 10 \), \( I_b = 5mA \), and \( LG = 2 \), the predicted phase noise will be around \( 106 \frac{dBc}{Hz} \) at a \( 600 \frac{kHz}{Hz} \) offset from the \( 3.8 \frac{GHz}{Hz} \) carrier frequency. Further reduction of phase noise can be achieved by the increase of \( Q \), \( I_b \), or \( L \). The freedom to increase \( Q \) has already been restricted by the specific CMOS process. Increasing \( I_b \) will increase power consumption, and may not be allowed in most low-power applications.

Increasing \( L \) implies that parasitic capacitance has to be reduced, which may not be feasible. For \( LG = 2 \), the required \( g_m \) for transistors \( M1 \) and \( M2 \) is around \( 6 \text{ mS} \). This yields a width of \( 30 \mu m \) for transistors \( M1 \) and \( M2 \). In the CMOS process used, the gate oxide capacitance is \( 10 \text{ fF} / \mu m^2 \). The total gate-oxide capacitance is therefore \( 0.054 \text{ pF} \). If we take all the other gate-drain, drain-bulk parasitic capacitance into consideration, the total parasitic capacitance resulting from \( M1-2 \) at each branch of the LC tank is below \( 0.1 \text{ pF} \). Putting together the parasitics from the inductor, parasitics from \( M1-2 \), and parasitics from \( M3-4 \) (much smaller that that of \( M1-2 \)), the parasitic capacitance is around \( 0.3-0.4 \text{ pF} \). and the remaining parasitic capacitance of \( 0.4 \text{ pF} \) can be allocated to each of the varactors, of which a tuning range of around \( 0.05 \text{ pF} \) is needed. This tells that we still have some space to increase either the size of \( M1-2 \) to obtain a larger loop-gain such that there is a larger margin for the start-up.
of the oscillator. Or we can use a slightly larger inductance such that phase noise can be reduced. However an increase in inductance will leave less capacitance allocation to the varactors, resulting in poor tuning performance.

6.3.3 Layout and simulation results

![VCO layout](image)

Figure 6.11: VCO layout

The layout of an oscillator for testing is shown in Figure 6.11. In the middle of the layout is the oscillator core, which is isolated by guard-rings from other part...
of the circuits. The whole VCO is also surrounded by a guard-ring to reduce the interference between the VCO and other circuits in the same die. The wires for ground and power supply are made fairly wide (30μm). The whole die area of the VCO is around 550 × 550μm², of which the two inductors take up most of the area. The tunable varactors are situated in the upper middle, which also takes up a rather large die area. The biasing circuitry and some buffer stages are placed in the lower middle of the die. Some probing pads are also placed in the die, and they are properly buffered.

![Oscillation start-up](image)

Figure 6.12: Oscillation start-up
The start-up process of the oscillator is shown in Figure 6.12. With a 1.5 V power supply, the peak-to-peak value of the oscillator output waveform is close to 3 V.

![Figure 6.13: Phase noise](image)

The simulated phase-noise performance is illustrated in Figure 6.13. At 600 kHz offset, the phase noise is 104.2 dBc/Hz, fairly close to the previous predicted value.
CHAPTER 7

IMPLEMENTATION OF A 5.8 GHz PLL

To demonstrate some of the ideals discussed in previous chapters, we have built a PLL for the 5.8 GHz wireless LAN application using the building blocks presented in this work. Figure 7.1 shows a testing chip. The chip is fabricated in a 0.18 μm 1-poly 6-metal 1.5/3.3 dual-voltage process twin-well Cu CMOS process. The chip area is 6 x 7 mm². In the upper-right corner of the chip is a testing PLL, which takes up only a very small fraction of the die area. In this chip we have enclosed some testing structures for spiral inductors and several of the key building blocks. In this chapter, we will give a discussion of the practical design issues, with some particular focus on fractional-N control. A simple but practical graphic method is demonstrated which can be used to obtain an optimal division ratio for the dual-modulus prescaler.

7.1 Some Practical Design Issues

7.1.1 Frequency-hopping

Because of the high operating frequency, 5.8 GHz WLAN transceivers were often implemented in GaAs or SiGe technology. In CMOS process, one of the biggest barriers to implement a PLL in the 5.8 GHz frequency band is the operating speed of frequency dividers. In this design, this is overcome by the use of the state-of-the-art
Figure 7.1: Testing chip
0.18 \mu m \text{ copper CMOS technology, and by the use of carefully designed high-speed}
dynamic logic TSPC circuits.

The PLL is tentatively designed for frequency-hopping systems in the 5725-5850
MHz ISM band. Under FCC regulation, a minimum 75 MHz of bandwidth should
be covered within the 5725-5850 MHz band [41]. Assume 1 MHz bandwidth for
each non-overlapping hopping channel, a minimum of 75 hopping channels would be
required. To synthesize the hopping frequencies, one needs to set up the reference
frequency and the division ratio with care, which is not so straightforward as it may
suggest. For instance, with a predetermined reference input of 20 MHz, one can
derive a division ratio of 287-291 to obtain an output frequency range of 5740-5820
MHz, which covers an 80 MHz bandwidth. To have all the 75 hopping frequencies
generated, fractional-X control is required. To facilitate the division ratio distribution
between the programmable divider and prescaler, a 4/5 dual-modulus prescalar is
used. The programmable divider ratio then takes integer values between 58-72.

The parameters in the above method are obtained through try-and-error. A better
way to derive the division ratio for the programmable M divider and the fractional-X
divider respectively, as well as the input reference frequency, is described below.

Firstly, one predetermines the output frequency range \(Y_1, Y_2\) and the dual mod­
ulus division ratio \(N (N = 2^b)\). It is obvious that there are only a few feasible values
for \(N\).

Next, for arbitrary \(Y_1, Y_2\), one can have the following expressions:

\[Y_1 = f_{ref} \times M_1 \times (N + \alpha_1)\]  
(7.1)

\[Y_2 = f_{ref} \times M_2 \times (N + \alpha_2)\]  
(7.2)

where \(\alpha_i \in [0, 1]\).
Let $y = f_{ref} \times M$, one can obtain:

\begin{align*}
    y &= Y_1/(N + \alpha) \\ 
    y &= Y_2/(N + \alpha)
\end{align*}

(7.3)  
(7.4)

These two functions specify two bounding curves (Figure 7.2). It follows that each hopping frequency will correspond to a curve which lies between the two bounding curves, and is in parallel to them. The curves corresponding to all the hopping frequencies are supposed to distribute between the two bounding curves evenly. On each curve, one can find an infinite set of $\alpha$ and $y$ ($y = f_{ref} \times M$) to generate the corresponding hopping frequency. However, since $M$ can only adopt an integer value.
and \( f_{ref} \) has to be large enough and be unique for all the hopping frequencies. The solution space for \((f_{ref}, M, \alpha)\) is very well constrained and can be easily obtained on the graph as demonstrated below.

![Graph showing solution space for \((f_{ref}, M, \alpha)\) with marked channels not covered.](image)

**Figure 7.3: N=128**

To illustrate the idea, we turn to a real problem. In Figure 7.3, we have chosen \( Y_1 = 5750 \text{MHz} \), \( Y_2 = 5825 \text{MHz} \), and \( N = 128 \). Notice that in this specific case, if we choose \( f_{ref} \) to be greater than 1 MHz, no matter what integer value \( M \) may take, only curves between the two dashed-lines will intersect with the horizontal line \((y=f_{ref} \times M)\). It will be impossible to generate the hopping frequencies corresponding to curves beyond the two dashed-lines. Of course, if small loop bandwidth is not
a issue, one can choose a smaller $f_{ref}$ (e.g., 0.1 MHz) such that whole 75 MHz bandwidth is covered with the use of several $M$ values.

By choosing $N = 64$, we obtain the bounding curves in Figure 7.4. This time, any horizontal lines between the two dashed-lines will be able to cross over all the hopping curves. It is obvious that at $y = 89.8$ (e.g., one can choose $f_{ref} = 22.45 MHz$ and $M = 4$), and a sweep of $\alpha$ from 0 to 1 will cover the desired bandwidth. It becomes clear that with smaller $N$ (32, 16, ...), one can also find the corresponding parameter set.

We are interested in knowing what $N$ value is better. This has something to do with the fractional-N controller. For a simple fractional-N controller using an

Figure 7.4: $N=64$
accumulator. the resolution of the hopping frequencies is inversely proportional to the bit-length of the accumulator. Assume the bit-length of the accumulator is $B$. the output frequency resolution is given by

$$\Delta f = f_{ref} \times M \times \frac{1}{2^B} \approx \frac{f_{out}}{N}$$  \hspace{1cm} (7.5)$$

With $N = 4$ and $f_{\text{ref}} = 5.8\text{GHz}$, an 18-bit accumulator will be needed in order to generate less than 1% frequency deviation (for 1 MHz channel spacing). This indicates that the accumulator will consume large die area and power. By increasing $N$, the accumulator bit-length required for the same accuracy becomes smaller. It should become obvious that the save in die area and power consumption is more obvious when $\Delta\Sigma$ modulator is used as the controller, due to the complexity of $\Delta\Sigma$ modulators.

According to the above analysis, $f_{\text{ref}}$ is chosen to be 22.45 MHz, along with $N = 64$, and $M = 4$ respectively. All the 75 channels are covered by changing $\alpha$, i.e., by properly controlling the dual-modulus divider.

### 7.1.2 Loop parameters

A nominal charge-pump current of 1 mA is chosen, and the VCO gain is designed to be $85 \text{MHz/V}$ to cover the 75 MHz band with only 1 V voltage swing at the charge-pump output. The loop bandwidth is chosen to be 2.3 MHz. With a phase margin of 55° degree, the loop filter parameters of a 3rd-order PLL are calculated according to a nominal division ratio of 257.5, which gives $C1 = 0.50pF$, $C2 = 4.5pF$, and $R2 = 48.6k\Omega$.

Once the loop filter parameters are settled, one may worry that during frequency-hopping the loop dynamics would be influenced by the variation of the division ratio.
In this application, this isn't a problem, as the change of the division ratio over the whole 75 MHz bandwidth is less than 1%.

7.1.3 Settling time

Notice that the settling time of the system is greatly reduced due to the use of large loop-bandwidth. In specifying the requirement for the settling time, one needs to take the frequency-hopping rate of the synthesizer into consideration. In addition, the fractional-X controlling process will slow down the settling behavior.

With regard to frequency-hopping rate, first of all, the transition time (settling time) of frequency-hopping will influence data transmission efficiency. The faster the hopping rate is, the more time will be spent on the non-transmitting hopping phase, then the lower the transmission efficiency will be. In general, a fast enough hopping rate is maintained to reduce the odds when a channel is instantaneously jammed by interference. On the other hand, faster hopping rate will also generate more interference to neighboring wireless network operating in the same frequency band. For these reasons, in current frequency-hopping systems the maximum hopping rate is often chosen according to specific engineering situations, and typical hoping rates are around or much less than 1000 hops/sec. In this scenario, the time budget left for settling does not seems to be tight, since in general the settling process of a PLL can be finished in a few reference signal cycles. With input reference frequency of higher than several MHz, the settling process takes only a few micro seconds. The use of fractional-X control to achieve frequency-hopping may underscore the statement made above. Refer to Figure 7.5. a brief observation is made to the reference frequency signal and the output frequency signal (Figure 7.6). Note that the output frequency

135
\( f_s \) is an average effect of \( N \cdot f_{ref} \) and \( (N + 1) \cdot f_{ref} \) within a period of \( T_N + T_{N-1} \).

When the output frequency hops from \( f_1 \) to \( f_2 \), the settling of the frequency will then need to expand over the \( T_N + T_{N-1} \) time period. In the worst case, \( T_N + T_{N-1} = 2^B \), where \( B \) is the bit-length of the accumulator. With \( B = 12 \), and \( f_{ref} = 22.45 \text{MHz} \), the worst case settling takes a significant amount of time of around 180 \( \mu \text{s} \). To reduce settling time, it is then necessary to make use of the programmable charge-pump to switch the PLL to larger loop bandwidth temporarily during the transition.

### 7.1.4 Reference spurs

While the mismatch in charge-pump currents can be partially compensated by a programmable charge-pump, the resolution of the programmable current sources determines how well the mismatch can be cancelled. In addition to current mismatch, the fractional-\( N \) controlling method also causes reference spurs. This type of spurs is
referred as "fractional spurs". The mechanism of how fractional spurs are generated is illustrated in Figure 7.7. When the divider is in divided-by-N mode, the instantaneous $f_{div}$ frequency is greater than $f_{ref}$. The phase difference between $f_{div}$ and
$f_{ref}$ will become incrementally larger, which is reflected by the negative $I_{CP}$ pulse. As a result of the negative current pulses generated by the charge-pump, the VCO output frequency will approach $N \cdot f_{ref}$ (assume no programmable divider). Then, at the instant when the dual-modulus divider is put in the divided-by-$N+1$ mode, its instantaneous output frequency ($f_{dv}$) will be smaller than $f_{ref}$, and a positive $I_{CP}$ pulse will be generated. The pulses in the charge-pump output current $I_{CP}$ cause voltage ripples to the VCO control input. The ripples cause spurs in the vicinity of the VCO center frequency. The distances between the spurs and the center frequency are $n \cdot \frac{x}{2^{i}} \cdot f_{ref}$, where $n = \pm 1, \pm 2, ...$ and $x$ is the accumulator control word.

$\Delta \Sigma$ modulators can be used as the fractional-$N$ controller. Fractional spurs will be reduced by the randomization and noise shaping capability of $\Delta \Sigma$ modulator [42].

7.1.5 Process variation related issues

It is important to observe that each block in the PLL and the whole loop are sensitive to process variation. The charge pump, for example, when subject to device mismatch or process corner variation, will make the pumping current mismatch even worse. Process variation can cause VCO frequency to deviate from its design value, and therefore some trimming techniques are needed to ensure that a VCO operates in its normal operation region.

Consider another case at the system level: At the instant when the power is turned on, the VCO output frequency is driven to its upper limit. In normal conditions, the VCO may never operate at this upper limit, or at least the maximum frequency is well affordable by the dividers. It is possible that this output frequency is so high that it exceeds the maximum operating frequencies of the dividers. In this situation,
the dividers will be overdriven by the VCO. The consequence is that there is no AC signals coming out from the dividers, and the feedback signal will be put in saturation to either of the power supply rails. Meanwhile, the phase detector detects that the feedback signal frequency is smaller than the reference frequency. accordingly, it will attempt to drive the VCO to an even higher frequency, which overdrives the divider even harder, and the PLL is just clamped and dead. In this scenario, there is no implication that the whole PLL is abnormal if the VCO starts from any valid output frequencies. A solution to this type of problem is to reset the VCO output frequency to its nominal center frequency at start-up.
CHAPTER 8

CONCLUSIONS AND FUTURE WORK

In this work, we have investigated the design issues and trade-offs systematically of implementing fully integrated phase-locked loops in standard CMOS technologies. Low-power, low voltage and high-frequency operation ICs are pursued in our design in accordance with the ever-growing demands in low-power low-cost consumer RF electronics.

CMOS technologies are continuing to downscale toward smaller feature sizes. However, the benefits of downscaling in traditional aluminum CMOS technologies are countered by the larger parasitic sheet resistance of thinner wires. New materials with lower resistivity are needed to replace aluminum interconnect in the mature Al CMOS technologies. This is made possible by the recent break-through in Copper CMOS technologies. It is anticipated that Copper CMOS will soon become the mainstream in the market. In this research, we used a 0.18 µm Copper CMOS technology with 1 poly and 6 metal layers. The merits of Copper interconnect are explored in the design of high-performance integrated PLLs.

The most important design considerations in PLLs are phase-noise performance and dynamic behavior, which are directly related to loop bandwidth. Some trade-offs are necessary in choosing the loop bandwidth to achieve good performance. First of
all. For noise suppression, the loop bandwidth needs to be chosen properly such that in-band noise and out-of-band noise are suppressed optimally. A large loop bandwidth facilitates fast settling behavior, but at the same time, will introduce higher level of reference spurs to the output spectrum. To reduce reference spurs caused by charge-pump current mismatch, programmable charge-pump can be used to compensate the mismatch. If properly designed, a programmable charge-pump can also be used to speed up the settling time.

Two other efforts are made to achieve low phase-noise. One is the design of high-Q inductors. In this effort, on-chip spiral inductors with Q-factor greater than 10 are obtained and applied to the design of a LC-VCO. Another effort we have made in this work is to minimize the dead-zone and blind-zone often existing in phase-frequency detectors. The phase noise of a free running VCO will directly appear at the PLL output unsuppressed when the phase difference falls within the dead-zone of the PFD. Blind-zone will reduce the linear phase detection range, and slow down the dynamic settling behavior of a PLL. Dead-zone and blind-zone often conflict with each other. In this work the reduction of dead-zone and blind-zone are achieved by the innovative design of a dynamic logic phase-frequency detector.

A programmable charge-pump is proposed that is capable of speeding up loop dynamic settling, compensating loop gain variation, and reducing reference spurs. It is pointed out that programmable charge-pumps are critical in developing PLLs for multi-standard applications in wireless communications.

The design of high-speed prescaler is also investigated. Through careful design and optimization, it is possible to implement dual-modulus dividers in CMOS that operate in the 5.8 GHz WLAN frequency band or at even higher frequencies.
Further, a graphic method for obtaining the division ratio in fractional-N PLLs is given. The result obtained will be optimum in terms of power consumption and loop-bandwidth.

The exploration of fully integrated high-performance gigahertz PLLs are to be continued, and challenges continue to exist. Some future work in this research area includes:

- With the aid of good circuit simulators \cite{43}, it is possible to implement behavior models that reflect most of the non-ideal characteristics of PLL building blocks. The dynamic behavior and possibly noise performance of PLLs can then be explored at system level. This will be beneficial for better trade-offs in PLL design.

- PLLs that use ∆Σ modulator as the fractional-N controller have some advantages such as fine frequency tuning, and lowered reference spurs. However, currently PLLs with ∆Σ modulator are not widely and commercially available. The major cause might be due to the complexity and the power consumption of the modulator. The dynamic settling behavior of ∆Σ modulator controlled fractional-N PLLs should also be carefully investigated.

- Process and temperature compensation is always a great interest of high-yield IC design. This will continue to be one of the major efforts of designing high-performance PLLs.
BIBLIOGRAPHY


145


