# Testing BiCMOS and Dynamic CMOS Logic

Siyad Chih-Hua Ma

## Abstract:

This technical report contains the text of Siyad Ma’s thesis “Testing BiCMOS and Dynamic CMOS Logic.”

## Funding:

This research was supported in part by the Innovative Science and Technology Office of the Strategic Defense Initiative Organization and administered through the Office of Naval Research under Contract No. N00014-92-J-1782, and by the National Science Foundation under Grant No. MIP-9107760.

---

<table>
<thead>
<tr>
<th>95-1</th>
<th>Center for Reliable Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CSL TR #95-669)</td>
<td>ERL 460</td>
</tr>
<tr>
<td>June 1995</td>
<td>Computer Systems Laboratory</td>
</tr>
<tr>
<td></td>
<td>Departments of Electrical Engineering and Computer Science</td>
</tr>
<tr>
<td></td>
<td>Stanford University</td>
</tr>
<tr>
<td></td>
<td>Stanford, California 943054055</td>
</tr>
</tbody>
</table>

Copyright © 1995 by Siyad Ma
All rights reserved, including the right to reproduce this report, or portions thereof, in any form
TESTING BiCMOS AND DYNAMIC CMOS LOGIC

Siyad Chih-Hua Ma

CRC Technical Report No. 95-1
(CSL TR 95-669)
June 1995

CENTER FOR RELIABLE COMPUTING
Computer Systems Laboratory
Departments of Electrical Engineering and Computer Science
Stanford University
Stanford, CA 94305-4055

Copyright © 1995 Siyad C. Ma
All Rights Reserved
TESTING BiCMOS AND DYNAMIC CMOS LOGIC

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
Siyad Chih-Hua Ma
June 1995
ABSTRACT

In a normal integrated circuit (IC) production cycle, manufactured ICs are tested to remove defective parts. The purpose of this research is to study the effects of real defects in BiCMOS and Dynamic CMOS circuits, and propose better test solutions to detect these defects. BiCMOS and Dynamic CMOS circuits are used in many new high performance VLSI ICs.

Fault models for BiCMOS and Dynamic CMOS circuits are discussed first. Shorted and open transistor terminals, the most common failure modes in MOS and bipolar transistors, are simulated for BiCMOS and Dynamic CMOS logic gates. Simulations show that a faulty behavior similar to data retention faults in memory cells can occur in BiCMOS and Dynamic CMOS logic gates. We explain here why it is important to test for these faults, and present test techniques that can detect these faults.

Simulation results also show that shorts and opens in Dynamic CMOS and BiCMOS circuits are harder to test than their counterparts in Static CMOS circuits. Simulation results also show that the testability of opens in BiCMOS gates can be predicted without time-consuming transistor-level simulations. We present a prediction method based on an extended switch-level model for BiCMOS gates.

To improve the testability of dynamic CMOS circuits, design-for-testability circuitry are proposed. Scan cell designs add scan capabilities to dynamic latches and flip-flops with negligible performance overhead, while design-for-current-testability circuitry allows quiescent supply current (IDDQ) measurements for dynamic CMOS circuits.
ACKNOWLEDGEMENTS

I express my gratefulness to my adviser, Prof. Edward J. McCluskey, for his constant support, guidance, and encouragement throughout my five years at CRC. His invaluable teachings and precious advice will guide me through my professional career.

I would also like to thank all current and former CRC members, especially to Dr. Piero Franco, Dr. Hong Hao, Dr. LaNae Avra, Samy Makar and Dr. Alice Tokamia for their useful comments and suggestions, and for making CRC enjoyable in my early RATS years. Thanks also to Dr. Nirmal Saxena for his professional assistance, to Nur Touba for sharing quals joy and deadline worries, to Jonathan Chang for stirring many fruitful discussions, to Rob Norwood, Shridhar Mukund, Wem-Yan Koe, Rong Pan, and to the TOPS gang: Dave Brokaw, Khader ‘KD’ Abdel-Hafez, Sunil Kosalge, Erin Kan, Vincent Lo, and Philip Shrivani. Special thanks to Siegrid Munda for her professional administrative support especially during BAST panic time.

I would like to thank Prof. Bruce Wooley, my associate adviser, for being on my orals committee, and for his patience in reading my dissertation. I would also like to thank Prof. Joseph Goodman for chairing my orals committee and reading my dissertation, and Prof. Oyekunle Olukotun for being on my orals committee.

Thanks to people at the Center for Materials Research (CMR) and the Geology Department of Stanford University, especially to Julie Paque and Brian Cross from Fisons Instruments, who provided me with financial support for my first two years.

Thanks to a lot of people for making my life enjoyable at Stanford: many members of the Stanford Chinese Culture Association and the Saturday basketball group.

Thanks to both my family and Mary’s family for their faith in me. My sincere respect to my brothers-in-law, whose moral and financial support have been a great motivation to complete my studies. I am also grateful to my parents, who supported my decision to continue my education during tough times. Finally, I express my deepest appreciation to my wife, Mary, who provided me with a peaceful mind to study, shared every moment of joy and tears, and always believed in me. To Mary, this dissertation is an accomplishment for you as much as it is for myself. My source of inspiration for the last few years came from my beautiful daughter, Eman, and my handsome son, Anan. Their strong learning ability taught me to never stop learning, for knowledge never ends.

This work was supported in part by the Innovative Science and Technology Office of the Strategic Defense Initiative Organization and administered through the Office of Naval Research under Contract No. N00012-92-J-1782, and by the National Science Foundation under Grant No. MIP-9 107760.
# TABLE OF CONTENTS

Abstract .................................................................................................................. i  
Acknowledgements ................................................................................................. ii  
Table of Contents ..................................................................................................... iii  
List of Illustrations .................................................................................................... v  
List of Tables ............................................................................................................ vii  

Chapter 1  
Introduction .......................................................................................................... 1  

Chapter 2  
Production Testing: Defects, Failures, and Fault Models .................................... 3  
  2.1 Defects and Failures ....................................................................................... 4  
  2.2 Fault Models .................................................................................................. 4  
    2.2.1 Stuck-At Fault Model ............................................................................ 5  
    2.2.2 Stuck-Open and Stuck-On Fault Models .............................................. 5  
    2.2.3 Bridging Fault Models ......................................................................... 7  
    2.2.4 Delay Fault Models ............................................................................. 8  
    2.2.5 Transistor-Level Fault Models ............................................................ 8  
    2.2.6 Layout-Level Fault Models ................................................................. 8  
  2.4 Contributions to Fault Modeling: Stationary Fault ...................................... 9  
  2.5 Test Pattern Generation ............................................................................... 10  
  2.6 Design-for-Testability .................................................................................. 10  

Chapter 3  
Failures in BiCMOS and Dynamic CMOS Logic .............................................. 12  
  3.1 BiCMOS and Dynamic CMOS Gate Structures ......................................... 12  
  3.2 Previous Work .............................................................................................. 13  
    3.2.1 Stuck-Open, Stuck-On, and Bridging Faults ........................................ 13  
    3.2.2 Resistive Shorts and Transistor Opens .............................................. 14  
    3.2.3 Layout-Level Fault Models ................................................................. 15  
  3.3 Contributions to Analysis of Resistive Shorts in BiCMOS Gates ................ 15  
  3.4 Contributions to Analysis of Transistor Opens in BiCMOS Gates ............. 20  
    3.4.1 Stuck and Non-stuck Nodes ................................................................. 21  
    3.4.2 Open Fault Testability Prediction for BiCMOS Gates ....................... 22  

iii
3.4.3 Extended Switch-Level Model for BiCMOS Gates .................................. 22
3.5 Contributions to Analysis of Resistive Shorts in Domino Gates .................. 23
3.6 Contributions to Analysis of Transistor Opens in Domino Gates ................. 28
3.7 Short and Open Testability in BiCMOS and Dynamic CMOS Logic .............. 28
3.8 Stationary Faults in CMOS and BiCMOS circuits ...................................... 29
3.9 Testing for Stationary Faults ...................................................................... 31

Chapter 4
Design-for-Testability ...................................................................................... 34
  4.1 Scan Cells for Dynamic Latches and Flip-flops ......................................... 34
  4.2 Design-for-Current-Testability (DFCT) for Dynamic CMOS Logic .......... 36

Chapter 5
Concluding Remarks ......................................................................................... 39

References ........................................................................................................ 41
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Production Test</td>
<td>3</td>
</tr>
<tr>
<td>2-2</td>
<td>Stuck-at Faults</td>
<td>5</td>
</tr>
<tr>
<td>2-3</td>
<td>CMOS NAND Gate</td>
<td>6</td>
</tr>
<tr>
<td>2-4</td>
<td>Bridging Fault</td>
<td>7</td>
</tr>
<tr>
<td>2-5</td>
<td>Comparison between fault models</td>
<td>9</td>
</tr>
<tr>
<td>2-6</td>
<td>Scan Chain</td>
<td>11</td>
</tr>
<tr>
<td>3-1</td>
<td>Conventional BiCMOS NAND Gate</td>
<td>12</td>
</tr>
<tr>
<td>3-2</td>
<td>CMOS Domino AND Gate</td>
<td>13</td>
</tr>
<tr>
<td>3-3</td>
<td>Simulated BiCMOS NAND gate</td>
<td>16</td>
</tr>
<tr>
<td>3-4</td>
<td>Oscillations in BiCMOS Gates</td>
<td>17</td>
</tr>
<tr>
<td>3-5</td>
<td>Physical defect causing stationary fault M4 gate-to-drain short</td>
<td>18</td>
</tr>
<tr>
<td>3-6</td>
<td>Stationary Faults in BiCMOS Gate M4 gate-to-drain short</td>
<td>19</td>
</tr>
<tr>
<td>3-7</td>
<td>Leakage Time vs. M4 Gate-to-Drain Short Resistance</td>
<td>20</td>
</tr>
<tr>
<td>3-8</td>
<td>Stuck and Non-stuck nodes in a BiCMOS NAND gate</td>
<td>21</td>
</tr>
<tr>
<td>3-9</td>
<td>Simulated domino AND circuit</td>
<td>23</td>
</tr>
<tr>
<td>3-10</td>
<td>Equivalent circuit for GS short on MP</td>
<td>24</td>
</tr>
<tr>
<td>3-11</td>
<td>MP gate-to-source short during precharge phase</td>
<td>25</td>
</tr>
<tr>
<td>3-12</td>
<td>MN gate-to-drain short during evaluation phase</td>
<td>26</td>
</tr>
<tr>
<td>3-13</td>
<td>MA gate-to-drain short during precharge</td>
<td>36</td>
</tr>
<tr>
<td>3-14</td>
<td>Broken feedback loop in a Static CMOS Latch</td>
<td>30</td>
</tr>
<tr>
<td>3-15</td>
<td>Short causing stationary fault in a dynamic CMOS latch</td>
<td>30</td>
</tr>
<tr>
<td>3-16</td>
<td>Short causing stationary fault in a domino AND gate</td>
<td>31</td>
</tr>
<tr>
<td>3-17</td>
<td>Leakage time for source-to-drain short on MB</td>
<td>31</td>
</tr>
<tr>
<td>4-1</td>
<td>Transmission gate latch</td>
<td>35</td>
</tr>
<tr>
<td>4-2</td>
<td>Scan cell design for Transmission Gate Latch</td>
<td>35</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>4-3</td>
<td>Static Scan Transmission Gate Latch</td>
<td>36</td>
</tr>
<tr>
<td>4-4</td>
<td>Domino AND Gate with bleeder circuit</td>
<td>37</td>
</tr>
<tr>
<td>4-5</td>
<td>Domino AND Gate with DFCT circuitry</td>
<td>38</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Common Failure Mechanisms</td>
<td>4</td>
</tr>
<tr>
<td>3-1</td>
<td>Effect of Resistive Shorts on Domino AND Gate</td>
<td>23</td>
</tr>
<tr>
<td>3-2</td>
<td>Testability of short and opens in various gate structures</td>
<td>29</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

The demand for high performance digital products has led to the use of circuit designs that were considered as academic research only a few years ago. Many high speed microprocessors that were announced recently contained either BiCMOS [Colwell 95] or dynamic CMOS logic [Bearden 95][Bowhill 95][Chamas 95] to enhance their performance. BiCMOS logic drives high capacitive loads faster than static CMOS, while dynamic CMOS logic has smaller input capacitance, allowing faster switching speeds.

Integrated circuit (IC) testing has traditionally targeted failures that typically occur in transistor-transistor logic (TTL), which dominated most ICs two decades ago. However, failures that occur in static CMOS circuits behave differently from failures in TTL circuits. Although this difference has been known in test research for more than a decade, it is only recently that some test methods that target static CMOS failures were used by industrial vendors to achieve high quality. For BiCMOS and dynamic CMOS circuits, the tests used were still targeting failures in either TTL or static CMOS circuits, which may not be adequate to achieve high quality.

Most VLSI ICs are implemented with static CMOS logic nowadays. Both static and dynamic CMOS circuits are manufactured with the same processing steps, while BiCMOS circuits are manufactured with a process similar to CMOS circuits with a few additional steps. Therefore, the defects that occur in static CMOS circuits can also occur in BiCMOS and dynamic CMOS circuits.

There are also some structural similarities between static CMOS, dynamic CMOS, and BiCMOS gates. For each case, parallel NMOS transistors are used to construct a NOR gate, while series NMOS transistors are used to construct a NAND gate. In dynamic logic such as the domino logic, additional MOS transistors provide precharging and output buffering, whereas in BiCMOS logic, additional bipolar transistors drive the output of the gate. Due to these structural similarities, failures that occur in static CMOS circuits can also occur in BiCMOS and dynamic CMOS circuits. Therefore, to study BiCMOS and dynamic CMOS circuit testing, a thorough understanding of failures in static CMOS circuits and how to test for them is necessary.

This dissertation is a digest summarizing my work in testing BiCMOS and dynamic CMOS circuits. Detailed descriptions of results are found in the appendices, which are reprints of papers published in journals and conferences.
Basic test terminology is defined in Chapter 2. Also discussed are how circuits fail, what fault models are used to model these failures, and how circuits are tested. A new fault model called a stationary fault is described in Chapter 2.

Resistive shorts and transistor opens in BiCMOS and dynamic CMOS circuits are discussed in Chapter 3. We show that some shorts and opens can produce stationary faults in BiCMOS and dynamic CMOS circuits. We also show that these faults may escape stuck-at and delay tests. To test for stationary faults, it is necessary to either apply the test vectors slowly and sample the output voltage, or measure the quiescent supply current (IDDQ). Both slow speed and IDDQ testing are generally not feasible for dynamic CMOS logic. To address these issues, we propose several design-for-testability circuitry for dynamic CMOS logic in Chapter 4.

Concluding remarks are given in Chapter 5.

Appendix A is a reprint of a paper published at International Test Conference 1992. This paper gives a detailed analysis of resistive shorts in BiCMOS gates. Stationary faults are first introduced in this paper. Detailed analysis of opens in BiCMOS gates is given in Appendix B, which a reprint of a paper published at VLSI Test Symposium 1994. We showed that some opens in BiCMOS gates cannot be detected by stuck-at or stuck-open tests; delay tests with accurate timing are required. Based on the results of Appendix B, a procedure to predict the testability of an open in a BiCMOS gate was presented in the May 1995 issue of IEEE Transaction on Computer-Aided Design. This procedure is based on a switch-level model that was extended to include bipolar transistors. A reprint of this journal paper is included as Appendix C.

Design-for-testability for dynamic CMOS circuits is addressed in the next two Appendices. Scan cell designs for dynamic latches and flip-flops are proposed in Appendix D, which is a paper submitted to the Journal of Solid-State Circuits. Some of these scan cells have negligible performance overhead compared to non-scannable dynamic latches. Appendix E is a Center for Reliable Computing Technical Report that explains the problems of applying IDDQ tests to dynamic CMOS circuits. To solve these problems, design-for-current-testability circuitry are presented.

(Note: The Appendices are not included in this Technical Report.)
Chapter 2
Production Testing: Defects, Failures, and Fault Models

The purpose of production testing is to remove any manufactured ICs that do not meet their specifications. This is done by applying a set of input patterns to the circuit-under-test (CUT) and observing its output response. A set of expected output patterns is compared with the actual output patterns, and any discrepancy is recorded as an IC failure. An IC that passes production tests is assumed to be non-defective and shipped to the customer whereas a failing IC is assumed to be defective and therefore discarded (see Fig. 2-1). Since it is impossible to test for every defect in a complex VLSI circuit, some defective ICs may pass production testing. These undetected defective ICs are called test escapes. On the other hand, some ICs that are not defective may fail production test. These are commonly known as type I errors. The main causes of type I errors are errors in test software and hardware, test conditions, or human error [Williams 92].

The goal of test research is to reduce both test escapes and type I errors. Quality level is a measure for how good a test is. Quality level, measured in defects-per-million (DPM), is defined as the proportion of shipped parts that are bad. Therefore, to improve the quality of a test actually means to lower its quality level (or DPM), which means reducing the number of test escapes.

This chapter discusses the cause of failures in ICs, how these failures are modeled, and how test inputs are generated.
2.1 Defects and Failures

Failures occur in CMOS and BiCMOS ICs for many reasons; some are due to manufacturing defects, while others are due to wearout mechanisms whose effects are accumulated over time. External disturbances such as heat, radiation, electrical and mechanical stress also produce failures. A defect is an imperfection that causes the IC to fail. Failure mechanisms are the physical causes of failure. For BiCMOS circuits, failure mechanisms in both MOS and bipolar transistors can produce circuit failures. Common failure mechanisms are discussed in several references [Amerasekera 87] [Woods 86] [Fantini 85] [Hu 89]. These failure mechanisms are listed in Table 2-1.

<table>
<thead>
<tr>
<th>Failure Mechanism</th>
<th>Failure Mode</th>
<th>Affected Locations</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESD, EOS</td>
<td>shorts, opens</td>
<td>interconnects, MOS, bipolar</td>
</tr>
<tr>
<td>gate oxide breakdown</td>
<td>leakage, shorts</td>
<td>MOS</td>
</tr>
<tr>
<td>hot carrier effects</td>
<td>threshold voltage shift</td>
<td>MOS</td>
</tr>
<tr>
<td>ionic contamination</td>
<td>threshold voltage shift</td>
<td>bipolar</td>
</tr>
<tr>
<td>piping, EED</td>
<td>shorts</td>
<td>bipolar</td>
</tr>
<tr>
<td>electromigration</td>
<td>opens, shorts</td>
<td>interconnects</td>
</tr>
<tr>
<td>contact migration</td>
<td>opens, shorts</td>
<td>contacts</td>
</tr>
<tr>
<td>corrosion</td>
<td>opens</td>
<td>interconnects</td>
</tr>
<tr>
<td>spot defects</td>
<td>shorts, opens</td>
<td>interconnects, MOS, bipolar</td>
</tr>
<tr>
<td>latchup</td>
<td>short circuit</td>
<td>CMOS</td>
</tr>
<tr>
<td>radiation</td>
<td>soft errors</td>
<td>memory elements</td>
</tr>
<tr>
<td>crosstalk</td>
<td>signal disturbance</td>
<td>interconnects</td>
</tr>
</tbody>
</table>

The electrical effects of a failure are called failure modes. The most dominant failure modes in CMOS and bipolar ICs are shorts and opens [Fantini85] [McEuen93]. To simulate these failures, fault models are used to model the important electrical behavior of the failures. The most common fault models are discussed next.

2.2 Fault Models

Important electrical properties of failures are modeled using fault models to ease test generation and fault simulation. Fault models can be defined at several levels, which can be:

- Architectural level: where the basic constructs are functional units such as adders, multiplexers, and registers.
. Gate level: where the basic constructs are logic gates such as AND, OR, XOR, etc.
. Switch level: where the basic constructs are signal switches and storage nodes, with discrete signal and drive strengths.
. Transistor level: where the basic constructs are transistors, resistors, capacitors, etc.
. Layout level: where the basic construct are diffusion area, polysilicon and metal, etc.

The complexity of simulation increases rapidly from one level to the next level. The most common fault models are described below.

2.2.1 Stuck-At Fault Model

In the stuck-at fault model, a faulty net is permanently connected to either the low logic level, called stuck-at-0, or the high logic level, called stuck-at-1. For example, Fig. 2-2 shows a multiplexer circuit with input A stuck-at-1. A single stuck-at fault model assumes that only one net can be faulty at a time, whereas a multiple stuck-at fault model allows more than one net to be faulty concurrently.

![Figure 2-2 Stuck-at Faults](image)

Stuck-at test sets can be generated at different levels. At the architectural level, faults are assumed to occur on the inputs and outputs of the modules. These are also known as pin faults. For example, in the multiplexer circuit of Fig. 2-2, the pin faults are stuck-at faults on inputs A, B, S and output Z. At the gate level, faults can occur on the inputs and outputs of individual gates. For the case of the multiplexer, this includes stuck-at faults on leads h and g in addition to the pin faults. Stuck-at faults in the switch level translate to stuck-on, stuck-open and bridging faults, which will be discussed in the following sections.

2.2.2 Stuck-Open and Stuck-On Fault Models

In a switch-level representation of a CMOS circuit, MOS transistors are modeled as switches that conditionally transfer signals. The stuck-open fault model assumes that a faulty transistor never switches on (permanently disconnected), while a stuck-on fault
model assumes that a faulty transistor never switches off (permanently connected) [Wadsack 78]. For an NMOS transistor, stuck-open and stuck-on faults are equivalent to stuck-at-0 and stuck-at-1 faults on the gate of the transistor, respectively.

Detecting a stuck-open fault generally requires a two-pattern test. When a transistor is stuck-open, either the output cannot be pulled high (for a PMOS stuck-open fault) or cannot be pulled low (for an NMOS stuck-open fault) for certain test patterns. To test for a PMOS stuck-open, the output is first set to a low logic level in the first test pattern, while during the second test pattern, we attempt to pull the output to a high logic level through the PMOS transistor. If the PMOS transistor was stuck-open, the output stays at a low logic level, otherwise the output will be pulled high. For example, for the NAND gate shown in Fig. 2-3, to test for a stuck-open fault in transistor $P_1$, we apply the two-pattern test $AB=11, 01$. The fault-free output is 1 after applying the second pattern, while the faulty output is 0. Similarly, to test for a stuck-open fault in transistor $N_1$, we apply the two-pattern test $AB=00, 11$.

![2-3 2 Input NAND Gate](image)

Testing for stuck-on faults requires more knowledge of transistor and interconnect electrical characteristics. If a PMOS transistor was stuck-on, we can potentially detect this fault by turning the corresponding NMOS transistor on. For example, to detect a stuck-on fault on transistor $P_1$ in Fig. 2-3, we apply $AB=11$. If the PMOS transistor pull-up ($P_1$) is stronger than the NMOS pull-down (series transistors $N_1$ and $N_2$), the output is at a high logic level, and the fault is detected. This is called $p$-dominance. If the circuit was $n$-dominant, where the NMOS transistor pull-down is stronger than the PMOS transistor pull-down, then a stuck-on fault on a PMOS transistor cannot be detected by voltage
measurements at the output. Stuck-on faults can be detected by measuring the quiescent 
supply current (IDDQ), since a static current path is formed between the supply voltage and 
ground [Malaiya 84].

2.2.3 Bridging Fault Models

A bridging fault is an unintentional connection between signal lines [Mei 74]. A 
multiple bridge is a bridge involving more than two nets (Fig. 2-4a). A bridge can form a 
feedback loop, possibly transforming a combinational network into a sequential one (Fig. 2-4b).

![Figure 2-4 Bridging Fault (a) Non-Feedback (b) Feedback](image)

In TTL and ECL technologies, a bridge between two nets can be modeled by 
ANDing or ORing the logic values on the two nets [Mei 74]. This is called the wired-AND 
and wired-OR model, respectively. However, bridges in CMOS technologies cannot be 
modeled as simple wired-AND or wired-OR logic; voting models are more accurate. In the 
voting model [Acken 91], when two nets driven to different logic levels are bridged 
together, the resolved logic level depends on the relative strengths of the pull-up network 
(p-transistors) and pull-down network (n-transistors). Although the voltage level of the 
bridged nets is a voltage divider between the pull-up and pull-down networks, the logic 
gates driven by the bridged nets would normally interpret that intermediate voltage level as 
either a high or low logic level. For example, in Fig. 2-4(a), the output of G1 may be 
high, and the output of G2 may be low, resulting in an intermediate voltage on the bridge. If the logic thresholds of G3 and G4 are different, the output of G3 may not be equal to the 
output of G4.

Bridging faults may cause high quiescent supply current, since the two nets that are 
bridged together may be set to different logic levels. A pull-up path may attempt to charge
one net, while a pull-down path may attempt to discharge the other net. With these two nets bridged together, a static current path between supply and ground occurs, and the bridge can potentially be detected by IDDQ testing [Hawkins94].

2.2.4 Delay Fault Models

Delay fault models are used to model the effect of defects that causes the circuit to be slower than normal [Breuer74]. Two types of delay fault models have been proposed: the gate delay and the path delay fault. In the gate delay fault model, the propagation delay through the faulty gate is longer than expected. However, even if a single gate is slow, the performance of the network may not suffer if the gate is not on the critical path, i.e., the longest path from the primary input to the primary output of the network. In the path delay model, the fault is in a path that contains a series of gates. A path is faulty if the propagation delay through that path is longer than expected.

As in the case of stuck-open faults, two-pattern tests are needed to detect delay faults. To test for a delay fault, we need to launch a transition at the inputs and measure the propagation delay of that transition at the output. Launching a transition is equivalent to applying two patterns at the inputs.

2.2.5 Transistor-Level Fault Models

Transistor-level fault models are used to accurately study the analog behavior of some real defects. The most important fault models at the transistor level are resistive shorts [Hao91] and capacitive opens [Maly88]. Many physical defects such as gate-oxide shorts [Hawkins85] and collector-emitter pipes [Amerasekera87] can produce shorts with high resistance values. The faulty behavior of a defective static CMOS gate is found to be strongly dependent on the resistance of shorts [Hao91] and capacitances coupled with open gates [Maly88].

Accurate transistor-level fault models require time consuming SPICE-like simulations. Therefore, their usefulness is limited to small portions of a large circuit. Our approach to study failures in BiCMOS and dynamic CMOS logic starts with injecting accurate transistor-level faults into simple gates, simulating the faulty circuits at the transistor level, and finally modeling their faulty behavior at the switch-level or gate-level.

2.2.6 Layout-Level Fault Models

Very accurate fault models can be used at the layout level, where physical faults such as possible extra or missing parts of different layers and contacts are considered. These fault models are used extensively in inductive fault analysis, where the effects of
layout level faults are simulated, and their behavior is extracted to gate level faults [Ferguson 88].

Note that the goal of fault modeling is not accuracy, but rather adequacy in achieving a target quality level with the least computational effort and the shortest test time. For example, if we are targeting a product with moderate quality demand, a high single stuck-at fault coverage may be adequate to achieve our goal. Using a fault model with higher complexity such as bridging fault or delay fault may translate into more computational effort for both test pattern generation and fault simulation, and may yield longer test sets that require longer tester time, leading to higher overall testing costs.

2.4 Contributions to Fault Modeling: Stationary Fault

A new gate-level fault model called a *stationary fault* is defined here.

**Definition**

A circuit is said to contain a *stationary fault* if the output initially switches correctly, but settles to an incorrect steady-state value.

Stationary faults are similar to data retention faults in memory cells, where the contents of a memory cell may leak away before the cells are refreshed. Stationary faults, however, effect logic gate outputs rather than memory cells.

A comparison between the output waveforms of a stuck-at, delay, stationary fault, and a static hazard is shown in Fig. 2-5. In each case, the solid line indicates the fault free output, while the dashed line indicates the faulty output. For the stuck-at fault, the output is incorrect initially and stays incorrect. For delay faults and static hazards, the output is incorrect initially but settles to a correct steady-state logic value.
An important attribute of a stationary fault is its \textit{leakage time}. The leakage time is defined to be the time from the correct transition to the output switching to an incorrect steady state value. Longer leakage time means that the stationary fault is harder to detect.

Stationary faults are found in BiCMOS and dynamic CMOS circuits. For dynamic CMOS circuits, the failure mode is the same as retention faults in memory cells, which is excessive leakage current. However, for BiCMOS gates, stationary faults occur because of different switching times of pull-up and pull-down paths in the presence of a defect. This will be explained further in Chapter 3.

\section*{2.5 Test Pattern Generation}

Test patterns can be generated with or without using a fault model. An example of a fault model used for test generation is the single stuck-at fault model, where only one stuck-at fault can happen in a faulty circuit. The single stuck-at fault model is the most widely used model to generate test patterns.

To generate a test, the fault must be provoked, and the faulty behavior must be propagated to the outputs. To provoke a stuck-at fault on a line, we force the line to be a logic value other than the stuck-at logic value. For example, to test for a line stuck-at-\textit{O}, the line is forced to \textit{1}. A logic \textit{1} value will appear on a fault-free line, whereas a logic \textit{0} value will appear on a faulty line. The effect of this faulty \textit{0} must be propagated to the outputs to detect the fault.

The advantage of using fault models to generate test sets is that it is very easy to compare among generated test sets by comparing the proportion of faults that are detected in each test set. The disadvantage is that a fault model cannot model every possible fault in the circuit; therefore, some unmodeled faults may not be detected.

An example of a test that is generated without a fault model is \textit{IDDQ} test \cite{Soden93}. An \textit{IDDQ} test is a test where the quiescent supply current is measured. The premise of \textit{IDDQ} tests is that the steady-state leakage current in a defect-free static CMOS circuit is negligible. Some defects can cause the steady-state leakage current to be higher by several orders of magnitude.

\section*{2.6 Design-for-Testability}

Without careful design of complex VLSI circuits, it is impossible to test for every fault in the circuit, especially for sequential circuits. This is due to both the inaccessibility of internal nodes in complex circuits, and the limited test time because modern automatic
test equipment (ATE) is very expensive. To ease testing, special circuitry, called design-for-testability circuitry, are added to the original circuit [McCluskey 86].

It is generally much more difficult to generate test patterns to test for faults in sequential circuits than combinational circuits. To test for the combinational logic block (CLB) in a sequential machine (Fig. 2-6), test patterns must be applied at the inputs to the CLB. However, not all inputs to the CLB are accessible from the primary inputs; some of them are provided by internal latches. Setting the desired test pattern on the internal latches may require several input patterns to be applied. Furthermore, not every output of the CLB is observable from the primary outputs; some outputs of the CLB are stored in internal latches. To observe the faulty output of the CLB, several clock cycles may be required to propagate the fault to the primary outputs.

A standard design-for-test method to solve this problem is to use special scannable memory elements (latches or flip-flops), and link these memory elements into a scan chain (Fig. 2-6) [McCluskey 86]. During normal operation, these scannable bistables act as normal bistables, whereas during scan operation, these bistables are configured as a long shift register, with a special primary input called the scan input (or scan-in), and a special primary output called the scan output (or scan-out).
Testing BiCMOS and Dynamic CMOS logic is discussed here. Resistive shorts and transistor opens are injected into BiCMOS and Dynamic CMOS gates, the defective gates are simulated using a transistor level simulator (HSPICE), and their faulty behavior is analyzed and modeled at switch or gate-level. We will show the implications of this faulty behavior on the testability of BiCMOS and dynamic CMOS logic.

### 3.1 BiCMOS and Dynamic CMOS Gate Structures

There are several different gate structures for both BiCMOS and dynamic CMOS logic. The discussion here will focus on the Conventional BiCMOS and the CMOS domino gate structures. The Conventional BiCMOS NAND gate and the CMOS domino AND gate are shown in Figs. 3-1 and 3-2, respectively. For details on the operation of these two gates, refer to [Weste 93]. The results of this study, however, are applicable to other BiCMOS structures, such as BiNMOS [El-Gamal 89] and full-rail BiCMOS [Embabi 93], as well as other precharged dynamic CMOS logic, such as NORA [Goncalves 83].

![Figure 3-1 Conventional BiCMOS NAND Gate](image)
3.2 Previous Work

BiCMOS and dynamic CMOS failures, faults, and testing were addressed in several papers. A brief survey of these papers is presented here.

3.2.1 Stuck-Open, Stuck-On, and Bridging Faults

Stuck-open, stuck-on and bridging faults in BiCMOS gates were studied by Menon et al. [Menon92]. They showed that the majority of stuck-open and stuck-on faults in a BiCMOS NAND gate cause delay faults, and only a few stuck-open and stuck-on faults cause stuck-at faults. Design-for-testability circuits targeting stuck-open faults in BiCMOS gates were proposed in [Menon93]. With the addition of two transistors, most stuck-open faults in the BiCMOS NAND gate were detectable with single-pattern tests instead of two-pattern tests (see Sec. 2.2.2). We showed that this scheme, however, cannot detect every transistor open in the BiCMOS NAND gate [Ma94a].

A more detailed study of the behavior of bridging faults in BiCMOS gates was done by Favalli et al. [Favalli93]. Because of the strong current drive of bipolar transistors, a resistive bridge between the output of two BiCMOS gates was generally harder to detect than CMOS gates. On the other hand, bridges in BiCMOS gates drew much higher supply current than CMOS gates, especially at low bridge resistance values (<1kΩ).

Stuck-open and stuck-on faults in CMOS domino logic were discussed in [Oklobdzija84] [Wunderlich86] [Singh88]. In general, stuck-open faults in domino logic are easier to detect than static CMOS logic, mostly requiring only single-pattern tests. This is because the precharge phase in domino logic acts as the first pattern in a two-pattern test.

Stuck-open and stuck-on fault models assume that the faulty transistor is either permanently off or on. These faults are equivalent to an infinite resistance open and a zero resistance short between the source and drain of the faulty transistor, for the case of stuck-
open and stuck-on faults, respectively. However, experimental evidence showed that short resistances can be fairly large [Hawkins85], and transistors with open gates may not turn off permanently [Maly 88]. Therefore, for more accurate failure mode analysis, short resistances and open capacitances cannot be ignored.

### 3.2.2 Resistive Shorts and Transistor Opens

In CMOS circuits, resistive shorts can produce stuck-at, stuck-on, delay faults, and reduced noise margins [Hao 91]. In general, functional failures occur for shorts with resistance values smaller than some critical resistance [Hawkins 94]. A functional failure is a failure that changes the function of the circuit regardless of the operation speed. For shorts with resistance values greater than the critical resistance, delay faults may occur. The amount of delay increases as the short resistance approaches the critical resistance. The critical resistance of a short is a function of several parameters, including process variations, circuit design, and logic input thresholds.

Several simulation studies have been reported on resistive shorts and transistor opens. Levitt et al. analyzed shorts and opens in several BiCMOS gate structures, including the Conventional BiCMOS gate structure [Levitt 94]. Each short was simulated with three different resistance values: \(100\Omega, 1k\Omega\) and \(10k\Omega\). Results of their simulations showed that most resistive shorts and transistor opens did not behave as stuck-at faults. Many shorts and opens were detectable by current measurements; however, there were some shorts and opens that can be detected only by delay tests. Some gate structures such as the full-rail BiCMOS gate had opens and shorts that were undetectable by either delay IDDQ tests.

Similar results were obtained by Stewart et al. [Stewart 91] and Salama and Elmasry [Salama 92]. However, the number of shorts and opens that were detectable by current measurements or delay tests varied. Variations in the detectabilities reported in the literature were due to differences in process technology, circuit parameters, and how the faulty shorts and opens were modeled. To enhance the resolution of supply current measurements, several design-for-testability circuits were proposed in [Levitt 94] [Salama 92] and [Osman 94].

Here are some open areas that were not discussed in previous work:

1) All previous studies simulated shorts with a few resistance values. In general, faulty conditions occurred at low and intermediate short resistance values, and no fault occurs when the short had a large resistance. There were no previous studies on the critical short resistance for BiCMOS circuits.
2) For transistor opens, there were no general explanations for why some opens were detectable by delay tests only.
3) There were no studies on resistive shorts and transistor opens in dynamic logic circuits.

The next few sections will address these three issues.

3.2.3 Layout Level Fault Models

Denner et al. performed an inductive fault analysis on a BiCMOS sea-of-gate array [Denner 91]. They found out that most layout level defects translated to shorts and opens at the transistor level, with a few defects producing stuck-open or stuck-on faults.

There were no studies found on layout-level fault analysis for dynamic CMOS logic. However, due to similarities in the layouts of static and dynamic CMOS, layout level defects in dynamic CMOS are expected to behave as shorts and opens, as in the case of static CMOS [Ferguson 88].

The discussion above validates our approach of injecting resistive shorts and transistor opens into BiCMOS and dynamic CMOS gates to analyze their faulty behavior.

3.3 Contributions to Analysis of Resistive Shorts in BiCMOS Gates

To study the effect of resistive shorts on BiCMOS circuits, a circuit with a BiCMOS NAND gate of Fig. 3-3 was simulated using SPICE. Process parameters of the Stanford BiCMOS process [Schott 90] were used. The inputs to the BiCMOS NAND gate were driven by CMOS inverters to shape the input waveform, and the BiCMOS output logic level was restored through CMOS inverters. The output of the BiCMOS gate was also loaded with a 1 pF capacitance. Shorts were injected between each pair of transistor terminals: the source, drain, and gate. Shorts were modeled as a resistor between the shorted terminals. The resistance of each short was varied from 2 kΩ down to OR. Special attention was given to short resistances that produced a BiCMOS output close to the switching threshold of the output CMOS inverters (2.5 V in our case).

Simulation results are summarized below. For more details, see [Ma 92].
**Critical Resistance**

Critical resistances beyond which a short can be detected as a functional failure is fairly low, ranging from $50\Omega$ to $900\Omega$. For some shorts, the transition from a functional failure to a delay fault as a function of short resistance is not as smooth as in the CMOS case. Shorts with resistances close to the critical resistance may produce oscillations or stationary faults (see below).

**Base-emitter (BE) shorts**

BE shorts in bipolar transistors cannot produce output stuck-at faults; they can only produce delay faults, since the base-emitter junction acts as a diode that allows current to flow from base to emitter (NPN transistor). Injecting a BE short is equivalent to adding a parallel conduction path to the diode.

**Collector-emitter (CE) shorts**

CE shorts can cause stuck-at faults at the output, since shorting the collector and emitter of Q1 is equivalent to shorting the BiCMOS output to $V_{DD}$, and shorting the collector and emitter of Q2 is equivalent to shorting the BiCMOS output to ground.

**Base-collector (BC) shorts**

BC shorts can produce stuck-at faults at the output, since there is an inherent diode connection between the base to emitter. Therefore, there exists a current path from the collector to the emitter, which, as discussed above, can generate output stuck-at faults.
**Input-Output Shorts**

A short between the output and an input of a BiCMOS gate can cause the output to oscillate. A BiCMOS NAND gate with a gate-to-drain short on M6, which is equivalent to a short between input B and the BiCMOS output, was simulated using SPICE. The output of the BiCMOS gate and its restored output are shown in Fig. 3–4a and b, respectively. A short resistance of \(300\Omega\) produces oscillations at the restored output. The reason why this short causes the output to oscillate is because it provides a permanent pull-up path for the output when B is high (Fig. 3-3), while transistor Q2 alternates between on and off states [Ma 92]. The range of short resistances that can cause the restored output to oscillate is very small. However, oscillations are hard to detect, and can escape both stuck-at and delay tests.

![Graph](image1.png)

**Figure 3-4 Oscillations in BiCMOS Gates (a) BiCMOS Output (b) Restored Output**

© IEEE [Ma 92]

**Q1 Base Shorts**

A short from the base of Q1 to ground, either directly or indirectly, can produce a stationary fault. An example of a direct short is a gate-to-source short on M7, whereas an example of an indirect short is a source-to-drain short on M4 when the BiCMOS inputs \(AB=10\), as shown in Fig. 3-5.

In Fig. 3-5, a stationary fault can occur when the inputs to the CMOS inverters are switched from 00 to 01, i.e., the inputs to the BiCMOS NAND gate are switched from 11 to 10. A rising transition occurs at the BiCMOS output of a fault-free circuit, while stationary faults can occur in a faulty circuit. The following steps explain how a stationary fault occurs:
1) **Initial conditions:** Initially, when the inputs are 00, transistor MP of the CMOS inverter is conducting, as well as M3 and M4. The gate-to-drain short on M4 forms a static path from \( V_{DD} \) to ground (through channel of MP, gate-to-drain short on M4, channel of M4, and base-emitter junction of Q2). Hence, input B is held at an intermediate voltage, and transistor M2 is slightly conducting. This means that node h (base of Q1) is also at an intermediate voltage.

2) **Correct transition:** When the inputs are switched to 01, transistor MP turns off while MN turns on. This transition will cause M2 to conduct stronger than before, allowing more current to flow through the base-emitter junction of Q1. Due to the bipolar current amplification, Q1 strongly charges the output capacitor to a high voltage. Hence the correct transition occurs without delay.

3) **Incorrect steady-state value:** When h is charged to a high voltage, M7 turns on. Also, when the voltage on h is higher, the voltage on input B is also higher due to the static current path shown in Fig. 3-5. Therefore, transistors M4 and M6 are slightly on, allowing the output to discharge slowly through M5, M6 and M7. When sufficient charge on the output is discharged, the output switches to a low logic value, and a stationary fault occurs.

An initial study showed that the short resistance range for which a stationary fault occurs was fairly small, ranging from 600\(\Omega\) to 650\(\Omega\) [Ma 92]. Also the maximum leakage
time was approximately 12ns. The small resistance range signified that the probability of a stationary fault occurring in a BiCMOS gate was very low. However, the short resistance range for which a stationary fault can occur depends strongly on circuit design (internal capacitances and drive strengths) and logic thresholds of the output restoring CMOS inverters.

To illustrate the sensitivity of stationary faults to circuit design, another BiCMOS gate was laid out with MAGIC [Mayo 90]. All NMOS transistors were $24\mu/2\mu$, and all PMOS transistors were $18\mu/2\mu$. The input threshold of the CMOS inverters was slightly higher than $V_{DD}/2$ (approximately 2V). SPICE files were extracted from the layout (except for bipolar transistor, which the extraction tool could not extract), and the circuit was simulated using HSPICE [Meta 95]. The result of SPICE simulations for inputs switching from 00 to 01 is shown in Fig. 3-6. The short resistance that can cause a stationary fault ranges from less than $500\Omega$ to approximately $710\Omega$. The maximum leakage time at $710\Omega$ is approximately 36ns. Figure 3-7 shows the relationship between the short resistance and the leakage time:-

![Figure 3-6 Stationary Faults in BiCMOS Gate: M4 gate-to-drain short](image)
3.4 Contributions to Analysis of Transistor Opens in BiCMOS Gates

The BiCMOS NAND circuit of Fig. 3-3 was simulated with transistor opens. Detailed analysis of simulation results is given in [Ma 94a]. Here is a summary of the results.

**M1, M2 Opens**

An open in either transistor M1 or M2 produce stuck-open faults. This means that if there is an open on M1, the output can be charged to a high voltage only if M2 is turned on; the output cannot be charged to a high voltage using M1. This is also true for M2 opens.

**Q1 Base and Emitter Opens**

Since the output can be connected to the supply voltage only through these two nodes (no Q1 collector current if there is no base-emitter current), an open in these two nodes will produce a stuck-at-0 fault at the output.

**M5, M6 Opens**

M5 and M6 are responsible for discharging the output to ground through the base-emitter junction of Q2, and therefore allowing Q2 collector current to further discharge the output during a high-to-low transition. If either M5 or M6 is open, the output will be stuck-at-1, since both output discharge paths (M5 and M6 path, and Q2 collector path) are not conducting.
Other Transistor Opens

Opens on all other transistor nodes do not affect the function of the BiCMOS NAND gate. An open in these nodes will only cause the performance to degrade, and a delay fault can occur.

3.4.1 Stuck and Non-stuck Nodes

Definition

A node is open fault testable if and only if an open on the node can be detected using stuck tests (stuck-at, stuck-open, or stuck-on tests). An open fault testable node is also called a stuck node. A node is a non-stuck node if an open on the node cannot be detected by stuck tests.

Stuck and non-stuck nodes in a Conventional BiCMOS NAND gate are shown in Fig. 3-8 as dashed and solid lines, respectively. Notice that the structure of stuck nodes resembles the structure of a CMOS NAND gate. Based on our SPICE simulations for opens in other BiCMOS gate structures, such as the BiNMOS [El-Gamal 89], full-rail BiCMOS [Embabi 93], and the CBiCMOS [Hiraki 92] NAND gates, we found that the structure of the stuck nodes resembled a CMOS NAND gate in each case [Ma 94a]. The only exception is the full-rail BiCMOS gate with collector-emitter (CE) shunt, since this BiCMOS gate consists of a BiCMOS and a CMOS NAND gate connected in parallel. The parallelism in the gate structure provides full functional redundancy, such that if one part of the gate is faulty, the other part will mask out the failure. Therefore, transistor opens in the full-rail BiCMOS gate with CE shunt cannot produce stuck faults [Ma 94a].

Figure 3-8 Stuck (dashed lines) and Non-stuck nodes (solid lines) in a BiCMOS NAND gate

© IEEE [Ma 95]
3.4.2 Open Fault Testability Prediction for BiCMOS Gates

Using the previous observation that the structure of stuck nodes in a BiCMOS gate resembles the structure of a CMOS gate with the same functionality, we derived procedures to predict whether a node in a BiCMOS gate is a stuck node. These procedures were based on switch-level models for CMOS circuits; with extensions for bipolar transistors. The extended switch-level model is discussed next. Using the extended switch-level model, we showed that the open fault testability of all nodes in a BiCMOS gate can be derived by a few simple set manipulations [Ma 95a]. This provides a very rapid way to assess the testability of each open fault in a BiCMOS gate without running time consuming transistor level simulations.

3.4.3 Extended Switch-Level Model for BiCMOS Gates

A switch-level model is a model between gate-level and transistor-level models, where each transistor is modeled as a switch. Switch-level models provide higher accuracy than gate-level models, yet allow simulation speeds faster than transistor-level models. Switch-level algorithms were developed for several important tasks such as fault simulation, automatic test pattern generation, formal verification, and timing analysis [Bryant 87].

Switch-level models were typically used for digital MOS circuits, since each MOS transistor can be modeled as a voltage switch. There were also several proposed switch-level models for emitter-coupled logic (ECL) circuits by modeling each bipolar transistor as a current switch [Yang 93]. The switch-level model for MOS transistors is extended here to include bipolar transistors.

In the switch-level model, MOS transistors are modeled as switches controlled by the voltage level of its gate terminal, conditionally allowing a signal to propagate between the source and drain terminals. MOS transistors are bi-directional elements; current can flow from source to drain or vice versa. In our extended switch-level model, each bipolar transistor is modeled as two switches. The first switch is controlled by the base-emitter voltage level, which determines whether a signal can propagate from base to emitter, and the other switch is controlled by both the collector-emitter voltage and whether the first switch is on, which determines whether a signal can propagate from collector to emitter. For more details, refer to [Ma 95a].

Although the bipolar transistor is actually a current switch, it is modeled as two voltage switches in the extended switch-level model to be consistent with switch-level models for MOS transistors. This approximation is adequate for the digital BiCMOS gate structures analyzed in [Ma 95a].
3.5 Contributions to Analysis of Resistive Shorts in Domino Gates

The domino AND circuit shown in Fig. 3-9 was laid out using MAGIC [Mayo 903, and SPICE files were extracted. Resistive shorts were injected into the domino AND gate and simulated using HSPICE [Meta95], with MOSIS 1.0\(\mu\)m N-well process parameters [Pi]. Notice that the Clk input is driven by a static CMOS buffer (not shown in Fig. 3-9), which consists of two cascaded static CMOS inverters. Results are summarized in Table 3-1, and explained briefly below. For each resistive short, the shorted nodes were listed in Table 3-1, as well as the faulty behavior when the short resistance was either high or low.

SA 1, SAO, SF, SH, D, and NM denote stuck-at- 1, stuck-at-O, stationary fault, static hazard, delay, and reduced noise margins at the output, respectively. A/SA1 and B/SA 1 denote stuck-at-1 at the input A and input B, respectively. \(\text{Clk}\) indicates that the output follows the complement of the Clk input, whereas, \(\overline{\text{Out}}\) means that the faulty output is the complement of the fault-free output.

![Figure 3-9 Simulated domino AND circuit](image)

**Table 3-1 Effect of Resistive Shorts on Domino AND Gate**

<table>
<thead>
<tr>
<th>Transistor</th>
<th>Gate-Drain</th>
<th>Gate-Source</th>
<th>Source-Drain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nodes</td>
<td>Low R</td>
<td>High R</td>
</tr>
<tr>
<td>MP</td>
<td>Clk-Z</td>
<td>Clk</td>
<td>(\text{Clk})</td>
</tr>
<tr>
<td>MN</td>
<td>Clk-g</td>
<td>SA0</td>
<td>D</td>
</tr>
<tr>
<td>MA</td>
<td>A-Z</td>
<td>SF, SH</td>
<td>D</td>
</tr>
<tr>
<td>MB</td>
<td>B-h</td>
<td>SF</td>
<td>D</td>
</tr>
<tr>
<td>MI</td>
<td>Z-Out</td>
<td>Out</td>
<td>NM</td>
</tr>
<tr>
<td>M2</td>
<td>Z-Out</td>
<td>Out</td>
<td>NM</td>
</tr>
</tbody>
</table>
In general, an excessive leakage on dynamic node \( Z \) will produce a stationary fault. Excessive leakage can be a result of either a direct or indirect short between node \( Z \) and ground.

**Transistor MP shorts**

*Gate-to-source short:* With a gate-to-drain short, the equivalent Clk input circuit is shown in Fig. 3-10. \( R_{sh} \) is the resistance of the short between the source and gate of \( MP \). The capacitance \( C \) is a lumped capacitance that includes gate capacitances of MP and MN and interconnect capacitance of the Clk wire. A short with small resistance value (<50\( \Omega \)) behaves as Clk stuck-at-1. This means that the precharge transistor MP of the domino AND gate of Fig. 3-9 never switches on, and \( Z \) is never precharged high. Hence, the output of the domino AND gate is stuck-at-1. For shorts with higher resistance, the precharge time is longer than normal. If the precharge phase is not long enough to fully charge node \( Z \) to \( V_{DD} \), then during the evaluation phase, the leakage on node \( Z \) may eventually lead to an incorrect logic level at the output. Therefore, stationary faults may occur.

![Figure 3-10 Equivalent circuit for GS short on MP](image)

*Gate-to-drain short:* A gate-to-drain short on MP will create a conduction path between Clk and node \( Z \) of the domino AND gate. For small resistance values, node \( Z \) follows the Clk input regardless of other inputs, hence the output behaves as \( \overline{\text{Clk}} \). For higher resistance values, node \( Z \) cannot be fully charged to \( V_{DD} \) during precharge, due to the static current path shown in Fig. 3-11. The reduced voltage level on \( Z \) will not produce a stationary fault during the evaluation phase, since the Clk input can charge node \( Z \) to \( V_{DD} \) through the gate-to-drain short when either input A or B is low. However, when both A and B are high, the short resistance will oppose \( Z \) from discharging, and either an output stuck-at-0 fault or a delay fault may occur, depending on the short resistance.
Source-to-drain short: A source-to-drain short on MP has no effect on the precharge phase. If the short resistance is small, the short prevents node Z from discharging to a low logic value during the evaluation phase, producing an output stuck-at-0 fault. For higher short resistances, the short will resist node Z from discharging, therefore a delay fault may occur.

Transistor MN shorts

Gate-to-drain short  A gate-to-drain short on MN has no effect on the precharging. During the evaluation phase, a static current path exists from $V_{DD}$ to ground as shown in Fig. 3-12. For small short resistances, the voltage on Clk is at an intermediate value between $V_{DD}$ and ground, and MP is on, preventing node Z from discharging to a low logic value even when both gate inputs A and B are high. Therefore, an output stuck-at-0 may occur. For higher resistance values, transistor MP may be non-conducting, however, the short will resist node Z from discharging to ground, and a delay fault occurs.

Gate-to-source short  A gate-to-source short on MN has no effect on precharging. For small short resistance, the short behaves as Clk stuck-at-0. Therefore, the output is stuck-at-0. For higher resistances, both transistors MP and MN are on during evaluation, and a delay fault may occur.

Source-to-drain short  A source-to-drain short on MN does not affect the operation of the domino AND gate. During precharge, both inputs A and B are low. A short on MN does not affect the precharging of node Z. During evaluation, transistor MN is on, and the source-to-drain short acts as a conduction path parallel to the channel of MN.
Transistor MA shorts

Gate-to-drain short  A gate-to-drain short on MA connects input A to node Z. For a small short resistance, the output behaves as $\overline{A}$, which means that the output is high during precharge (instead of being low). For higher short resistances, $Z$ may be precharged to an intermediate voltage between $V_{DD}$ and ground, due to the static current path shown in Fig. 3-13. This intermediate may be sufficiently high that the output of the domino AND gate is low during precharge; however, once evaluation starts, $M_2$ is turned off, and node $Z$ is discharged through the resistive short. If the discharging process is slow, a stationary fault occurs.
Notice that if input A switches high after some delay (with B still low), node Z may start charging high again, restoring the correct logic at the output. In this case, a static hazard occurs on the output, which may cause errors in a domino CMOS logic circuits, since domino logic is designed to be a hazard-free logic. If both A and B are high, the output may switch correctly if the gate-to-drain short resistance on MA is high, however, it will take a longer time to discharge node Z, and a delay fault may occur.

*Gate-to-source short* A gate-to-source short on MA will generally cause no functional errors, but may cause delay faults during evaluation.

*Source-to-drain short* A source-to-drain short on MA will cause node Z to discharge if input B is high during the evaluation phase. If the discharge time is long, a stationary fault occurs.

**Transistor MB shorts**

*Gate-to-drain short* A gate-to-drain short on MB has no effect during precharge, since input A is low. During evaluation, a fault may occur only if input A switches high, since transistor MA isolates the dynamic node Z from the resistive short whenever input A is low. Node Z starts discharging when A becomes high, even if B is kept low, and a stationary fault occurs. The output behavior is similar to a gate-to-source short on MA if both inputs A and B are switched high.

*Gate-to-source short* When input A is high, the output behavior of a gate-to-source short on MB is similar to a gate-to-source short on MA with input B high. When input A is low, the output behaves correctly.

*Source-to-drain short* A source-to-drain short on MB behaves as a source-to-drain short on MA, causing a stationary fault when input A is high and input B is low.

**Transistor MI shorts**

*Gate-to-drain short* A gate-to-drain short on MI connects the output to node Z. The output follows node Z (complement of fault-free output) when the resistance of the short is low. For higher short resistances, the gate functions correctly, but at a lower speed and reduced noise margin. Reduced noise margins may not be acceptable for dynamic logic design, since they may allow subthreshold currents to flow, causing the dynamic nodes of subsequent gates to discharge faster.

*Gate-to-source short* A gate-to-drain short on MI behaves as a source-to-drain short on MP.
Source-to-drain short A source-to-drain short will produce an output stuck-at-1 fault for low short resistances. For higher resistances, reduced noise margins may produce errors on subsequent gates.

Transistor M2 shorts

Gate-to-drain short A gate-to-drain short on M2 behaves as a gate-to-drain short on M1. Gate-to-source short A gate-to-source short on M2 connects node Z to ground. If the short resistance is low, node Z cannot be charged high, and an output stuck-at-1 fault occurs. For higher short resistances, node Z is precharged to an intermediate value, and discharging takes place once evaluation starts. Hence a stationary fault may occur. Source-to-drain short A source-to-drain short produces an output stuck-at-0 fault for low short resistances. For higher resistances, reduced noise margins may produce errors on subsequent gates.

3.6 Contributions to Analysis of Transistor Opens in Domino Gates

In general, the behavior of source and drain terminal opens is identical to a transistor stuck-open fault. For gate terminal opens, the behavior depends on surrounding coupling capacitances. If the coupling capacitances keep the open gate at an intermediate voltage level, the behavior of the open gate fault is similar to a resistive source-to-drain short on the faulty transistor, since the transistor is partially on.

3.7 Short and Open Testability in BiCMOS and Dynamic CMOS Logic

Most automatic test pattern generation tools generate tests using a gate-level description of the circuit. The most commonly used fault model in test generation is the single stuck-at fault, as mentioned in Sec. 2.4. A test that detects all single stuck-at faults in a two input NAND (or AND) gate is 01, 11, 10. The order of the vectors in a single stuck-at test is not important, since detecting stuck-at faults requires only single-pattern tests.

To detect all stuck-open faults in a two input CMOS NAND gate, the following test can be applied: 11, 01, 11, 10. Notice that the ordering of test vectors is important, since each pair of patterns in the test detects some stuck-open faults (see Sec. 2.2.2). No assumptions are made on how fast the test patterns must be clocked, since the basic assumption is that for a stuck-open fault, the output is incorrect for a very long time.
Hence, the faulty output should be observable even when the inputs are applied at low speed.

A test that detects every delay fault in a CMOS NAND gate has the same input patterns as the stuck-open test, but the input patterns are applied with accurate timing. Faults are detected by sampling the output at rated speed.

To compare the effectiveness of each of these tests in detecting resistive shorts and transistor opens, we simulated a static CMOS NAND gate, the Conventional BiCMOS NAND gate of Fig. 3-1, and the domino AND gate of Fig. 3-2, injecting all possible resistive shorts and transistor opens. For each fault, the single stuck-at test, the stuck-open test, and the delay test were applied. Table 3-2 shows the proportion of shorts and opens that were detected for each circuit under each test.

<table>
<thead>
<tr>
<th>Gate Structure (NAND, AND)</th>
<th># shorts and opens</th>
<th>Test Set</th>
<th>Stuck-At</th>
<th>Stuck-Open</th>
<th>Delay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Static CMOS</td>
<td>24</td>
<td>79.2%</td>
<td>91.7%</td>
<td>100%</td>
<td></td>
</tr>
<tr>
<td>Domino CMOS</td>
<td>36</td>
<td>86.1%</td>
<td>86.1%</td>
<td>91.7%</td>
<td></td>
</tr>
<tr>
<td>BiCMOS</td>
<td>54</td>
<td>42.6%</td>
<td>55.5%</td>
<td>79.6%</td>
<td></td>
</tr>
</tbody>
</table>

The total number of shorts and opens shown in Table 3-2 also indicates the number of transistors in each gate structure. In general, more transistors in a gate mean higher functional redundancy, and therefore stuck-at and stuck-open tests detect fewer shorts and opens. For BiCMOS gates, it is clear that stuck-at test would not be sufficient to achieve high quality. The detectability of shorts and opens may be even worse for more complicated BiCMOS gate structures, such as full-rail BiCMOS [Embabi 93] and CBiCMOS [Shin 90].

### 3.8 Stationary Faults in CMOS and BiCMOS circuits

From the simulation results shown in Sec. 3.7, many BiCMOS shorts and opens are not detectable by delay tests. Most of the undetected shorts cause stationary faults, which are undetectable by delay tests.

Stationary faults can also occur in static and dynamic storage elements (latches and flip-flops). These faults are similar to data retention faults in random-access memories (RAMs) [VanDeGoor 91].
Stationary Faults in Static CMOS Latches and Flip-flops

A broken feedback loop in a static CMOS latch can cause a stationary fault. This is because the broken feedback loop transforms a static latch into a dynamic latch as shown in Fig. 3-14.

![Figure 3-14 Broken feedback loop in a Static CMOS Latch](image)

 Normally, logic values are stored in a static CMOS latch with a cross-coupled latch, and no refreshing is required. Without the feedback, the logic level is stored as charge on the input capacitance of the inverter shown in Fig. 3-14, and this charge can leak if not refreshed periodically. If the stored charge leaks before the latch is clocked, a stationary fault can occur. Our SPICE simulation results show that the leakage time for a stationary fault to occur in static CMOS latch can be as long as $2\mu s$.

Stationary Faults in Dynamic CMOS Circuits

Stationary faults occur in dynamic CMOS circuits when there is an excessive leakage on the dynamic storage nodes. Figure 3-15 shows an example of a short that can cause a stationary fault in a dynamic CMOS latch. The resistive short prevents the storage node A from fully charged to $V_{DD}$, and provides a discharge path for node A when Clk=0.

![Figure 3-15 Short causing stationary fault in a dynamic CMOS latch](image)

In Sec. 3.5, several shorts in a domino CMOS circuit are shown to cause stationary faults. The source-to-drain short shown in Fig. 3-16 was simulated using HSPICE, and the relationship between the short resistance and the leakage time is plotted in Fig. 3-17. In
Fig. 3-17, the propagation delay from a rising Clk input to a rising domino gate output is also plotted as a horizontal line (0.492 ns). Theoretically, the clock can run as fast as this propagation delay. If the clock is run such that its period is as fast as 0.5 ns, a short with resistance greater than 1 kΩ resistance will not be detected.

Stationary Faults in BiCMOS Gates

Stationary faults in BiCMOS gates were thoroughly explained in Sec. 3.3.

3.9 Testing for Stationary Faults

Stationary faults occur only when a circuit is clocked at a relatively slow speed. Although one may argue that the circuit functions correctly at high speeds, and hence tests for stationary faults are not necessary, there are several reasons why we still need to detect the defects that can cause stationary faults.

The resistances of some physical shorts such as gate-oxide shorts may change with time [Hawkins 85]. Consequently, a short that causes a stationary fault during testing may
degrade further in the field, causing functional failures. Therefore, stationary faults should be removed before causing catastrophic failures in the field.

In modern high speed circuits, not every latch or flip-flop is necessarily clocked all the time. One example is in low power designs, where the clocks to some latches or flip-flops are disabled during a ‘sleep’ mode to save power. The assumption is that for a static CMOS, the state is maintained in the cross-coupled inverters. However, if a stationary fault occurs, the contents of the faulty latch are lost.

Stationary faults can be tested by either output voltage sampling or supply current monitoring (IDDQ testing). The issues concerning these two test methods are presented below.

**Testing for Stationary Faults by Output Voltage Sampling**

To detect a stationary fault on the output of a circuit, the fault must be provoked. To provoke a stationary fault, a transition is launched at the circuit inputs, and the inputs should be held constant for a long time to observe whether the output switches to an incorrect logic. The time during which the inputs are held constant depends on the leakage time of stationary faults.

Circuits that contain dynamic storage elements cannot be clocked at very low speeds, otherwise we will risk losing the stored values on the dynamic storage elements. Therefore testing a circuit with dynamic storage at very slow speeds may cause type I errors. The other limitation of testing at very slow speeds is tester time. Longer tester time translates to higher testing cost.

**Testing for Stationary Faults by Supply Current Monitoring (IDDQ Tests)**

Stationary faults in BiCMOS and dynamic CMOS logic gates are mostly caused by shorts. These shorts can also create static current paths from $V_{DD}$ to ground. The quiescent supply current of a BiCMOS gate with a gate-to-drain short on transistor M4 (Fig. 3-1) was found to be two orders of magnitude larger than the fault-free gate [Ma 92]. This abnormally high quiescent supply current can potentially be detected by IDDQ tests.

Some problems of applying IDDQ tests to dynamic logic are:
1) Supply current measurements are slow. Using the precision measurement unit (PMU) on a tester, each IDDQ measurement may take milliseconds. Dynamic nodes may lose their charges before any measurements are made.
2) Many shorts in a domino circuit cannot be detected by IDDQ tests. An example is the source-to-drain short on transistor MB (Fig. 3-2) [Ma 94b].
To allow IDDQ testing for dynamic circuits, a special design-for-current-testability (DFCT) circuit is proposed. This DFCT circuit is discussed in Chapter 4.
Chapter 4
Design-for-Testability

Dynamic storage elements (latches and flip-flops) are often used to temporarily store the results of dynamic logic blocks. Dynamic storage elements are faster and more compact than static storage elements. Many scan cell designs for static storage elements have been previously proposed [McCluskey 86][Funatsu 89][Zasio 85]. However, there were no scan cell designs for dynamic storage elements in the literature.

IDDQ testing can detect many defects in static CMOS circuits that are undetectable by voltage sampling [Hawkins 94]. However, IDDQ tests cannot be applied to dynamic circuits without some circuit modifications.

In this chapter, we present scan cell designs for dynamic latches. We also show circuit modifications to make dynamic logic IDDQ-testable.

4.1 Scan Cells for Dynamic Latches and Flip-flops

The advantages of using dynamic latches and flip-flops over their static counterparts are higher performance and smaller area. For modern high speed designs, performance is the most important design parameter. Therefore, scan cells for dynamic latches and flip-flops should be designed to have low performance overhead. Other important properties for scan cell designs are low area overhead and flexibility when incorporating the cell into a scan chain.

Three common dynamic latch structures were investigated: the transmission gate latch, the C2MOS latch, and the true single-phase clock latch [Ji-Ren 87]. We will present a scan cell design for the transmission gate latch here. Other scan latches are described in [Ma 95b].

Transmission Gate Latch

A transmission gate latch consists of a transmission gate followed by an inverter as shown in Fig. 4-1. Logic values are stored as charge on the storage node A, which is the input capacitance of the inverter. When Clk is high, input D charges or discharges the storage node, storing a logic 1 or 0. When Clk is low, node A is tristated, and charge on node A is trapped. Charge is kept on node A except for small leakage currents.
Scan Cell Design for Transmission Gate Latch

Under normal operation, the critical path of the transmission gate latch includes input D, the storage node A, and output $\overline{Q}$. Therefore, to minimize the performance overhead, the capacitances on nodes D, A, and $\overline{Q}$ need to be minimized when adding scan.

![Figure 4-1 Transmission gate latch](image)

To add scan capabilities to a latch, another input that can store charge on the storage node A must be added. This is done by adding another transmission gate, shown as transistors M5 and M6 in Fig. 4-2, which is controlled by a separate scan clock (Sclk). Notice that the source regions of transistors M5 and M6 can be merged with the source regions of M1 and M2, hence there need not be extra capacitance on node A.

A transmission gate and an inverter are added to the output of the transmission gate latch to isolate the possibly long capacitive interconnect of the scan out signal (SO) from the normal data output $\overline{Q}$. Notice again that the drain regions of M7 and M8 can be merged with the drain regions of M3 and M4, adding no extra capacitance on $\overline{Q}$.

![Figure 4-2 Scan cell design for Transmission Gate Latch](image)

The scannable transmission gate latch of Fig. 4-2 was laid out using MAGIC [Mayo 90] and SPICE files were extracted. The cell area is almost twice as large as the
non-scannable transmission gate latch of Fig. 4-1. The cell was simulated using HSPICE [Meta95] with MOSIS 1.0mm N-well technology [Pi]. The performance overhead of the scannable transmission gate latch was negligible compared to the non-scannable transmission gate latch of Fig. 4-1.

**Static Scan Dynamic Latch**

Dynamic latches store logic values as charge on capacitive nodes. If the latch is not clocked at high speed, the contents of the latch are lost. To perform IDDQ testing, however, all nodes in the circuits should be stable for a relatively long time. To maintain the stored logic values on the latch, a dynamic latch that can be transformed into a static latch under scan mode is designed. The latch shown in Fig. 2-5 is a dynamic latch when SCAN=0 and SClk=1. During scan operation, SCAN=1, and a feedback loop is formed. Therefore, the latch becomes static when clocked using SClk. Note that Clk should be held low when SCAN=1.

![Figure 4-3  Static Scan Transmission Gate Latch](image)

**Other Scan Latches and Flip-Flops**

Scan cells for the C²MOS latch, true single-phase clock latch, and various dynamic flip-flops have been designed. Area and performance overhead, as well as scan chain clocking and arrangements for each scan latch were discussed [Ma 95b].

**4.2 Design-for-Current-Testability (DFCT) for Dynamic CMOS Logic**

IDDQ testing can detect some defects that cannot be detected by functional tests [Hawkins 94]. Stuck-open faults, excessive leakage, shorts and bridging faults in dynamic circuits are shown to be detectable using IDDQ testing [Homing 87] [Jacomino...
Vandris showed that stuck-on faults can always be detected by either functional or IDDQ tests [Vandris 91].

Some defects in dynamic CMOS gates are undetectable if the gate inputs are not fully controllable [Ma 94b]. For example, during precharge, all inputs to domino logic gates are held at 0; an input of a domino gate cannot be set to 1 when the clock is 0. Therefore, a defect cannot be detected if its only test requires both an input to be 1 and clock to be 0. Furthermore, IDDQ is typically measured at very slow speeds (10 to 100kHz) by external testers [Hawkins 89]. Dynamic nodes may lose their charges before actual measurements are made.

To solve these problems, a design-for-current-testability circuit is proposed for domino CMOS logic. One solution is to add a bleeder circuit [Weste 93] to each domino gate as shown in Fig. 4-4. Notice that the gate terminal of the bleeder transistor M3 must be connected to the output Out, and cannot be tied to ground. Connecting the gate of M3 to ground will create a static current path from VDD to ground during evaluation for a fault-free gate, hence increasing the quiescent supply current significantly.

![Figure 4-4 Domino AND Gate with bleeder circuit](image)

The bleeder circuit of Fig. 4-4 solves all the issues mentioned above [Ma 94b]. However, M3 opposes the dynamic node Z from discharging during evaluation; hence the output rise time is longer. Our simulations show that the low-to-high propagation delay \( t_{ph} \) can increase by 24%. To minimize this performance penalty, we conditionally switch off the bleeder circuit M3 during normal operation, as shown in Fig. 4-5. During normal operation ITEST=0, and M3 is turned off. ITEST is switched to 1 during IDDQ testing, and the bleeder transistor M3 turns on whenever the output is low.

The DFCT circuit shown in Fig. 4-5 has negligible performance penalty compared to the domino gate without the bleeder circuit [Ma 94b].
Figure 4-5 Domino AND Gate with DFCT circuitry
Chapter 5
Concluding Remarks

BiCMOS and dynamic CMOS logic offer high-speed circuit design solutions for modern microprocessors. Without careful considerations in reliability and testing, some circuits are likely to fail in the field. This dissertation covers many issues in BiCMOS and dynamic CMOS testing that were never addressed before.

First, a new fault model called a stationary fault was introduced. This fault can escape both stuck-at and delay tests, and can result in functional failures in the field.

We analyzed some real defects that can occur in BiCMOS and dynamic CMOS circuits, and found that these defects can cause faults that occur in static CMOS circuits, as well as other faults such as stationary faults and oscillations. For dynamic circuits, static hazards caused by physical shorts pose a problem for hazard-free designs.

Analysis of opens in BiCMOS gates showed that the behavior of each open depends on which path is open. Opens in functional paths caused functional failures, whereas opens in performance paths caused performance degradation only. The structure of functional paths in a BiCMOS gate was shown to resemble the structure of a CMOS gate of the same function.

Switch-level models for CMOS gates were extended to include bipolar transistors, allowing efficient and accurate switch-level simulations for BiCMOS circuits. The extended model was used to predict whether an open in the BiCMOS gate was detectable with functional tests.

Based on our simulations, we found that stuck-at tests can detect only half of the resistive shorts and transistor opens in a BiCMOS gate. Even delay tests could not detect all shorts and opens in a BiCMOS or dynamic CMOS gate, mainly because of stationary faults and redundancy.

After we established the importance of stationary fault testing, two methods to test for stationary faults were presented. Running tests at slow speeds would detect stationary faults; however, there are limitations on how slow we can run a dynamic circuit. Furthermore, slower test speed translates into longer tester time and hence higher test costs. The other method to detect defects that can cause stationary faults was by measuring the quiescent supply current. Quiescent supply current measurement is also a very slow process and cannot be correctly applied to dynamic circuits without precautions.

To solve the problems of slow speed operation and supply current measurement in dynamic circuits, design-for-testability circuits were proposed in Chapter 4. Dynamic
latches and flip-flops were made scannable, allowing high controllability and observability for circuits designed with dynamic storage elements. A scan cell design that transforms a dynamic latch into a static latch during scan operation offered a simple solution to slow speed testing for dynamic circuits.

To allow supply current measurement tests for dynamic logic, a design-for-current-testability circuit based on the bleeder circuit was presented.

Many future directions in BiCMOS and dynamic CMOS logic are possible. Many BiCMOS and dynamic CMOS gate structures have been proposed in the literature. A complete study of each structure should be done before each gate can be reliably used.

Inductive fault analysis (IFA) can be performed on BiCMOS and dynamic CMOS layouts to investigate whether other non-conventional faults may occur. From IFA results, testable structures at the layout level can be derived.

Finally, experimental results on real defects are needed for BiCMOS and dynamic CMOS circuits. The experiments should show the effectiveness of different test methodologies in detecting failures in BiCMOS and dynamic CMOS circuits. Such experiments will help us understand how to achieve high quality in BiCMOS and dynamic CMOS circuits.
REFERENCES


