POLYPUBLIE

High-speed CMOS design techniques for multi-gigahertz transceivers

Hung Tien Bui

Document en libre accès dans PolyPublie
Open Access document in PolyPublie

Date	2006
Type	Mémoire ou thèse / Dissertation or Thesis Thèse de doctorat/Ph.D. thesis
URL de PolyPublie PolyPublie URL	https://publications.polymtl.ca/7740/
Conditions d'utilisation Terms of use	© Tous droits réservés / All rights reserved
Référence Citation	Bui, H. T. (2006). High-speed CMOS design techniques for multi-gigahertz transceivers [Thèse de doctorat, École Polytechnique de Montréal]. PolyPublie. https://publications.polymtl.ca/7740/

Détails de la thèse
Thesis details

Directeurs de recherche Advisors	Yvon Savaria
Programme Program	Non spécifié

Ce fichier a été téléchargé à partir de PolyPublie, le dépôt institutionnel de Polytechnique Montréal
This file has been downloaded from PolyPublie, the institutional repository of Polytechnique Montréal
https://publications.polymtl.ca

UNIVERSITÉ DE MONTRÉAL

HIGH-SPEED CMOS DESIGN TECHNIQUES FOR MULTI-GIGAHERTZ TRANSCEIVERS

HUNG TIEN BUI

DÉPARTEMENT DE GÉNIE ÉLECTRIQUE

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION

DU DIPLÔME DE PHILOSOPHIAE DOCTOR (PH.D.)

(GÉNIE ÉLECTRIQUE)

FÉVRIER 2006

© Hung Tien Bui, 2006

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-17973-4
Our file Notre référence
ISBN: 978-0-494-17973-4

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée:

HIGH-SPEED CMOS DESIGN TECHNIQUES FOR MULTI-GIGAHERTZ TRANSCEIVERS

présentée par: BUI Hung Tien

en vue de l'obtention du diplôme de: Philosophiae Doctor

a été dûment acceptée par le jury d'examen constitué de:

M. SAWAN Mohamad, Ph.D., président

M. SAVARIA Yvon, Ph.D., membre et directeur de recherche

M. AUDET Yves, Ph.D., membre

M. AL-KHALILI Asim, Ph.D., membre externe

Dedications

To the memory of my grand-parents.

To my family.

Acknowledgments

First and foremost, I would like to thank my research supervisor Dr. Yvon Savaria who supported me technically, morally and financially through this process. He took me under his wing and showed me a whole new level of excellence. His guidance has helped me evolve both as a researcher and as a person. He is truly an inspirational man.

I would also like to thank Dr. Mohamad Sawan, Dr. Yves Audet and Dr. Asim Al-Khalili for taking time off their busy schedule to be part of my jury. Each of these people has had significant accomplishments and I am honored to have them on my committee.

This thesis would not have been completed in time had it not been for the participation of several key contributors. A heartfelt "thank you" goes out to my proof-reading team that consists of the following people (in alphabetical order): Andria Alter, Robert Grou-Szabo, Bill Pontikakis, Bruno Tanguay and Guillaume Wild.

My research was supported financially by FQRNT and NSERC. The CAD tools and technologies used to complete the project were supplied by the CMC. Their contributions are much appreciated.

The years I spent at the GRM were made so pleasant due to the great office support from Mrs. Ghyslaine Carrier and to the flawless technical support provided by Mr. Rejean Lepage. I also want to thank everyone else at the GRM for creating such a great working environment.

Finally, where would I be without my family? I owe everything to my parents who have supported me in all imaginable ways while I was busy completing my degree. I also

want to thank my brother and my sister-in-law who have provided some much-needed distractions and who have helped me stay grounded. A final "thank you" goes out to my niece Minh Nhi, who always knew how to put a smile on my face.

Thank you all. None of this would have been possible without you.

Résumé

La croissance continuelle de l'Internet a provoqué une demande grandissante pour la bande passante. Cette demande a amené le développement de nouvelles normes en télécommunication telles que le OC-192 et le OC-768 qui spécifient des vitesses de transmission de données allant jusqu'à 10 Gb/s (OC-192) et 40 Gb/s (OC-768). Pour pouvoir mettre à profit ces liens, il est important de développer des circuits pouvant fonctionner à ces vitesses.

Le premier problème rencontré dans la conception de ces circuits est celui de la bande passante. Pour augmenter celle-ci, plusieurs utilisent la technique de la pointe-inductive avec des inductances en spirales. Cependant, les inductances en spirale ont une grande superficie et ne sont pas facile à optimiser. Pour résoudre ce problème, nous proposons d'utiliser des inductances actives. En simulation, avec les inductances actives, il était possible d'augmenter la bande passante des portes logiques jusqu'à 17%. Le défaut majeur des inductances actives est la sensibilité aux variations de procédés et de température. Pour pallier ce problème, nous proposons une nouvelle structure d'inductances et un système de mesure et de compensation de ces variations. Pour tester la structure de mesure de variations de procédés et de température, nous avons conçu et fait fabriquer une puce. Les mesures expérimentales démontrent que le système fonctionne tel que spécifié.

Bien que certaines portes logiques CMOS en mode courant soient capables de fonctionner à des vitesses de plusieurs GHz, elles ont très souvent la caractéristique d'être

asymétriques. Ceci est d'autant plus important que plusieurs transcepteurs utilisent des portes XOR pour aligner le signal d'horloge aux données. Cette thèse propose trois approches pour concevoir des structures symétriques qui éliminent théoriquement la gigue en sortie. En utilisant ces solutions, la gigue peut être réduite jusqu'à 95% à basse fréquence et 75% à une vitesse de 10 Gb/s. Pour confirmer les résultats de simulation, une puce a été conçue et fabriquée. Les mesures démontrent clairement que la structure fabriquée est en fait symétrique.

Nous savons que les détecteurs de phases ne sont pas capables de distinguer des différences de fréquences en entrée. Il est donc attendu que, lorsque la différence en fréquence entre les données et l'horloge est trop élevée, le transcepteur ne soit pas capable de converger. Une technique pour aider la convergence est d'utiliser une boucle d'asservissement de fréquence. Pour ce faire, nous proposons un nouvel algorithme et des nouveaux circuits pour créer un convertisseur fréquence-tension. Avec ce convertisseur, nous avons conçu une boucle d'asservissement de fréquence qui fonctionne à 5 GHz en simulation. Le circuit a aussi été implanté sur une puce et les résultats de mesures montrent que le circuit opère bien à des vitesses de jusqu'à 3.65 GHz. Une étude plus poussée révèle que cette différence en fréquences est probablement due aux variations de procédés.

Une technique populaire pour la détection de phases utilise une horloge à demie fréquence. Avec cette technique, pour bien récupérer les données, il faut que le rapport cyclique de l'horloge soit de 50%. Pour ce faire, nous proposons une nouvelle boucle de contrôle de rapport cyclique. Ce circuit, qui se sert du convertisseur fréquence-tension,

transforme un signal déformé en un signal d'horloge à rapport cyclique de 50%. Même à une vitesse de 5 GHz le circuit a été capable de transformer une horloge ayant un rapport cyclique de 25% et un déphasage de 49 ps en un signal d'horloge utilisable.

Les techniques proposées servent à améliorer différents aspects des transcepteurs de haute vitesse. En utilisant ces méthodes, il est possible d'améliorer la performance des transcepteurs sans toutefois être exposé aux difficultés associées aux approches conventionnelles.

Abstract

With the development of standards such as OC-192 and OC-768, the design of transceivers has become increasingly difficult. Designers are faced with the task of making circuits operate at speeds of multiple GHz which is not a trivial task. Numerous factors are responsible for the difficulty in designing high speed transceivers.

The main problem with multi-GHz design is the bandwidth. In an effort to resolve the bandwidth limitation issue, this thesis proposes the use of active inductors in the design of shunt-peaked MOS current-mode logic (MCML) gates. Conventionally, shunt-peaking is done with spiral inductors. However, they tend to be very large and optimizing shunt-peaked circuits can be long and difficult. For bandwidth improvement without much area penalty and design effort, we propose the use of active inductors. Results show that it can improve the performance of the MCML gates by up to 17%. One drawback to this approach is that active inductors are sensitive to process and temperature variations. To solve this problem, a new active inductor topology is proposed to allow the inductance value to be changed. The required changes in inductance are determined by a new method of measuring process and temperature variations on chip. This measurement technique has been implemented on a chip and has shown good measured results.

At high speeds, recovered signal quality becomes an issue. The second contribution of this thesis addresses this issue by proposing techniques for the design of high speed and low jitter XOR gates. XOR gates are crucial in the design of many clock and data recovery circuits as they are used to measure the precise phase alignment of the clock.

The problem is that the conventional XOR gate is asymmetric and generates output jitter. To resolve this, three design techniques have been suggested for the design of high speed symmetric XOR gates. Results show that, at low speeds, the jitter could be reduced from 26% to 95%, while at speeds of 10 Gb/s, the jitter can be reduced by up to 75%. A chip containing one of the proposed topology was made to demonstrate its functionality. Measurement results show that the implemented device is symmetric and works as described.

The last contributions of this thesis revolve around a novel frequency-to-voltage converter. Frequency-to-voltage converters can be used in transceivers for functions such as frequency acquisition and clock duty cycle adjustment. To implement this converter, a novel algorithm is proposed as well as new circuits techniques. Simulation results show that the resulting circuit is able to operate at speeds of 5 GHz.

Using this converter, a novel pulse width control loop has been designed. This circuit is useful in half-rate phase detectors which are becoming increasingly popular. Since these phase detectors use both edges of a clock, they require that the clock signal have 50% duty cycle. While the oscillator can be designed to have this characteristic, process variations and layout mismatches can change the signal's symmetry. To restore 50% duty cycle, a new differential pulse width control loop has been designed and simulations show that it can operate up to 5 GHz.

To improve the frequency acquisition range of a transceiver, a frequency locked loop based on this converter has been designed and fabricated. Simulation results show that it can operate at 5GHz. After fabrication, however, measurements show that the

implemented frequency locked loop only operates reliably up to 3.65 GHz. Further investigations indicated that this difference in expected and measured frequencies is probably due to process variations.

The design techniques proposed throughout this thesis help increase the bandwidth, improve the quality of the recovered signals and proposes new approaches to designing multi-GHz transceivers. It is the author's belief that these techniques will help design high speed circuits with less silicon area, better signal quality and with less design effort.

Condensé en Français

La croissance continuelle de l'Internet a provoqué une demande grandissante pour la bande passante. Cette demande a amené le développement de nouvelles normes en télécommunication telles que le OC-192 et le OC-768 qui spécifient des vitesses de transmission de données allant jusqu'à 10 Gb/s (OC-192) et 40 Gb/s (OC-768). Pour pouvoir mettre à profit ces liens, il est important de développer des circuits pouvant fonctionner à ces vitesses. La conception de ces circuits a traditionnellement été faite avec des technologies telles que le GaAs et le SiGe, qui permettent d'obtenir des transistors plus rapides. Cependant, ces technologies sont coûteuses et consomment beaucoup de puissance.

Dans le passé, la technologie CMOS n'était pas communément utilisée dans la conception de circuits de haute performance puisqu'elle était considérée comme étant trop lente. Cependant, avec la réduction à l'échelle, il est désormais possible d'atteindre des vitesses de plusieurs GHz en technologie CMOS 0.18µm. Les chercheurs s'intéressent à la possibilité de réaliser des transcepteurs en technologie CMOS, puisque celle-ci coûte moins cher, consomme moins de puissance et peut s'intégrer aux autres circuits CMOS sur une même puce.

Un lien de communication comprend trois parties importantes: un transmetteur, un canal et un récepteur. Le transmetteur envoie des bits de données à travers un canal et ces bits sont ensuite récupérés par le récepteur. La transmission et la réception doivent se faire à la même vitesse. De plus, l'échantillonnage des données doit être faite au milieu

du bit pour réduire les possibilités d'erreurs. Pour ce faire, les systèmes de réception doivent utiliser un circuit de recouvrement d'horloge et de données (RHD).

Le circuit RHD, dont le diagramme bloc est présenté à la Figure 0.1, est un des éléments qui posent les plus grand défi dans la conception de transcepteurs de haute performance. La raison est que ce circuit doit être précis et doit également être en mesure d'opérer à des vitesses de plusieurs GHz.

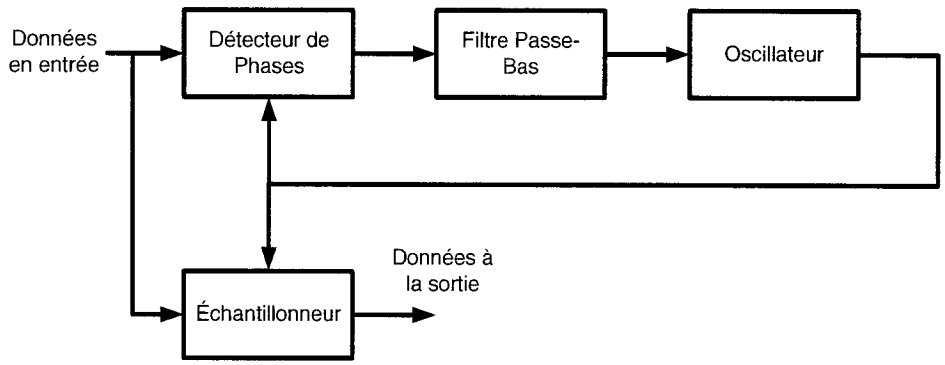


Figure 0.1. Diagramme bloc d'un circuit RHD

Le détecteur de phase est utilisé pour mesurer la différence entre la position d'échantillonnage optimale et la position d'échantillonnage actuelle. Ce signal est filtré pour éliminer le bruit de haute fréquence avant d'être envoyé à l'oscillateur. En réponse au signal de contrôle, ce dernier déplace la phase de l'horloge dans la direction qui permet à la boucle de converger vers l'échantillonnage optimal.

Il y a plusieurs problèmes liés à la conception de circuits RHD. Le premier problème est celui de la bande passante. Les techniques de conception conventionnelles sont

poussées au-delà des limites faisables avec une technologie donnée. Quand un système opère à une vitesse de quelques GHz, la qualité des signaux devient un autre problème. Plusieurs circuits importants ont des propriétés asymétriques. Sachant que l'asymétrie est une propriété indésirable et que la qualité des résultats est directement liée à ces circuits, il est important de les améliorer. Un troisième problème est celui de l'acquisition de fréquence. Les détecteurs de phase sont normalement incapables de signaler une différence de fréquences. Or, si les fréquences des deux entrées sont trop éloignées, le RHD peut perdre la capacité de converger. Ce problème, de même que les autres problèmes mentionnés ci-dessus, sera traité tout au long de cette thèse.

Lorsque la vitesse d'opération atteint quelques GHz, la logique complémentaire CMOS n'est plus en mesure de fonctionner. Le temps requis pour effectuer des transitions entre VDD et VSS devient trop long. Pour les circuits de haute vitesse, les concepteurs adoptent une technique appelée la logique en mode courant CMOS (LMCC). La technique est basée sur des paires de transistors connectées à la source qui commutent le courant d'une branche à l'autre. Le courant passe presque exclusivement dans une des branches quand le signal différentiel à l'entrée est plus élevé que :

$$\Delta V = \sqrt{\frac{2ISS}{\mu_N C_{OX}(W/L)}} \qquad (0.1)$$

Quand le courant passe dans une branche, une chute de tension égale à $ISS*R_L$ est observée à la sortie. L'autre sortie, qui théoriquement ne devrait être traversée par aucun courant, voit son potentiel s'élever vers VDD par l'intermédiaire de la résistance. En ajustant la valeur du courant et de la résistance, il est donc possible de régler le niveau de

l'excursion de tension en sortie. Avec une plus petite transition, il est possible d'atteindre des vitesses plus élevées.

Cette technique peut être généralisée pour concevoir n'importe quelle porte logique complexe. Celles-ci peuvent être conçues en empilant plusieurs paires de transistors les unes sur les autres.

Inductances actives

La technique LMCC trouve ses limites avant de satisfaire les besoins des applications courantes quand elle est utilisée seule dans le but de concevoir des transcepteurs à haute vitesse avec une technologie populaire. Par exemple, ses limites pratiques se trouvent à quelques GHz en technologie CMOS 0.18µm. En 1996, Mohan et al. ont proposé une méthode pour augmenter la bande passante dans les amplificateurs. Cette technique, appelée la pointe-inductive, consiste à placer une inductance en parallèle avec la sortie. Cette approche peut, en théorie, augmenter la bande passante jusqu'à 85%. Récemment, la pointe-inductive a aussi été appliquée à la conception de portes LMCC de haute performance.

Bien que cette technique offre des améliorations significatives, elle a ses défauts. À l'intérieur des puces, les inductances sont fabriquées en forme de spirales tel qu'illustré à la Figure 0.2.

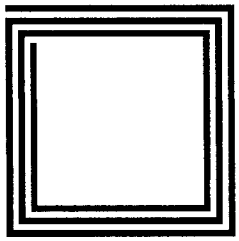


Figure 0.2. Inductance en Spirale

La taille de ces spirales est rarement plus petite que 100µm par 100µm. De plus, l'espace autour de ces inductances doit être libéré afin de limiter les risques d'interférence électromagnétique. Un autre problème lié aux inductances en spirales est la difficulté avec laquelle on conçoit les circuits comportant ces éléments. En effet, la conception d'inductances doit être faite dans un environnement spécialisé spécifique à la conception de circuits radiofréquences tandis que la conception de circuits intégrés doit être faite dans un environnement spécifique à cette tâche. Par conséquent, l'optimisation, qui est un processus itératif, peut s'avérer très longue.

Pour résoudre ces problèmes, nous proposons d'utiliser des inductances actives. Ces inductances, qui sont faites à partir d'au moins un transistor, imitent le fonctionnement des inductances pour un modèle linéaire petit-signal. Cependant, les améliorations en terme de vitesse apportées par l'utilisation d'inductances actives ne sont pas aussi importantes que celles offertes par les inductances en spirales. L'amélioration en performance varie d'une application à l'autre. Dans les portes LMCC, il était possible d'augmenter la bande passante jusqu'à 17%. De plus, le temps requis pour la conception tout comme la taille du circuit résultant sont réduits.

Un des grands défauts des inductances actives est leur sensibilité aux variations de procédés et de température. Cette sensibilité a pour effet de changer les caractéristiques des inductances actives lorsqu'il y a un changement au niveau de la température ou des procédés de fabrication. Pour résoudre ce problème, nous proposons un système de mesure et de compensation. Ce système a pour but de quantifier les changements de température et de procédés et d'ajuster les paramètres pour en annuler les effets. Afin de s'assurer que le système fonctionne, une partie du circuit a été conçue et fabriquée en technologie CMOS 0.18μm. Les résultats expérimentaux confirment que le système de mesure des variations de procédés et de températures se comporte bien.

Porte XOR symétrique

Bien que les portes logiques de type LMCC soient capables de fonctionner à des vitesses de plusieurs GHz, elles ont la caractéristique d'être asymétriques. Ceci est d'autant plus important que plusieurs circuits RHD utilisent des portes XOR de type LMCC pour aligner le signal d'horloge aux données. Lorsqu'une gigue est présente à la sortie de la porte XOR, la qualité du signal d'horloge est réduite. Puisque la qualité des données récupérées est déterminée par la qualité du circuit RHD, il est important de résoudre le problème d'asymétrie lié aux portes XOR LMCC.

Une analyse théorique du premier ordre peut aider à trouver les tailles des transistors qui permettent de concevoir une porte XOR sans gigue. Cependant, les tailles qui éliminent la gigue ne sont pas nécessairement les tailles qui augmentent la vitesse de ces

portes. Si seul l'ajustement de la taille des transistors est fait, il n'est généralement pas possible d'optimiser la vitesse et la gigue en même temps.

Une des techniques proposée est d'ajuster les circuits qui contrôlent les entrées en plus d'ajuster la taille des transistors du XOR LMCC. De cette façon, il y a plus de degrés de liberté et l'optimisation en terme de gigue et de vitesse peut se faire indépendamment.

Un des problèmes avec cette approche est que la performance de ces portes est étroitement liée aux modèles du premier ordre et à la précision des procédés de fabrication. Il y a plusieurs effets qui ne sont pas inclus dans le modèle et qui peuvent affecter la performance de la porte XOR. De plus, même si le XOR est conçu pour n'avoir aucune gigue, les variations de procédés peuvent dégrader la performance de ces portes.

L'approche favorisée est celle qui utilise des structures symétriques. De cette façon, même les effets qui ne sont pas modélisés sont compensés. Nous proposons quatre circuits XOR symétriques qui éliminent théoriquement la gigue en sortie. En utilisant ces solutions, la gigue peut être réduite jusqu'à 95% à basse fréquence et 75% à une vitesse de 10 Gb/s.

Pour prouver l'efficacité des portes symétriques, un des circuits proposés a été fabriqué en technologie CMOS 0.18µm. Les résultats expérimentaux indiquent que le circuit fonctionne bien et qu'il a en fait des caractéristiques symétriques.

Convertisseur fréquence-tension

Nous savons que les détecteurs de phase ne sont pas capables de distinguer des différences de fréquences en entrée. Il est donc attendu que, lorsque la différence en fréquence entre les données et l'horloge est trop élevée, le circuit RHD ne soit pas capable de converger. Pour résoudre ce problème, nous avons recours à une boucle de verrouillage en fréquences (BVF). Cette boucle sert à rapprocher la fréquence de l'horloge de la fréquence des données. Lorsque la différence en fréquence est assez basse, le BVF se désactive et le RHD prend le contrôle de la boucle.

Pour réaliser le circuit BVF, on a recours à un circuit de conversion fréquence-tension. Un des circuits populaires de ce genre a été proposé par Djemouai et al. Ce circuit opère très bien à des vitesses modérées de l'ordre de centaines de MHz. Pour des applications dans les GHz, ce circuit a des caractéristiques qui l'empêchent de bien fonctionner.

Le premier problème est associé à l'usage de signaux de pleine amplitude (VSS à VDD). Nous avons vu précédemment que cette technique n'est pas recommandée pour des circuits de haute vitesse. Le deuxième problème est lié à l'utilisation de pulsations pour gérer les phases d'opération du circuit (*S1*, *S2* et *S3*). Le circuit utilise un signal oscillatoire en entrée pour générer ces pulsations. La première demie période du signal en entrée sert a générer la phase *S1*. Les phases *S2* et *S3* sont générées durant la deuxième demie période. Les phases *S2* et *S3* doivent être assez longues pour transférer des charges et doivent être de pleine amplitude. Puisque ces phases doivent en plus être

générées de façon mutuellement exclusive, il s'en suit que la période minimale supportée par ce circuit doit être plus grande que la durée de quatre pulsations de durée minimale. Ceci devient une contrainte importante lorsqu'un circuit doit opérer à plusieurs GHz.

Pour régler ces problèmes, nous proposons un nouvel algorithme de fonctionnement et de nouveaux circuits. Le nouvel algorithme proposé s'exécute en deux phases et peut être contrôlé directement par le signal en entrée sans l'intermédiaire de pulsations générées à l'interne. De plus, les techniques proposées au niveau circuit permettent au convertisseur fréquence-tension de fonctionner à des vitesses de 5 GHz. Le convertisseur a été utilisé pour créer une BVF qui opère à 5 GHz en simulation. Quand cette même BVF a été fabriquée en technologie CMOS 0.18µm, les résultats expérimentaux montrent qu'elle peut fonctionner jusqu'à 3.65GHz. Des tests additionnels ont montré que la différence en performance était probablement due aux variations de procédés.

Un des problèmes rencontré dans la conception de transcepteurs à haute vitesse est la conception d'un oscillateur. À haute vitesse, il n'est pas évident de parvenir à cumuler toutes les caractéristiques désirables telles que la gamme dynamique de fréquence, la stabilité de l'oscillation, etc. Pour cette raison, le concept de RHD à demie fréquence a été introduit. Les RHD conventionnels utilisent une horloge rapide qui n'a qu'un seul front actif. Dans le cas du RHD à demie fréquence, les deux fronts de l'horloge sont mis à profit. Pour s'assurer de bien récupérer les données, il faut que le rapport cyclique de l'horloge soit de 50%. Pour ce faire, une nouvelle boucle de contrôle de rapport cyclique a été proposée. Ce circuit se sert du convertisseur fréquence-tension pour transformer un signal déformé en un signal d'horloge à rapport cyclique de 50%. Même à une vitesse de

5 GHz, le circuit a été capable de transformer une horloge ayant un rapport cyclique de 25% et un déphasage de 49 ps en un signal d'horloge utilisable. Dans ce cas extrême, le signal résultant est une horloge avec un rapport cyclique de 48% qui a une légère disparité dans le niveau DC de ses composantes différentielles. Ce problème peut être réglé simplement en ajoutant des étages de gain pour restituer le signal.

Les techniques proposées tout au long de cette thèse servent à améliorer différents aspects des transcepteurs de haute vitesse. Puisque ces techniques sont orthogonales, elles peuvent être combinées. En utilisant ces méthodes, il est possible d'améliorer la performance des transcepteurs sans toutefois être exposé aux difficultés associées aux approches conventionnelles.

Table of Contents

DEDICATIONS....iv

ACKNOWLEDGMENTS....v

RÉSUMÉ....vii

ABSTRACT....x

CONDENSÉ EN FRANÇAIS....xiii

TABLE OF CONTENTS....xxiii

LIST OF FIGURES....xxvii

LIST OF TABLES....xxxii

LIST OF ABBREVIATIONS....xxxiii

CHAPTER 1: INTRODUCTION....1

1.1. CONTEXT....2

1.2. ORGANIZATION....4

1.3. CONTRIBUTIONS....5

CHAPTER 2: BACKGROUND....8

2.1. INTERCONNECTIONS....9

2.2. CLOCK AND DATA RECOVERY....12

2.3. FREQUENCY ACQUISITION....19

2.4. MOS CURRENT-MODE LOGIC....21

2.4.1. SYMMETRIC XOR....24

2.5. BANDWIDTH IMPROVEMENT TO MCML GATES....26

2.6. SUMMARY....34

CHAPTER 3: BANDWIDTH ENHANCEMENT USING ACTIVE INDUCTIVE LOADS IN MCML GATES....35

3.1. THEORETICAL ANALYSIS OF ACTIVE INDUCTORS....37

3.2. SIMULATION RESULTS....44

3.3. PROCESS VARIATIONS....47

3.3.1. PROCESS/TEMPERATURE VARIATIONS AND MEASUREMENT....47

3.3.2. ACTIVE INDUCTOR COMPENSATION....52

3.3.3. SIMULATION RESULTS AND ANALYSIS....53

3.3.4. EXPERIMENTAL RESULTS....55

3.4. CONCLUSIONS....59

CHAPTER 4: HIGH-SPEED AND LOW JITTER DESIGN TECHNIQUES FOR MCML XOR GATES....61

4.1. JITTER ANALYSIS OF THE MCML XOR GATE....61

4.1.1. UNEVEN DRIVE STRENGTH....63

4.1.2. UNEVEN INPUT LOADING....66

4.1.3. GLITCHES....68

4.1.4. TOTAL DETERMINISTIC JITTER....72

4.1.5. JITTER COMPENSATION THROUGH SIZING....72

4.1.6. JITTER COMPENSATION THROUGH CONTROLLED IMBALANCED DRIVERS....73
4.2. SYMMETRIC GATES....74
4.2.1. CRISS-CROSSED XOR GATES....75
4.2.2. 4-BRANCH SYMMETRIC XOR....77
4.2.3. 4-BRANCH SYMMETRIC XOR WITH RESET....80
4.3. SIMULATION RESULTS....83
4.4. EXPERIMENTAL RESULTS....90
4.5. CONCLUSIONS....93
CHAPTER 5: DESIGN OF A HIGH-SPEED DIFFERENTIAL FREQUENCY-TO-VOLTAGE CONVERTER AND ITS APPLICATIONS....95
5.1. SINGLE-PHASE ARCHITECTURE....97
5.2. CIRCUIT-LEVEL IMPLEMENTATION....104
5.3. SAMPLE APPLICATION: PULSE WIDTH CONTROL LOOP....109
5.3.1. PROPOSED PWCL....112
5.3.1.1 DUTY CYCLE ADJUST....114
5.3.1.2 FVC....115
5.3.2. SIMULATION RESULTS....116
5.4. FREQUENCY LOCKED LOOP: A Second Sample Application....119
5.4.1. ANALYTICAL MODEL....120
5.4.2. BUILDING BLOCKS....127
5.4.2.1. ICO....127

5.4.2.2. COMPARATOR....128

5.4.2.3. V/I....128

5.4.2.4. DIVIDE-BY-64....129

5.4.3. TOP-LEVEL SIMULATIONS AFTER EXTRACTED LAYOUT....130

5.4.4. EXPERIMENTAL RESULTS....131

5.5. CONCLUSIONS....135

CHAPTER 6: CONCLUSIONS....137

6.1. FUTURE WORK....139

REFERENCES....141

List of Figures

Figure 0.1. Diagramme Bloc d'un RHD.....xiv.

Figure 0.2. Inductance en Spirale.....xvi

Figure 1.1. Typical Communications Link.....2

Figure 2.1. Sample Eye Diagram Showing Non-Ideal Sampling Instants.....11

Figure 2.2. Block Diagram of a CDR.....13

Figure 2.3. Determination of Phase Relations in 2X Oversampling.....14

Figure 2.4. Hogge's Phase Detector.....15

Figure 2.5. Waveforms Showing Different Clocking Schemes.....16

Figure 2.6. Savoj's Half-Rate Phase Detector.....17

Figure 2.7. Example of 4-Phase Time-Interleaving CDR.....18

Figure 2.8. Ideal Model of Djemouai's Frequency-to-Voltage Converter.....19

Figure 2.9. Input Clock Showing the Different Phases for Frequency-to-Voltage Conversion.....20

Figure 2.10. MCML Buffer.....22

Figure 2.11. MCML XOR Gate.....23

Figure 2.12. Single Ended Symmetric XOR.....25

Figure 2.13. Buffer with Compensated Miller C_{GD}.....27

Figure 2.14. Differential f_t Doubler.....28

Figure 2.15. Buffer with Negative Feedback.....29

Figure 2.16. Representation of a Spiral Inductor.....30

Figure 2.17. Shunt-Peaked Buffer.....30

Figure 2.18. Lumped Model of a Spiral Inductor.....32

Figure 2.19. Active Inductors.....33

Figure 3.1. Active Inductors.....36

Figure 3.2. MCML Buffers Loaded with Active Inductors.....36

Figure 3.3. Small Signal Model of the Buffer Loaded with an NMOS Active Inductor.....37

Figure 3.4. a) Bode Plot and b) Transient Step Response for Different Pole-Zero Positions.....41

Figure 3.5. Simulation Results for a) the Output of a Buffer with Active Shunt-Peaking and b) the Corresponding Voltage at the Gate of the Active Inductor's Transistor.....43

Figure 3.6. Simulation Results for the Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked Buffers.....45

Figure 3.7. Simulation Results for the Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked AND Gates.....46

Figure 3.8. Simulation Results for the Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked XOR Gates.....46

Figure 3.9. a) N-Type and b) P-Type Process and Temperature Measurement Circuits...48

Figure 3.10. I_D vs. V_{GS} for Different Temperatures.....50

Figure 3.11. Measurement Output for Different Temperatures in a Typical Process.....52

Figure 3.12. Adjustable Active Inductor.....53

Figure 3.13. Frequency Response of Active Shunt-Peaked Buffer in Presence of Process Variations....54

Figure 3.14. Frequency Response of Active Shunt-Peaked Buffer after Process Variation Compensation....55

Figure 3.15. Process and Temperature Measurement Circuit and Test Circuit....56

Figure 3.16. Sample Output of the Test Circuit....57

Figure 3.17. Output Voltage as a Function of Oscillator Frequency....58

Figure 4.1. MCML XOR Gate....62

Figure 4.2. MCML XOR Gate Output Branch....64

Figure 4.3. Small-Signal Model for XOR Gate with Transition on a) VIN1 and b) VIN2....64

Figure 4.4. Possible Change in Current Path Causing Glitches....69

Figure 4.5. Jitter Measurement....69

Figure 4.6. Gate Level Criss-Crossed Symmetric XOR....75

Figure 4.7. Criss-Crossed Symmetric XOR Gate....76

Figure 4.8. Small-Signal Model for the Criss-Crossed Symmetric XOR Gate....76

Figure 4.9. 4-Branch Symmetric XOR Gate....78

Figure 4.10. Single Branch of the 4-Branch Symmetric XOR Gate....78

Figure 4.11. Symmetric XOR Model....79

Figure 4.12. Intermediate Nodes of the 4-Branch Symmetric XOR Gate with Reset when Inputs are a) $VIN1 - VIN2 = 11 \rightarrow 00 \rightarrow 11$ and b) $VIN1 - VIN2 = 01 \rightarrow 00 \rightarrow 11$....81

Figure 4.13. 4-Branch Symmetric XOR Gate with Reset....82

Figure 4.14. Voltage vs. Time Eye Diagram for a (π/2) Phase Offset Periodic Input 85

Figure 4.15. Voltage vs. Time Eye Diagram for a Random Input.....87

Figure 4.16. Jitter as a Function of Bit Period.....89

Figure 4.17. Micrograph of the 4-Branch Symmetric XOR Gate.....90

Figure 4.18. Experimental Setup for the XOR Gate.....91

Figure 4.19. Chip Output in Response to π/2 Phase Offset Square Waves.....92

Figure 4.20. DC Measurements of the 4-Branch Symmetric XOR Gate.....93

Figure 5.1. Counter-Based Conversion.....95

Figure 5.2. Integrator-Based Conversion.....96

Figure 5.3. Model of Single Phase Algorithm.....97

Figure 5.4. Steps in Single Phase Operation.....98

Figure 5.5. Ideal Simulation Waveform of Single Phase Algorithm.....99

Figure 5.6. MATLAB Simulation with Two Different Values of C2.....103

Figure 5.7. Charging/Discharging Switch.....105

Figure 5.8. Charge Transfer Switch.....106

Figure 5.9. Valid Input Voltages.....107

Figure 5.10. Frequency-to-Voltage Conversion at 5 GHz.....108

Figure 5.11. Outputs when Input Signal Periods are 199, 200 and 201 ps.....108

Figure 5.12. Conventional PWCL.....111

Figure 5.13. Block Diagram of PWCL.....113

Figure 5.14. Ideal Waveform for PWCL.....113

Figure 5.15. Differential Duty Cycle Adjust.....114

Figure 5.16. Input Signal with Uneven Loading Condition....117

Figure 5.17. Output Signal with Uneven Loading Condition....117

Figure 5.18. Input Signal with 25% Duty Cycle and 49ps Skew....118

Figure 5.19. Output Signal when Input has 25% Duty Cycle and 49ps Skew....119

Figure 5.20. Block Diagram of FLL....120

Figure 5.21. Linearized Model of a PLL....121

Figure 5.22. Block Diagram of FLL with Transfer Function of Each Block....122

Figure 5.23. Root Locus Plot of FLL....126

Figure 5.24. Biasing Scheme and Single Delay Element....127

Figure 5.25. Fully Differential Two Stage Amplifier....128

Figure 5.26. V/I Circuit....129

Figure 5.27. Full Speed Oscillators Before Lock....130

Figure 5.28. Full Speed Oscillators after Lock....131

Figure 5.29. Chip Micrograph of the FLL....132

Figure 5.30. Experimental Setup for FLL....133

Figure 5.31. Output Signals Before Lock....134

Figure 5.32. Output Signals After Lock....134

List of Tables

Table 3.1. Summary of Different Design Parameters for NMOS Active Inductors.....42

Table 3.2. Summary of Performance.....46

Table 3.3. Measurement Output for Process Corners.....49

Table 3.4. Measurements of Process Variation Chips.....57

Table 4.1. XOR Truth Table.....63

Table 4.2. Description of Simulated XOR Gates.....84

Table 4.3. Eye Diagram Measurements for (π/2) Phase Offset Input (100ps output bit time).....86

Table 4.4. Eye Diagram Measurements for Random Input.....88

List of Abbreviations

Abbreviations

AHDL	Analog Hardware Description Language
CDR	Clock and Data Recovery
CMOS	Complementary MOS
CQFP	Ceramic Quad-Flat Package
DC	Direct Current
DDR	Double Data Rate
ECL	Emitter-Coupled Logic
FLL	Frequency Locked Loop
f_t	Frequency at which current gain of a transistor is 1 (Transition Frequency)
FVC	Frequency to Voltage Converter
GaAs	Gallium Arsenide
IC	Integrated Circuits
ICO	Current Controlled Oscillator
L	Channel Length
MCML	MOS Current Mode Logic
MOS	Metal Oxide Semiconductor
NMOS	MOS transistor having electrons as majority carriers (N-Type)
PCB	Printed Circuit Board

PLL	Phase Locked Loop
PMOS	MOS transistor having holes as majority carriers (P-Type)
PWCL	Pulse Width Control Loop
RAM	Random Access Memory
SCL	Source-Coupled Logic
S/H	Sample and Hold
SiGe	Silicon Germanium
VCO	Voltage Controlled Oscillator
VDD	Power Supply Voltage
V/I	Voltage to Current Converter
VSS	Ground Voltage
VTH	Threshold Voltage of a Transistor
W	Channel Width

Chapter 1

Introduction

The increasing demand for high speed communications has led to the development of multi-GHz transceivers. Modern optical transmission standards such as OC-192 and OC-768 specify links operating at speeds of 10 Gb/s and 40 Gb/s respectively. To implement these links, transceiver circuitry must be able to support these data rates.

Multi-GHz transceivers have traditionally been designed in technologies such as GaAs and SiGe, because of their fast intrinsic device speed. Within the last few decades, CMOS technology has improved tremendously and has been used successfully in the design of high speed circuits. There are many benefits to using CMOS in the design of transceivers, including lower power consumption and reduced cost. In addition, CMOS is a mature technology for which many tools and design techniques have been developed. Due to these characteristics, there is widespread interest in CMOS technology which has been driving the aggressive transistor scaling and performance improvements seen over the years. Since CMOS provides higher levels of integration, lower cost and lower power consumption, it is an appealing solution for the design of high speed transceivers.

The goal of this research is to provide techniques for improving the performance of high speed interface circuits in CMOS technology. As much as possible, the developed techniques aim to be orthogonal, in that they can be combined to achieve even better

performance. The contributions of this thesis are made at the algorithmic, architectural and at the circuit level.

1.1. Context

The work presented in this thesis is focused on ways of improving communications links through the design of better transceivers. A communications link primarily consists of three parts: a transmitter, a channel and a receiver (Figure 1.1). Data from the transmitter side are sent as a series of bits through a channel. These bits are then recovered at the receiver end.

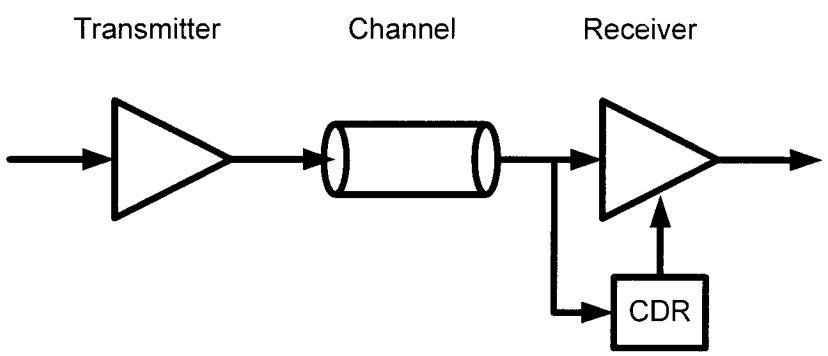


Figure 1.1. Typical Communications Link

The transmitter sends data at a given frequency through the channel and these data need to be recovered at the same speed. In addition to having the right frequency, there are also instances during the period where data recovery is less prone to errors. It is

highly desirable to sample data at those specific moments. Consequently, receivers need to have a clock that oscillates with the right frequency and also needs to have the right phase.

Receivers typically have a local oscillator that operates at about the same frequency as the transmitter's clock. However, the required phase information is not known. To solve this problem, some systems use a source-synchronous approach, where a clock signal is transmitted along with the data. The transmitted clock is then used to sample the data. This technique greatly facilitates the design of a receiver. However, since source-synchronous approaches are not always possible, as in the case of optical links, receivers often require a clock and data recovery circuit (CDR). CDRs typically use the transient characteristics of the incoming data to extract the phase information.

The CDR is perhaps the most challenging part of the transceiver. It needs to derive timing information from the incoming data which can arrive at rates of multi-Gb/s. The task becomes even more difficult because the channel attenuates and distorts the transmitted signals. While some solutions have been proposed to recover clock and data signals at 10 and 40 Gb/s [13][37][64], the circuits are large and these results are not easily reproducible. Integration of a number of these devices onto a single chip can be difficult and costly.

This thesis proposes a series of algorithmic, architectural and circuit-level techniques to provide more efficient solutions for high speed CDRs. As it will be shown throughout, these techniques help reduce silicon area, ease design efforts and improve the quality of the recovered signals.

1.2. Organization

To understand the contributions of this thesis, it is essential to explore the significant research that has been done in the field. A thorough literature survey is provided in Chapter 2. It explains the theoretical background required to understand the remaining parts of the thesis. It describes CDRs and how their different components are pieced together. It also examines the logic style and design techniques that are typically used in these applications.

One of the most difficult aspects of transceiver design for OC-192 and OC-768 is making circuits that can respond to the specified data rates. The simplest way of making a transceiver operate faster is to increase its clock speed. Unfortunately, many circuits cannot operate reliably when the clock speed is increased too much due to large parasitic capacitances. This problem is addressed in Chapter 3, where a method of increasing the bandwidth of conventional logic gates is proposed. The technique allows for devices to operate faster with minimal area penalty and minimal design effort.

As will be shown in Chapter 2, device symmetry is a very important issue especially at high speeds. Typical CDR circuits use XOR gates to align the local clock to the incoming data. Since the XOR gate directly controls this alignment, it is important to have a well-behaved XOR gate that can operate at high speeds. Chapter 4 proposes several techniques to design these gates.

It is known that communications are facilitated when the clock frequency of the transmitter and receiver are the same. In practice, however, this is usually not the case,

thus the need for CDR circuits. When the difference between these frequencies is too large, typical CDRs will not be able to lock because they do not have enough frequency acquisition range. To address this problem, Chapter 5 introduces a differential high speed frequency-to-voltage converter (FVC). As will be shown, this FVC can be applied to the design of a frequency-locked loop (FLL) to increase the frequency acquisition range of a CDR.

An important class of CDRs uses a concept called half-rate clocks to facilitate the design of the local oscillator. These CDRs require a clock with 50% duty cycle. As will be shown in Chapter 5, the proposed FVC can also be used in the design of a pulse-width control loop (PWCL) to restore a 50% duty cycle clock from distorted signals.

Finally, Chapter 6 summarizes the results that have been brought forth in this thesis. In addition, a list of future research directions is proposed.

1.3. Contributions

The goal of this thesis is to help improve the design of multi-GHz transceivers. A number of contributions have been made throughout this research and they are summarized as follows:

1. Application of shunt-peaking to non-linear circuits. By using shunt-peaking with MCML gates, it was possible to design a half-rate clock and data recovery circuit operating at 20 Gb/s [5].

2. Implementation of shunt-peaking using active inductors. The use of active inductors in shunt-peaking helps improve the bandwidth of MCML gates with little penalty in terms of area and design effort [7].

3. Design and analysis of high speed and low jitter XOR gates. We provided a model for deterministic jitter in XOR gates and proposed several design techniques to reduce this jitter [6]. A test chip has also been successfully fabricated and tested in 0.18μm CMOS technology.

4. Design of a 5 GHz differential frequency-to-voltage circuit [8]. Frequency-to-voltage converters can have numerous applications in multi-GHz transceivers such as controlling the duty-cycle of clock signals [11]. A correct duty cycle is crucial in half-rate clock and data recovery circuits.

5. Design, fabrication and testing of a 5 GHz frequency-locked loop. The frequency-to-voltage converter proposed in [8] was also used to implement a 5 GHz frequency-locked loop in 0.18 μm CMOS technology.

6. Development of a process and temperature measurement and compensation technique [10]. The effects of process and temperature variations are important and need to be considered in the design of circuits. By using this method, fabricated circuits can have predictable and reliable performance despite process and temperature variations. The

measurement technique has been implemented and tested in 0.18μm CMOS technology with success.

7. Design of a 10 GHz PLL. Using some of the techniques previously described, a 10 GHz PLL has been designed [9].

Chapter 2

Background

While advances in silicon technology have led to a rapid increase in speed and scaling of current integrated circuits (IC), the bandwidth of interconnections between systems has not improved as much. For many applications, such as multi-processor and processor-memory interconnections, the bottleneck lies in the links. Even systems that can process data at multiple Gb/s may not be able to transmit or receive data at that same rate. While the issue can be partly addressed by improving the packaging or the channel, this thesis is mainly concerned with the design of electronic transceivers that are often the dominant bottleneck in high bandwidth interconnections.

The design of transceivers and links involves a number of architectural decisions. The interconnection could be point-to-point or in the form of a bus, serial or parallel. Such interconnections may be source-synchronous or have an embedded clock, and signals may carry a single-bit or multiple bits. These choices must be made carefully in order to maximize the performance of a given system within the bounds of the given specifications.

2.1. Interconnections

The bus paradigm is based on a communication medium that is shared by various modules. The transaction between these modules is scheduled by an arbiter to ensure that bus contention does not occur while maximizing the efficiency of the bus. The bus is very popular with systems-on-chip as it facilitates design-reuse. To make a design easily reusable for different applications, it is necessary to have a known interface. It is common to see intellectual property (IP) blocks being developed for a given bus type. Though the bus eases design reuse and facilitates the interconnection between modules, it does have its drawbacks. Since multiple modules are competing for bandwidth, bus latency is inevitable. In a bus with more than one master, a module will have to wait for a bus-grant from the arbiter before any type of transaction can be initiated. The second drawback lies in the fact that the bus is a shared medium. Since the bus is connected to multiple modules, the capacitance associated to the lines tends to be large. This could contribute to slowing down data transmission. Using a point-to-point link increases the attainable bandwidth by effectively removing the two main drawbacks of the bus. Given that the link only connects two devices, the load is reduced and arbitration becomes unnecessary, and it becomes easier to use well defined and terminated transmission lines.

To increase the bandwidth of communications links, it is common practice to have multiple bits transmitted through the medium. One way of transmitting multiple bits is to send them in parallel [67]. Most parallel links are source-synchronous, which means that

the clock signal is sent along with the parallel data word. Since the clock signal is provided, it easily allows the module at the receiving end to recover the data. At high-speeds, however, several design aspects need to be considered. When designing the parallel link, it is often difficult to match the transmission lines through which individual bits of the link are transmitted. If the electrical lengths are not matched, the arrival time of the individual bits could be different. In severe cases, this could lead to incorrect data transmission. To address some of the skew problems, several deskewing circuits have been proposed in the recent years [2]. In an effort to reduce the matching problem, parallel bits of a link are usually routed next to each other so that they would have approximately the same electrical length. However, having the bit lines in close proximity could cause cross-talk and inductive coupling between the bits, which could have unmodeled and unexpected pattern-dependent behavior. When designed carefully, one of the advantages for such source-synchronous systems is that they do not require complicated clock and data recovery (CDR) circuits.

An alternative to parallel links is the single wire serial link [38][80]. Using this method, the data are serialized at the transmitting end before being sent through the channel. The circuit at the receiving end deserializes the stream and recovers the data. The system is plesiochronous, which means that the transmitter and the receiver use separate local clocks that oscillate at the same frequency, but not with a predefined phase relationship.

Having a plesiochronous system also implies that the sampling edge of the receiver clock and the incoming data are not necessarily aligned. Without this proper alignment,

sampling of the data could be prone to errors. This can be seen in a sample data eye shown in Figure 2.1. For instance, if the sampling edge of the clock occurred close to a data transition, the recovered data could be incorrect or even metastable. In order to avoid errors, it is usually recommended to sample the data at the middle of the bit time. To find this middle, extra circuitry is required. The action of aligning the active transition of the local clock to the middle of the data bit is known as clock recovery. Despite the extra design requirements, serial links could inherently transfer data more rapidly. Since serial links do not have multiple wires to handle, it does not suffer from skew problems.

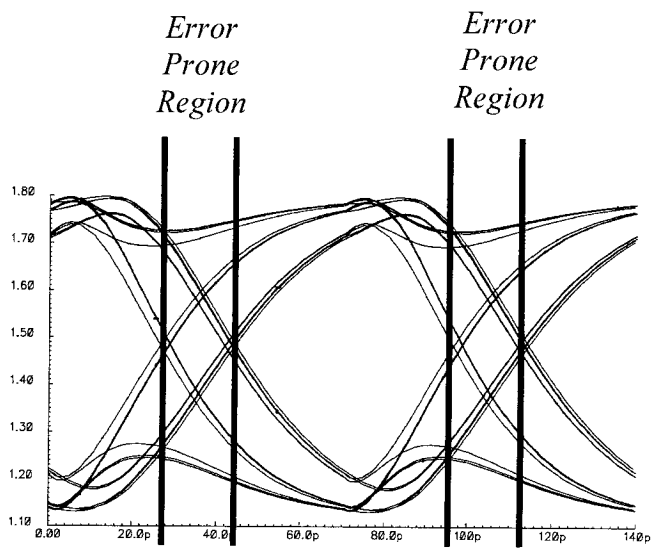


Figure 2.1. Sample Eye Diagram Showing Non-Ideal Sampling Instants

It is known that parallel links offer a way of sending multiple bits every clock cycle. In recent years, another technique has been proposed: multi-level signaling [22][72].

Static CMOS systems use two discrete levels of logic, which associate VDD to logic '1' and VSS to logic '0'. Multi-level signaling is a system that uses intermediate values between VDD and VSS to increase the amount of information per bit time. For example, PAM-4 divides the voltage range into equal sections so as to allow for four distinct levels of logic (two bits) to be transmitted at a time. The advantage of this method is that more information can be sent in a single bit time. The drawbacks are the added complexity in the design of the receiver and the reduced noise margin. The receiver needs to use high-performance analog-to-digital converters (ADC) and novel CDR algorithms. Another drawback to multi-level signaling is that it cannot be used for optical links.

2.2. Clock and Data Recovery

CDRs have been the topic of many works over the years [1][5][13][37][59][64][70]. While numerous different architectures have been proposed, all CDRs accomplish two main tasks: clock recovery and data recovery.

As discussed previously, the clocks of the receiver and of the transmitter oscillate at about the same frequency without any predefined phase relationships. Clock recovery is the task of deriving the required clock phase and frequency to correctly retrieve the data. Data recovery uses the recovered clock to sample the data and retime the bits so that they can be used in a synchronous manner by the receiver. The block diagram of a simple CDR is shown in Figure 2.2.

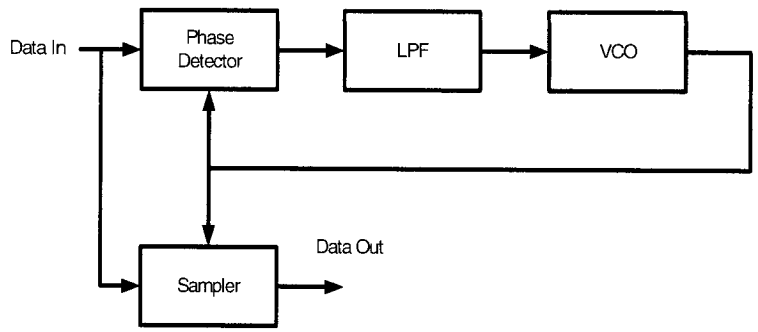


Figure 2.2. Block Diagram of a CDR

In the figure, the upper loop consists of the modules required for clock recovery. It consists of a phase detector, a low pass filter (LPF) and a voltage controlled oscillator (VCO). The sampler located at the bottom of the Figure is shown here for illustrative purposes, since it is generally incorporated into the phase detector.

Clock recovery typically uses the transitions in the incoming data pattern to help align the active transition of the local clock to the middle of the data bit, where it is least prone to errors. To accomplish this task, the CDR relies on a phase detector, which is a module that checks the position of the active clock edge with respect to the data. When the clock is misaligned, the output of the phase detector helps the CDR adjust the clock phase.

Phase detectors can generally be classified into two categories: binary (bang-bang) and linear [4]. A binary phase detector is a circuit that indicates whether the clock transition occurred early or late with respect to the data transition. If the data are sampled before the middle of the bit time, they are considered early. Similarly, when the sampling

is done after the middle, it is considered late. An example of binary phase detection is a double-rate (2X) oversampling scheme, which is shown in Figure 2.3. Using this technique, each bit is sampled twice: once in the middle of the data bit and once during data transition. The dotted lines in the Figure show the location of the ideal sampling points. Let the mid-transition sample be known as *Sample 1* and the mid-bit sample be known as *Sample 2*. Assuming that the oscillator frequency is correct, when the values of *Sample 1* and *Sample 2* are different, it indicates that *Sample 1* occurred before mid-swing. *Sample 2* would therefore occur before the middle of the bit time. In this case, the clock is considered early. Conversely, when the values of *Sample 1* and *Sample 2* are the same, the clock is considered late. Assuming the sampler does not become metastable, the binary phase detector will continuously detect early or late clocks and will keep adjusting its control signal. This can cause rippling on the VCO control line.

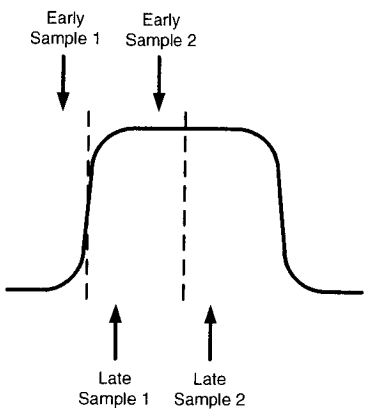


Figure 2.3. Determination of Phase Relations in 2X Oversampling

As previously discussed, binary phase detectors operate by indicating whether a clock is early or late. However, it does not give information as to how early or how late the clock actually is. On the other hand, linear phase detectors give a response that is proportional to the data-clock misalignment. A classic way of providing linear phase detection is to measure the time between a data transition and a clock transition. The goal is to adjust the clock signal until the time difference between the transitions becomes equal to half of a bit time. Hogge's phase detector [29], shown in Figure 2.4, is a linear phase detector that makes use of flip-flops and XOR gates to accomplish this task. The flip-flop and the XOR gate on the left of the Figure are used to measure the time difference between the data transition and a clock transition. It generates a pulse with a duration that is proportional to the time difference. The right-most XOR gate, combined with the two flip-flops, operates as an edge-detector. It generates a pulse of duration equal to half of a bit time when a transition is detected on the data line. When the clock is properly aligned with the data, both pulses should have equal width.

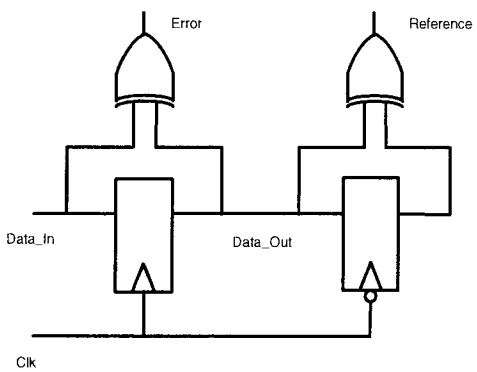


Figure 2.4. Hogge's Phase Detector

Hogge's phase detector, along with many other phase detectors, is classified in a group known as full-rate circuits. Full-rate circuits operate using a single active-edge of a clock during any given period. That means that the full-rate clock needs to make two transitions during a single data bit time. When the data rate is fast, it is difficult to design a VCO that functions at the full-rate. In an effort to alleviate this constraint, the concept of half-rate phase detection has been proposed [65][70]. Using such architecture allows the clock to function at a speed equal to half of the full-rate clock. Figure 2.5 illustrates the different clocking schemes.

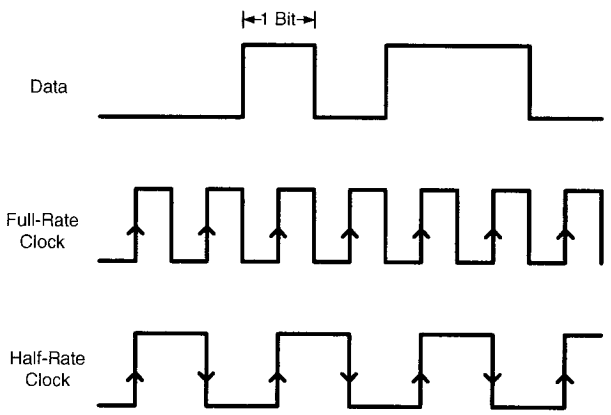


Figure 2.5. Waveforms Showing Different Clocking Schemes

Perhaps the most significant half-rate phase detector in recent years has been proposed by Savoj (Figure 2.6). The structure of the phase detector can be seen as consisting of two of Hogge's phase detectors working in parallel: one operating on the rising clock edge and one operating on the falling clock edge. Note that each of the

sampling devices in Figure 2.6 is a level-sensitive latch as opposed to the conventional edge-triggered flip-flop. In addition to operating as a phase detector, the circuit is also used to sample and retime the incoming data. The data retrieved through the top and bottom portions are combined using a multiplexer. While the half-rate concept is useful in reducing constraints on the VCO, the main drawback is that it requires a 50% duty cycle clock.

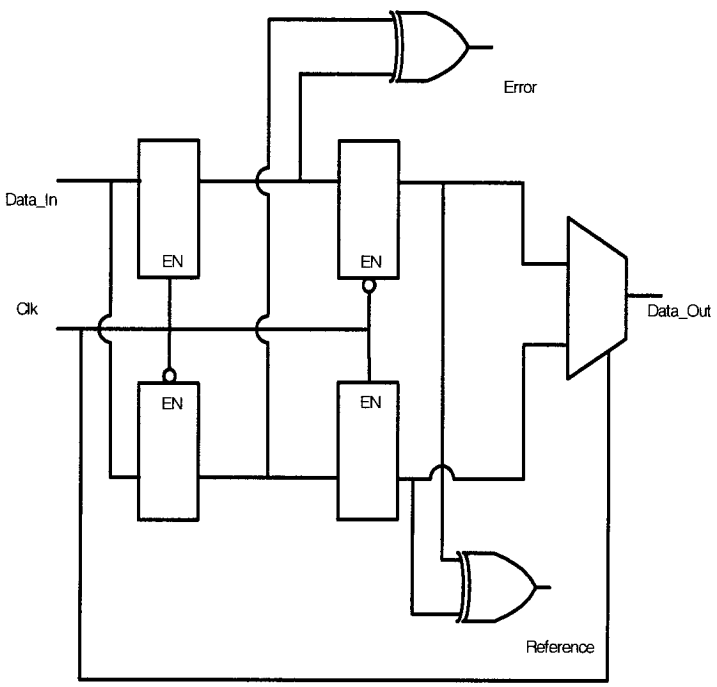


Figure 2.6. Savoj's Half-Rate Phase Detector

Another way of reducing the constraint on the VCO and on the samplers is to use the concept of time-interleaving [69]. Time-interleaving consists of using N phases of a clock

that are separated by T/N, where T is the clock period. The incoming data are then retrieved using N samplers, each of which only needs to operate with clock period T. The benefits of using this technique are that, for a given data rate, the VCO only needs to operate at a speed of $1/N$ of a full-rate clock and the samplers can operate N times slower. This, of course, assumes that the samplers can handle the bandwidth of the incoming data. A drawback to this approach is the increased number of devices connected to the input data line. This can potentially add large amounts of load capacitance to the driver of the data line.

An example of time-interleaving using four clock phases is shown in Figure 2.7.

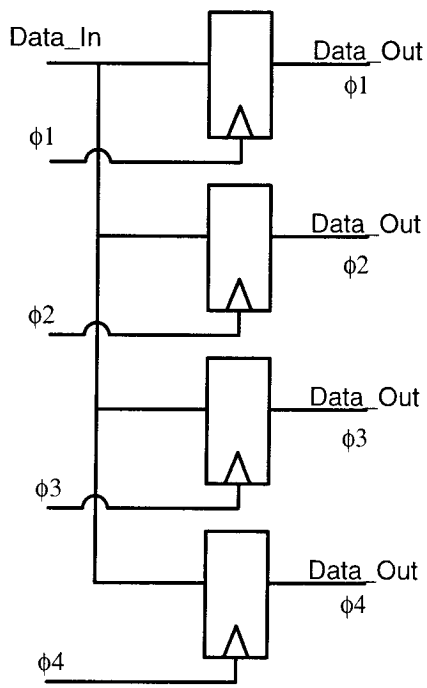


Figure 2.7. Example of 4-Phase Time-Interleaving CDR

2.3. Frequency Acquisition

CDRs do not usually have a good frequency acquisition range due to the fact that they rely on phase detectors. When the frequency difference is large, the CDR may fail to lock. In order to increase the acquisition range, it is customary to include a module, which brings the VCO close to the target frequency. As soon as the frequency difference is small enough, it lets the phase detector take control. The frequency-adjusting module can be seen as providing coarse-tuning for the CDR, whereas the phase detector provides fine-tuning. Many frequency-measuring circuits exist in literature, although most of them cannot operate in the GHz range. An important class of frequency measurement circuits is known as integrator-based frequency-to-voltage converters (FVCs). A significant architecture, shown in Figure 2.8, was proposed by Djemouai [20].

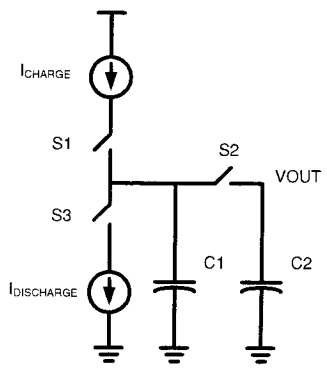


Figure 2.8. Ideal Model of Djemouai's Frequency-to-Voltage Converter

Its principle of operation relies on three control phases. These three phases controlled by switches S1, S2 and S3, are generated in response to the input waveform. The relationship between the input signal and the control signals are shown in Figure 2.9.

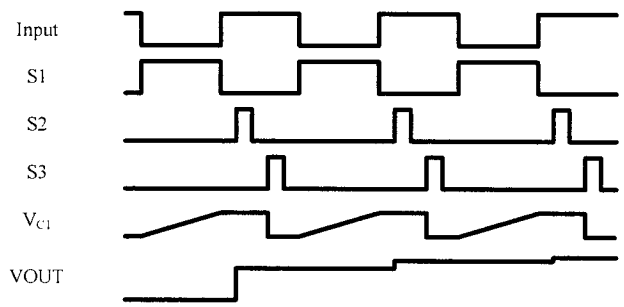


Figure 2.9. Input Clock Showing the Different Phases for Frequency-to-Voltage Conversion

During the first half period, S1 conducts and capacitor C1 works as an integrator by accumulating charges from a constant current source. The voltage at the capacitor node is proportional to the duration of the half-period. During the second half of the period, the accumulated charges are transferred to a second capacitor through switch S2 and the integrating capacitor C1 is discharged through S3. Switches S2 and S3 are controlled by pulses that are generated internally when the first phase is complete. After a few cycles, the output voltage stabilizes to a constant value representing the associated frequency.

Although the circuit operates well at moderate frequencies, several factors limit its performance in the GHz range. The first problem lies in the fact that the circuit makes use

of two pulses that are generated internally. The minimum duration of a half-period is therefore two pulses. In addition, it relies on these pulses to control NMOS switches: the switches typically require full-swing pulses with a duration that is long enough for charge transfer. These constraints set the minimum clock period to be equal to four full-swing pulses. Even with its limits, the FVC is able to function properly at speeds of 200 MHz. Provided its operating speed can be increased, the use of such FVC circuit can contribute to the design of a frequency-acquisition aid that would extend the frequency range of the CDR.

2.4. MOS Current-Mode Logic

These days, transceivers operate at frequencies that are in the multi-GHz range. This operating speed cannot be achieved with conventional design techniques using static CMOS gates. At such frequencies, circuits requiring rail-to-rail output transitions are not feasible with the 0.18 μm technology used in our work. At such operating speeds, MOS current-mode logic (MCML) is the preferred design technique [79]. The principle of operation of MCML gates is best described using an example. Consider the buffer/inverter circuit, which is the simplest MCML gate, shown in Figure 2.10.

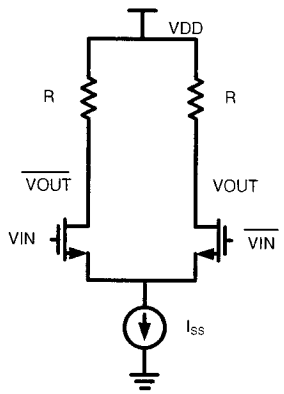


Figure 2.10. MCML Buffer

Its functionality essentially relies on a differential pair that steers a tail current according to the input voltage. When the same voltage is applied to VIN and $\overline{VIN}$, the tail current is split equally between the branches. If the voltage at VIN is increased while decreasing the voltage at $\overline{VIN}$, more current will start flowing through the VIN branch. When the differential input voltage, VIN - $\overline{VIN}$, is larger than a given ΔV, the tail current is completely steered to one branch, leaving the output voltage of the other controlled by a passive pull up. This ΔV is approximated by the following equation [28]:

$$\Delta V = \sqrt{\frac{2ISS}{\mu_N C_{OX}(W/L)}} \tag{2.1}$$

In this case, the output voltage on the current-carrying branch drops by $I_{SS}*R_L$, whereas the other branch goes to VDD. By adjusting the values of I_{SS} and R_L, the swing of the gate can be adjusted, which could lead to higher operating speeds.

More complex gates can be designed simply by connecting a number of these differential pairs together. For instance, the XOR gate shown in Figure 2.11 can be

implemented with two parallel differential pairs stacked on top of a third differential pair. The tail-current is first steered through the bottom transistors and is then switched between the top transistors. More complete discussions on the design of MCML gates can be found in [50].

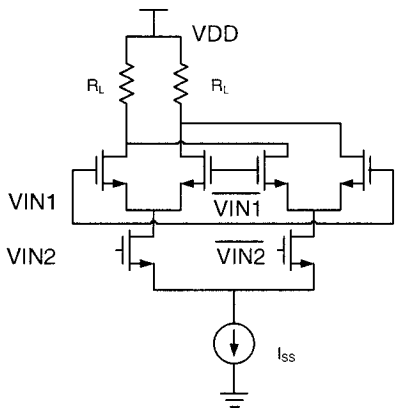


Figure 2.11. MCML XOR Gate

Besides its potential for achieving higher speed, MCML has several other advantages. The current drawn by MCML gates does not vary much with frequency and consequently, the power dissipated is also nearly constant. On the other hand, CMOS gates consume more power as the switching activity increases. It can be shown that in the gigahertz range, MCML gates often dissipate less power than their CMOS counterpart. In addition to consuming less power at high frequencies, since the current is nearly constant, the L (di/dt) noise is minimized [50], which greatly alleviates signal integrity issues that often become the bottleneck in high speed systems. Finally, the fully differential topology of the MCML gates reduces its sensitivity to common-mode noise.

2.4.1 Symmetric XOR

One of the problems encountered when using multiple-input MCML gates in high speed applications, such as CDRs, is gate asymmetry. Gate asymmetry gives rise to some uncertainty regarding the time an output transition is generated in response to switching inputs. While this timing variation is a negligible jitter in most low-speed applications, its effect becomes more significant as the speed increases. In applications such as CDRs, the XOR gate is often used to measure the precise phase difference between two signals [29][64]. An asymmetric XOR would introduce jitter that generates errors in the phase measurements, ultimately causing data recovery errors. It is therefore important to reduce the jitter introduced by the XOR gate.

While most of the work in recent years has dealt with speed and power issues, only a limited number of designs have addressed jitter. For instance, the work in [19][52] perform a theoretical jitter analysis on the use of XOR gates in CDR applications. In an attempt to reduce output jitter in XOR gates, several works have proposed symmetric circuit topologies. The first symmetric XOR gate was proposed in 1990 [66]. The authors in [66] suggest a criss-crossed connection of bipolar emitter-coupled logic (ECL) XOR gates, which would eliminate any asymmetry in the output signal. This topology has recently been adapted to CMOS technology in [58], where the authors implemented the circuit using MCML [79], or source-coupled logic (SCL) as it is sometimes known [16].

Another significant XOR gate is the one proposed by Savoj and Razavi in [64] (Figure 2.12). This circuit is a symmetric XOR gate with symmetric drive strength and

symmetric loading. It operates as follows. When *VIN1* and *VIN2* have different values, both input transistors on one branch carry the tail current. On the other branch, none of the input transistors are selected and consequently, the tail current flows through one of the V_B-biased transistors. The current through that V_B-biased transistor is then copied through the current mirror which brings the output node high. In the case where inputs *VIN1* and *VIN2* are equal, only one input transistor on each branch carries its respective tail current. Since there is no current flow through the V_B transistors, no extra current is injected at the output node and the resistor pulls the output low. The current source located at the output node is used to set the DC level of the XOR gate.

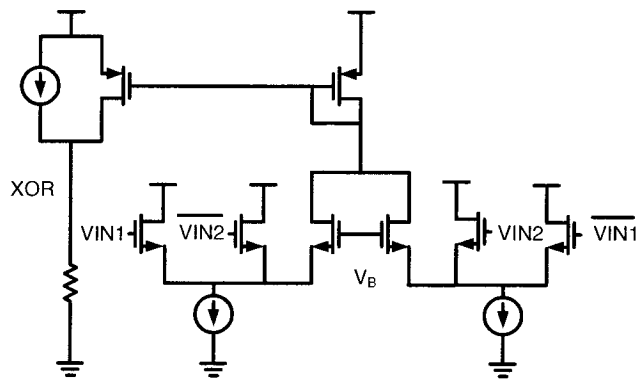


Figure 2.12. Single-Ended Symmetric XOR

The benefits of this topology are that input drivers are loaded equally and only have to drive a single transistor. The main drawback of this circuit is its single-ended output. While it performs the required functionality in the CDR presented in [64], its single-ended output limits its versatility and the gate cannot be used in many applications. Therefore, it will not be considered any further in this thesis.

2.5. MCML Gate Bandwidth Improvement

Although MCML provides the highest achievable speed compared to most other CMOS design techniques, it is quickly reaching its limits. In 0.18 μm CMOS, the gain of the fastest MCML gate, namely the buffer, drops below unity soon after 10 GHz. Most other MCML gates cannot operate reliably at speeds beyond 5 GHz. To push back the limits of these MCML gates, numerous techniques have been proposed in recent years. This section will provide a brief overview of the significant developments.

As stated previously, the structure of MCML gates is based on differential amplifiers. It is also known that differential amplifiers and other inverting amplifiers are subject to Miller capacitance, which can potentially reduce the maximum bandwidth that can be achieved by the circuit. Due to the Miller effect, part of the capacitance (C_{GD}) seen at the input node is effectively multiplied by the gain of the amplifier. Since MCML gates are based on differential pairs, they are also affected by the Miller capacitance.

In an effort to extend the maximum bandwidth, researchers have attempted to cancel the Miller effect by including a series capacitor between the non-inverting input and the drain node, as shown in Figure 2.13 [28].

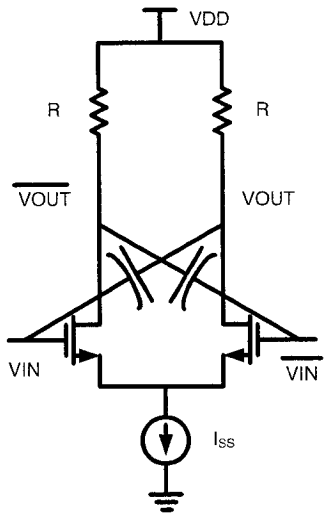


Figure 2.13. Buffer with Compensated Miller C_{GD}

Several factors affect the speed of operation of a circuit. These factors include the drive strength of a circuit and its output load. This means that a circuit could possibly function at a higher speed if the drive strength is increased or if the output load is decreased. Since the input capacitance of one gate is the output capacitance seen by the driving stage, it is also beneficial to reduce the input capacitance. In the literature, there are numerous structures that are known as f_t doublers [39][56]. While their topologies differ, they all have a similar characteristic: doubling the ratio of the transconductance gain to the input capacitance. The differential f_t doubler shown in Figure 2.14 reduces the input capacitance by half while maintaining the gain constant. The drop in input capacitance is due to the fact that the differential inputs are not connected to the same source-coupled transistors [62]. The smaller capacitance reduces the load of the driving stage, thereby allowing it to operate faster. The drawbacks from this technique include

the increase in power consumption and the increased output capacitance due to the extra drain capacitance.

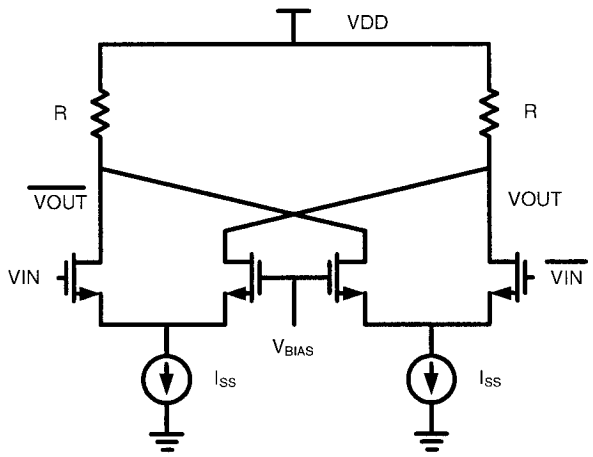


Figure 2.14. Differential f_t Doubler

It is known in control systems theory that negative feedback in a system could provide larger bandwidth in exchange for reduced gain. This feedback reduces the lower frequency gain, which remains stable for a larger part of the frequency spectrum. Some authors have applied this feedback theory to MCML gates. They report that higher bandwidth can be achieved in MCML gates by connecting a resistor between the drain and the gate of the input transistors [71]. An example of such feedback on a buffer is illustrated in Figure 2.15. This effectively creates a negative feedback, which explains the resulting bandwidth extension.

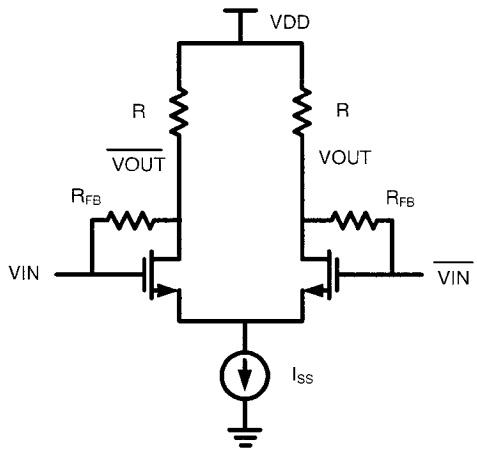


Figure 2.15. Buffer with Negative Feedback

Another insight into improving the bandwidth of MCML gates can be obtained by examining their transfer functions. It is known that a main factor that typically limits the maximum operating frequency of MCML gates is the capacitance seen at the output node. On the other hand, control systems theory states that the effects of a pole in a transfer function can be effectively cancelled if a zero can be placed at the same location on the s-plane. This means that canceling the dominant output pole could potentially increase the maximum bandwidth.

At the circuit level, a zero can be created either by connecting a series capacitor or a parallel inductor to the output node. The latter technique is known as shunt-peaking [47]. When using ideal inductors, the bandwidth can be increased by up to 85%. In practical situations, inductors can be implemented using planar spiral inductors. A representation of the spiral inductor is shown in Figure 2.16.

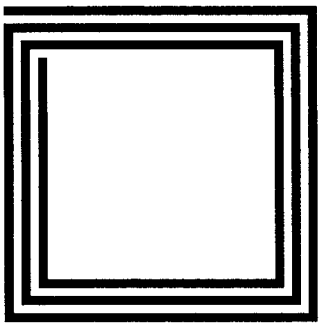


Figure 2.16. Representation of a Spiral Inductor

With this choice of implementation, the various parasitics limit the possible bandwidth enhancement to about 40-50% [37]. The technique has shown its effectiveness in analog applications and has recently been introduced to the digital world [5][33]. It has been shown that shunt-peaking could be applied to any given MCML gate to provide extra bandwidth. A shunt-peaked buffer is shown in Figure 2.17.

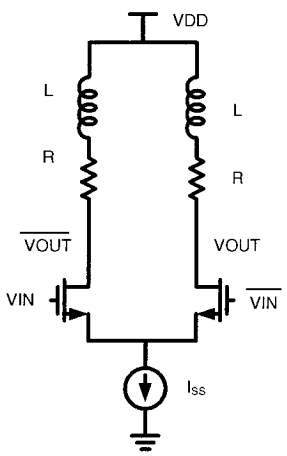


Figure 2.17. Shunt-Peaked Buffer

Even though planar spiral inductors are a widely accepted technique for shunt-peaking, they have several drawbacks. The first, and perhaps the most significant, is their large area. The inductance of a spiral inductor comes from two sources: the self-inductance of each branch and the mutual inductance between the branches. In order to get a large inductance value, the branches need to be long so that enough self-inductance can be generated. This constraint sets the minimum value on the size of the spiral inductor. Typically, spiral inductors are rarely any smaller than 100μm X 100μm. While technology scaling allows for thinner metal lines and smaller wire spacing, this increases series resistance, causing losses, and internal capacitances limiting the inductive reactance. Thus the size of inductors is not scaled as rapidly as that of active devices. Therefore, the relative space occupied by spiral inductors of fixed inductance will increase as a result of technology scaling.

In addition to their large size, spiral inductors need to be isolated from other structures to avoid interference (e.g.: eddy currents). This prevents any other structures from being placed too close to inductors. Such restriction increases the total silicon area associated to the design of spiral inductors.

Over the years, numerous improvements to the planar spiral inductor have been proposed. To make use of the multiple metal layers offered in new technologies, some have proposed the use of multiple layer inductors [46]. The additional layers can help increase the inductance provided by a structure for a given area. The drawback to this approach is the increased capacitance, which reduces the self-resonance frequency.

Another problem related to spiral inductors is the modeling of their characteristics. Perhaps, the most popular way of designing circuits with spiral inductors is to use the equivalent lumped model shown in Figure 2.18 [46].

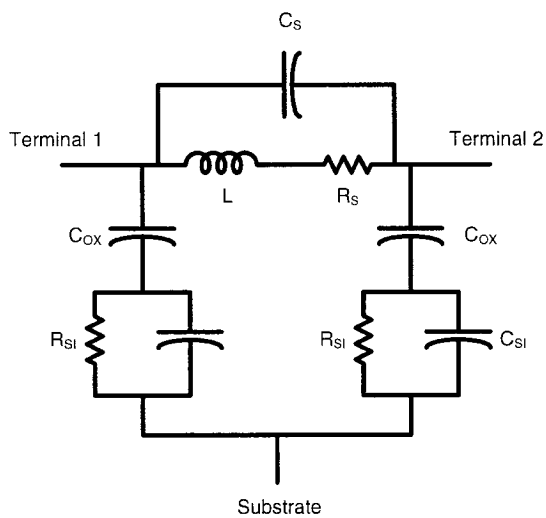


Figure 2.18. Lumped Model of a Spiral Inductor

To accurately determine the values of the lumped model parameters, it is necessary to perform electromagnetic simulations using tools such as Agilent's ADS tool suite. The resulting model is then included in a netlist and used in a SPICE-like simulator to verify the performance. This process is time-consuming and only yields a model that is valid over a limited bandwidth. Optimization of such circuits requires multiple iterations, which could be a lengthy process because it requires switching between design environments.

The implementation of shunt-peaking can also be done using bondwire inductors. Bondwire inductors have the benefit of having higher quality factors (Q) than spiral

inductors. For shunt-peaking, however, Q does not affect the performance of the gate, since the value of the series load resistor can be adjusted to compensate for a different Q. In terms of silicon area, bondwire inductors require the space of at least two pads. In addition, the area surrounding this structure should also be kept empty to prevent excessive interference, similar to the situation experienced with spiral inductor.

One way of implementing an inductor without excessive silicon area and design complications is to use active inductors. Active inductors are circuits, consisting of at least one transistor, that behave like an inductor for small signals. These devices have been found in numerous applications including active filters, active phase shifters and VCOs. It may be possible to use these active inductors for shunt-peaking. Typically, for high speed operation, the number of components in the design should be minimized. The active inductors with the least number of elements found in the literature consist of a transistor and a resistor. These inductors are shown in Figure 2.19.



a) b)

Figure 2.19. Active Inductors

2.6. Summary

This chapter provided an overview of the research done in recent years that relates to the design of multi-GHz transceivers. It described the basic concepts used in communication links and examined some of the components that are typically found in transceivers. It also discussed some of the problems and challenges that involved in high-speed transceivers.

While MCML can be used to design high speed circuitry, its performance is limited by numerous factors including load capacitance, miller effect and gate asymmetry. Some of these problems have previously been addressed by several authors and their proposed solutions have been presented along with the benefits and deficiencies.

The remaining chapters in this thesis aim to improve existing design methodologies by helping reduce silicon area ease design efforts and improve the quality of the results. The following chapter discusses a new way of improving bandwidth in MCML gates using active inductors.

CHAPTER 3

Bandwidth Enhancement using Active Inductive Loads in MCML Gates

As the speed of operation in digital integrated circuits increases, the use of CMOS logic ceases to be a viable option. This is because the time required to make rail-to-rail transitions is too long. As an alternative, many designers choose to use MCML [79]. However, the increasing demand for high-speed circuitry is pushing MCML to its limits. In [47], Mohan et al. proposed the use of inductors to increase on-chip bandwidth of amplifiers. This technique, known as shunt-peaking, consists of placing an inductor in parallel with the output node. In recent years, it was found that shunt-peaking could also be used to increase the bandwidth of MCML gates.

In practice, shunt-peaking using planar spiral inductors can increase the bandwidth of certain circuits by 40% to 50% [37]. The main drawback, however, is that spiral inductors are very large in size compared to normal gates and transistor. In addition, the design of spiral inductors and the design of circuits are done using different sets of tools. On one hand. spiral inductor design requires an electromagnetic field solver such as Agilent's ADS tool suite. On the other hand, transistor-level circuits are often simulated using SPICE-like programs. Designing circuits containing inductors involves switching between working environments and optimizing these designs can be a long iterative process.

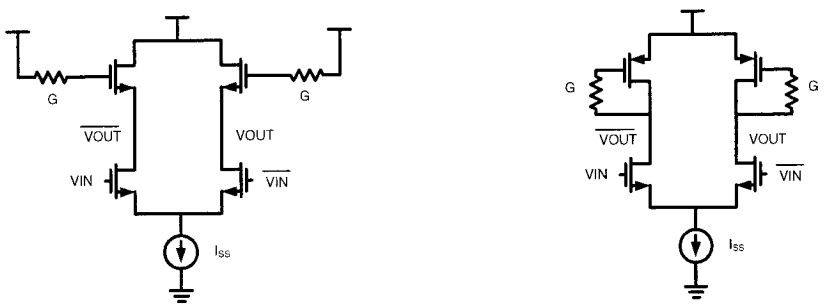
One way of implementing an inductor without excessive silicon area and design complications is to use active inductors. Figure 3.1 shows two active inductors that can potentially be used in shunt-peaking of MCML gates [1][77].



a) b)

Figure 3.1. Active Inductors

For this application, the resistive loads of MCML gates are replaced by the active inductors. Figure 3.2 illustrates how active inductors are used in the design of buffers [1][77].



a) b)

Figure 3.2. MCML Buffers Loaded with Active Inductors

3.1. Theoretical Analysis of Active Inductors

Active inductors can be applied to shunt-peaking of MCML gates. To quantify the performance improvements incurred, it is useful to analyze the small signal model of the circuit. The large signal model could also be considered as it is more accurate. However, its non-linear nature would make the analysis intractable. The small signal analysis is first made on a buffer loaded with the NMOS active inductor as shown in Figure 3.2 a). The small signal model of the relevant portion of the circuit is shown in Figure 3.3. It considers one branch of the buffer and models the input transistor as a voltage-controlled current source.

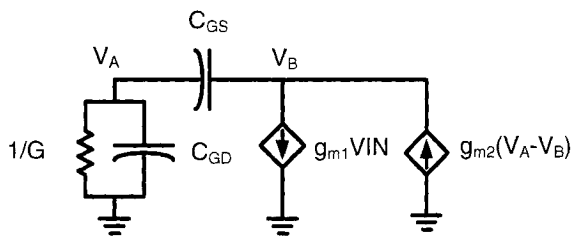


Figure 3.3. Small Signal Model of the Buffer Loaded with an NMOS Active Inductor

In the small signal model of the active inductor, only the dependent current source, the gate-drain and the gate-source capacitances are considered. It is assumed that the transconducance g_m of the active inductor is the same as that of the source-coupled transistors.

Since the voltage at the output node is V_B, the transfer function is found to be

$$TF_{ACTIVE}(s) = -\frac{g_{m1}(G+s(C_{GD}+C_{GS}))}{s^2C_{GD}C_{GS}+s(C_{GS}G+C_{GD}g_{m2})+(g_{m2}G)} \quad (3.1)$$

The transfer function has two poles and a zero and their positions on the s-plane can be determined analytically. Using the simplifying assumption that C_{GS} and C_{GD} are equal, the transmission zero can be found, by inspection, to be located at $(-\frac{G}{2C})$, where $C=C_{GS}=C_{GD}$. It can also be shown, assuming that $g_m=g_{m1}=g_{m2}$, that the poles are located at $(-\frac{g_m}{C})$ and $(-\frac{G}{C})$.

The goal of shunt-peaking is to provide a zero in the transfer function that can alter the effect of the dominant pole at the output node. This can push back the frequency at which the gain starts diminishing. The resulting increase in bandwidth can be quantified by comparing the –3db frequencies among different topologies.

Perhaps the simplest comparison can be made between an active NMOS inductor and a resistive NMOS load. The small signal model for the resistive NMOS load can be found by replacing the resistor in Figure 3.3 by a short-circuit ($G \to \infty$). The transfer function of this circuit is found to be:

$$TF_{NMOS}(s) = -\frac{g_m}{sC_{GS}+g_m} \quad (3.2)$$

It can be shown that the DC gain of a buffer using either NMOS loads is equal to 1, which implies that their –3db gain is $\frac{1}{\sqrt{2}}$. The –3db frequency of the resistively loaded buffer is found to be equal to $\frac{g_m}{C_{GS}}$. For the case of the inductive load, the –3db frequency can be found by solving for ω in the following squared magnitude equation:

$$\frac{1}{2}=\frac{G^2 g_m{}^2+4C^2 g_m{}^2\omega^2}{C^2\omega^2(G+g_m)^2+(G\cdot g_m-C^2\omega^2)^2} \tag{3.3}$$

The maximum bandwidth improvement occurs as G goes to 0. In this case, the –3db frequency becomes $\frac{\sqrt{7}g_m}{C}$, which is an improvement of roughly 2.65 over the resistive NMOS load. It is interesting to note that a similar analysis performed on a buffer loaded with a PMOS active inductor also yields $\frac{\sqrt{7}g_m}{C}$ as the -3db frequency. Although the -3db frequency is given by the same expression in both cases, it does not mean that both active inductors offer the same benefits. A major distinguishing factor lies in the fact that hole mobility is lower than electron mobility. This means that, for a PMOS to have the same transconductance g_m as an NMOS under similar biasing conditions, the W/L ratio of the PMOS needs to be larger. With a larger W/L ratio, the parasitic capacitance C is also larger, which indicates that the PMOS configuration would have a lower bandwidth than its NMOS counterpart. For this reason, the PMOS active inductor will not be considered any further in this thesis.

The previous analysis shows that the NMOS active inductor can improve the bandwidth by 2.65 times over its resistive counterpart. However, further analysis shows that, under these conditions, some frequencies will have a gain that is up to twice the DC gain. This phenomenon is known as "peaking" and may be undesirable in some applications.

For these applications, the design goal should perhaps be to obtain maximum bandwidth improvement with a flat frequency response. This is obtained as follows. It is

known that all derivatives of a constant function are equal to 0. Thus, if the gain of a circuit is constant throughout the whole frequency spectrum, all the frequency derivatives of $\|TF_{ACTIVE}\|$ would also be 0. This, however, is not possible since all known circuits have finite bandwidth. It can be shown that to obtain a maximally flat response, it is required to maximize the number of derivatives of $\|TF_{ACTIVE}\|$ that are equal to 0 at DC [39]. A good approximation is to find these parameters for the first three derivatives:

$$\frac{\partial \|TF_{ACTIVE}\|}{\partial \omega}\Big|_{\omega=0} = 0 \tag{3.4}$$

$$\frac{\partial^2 \|TF_{ACTIVE}\|}{\partial \omega^2}\Big|_{\omega=0} = 0 \tag{3.5}$$

$$\frac{\partial^3 \|TF_{ACTIVE}\|}{\partial \omega^3}\Big|_{\omega=0} = 0 \tag{3.6}$$

Since the value of (3.4) and (3.6) are always 0 when ω is 0, the only equation to solve is (3.5). The solution shows that G should be equal to $\sqrt{3}g_m$ and the resulting bandwidth improvement is equal to $\sqrt{2+\sqrt{7}} \cong 2.16$. It is interesting to note that a similar proof is used to demonstrate the maximally flat response of a Butterworth filter.

The previous results indicate that a value of G that is lower than $\sqrt{3}g_m$ would cause overshoot whereas a higher value of G would dampen the response. These situations are illustrated in Figure 3.4. Figure 3.4 a) shows the Bode plot generated by MATLAB for three cases: a maximally flat response, a dominant zero behavior and the case where the zero is either non-existent or far away. The corresponding step responses are shown in Figure 3.4 b).

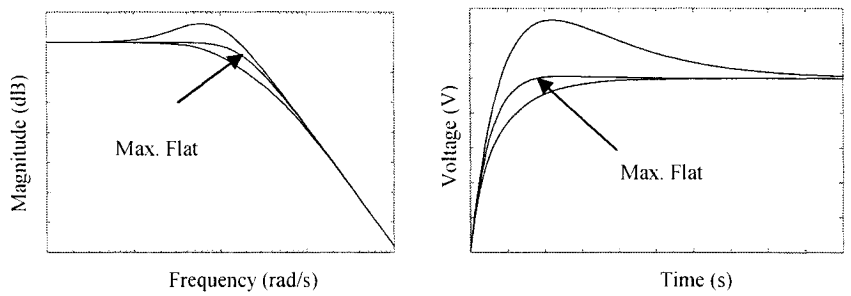


Figure 3.4. a) Bode Plot and b) Transient Step Response for Different Pole-Zero Positions

Obtaining a large bandwidth extension is beneficial since it allows circuits to operate faster. However, in many cases, the resulting output signal may be distorted. One such type of distortion, which has previously been discussed, is the peaking in the frequency response. This can cause overshoots in the transient response and may be undesirable. Another type of distortion is related to the phase of the different spectral components of the signal. To generate an output signal without phase distortion requires that all spectral components be delayed by the same amount of time. Since time delay is equal to $-\frac{d\phi}{d\omega}$, a frequency-independent constant delay means that the phase has to vary linearly with frequency.

From Eq. (3.1), the phase transfer function of the active shunt-peaked buffer can be found to be equal to:

$$\phi(\omega) = \arctan\left(\frac{-2C \cdot gm \cdot \omega}{-G \cdot gm}\right) - \arctan\left(\frac{C \cdot \omega (G+gm)}{G \cdot gm - C^2 \omega^2}\right) \tag{3.7}$$

In order to make the time delay constant for all frequencies, all the frequency derivatives of this time delay must be equal to 0. A good approximation can be obtained by making the first three derivatives of the time delay equal to 0 at DC. This situation is similar to the maximally flat response. As in this case, the value of the first and the third derivates are always equal to 0 when ω=0. Therefore, the only equation left to solve is the second derivative. Solutions show that minimal phase distortion can be obtained by setting G to $\sqrt[3]{7}g_m$. The resulting bandwidth extension is $\sqrt{\frac{\sqrt[3]{7^2}\left((\sqrt[3]{7}-1)+\sqrt{\sqrt[3]{7^2}+1-\frac{10}{\sqrt[3]{7^2}}}\right)}{2}}$, which is roughly equal to 2.05.

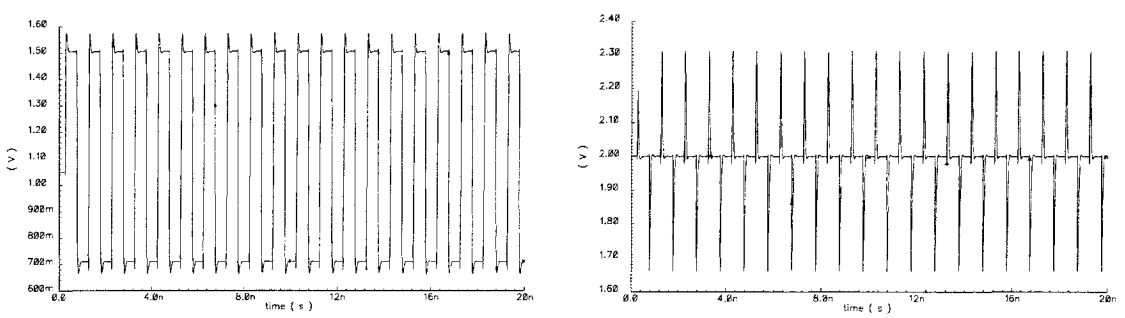
The analyses on the bandwidth of active shunt-peaked buffers show that lower distortion leads to having lower bandwidth. It indicates that a tradeoff can be made between signal quality and bandwidth extension. This decision is left to the designer. A summary of the different results is shown in Table 3.1.

Table 3.1. Summary of Different Design Parameters for NMOS Active Inductors

Desired Characteristic	Required G	Bandwidth Extension
Maximum Bandwidth	0	2.65
Maximally Flat	$\sqrt{3}g_m$	2.16
Minimal Phase Distortion	$\sqrt[3]{7}g_m$	2.05

To develop better insight into the functionality of the active inductor, a time-domain analysis of the structure can be beneficial. This analysis goes as follows. The functionality of the NMOS active inductor relies on two main elements, which are the

gate-source capacitance of the transistor and the resistor at the gate. When a transition occurs at V_{OUT}, charges are injected to the gate via capacitive coupling between the source (V_{OUT} node) and the gate. Due to the resistor at the gate node, the injected charges do not disappear immediately. This can affect the biasing of the transistor. If, for example, the transition at V_{OUT} is a rise, there would be positive transient voltage on the gate node. This positive bias contributes to a momentary increase in V_{GS}, which increases the current flow through the transistor. This accelerates the charging process of V_{OUT} and decreases the transition time. The behavior is illustrated in the simulation results shown in Figure 3.5. The left part of Figure 3.5 shows the voltage at the output node, whereas the right part shows the voltage at the gate of the active inductor's transistor. Every time there is a transition at V_{OUT}, a voltage spike appears at the gate of the transistor. When the duration of this spike is longer than the transition time, it causes an overshoot in the transient response.



a) b)

Figure 3.5. Simulation Results for a) the Output of a Buffer with Active Shunt-Peaking and b) the Corresponding Voltage at the Gate of the Active Inductor's Transistor.

3.2. Simulation Results

In order to test the effectiveness of active shunt-peaking, several gates have been designed and simulated in 0.18μm CMOS technology. Even though small signal analyses have been performed previously, simulations have been done using large signal. This is because large signal simulations yield more precise results for the case of MCML gates. The gates, which are namely AND gates, XOR gates and buffers, have input transistors with a width of 20μm and a tail current of 3mA. The loads of the MCML gates are adjusted to have equal output swings at low frequencies and a maximally flat magnitude response. In total, three sets of gates have been designed: one using resistive NMOS loads, one using NMOS active inductors and one using a resistor load.

To model realistic loading and driving conditions, the device under test is driven by a 20μm buffer with tail current of 3mA and is loaded with a similar circuit. To test the gates, a periodic $\pi/2$ phase-offset stimulus is used. It consists of having two periodic signals with the same frequency that are shifted by $\pi/2$ in phase. When applied to 2-input gates, all the possible input combinations are generated.

The first gate to be tested is the buffer. Figure 3.6 shows the simulation results when the circuits are stimulated by a square wave with a period of 1 ns. The signal with the fastest transition time is the active-peaked buffer. Its rise time is measured to be 32.4 ps compared to that of the resistor-loaded MCML gate which is 37.4 ps. The NMOS load requires 37.8 ps to transition between the 10% and 90% points of the full range. The

measurements show that the use of active inductors reduced the transition time by little under 14% when compared to the conventional resistor-loaded gate. It should be noted that, even though the gates have been designed for maximally flat magnitude response, the output in transient simulations still show some overshoot. This is because small-signal analysis is only an approximation when the input and output voltages affect the gate's biasing conditions.

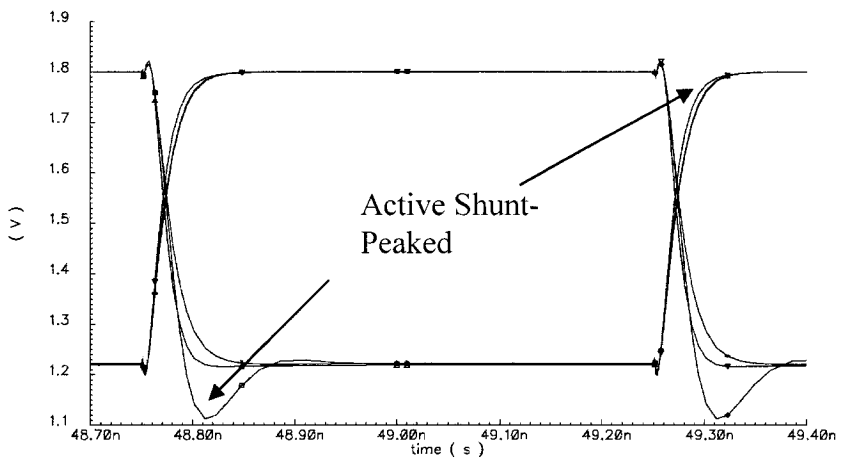


Figure 3.6. Simulation Results for the Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked Buffers

Simulations run on the AND and XOR gates show results that are similar to those of the buffer. In the case of the AND gate, the improvement in transition time is slightly less than 16% whereas the active shunt-peaked XOR gate showed an increase of more than 16% in performance. The measured results are presented in Table 3.2.

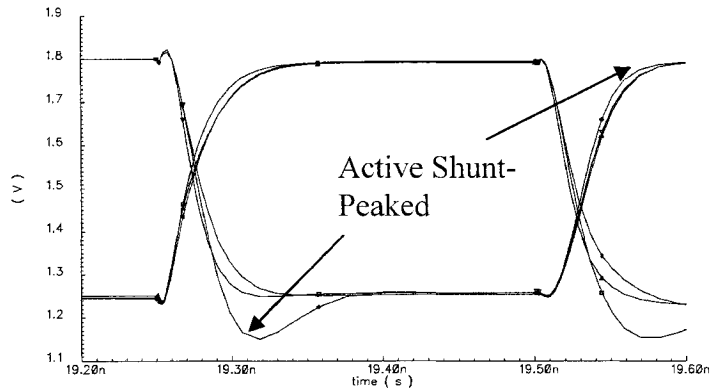


Figure 3.7. Simulation Results for Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked AND Gates

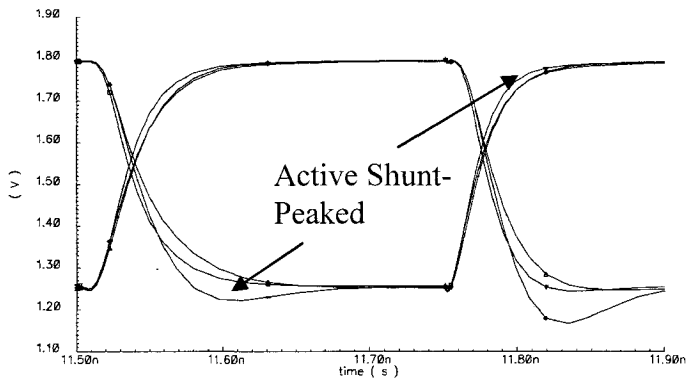


Figure 3.8. Simulation Results for Resistor-Loaded, NMOS-Loaded and Active Shunt-Peaked XOR Gates

Table 3.2. Summary of Performance

Load	Transition Time (ps)		
	Buffer	AND	XOR
NMOS	37.8	47.0	45.5
Resistor	37.4	48.7	46.5
Active	32.4	41.0	38.7

3.3. Process Variations

In the previous section, the use of active inductors was proposed for bandwidth enhancement. Active inductors are smaller in size and more easily designed than their passive counterpart. However, spiral inductors typically offer larger bandwidth extension than active inductors and are less sensitive to process and temperature variations. The sensitivity to process and temperature variations can be problematic since the designed circuit may not have the desired characteristics once fabricated.

This section proposes a way of quantifying process and temperature variations on a chip. The measurements are made by on-chip circuitry and the output can be used to compensate for these variations. While the measurement and compensation scheme is designed for active inductors, it can be used for any circuit.

3.3.1. Process/Temperature Variations and Measurement

In an integrated circuit, numerous parameters are affected by process and temperature variations. While it is difficult to keep track of the changes occurring in all parameters, it is possible to deal with them at a higher level of abstraction. From a circuit designer's perspective, one of the main parameters affected is the transconductance of the transistors. The method proposed here relies on this change to determine which process

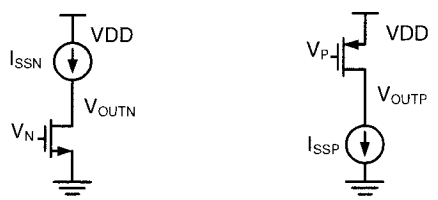
corner the chip is operating in. Using the same technique, it is also possible to measure temperature variations and their impact on circuit parameters.

To understand how process variations affect transconductance, it is useful to examine the transconductance equation of an NMOS transistor:

$$g_m = \mu_N C_{OX} (\frac{W}{L})(V_{GS} - V_{TH}) \tag{3.8}$$

Eq. (3.8) shows that for given V_{GS} and transistor size ratio, the transconductance is determined by μ_N, C_{OX} and V_{TH}. μ_N and V_{TH} are affected by both temperature and process variations, while C_{OX} is only affected by process variations. The fact that these parameters change in presence of process and temperature variations also causes the transconductance to be affected. Therefore, transconductance can be used to monitor how process and temperature change and affect circuit characteristics.

To measure process and temperature variations, the use of a common-source amplifier is proposed, as shown in Figure 3.9.



a) b)

Figure 3.9. a) N-Type and b) P-Type Process and Temperature Measurement Circuits

These circuits are designed to have a large-signal output of VDD/2 in typical process, with known current $I_{SSP}(I_{SSN})$ and known input voltages $V_P(V_N)$. The use of two

measurement circuits allows for the effects of process and temperature to be determined for NMOS and PMOS independently. When the transconductance of NMOS transistors is higher than the nominal case, the output voltage of the N-type measurement circuit will be lower than VDD/2. Similarly, when the transconductance of PMOS transistors is higher than the nominal case, the output voltage on the P-type measurement circuit will be higher than VDD/2. Thus, in a fast-fast corner case, the output of the P-type measurement circuit is high, whereas the output of the N-type measurement circuit is low. The remaining corner case input-output relations are listed in Table 3.3. It should be noted that, although the outputs are labeled 'HIGH' and 'LOW', it does not mean that they are binary: it simply states that the output signal is higher or lower than VDD/2, respectively.

Table 3.3. Measurement Output for Process Corners

Process Type (NMOS-PMOS)	N-Type	P-Type
Typical-Typical	VDD/2	VDD/2
Fast-Fast	LOW	HIGH
Fast-Slow	LOW	LOW
Slow-Fast	HIGH	HIGH
Slow-Slow	HIGH	LOW

This measurement technique has the advantage of being simple, providing a stable output and being able to measure the performance of the PMOS and the NMOS independently. This is in contrast to the common technique that consists of using the frequency of a ring oscillator to characterize the process variation.

While the effects of corner case process variations can simply be listed as shown in Table 3.3, variations in temperature affect the output of the measurement circuit in a more complex way. It is known that transconductance is affected by two temperature-sensitive parameters, which are namely threshold voltage and carrier mobility. These components have opposite effects on the net transconductance and could make its dependency on temperature a non-monotonic relation. As temperature increases, mobility is reduced because of lattice scattering (phonons) but the absolute value of the threshold voltage is also reduced [68]. This means that, in some conditions, different temperatures will yield the same output. This type of temperature aliasing may be useful to build temperature insensitive devices, but is undesirable with the present circuit that aims at detecting parametric shifts.

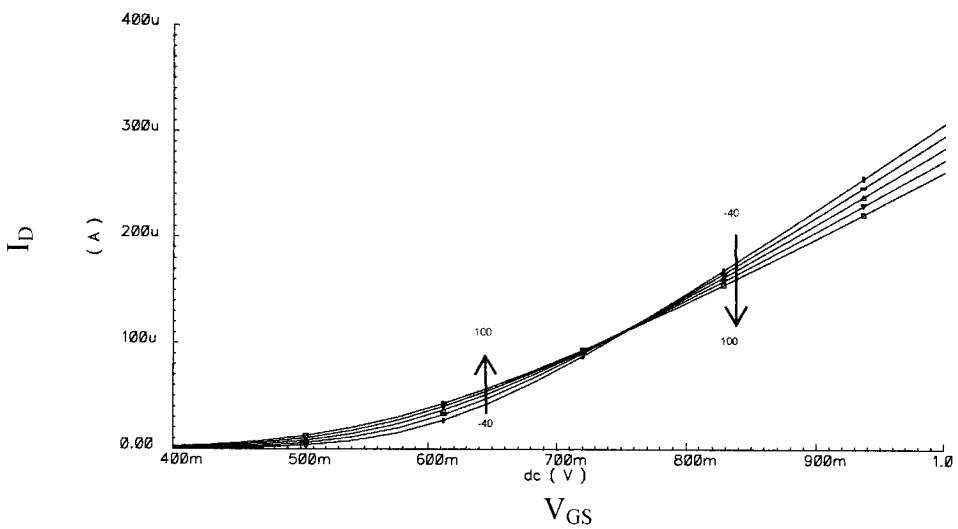


Figure 3.10. I_D vs. V_{GS} for Different Temperatures

Figure 3.10 shows V-I curves plotted at different temperatures between -40°C and 100°C. It shows three regions with different responses to a change in temperature. The region to the right, which is characterized by large V_{GS} and large drain current, has a transconductance that decreases with temperature. The reason for this behavior is that, with large V_{GS}, the effects of decreased $|V_{TH}|$ are much less significant than the effects of decreased mobility. The region to the left of Figure 3.10 has the opposite behavior where transconductance increases with temperature. There is a third region, where threshold voltage and mobility variations tend to compensate each other and where temperature sensitivity is reduced. Some designers have used it to make their circuits less sensitive to temperature variations [31].

When designing a measurement circuit, V_P and V_N should be chosen so that the region of operation is either in the left part or in the right part of Figure 3.10. This guarantees monotonic behavior and allows the measurement circuit to have a larger step sizes between temperatures.

To maintain a certain consistency with the effects of process variations, the right-side region is chosen. Figure 3.11 shows the output of the measurement circuits in a typical process as the temperature is swept between -40°C and 100°C with $|V_{GS}|$=VDD. The simulation is done in 0.18μm CMOS technology with $I_{SSP}=I_{SSN}$=150μA. The W/L ratio of the NMOS input transistor is 0.82μm/1μm whereas the input PMOS has a W/L of 4.1μm/1μm. The remaining components are simple current mirrors with transistor W/L ratio of 5μm/1μm. Simulation results show that, with the chosen biasing conditions and

temperature range, the outputs are monotonic and nearly linear. This circuit is therefore suitable for measurement of temperature and process variations.

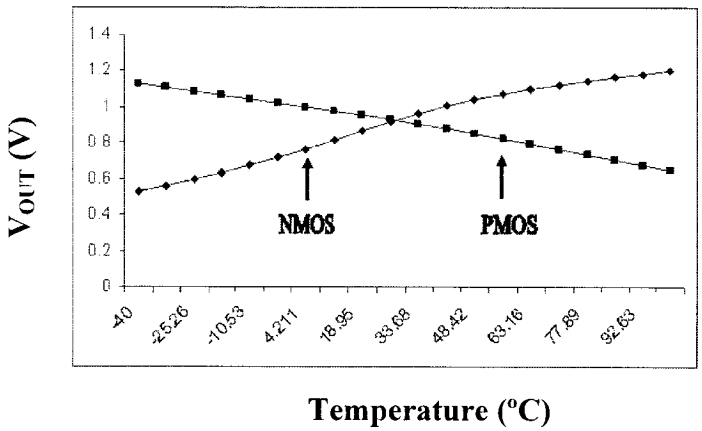


Figure 3.11. Measurement Output for Different Temperatures in a Typical Process

3.3.2. Active Inductor Compensation

To reduce the effects of process and temperature variations on circuits, the most straight-forward approach is to adjust circuit parameters to compensate for these variations. It is known that the performance of active inductors is mainly defined by the resistance connected to the gate and the characteristics of the transistor. With the proposed topology, to change the characteristics of the active inductor would imply replacing either one of its components. After a chip is fabricated, this cannot be done conveniently.

To make the inductance adjustable after chip fabrication, a novel structure for this active inductor is proposed. This new topology is similar to the active NMOS inductor

where the resistor is replaced by a PMOS. The value of the inductance can be modified by changing the voltage at the gate of the PMOS which adjusts the channel resistance. The DC level of the output can also be adjusted by changing the signal at the source of the PMOS. This is illustrated in Figure 3.12.

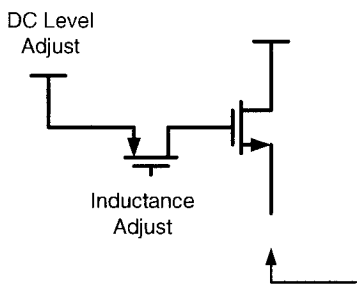


Figure 3.12. Adjustable Active Inductor

While the extra transistor introduces new parasitic capacitances, simulations show that their effects are negligible. Using this new topology, changes in characteristics due to process variations can be compensated.

3.3.3. Simulation Results and Analysis

To verify that the proposed adjustments can reduce the effects of process variations on active inductors, an active shunt-peaked buffer has been designed and simulated in 0.18μm CMOS technology. The buffer uses active inductors with adjustable inductance and DC bias level.

The frequency response under typical process and all four corner cases is shown in Figure 3.13. It shows that the gain in every corner case is different.

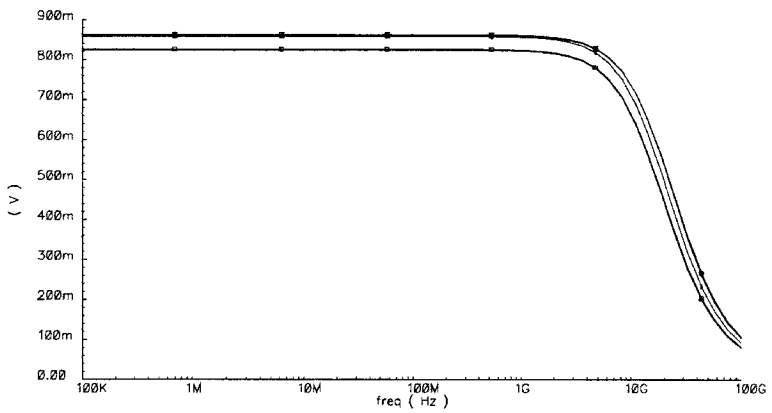


Figure 3.13. Frequency Response of Active Shunt-Peaked Buffer in Presence of Process Variations

When process variations are detected and measured, the parameters of the circuit can be adjusted so that the frequency response becomes similar to that obtained under typical conditions. This can either be done manually or automatically through a look-up table that associates the output of the measurement circuit to the required change in parameters.

To adjust the performance of the active inductor, the relevant parameters are the DC level and inductance effective value. Once these parameters have been changed according to the process corners under which they operate, the frequency response becomes more uniform. This is shown in Figure 3.14.

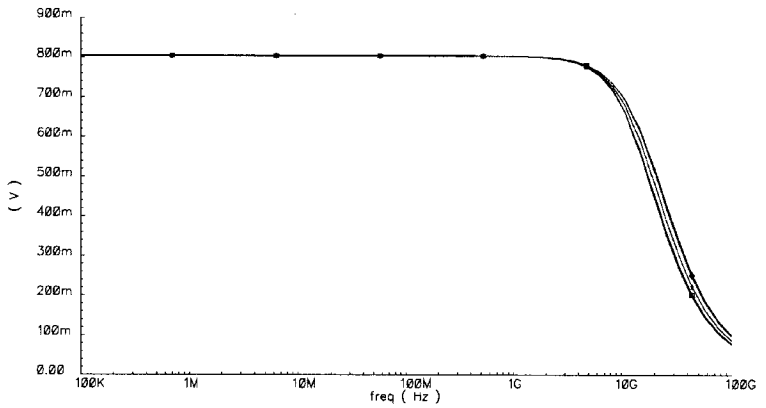


Figure 3.14. Frequency Response of Active Shunt-Peaked Buffer after Process Variation Compensation

It should be noted that, while the simulations presented here have been done for process variations, a similar approach can be used to compensate for temperature variations.

3.3.4. Experimental Results

The simulation results presented in the previous section show that the effects of process variations on active inductors can be reduced if they are detected and quantified. To show the effectiveness of the proposed process and temperature measurement system, a circuit has been designed and fabricated in 0.18μm CMOS technology.

The validity of the measurements made by the circuit is checked by including a ring oscillator on the chip. Since the frequency of the oscillator is affected by process and temperature variations, it can help confirm whether the outputs of the measurement

circuits are correct. For instance, the oscillation frequency in a fast-fast process should be faster than the frequency in a slow-slow process. Therefore, if the measurement circuit indicates a fast-fast process and the oscillator frequency is also fast, it would confirm that the measurement circuit operates properly.

The micrograph of the implemented circuits is shown in Figure 3.15. It includes the measurement circuit, the ring oscillator and the tapered buffers used to drive the signal off chip.

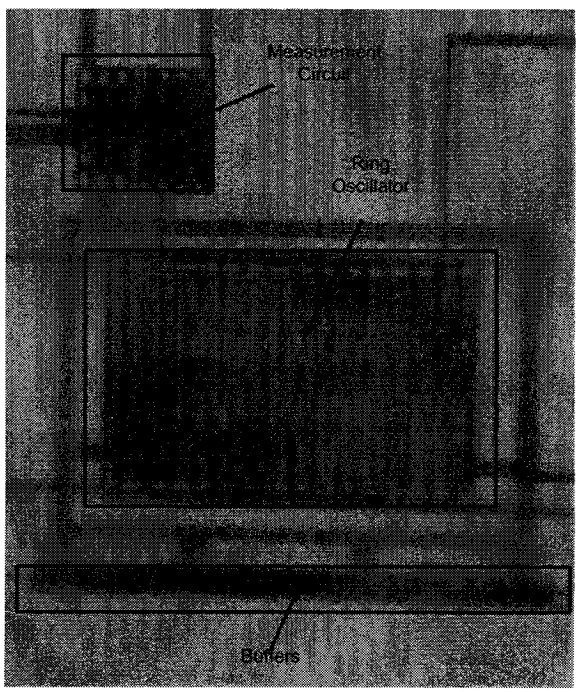


Figure 3.15. Process and Temperature Measurement Circuit and Test Circuit

The ring oscillator used in this application is a 15-stage current-starved ring oscillator. Its output waveform during one of the measurements is shown in Figure 3.16 where the oscilloscope resolution was set to 1 ns per square.

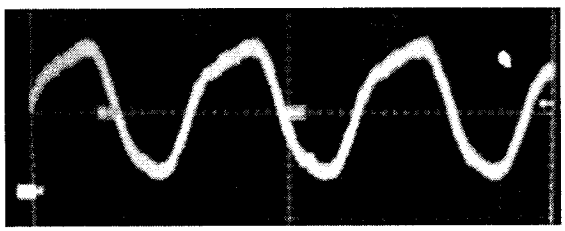


Figure 3.16. Sample Output of the Test Circuit

Each fabricated chip can provide information about its own environment. Therefore, for process variation measurements, a chip can only provide a single measurement. The results taken from our three available chips are shown in Table 3.4.

Table 3.4. Measurements of Process Variation Chips

Chip	P-Type Output	N-Type Output	Oscillator Frequency
#1	0.21V	1.44V	218 MHz
#2	0.32V	0.996V	273 MHz
#3	0.27V	0.38V	290 MHz

Data presented in Table 3.4 show that there is one chip operating under slow-slow corner case (#1), one chip that has typical NMOS and slow PMOS (#2) and one chip that has fast NMOS transistors and slow PMOS transistors (#3). The oscillation frequency observed for each case confirms the measurements made on the circuit. Indeed, the N-type and P-type measurements taken from chip #1 indicate that it operates in a slow-slow

process and the frequency of the oscillator is the slowest of the three cases. Similarly, measurements made on chip #3 show that it comparatively has the fastest transistors. These data are confirmed by its oscillation frequency which is also the fastest of the three.

To test the chip's temperature measurement capabilities, the temperature of the circuit has to be changed. One way of accomplishing this task is by heating the chip's ceramic package with a soldering iron. As the temperature increases, three different outputs are examined: the P-type output, the N-type output and the oscillation frequency. These outputs are shown in Figure 3.17.

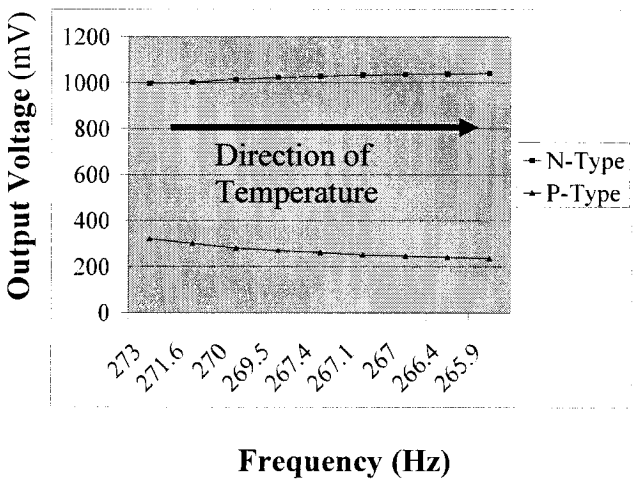


Figure 3.17. Output Voltage as a Function of Oscillator Frequency

The figure shows that the frequency of the oscillator decreases as the temperature of the chip increases. It also shows that the value of the N-type output voltage increases and

that of the P-type output decreases with such change in temperature. This situation follows a similar trend to the one seen in the simulations presented in Figure 3.11. These results confirm that the proposed circuit can be used to detect and quantify process and temperature variations.

3.4. Conclusions

This chapter proposed the use of active shunt-peaking in the design of MCML gates. It shows that a bandwidth improvement of nearly 17% can be obtained over conventional MCML gates with minimal penalty in terms of area and design effort. The small signal model of an active shunt-peaked gate has been analyzed and design equations have been found. Although bandwidth improvements are not as significant as in the case of passive inductors, the novel technique can be used as an alternative when the drawbacks of spiral inductors cannot be tolerated.

One of the problems associated to active inductors is that they are sensitive to process and temperature variations. Consequently, when fabricated, an active shunt-peaked circuit may not operate as expected. To alleviate this concern, improvements have been proposed to the active inductors allowing them to have their DC levels and inductance adjusted. It was shown through simulations that such adjustments can make all corner cases function alike.

In addition, a method of detecting and quantifying process and temperature variations has been developed. It uses two calibrated common-source amplifiers to indicate when

the transconductance of NMOS and PMOS transistors depart from typical values. This measurement circuit has been fabricated in 0.18μm CMOS technology and bench test results prove the effectiveness of the proposed solution.

Chapter 4

High-Speed and Low Jitter Design Techniques for MCML XOR Gates

The scope of this chapter is to explore techniques for the design of XOR gates for high speed and low-jitter applications. It discusses the previous circuits, analyzes their problems and proposes new solutions. Simple and intuitive transient models are developed to provide more insight into the jitter characteristics of the MCML XOR gate. These models are then analyzed and solutions are proposed to design XOR gates with better output jitter. To validate the models, simulations are presented and waveform measurements are made. In addition, one of the proposed XOR gates has been fabricated and tested in 0.18μm CMOS technology to demonstrate its performance.

4.1. Jitter Analysis of the MCML XOR Gate

The jitter problem in XOR gates stems from the fact that most multiple-input MCML gates exhibit asymmetric behaviors. These asymmetric behaviors can cause an output transition to be faster for some input patterns than for others. This gives rise to some uncertainty regarding the time an output transition is generated in response to switching inputs. While this timing variation is a negligible jitter in most low speed applications, its

effect becomes more significant as the speed increases. In applications such as CDRs, the XOR gate is often used to measure the precise phase difference between two signals [28][63]. An asymmetric XOR could introduce jitter that generates errors in the phase measurements, and ultimately causes data recovery errors. It is therefore important to reduce the jitter introduced by the XOR gate.

Jitter can generally be defined as being the maximum time difference between true output transitions and ideal output transitions. Let us define the ideal behavior of a gate as one where the time delay between input and output transitions is the same for all input patterns. However, most MCML gates do not behave that way. By examining the classic XOR circuit of Figure 4.1, it is possible to see that there are two main characteristics that may cause jitter: uneven loading and uneven drive strength. To explore the effects of these characteristics on jitter, it is necessary to understand the transient behavior of the circuit.

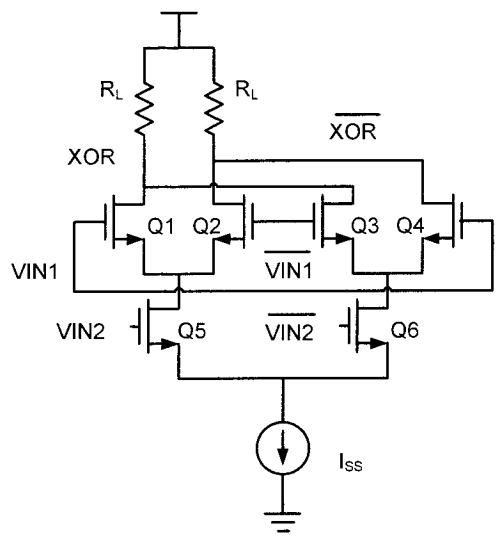


Figure 4.1. MCML XOR Gate

4.1.1. Uneven Drive Strength

In a given circuit, there are often multiple paths for a signal to travel from inputs to outputs. When these different paths have different time constants, it usually results in output jitter. A circuit having this characteristic is said to have uneven delay and drive strength. For example, in Figure 4.1, the transition time for $VIN1-VIN2=00\rightarrow 10$ (inputs VIN1-VIN2 switch from 00 to 10) is not expected to be the same as that for $VIN1-VIN2=00\rightarrow 01$ (inputs VIN1-VIN2 switch from 00 to 01). The latter input transition is expected to produce a longer delay because the intermediate node capacitance at the drain of transistor *Q5* needs to be charged or discharged.

Since jitter results from delay differences between output transitions. it is important to know when these transitions occur. According to the XOR truth table (Table 4.1), output transitions occur only when a single input changes. When both inputs change, the output remains the same.

Table 4.1. XOR Truth Table

A	B	XOR
0	0	0
0	1	1
1	0	1
1	1	0

To analyze the transient response of the XOR gate, it is not necessary to examine the whole circuit. Figure 4.2 shows the segment of the XOR gate through which the signal

passes during a given transition. It includes the two input transistors, the tail current source, the load resistance and a load capacitance.

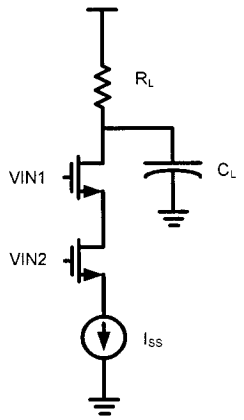
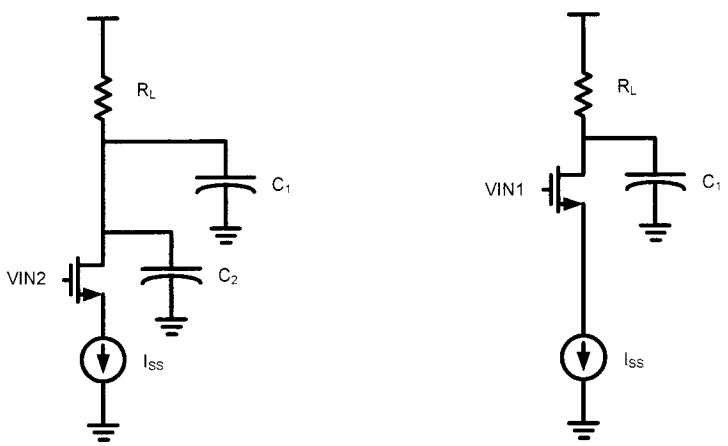


Figure 4.2. MCML XOR Gate Output Branch

When a transition occurs at the output, it can only result from a change on a single input: it can either be on the top transistor or the bottom transistor (Figure 4.3).



a) b)

Figure 4.3. Small-Signal Model for XOR Gate with Transition on a) VIN1 and b) VIN2

Figure 4.3 a) shows the situation where only *VIN2* changes, while *VIN1* remains high. In this case, the *VIN1* input transistor is modeled as being a capacitive load. To simplify the analysis, the channel resistance of the transistors is ignored. This assumption is reasonably accurate as long as the transistor in the current source does not saturate. Figure 4.3 b) shows the case where only *VIN1* changes, while *VIN2* remains high. Since *VIN2* is "ON" and is already biased, it can be considered as being part of a current source carrying the tail current.

In Figure 4.3, C_1 and C_2 represent lumped capacitances. C_1 includes the drain-side parasitic capacitance of the top transistors and the output load C_L. Capacitor C_2 represents the drain-side capacitances of the bottom transistor and the source-side capacitances of the top transistors.

If the timing jitter due to uneven drive strength is taken as the maximum time difference between signals at mid-swing, it can be written as:

$$\Delta t_{DRIVE} = t_{midswing2} - t_{midswing1} . \tag{4.1}$$

In Eq. (4.1), $t_{midswing1}$ is the time required for the output to reach mid-swing when a transition occurs on *VIN1*. Similarly, $t_{midswing2}$ is the time required for the output to reach mid-swing when a transition occurs on *VIN2*.

To calculate the mid-swing time, the first step is to find the transient voltage response of the circuit. This voltage can be found by summing the currents at the output node and solving for the output voltage V_{XOROUT}. The result is given by:

$$V_{XOROUT}(t) = VDD + I_{SS} R(e^{-t/R_L C} - 1) . \tag{4.2}$$

The desired parameter can be found by making V_{XOROUT} equal to the mid-swing voltage and solving for t. With the mid-swing voltage equal to $VDD - \frac{I_{SS}R_L}{2}$, the time required to reach mid-swing is $t_{midswing} = R_L C \ln(2)$, where C is the capacitance seen by the current source. The jitter, Δt_{DRIVE}, is then the difference between the mid-swing times:

$$\Delta t_{DRIVE} = R_L C_2 \ln(2) \tag{4.3}$$

As expected, this equation confirms that the uneven output transition time is mainly due to the intermediate node capacitance C_2 that needs to be charged or discharged.

4.1.2. Uneven Input Loading

The second main source of jitter is caused by uneven input loading. Uneven loading occurs when the two input signals to the XOR do not arrive at the same time due to different capacitances. In the case of the XOR gate, uneven input loading can result from the fact that the circuit driving $VIN1/\overline{VIN1}$ is connected to four transistors (*Q1-Q4*), whereas the one driving $VIN2/\overline{VIN2}$ is connected to two transistors (*Q5-Q6*). Assuming that the circuits controlling these inputs are similar and that the transistor sizes are also similar, the time required to charge and discharge the two inputs will not be the same. To analyze the effects of uneven load jitter, it is important to know how much capacitance is seen at each input and how this affects timing.

The driver connected to $VIN2/\overline{VIN2}$ is loaded with two transistors, which, for simplifying purposes, are both assumed to be operating in the saturation region. Each side

of the differential signal is therefore connected to one transistor. Since the gate-channel capacitance of a transistor in saturation region is roughly $(\frac{2}{3})C_{OX}WL$ [3], the capacitance seen by each line is $(\frac{2}{3})C_{OX}WL$. Similarly, the driver connected to $VIN1/\overline{VIN1}$ is loaded with four transistors. Each side of the differential signal is loaded with two transistors, which have a combined capacitance of $(\frac{4}{3})C_{OX}WL$.

Let us assume that the input drivers are MCML buffers with the same transistor size, resistive load and tail current as the XOR gate. In that case, the voltage at the output of the buffer would be given by the following approximation:

$$V_{BUFOUT}(t) = VDD - I_{SS}R_L(e^{-t/R_LC} - 1). \quad (4.4)$$

To be consistent with the previous section, timing jitter is measured at mid-swing, which is equal to $VDD - \frac{I_{SS}R_L}{2}$. By setting the output of V_{BUFOUT} to mid-swing and solving for t, the following is obtained:

$$t = R_LC\ln(2). \quad (4.5)$$

The analysis, up to this point, has assumed that all the XOR's transistors have the same dimensions. In practice, however, the size of the bottom input transistors is usually not the same as that of the top transistors. Crain and Perrott [15] suggest that the lower transistors be roughly 20% larger than the top transistors. To account for size variations and make the analysis more generic, let α be the channel width ratio of the bottom transistor to the top transistor (assuming that all transistors have minimum length). With this in mind, the time required to charge the gate of the top transistors to mid-swing becomes $t_1 = 2 \cdot R_L \cdot (\frac{2}{3}C_{OX}WL) \cdot \ln(2)$, and the time required to charge the bottom

transistors becomes $t_2 = \alpha \cdot R_L \cdot (\frac{2}{3} C_{OX} WL) \cdot \ln(2)$. The timing jitter contribution from uneven loading can therefore be written as the difference between the two mid-swing times:

$$\Delta t_{LOAD} = t_2 - t_1 = (\alpha - 2)(\tfrac{2}{3} C_{OX} WL) R_L \cdot \ln(2) \qquad (4.6)$$

The above equation shows that, if the bottom transistors are twice the size of the top transistors, the uneven input loading component of the jitter would be eliminated. It should be noted that this relation only holds for first order approximations.

4.1.3. Glitches

In addition to uneven loading and uneven drive strength, glitches are another element that could contribute to degrading the output signal quality of the XOR gate. In this context, a glitch is defined as an undesired momentary change in voltage that may, or may not, be large enough to cause a change in logic level. Glitches in logic gates can be generated when a transition occurs at the input and the expected output does not change. In the specific case of the XOR gate, this happens when the value of both input signals changes at the same time. During this transition time, the current flow switches from one branch to another and follows a different path. This change in path affects the distribution of current in the branches and results in a glitch. Consider the circuit in Figure 4.4. The figure shows an XOR gate whose input is $VIN1 - VIN2 = 11$ and whose initial current flow is indicated by the solid arrow on the left. If the input suddenly changes to

$VIN1 - VIN2 = 00$, the current would follow the path shown by the dotted arrow on the right. The time required to switch branches causes a glitch at the output.

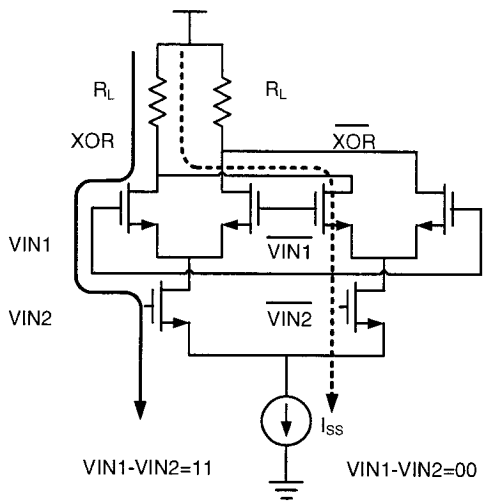


Figure 4.4. Possible Change in Current Path Causing Glitches

A glitch does not always contribute to jitter. To understand the relation between the glitch and jitter, let us refer to the eye diagram of Figure 4.5.

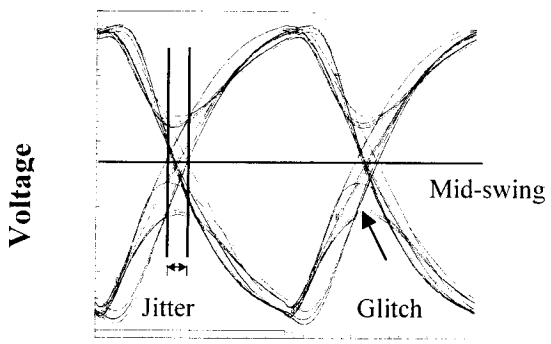


Figure 4.5. Jitter Measurement

On an eye diagram, the jitter can be seen as the thickness of the transition area measured at mid-swing. When the magnitude of the glitch is smaller than mid-swing, it does not contribute to jitter. This is the situation shown in Figure 4.5, where the glitches have no impact on jitter. However, as the magnitude of the glitch gets larger and as its duration gets longer, it will start having an impact on the quality of the eye diagram. When the magnitude of the glitch is at least equal to mid-swing, its contribution will be significant only if it enlarges the transition area.

For the purpose of this analysis, let us assume that the transitions on the input signals are instantaneous. Even though this assumption may seem unrealistic, it helps provide an intuitive model of the glitch and illustrate its underlying causes. In addition, the validity of the derived results is confirmed by simulations that show that the difference between glitches caused by instant transitions and typical transitions is less than 10%.

When these transitions occur, as shown in Figure 4.4, the tail current is transferred from one branch to another. The current-carrying bottom transistor immediately begins operating as a current source. The top transistor, however, does not yet have the required V_{GS} and V_{DS} to allow current flow. In order for this to happen, the intermediate node needs to get charged to the required voltage. As the input transitions are assumed to be instantaneous, the current shuts off instantaneously in one branch while the other branch does not immediately draw current. This allows the resistor to pull up the output node. This resistive pull-up creates the output glitch.

The magnitude and the duration of the glitch depend on the moment at which current flow begins on the destination branch. For proper current flow, the top transistor needs to

be biased properly. Since the bottom transistor can be approximated by a constant current source, the time required to charge the node is given by:

$$t_{GLITCH} = \frac{C_2 \cdot V_S}{I_{SS}} \tag{4.7}$$

In the previous equation, C_2 is the same capacitance as the one defined in Figure 4.3 and V_S is the required voltage change in the internal node to bias the top transistor to the right V_{GS}. During time t_{GLITCH}, it is assumed that the other transistors do not draw any current at the output node. The voltage glitch that occurs at the output is equal to:

$$V_{GLITCH} = I_{SS} R_L \cdot (e^{-t/(R_L \cdot C_1)} - 1) \tag{4.8}$$

Replacing t by the value of t_{GLITCH} yields

$$V_{GLITCH} = I_{SS} R_L \cdot (e^{-(\frac{C_2 \cdot V_S}{I_{SS} \cdot R_L \cdot C_1})} - 1) \tag{4.9}$$

To ensure that the glitch does not have a negative impact on the jitter, its magnitude needs to be kept below $\frac{I_{SS} R_L}{2}$. With some algebraic manipulation, the design requirement to produce acceptable glitches is given by:

$$C_1 \cdot I_{SS} R_L > \frac{C_2 \cdot V_S}{\ln(2)} \tag{4.10}$$

This equation shows that, in order to reduce the effects of glitch on jitter, the load capacitance and/or the output swing needs to be large enough. It also indicates that a small intermediate node capacitance and/or a low internal swing could help reduce glitches.

4.1.4. Total Deterministic Jitter

Provided the circuit does not have significant glitches, the total deterministic jitter can be calculated simply by adding its two components given by Eq. (4.3) and Eq. (4.6):

$$\Delta t_{TOTAL} = (\alpha - 2)(\tfrac{2}{3} C_{OX} WL) R_L \ln(2) + R_L C_2 \ln(2) \tag{4.11}$$

It should be pointed out that for $\alpha < 2$, the value of Δt_{LOAD} is actually negative. This is not a deficiency in the model. It simply states that, in this situation, the uneven load partially compensates for the uneven drive. The reason is that the uneven drive causes one output transition to be slower and the uneven loading causes that same transition to start earlier.

4.1.5. Jitter Compensation through Sizing

Since Δt_{LOAD} partially compensates for Δt_{DRIVE}, it may be possible to find parameters that would make the total jitter equal to 0. To find these parameters, Eq. (4.11) may be expanded as follows:

$$\Delta t_{TOTAL} = (\alpha - 2)\ln(2)(\tfrac{2}{3} C_{OX} WL) R_L + \ln(2) R_L (\tfrac{2}{3} C_{OX} WL + C_{DRAIN}) \tag{4.12}$$

In Eq. (4.12), C_2 is split into its two components: the source-side channel capacitance of the top transistor and the drain-side capacitance of the bottom transistor. To make the total jitter equal to 0, the value of α needs to follow this relation:

$$\alpha = \frac{(\tfrac{2}{3} C_{OX} WL) - C_{DRAIN}}{(\tfrac{2}{3} C_{OX} WL)} \tag{4.13}$$

Here, C_{DRAIN} represents the drain-side capacitance of the bottom transistor. This equation shows that to obtain 0 jitter, the value of α should be less than one. However, it is not possible to be more explicit about the value for α as it would require the process-specific value of the overlap and the depletion capacitance at the drain node.

4.1.6. Jitter Compensation through Controlled Imbalanced Drivers

The model that was used in the previous discussion assumed that drivers controlling the two inputs are identical. This constraint is unnecessary and only serves to simplify the analysis. By having drivers that differ in strength, it is possible to have a better control over the jitter resulting from an uneven load. In addition, it allows the optimization of jitter and other performance criteria to be done simultaneously. By increasing the drive strength of the bottom driver and decreasing the drive strength of the top driver, a larger load jitter can be obtained. This uneven load jitter can be used to compensate for the drive jitter and ultimately, reduce the total jitter to 0.

To analyze this new situation, let R_{TOP} be the load resistance of the circuit driving the top transistors and the R_{BOT} be the load resistance of the circuit driving the bottom transistors. The jitter due to uneven loading could be recalculated using these new parameters:

$$\Delta t_{LOAD} = (\tfrac{2}{3} C_{OX} WL)(\alpha R_{BOT} - 2R_{TOP})\ln(2) \tag{4.14}$$

The drive jitter equation, which is not affected by the controlled imbalanced drivers, can be rewritten in another form in order to show the individual components of C_2:

$$\Delta t_{DRIVE} = R_L(\tfrac{2}{3}C_{OX}WL + C_{DRAIN})\ln(2) \tag{4.15}$$

The total jitter can then be found by adding the load and the drive jitters, yielding

$$\Delta t_{TOTAL} = (\tfrac{2}{3}C_{OX}WL)(\alpha R_{BOT} - 2R_{TOP})\ln(2) + R_L(\tfrac{2}{3}C_{OX}WL + C_{DRAIN})\ln(2) \tag{4.16}$$

By setting the previous equation equal to 0 and by solving for α, it is possible to find the transistor size ratio that would tend to eliminate the jitter. The resulting value for α is found to be

$$\alpha = -\left(\frac{3C_{DRAIN}R_L + 2C_{OX}WL(R_L - 2R_{TOP})}{2C_{OX}WL \cdot R_{BOT}}\right) \tag{4.17}$$

Having input drivers of different sizes relaxes the constraint on the sizing of the XOR gate transistors. In addition, it allows for the transistors of the XOR gate to be optimized for better performance while still theoretically obtaining 0 output jitter.

4.2. Symmetric Gates

The first order analysis provided in the previous section shows that jitter in an MCML XOR gate can theoretically be removed if the transistors are sized properly. Although the proposed sizing methods improve output jitter, the design process can be complicated and the resulting gates can be sensitive to process variations. In this section, we propose three symmetric solutions that are simple and that, in principle, can be used to reduce the output jitter to 0. The proposed circuits are also much less sensitive to process variations.

4.2.1 Criss-Crossed XOR Gates

A known method for removing deterministic jitter from an asymmetric gate is to combine two such gates together using criss-crossed connections [26][56][64]. An example of this connection is shown in Figure 4.6. The driver connected to input *VIN1* on one gate is connected to input *VIN2* on the other gate and vice versa. The outputs of the two gates are then connected together, forming a symmetric XOR gate.

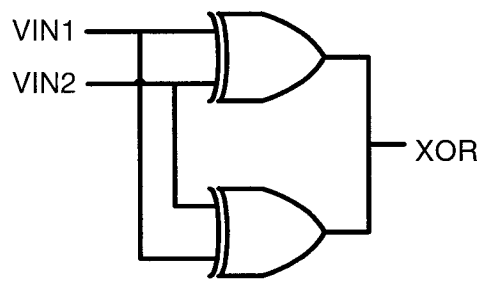


Figure 4.6. Gate Level Criss-Crossed Symmetric XOR

The criss-crossed MCML XOR gate implemented at the transistor-level is shown in Figure 4.7.

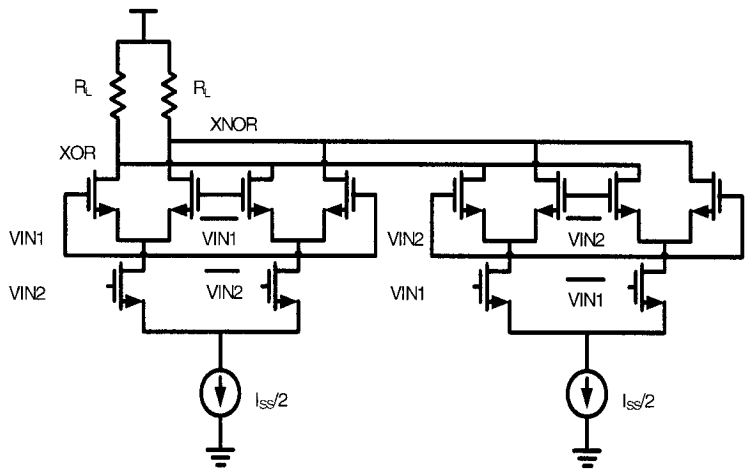


Figure 4.7. Criss-Crossed Symmetric XOR Gate

The small signal model for the circuit is shown in Figure 4.8. In this model, C_3 is the capacitance contributed by each of the top transistors, excluding the capacitive load. Similarly, C_4 is the capacitance seen by the bottom transistor. The situation described in Figure 4.8 is the case where *VIN1* switches from 0 to 1 while *VIN2* remains high ($VIN1 - VIN2 = 01 \rightarrow 11$).

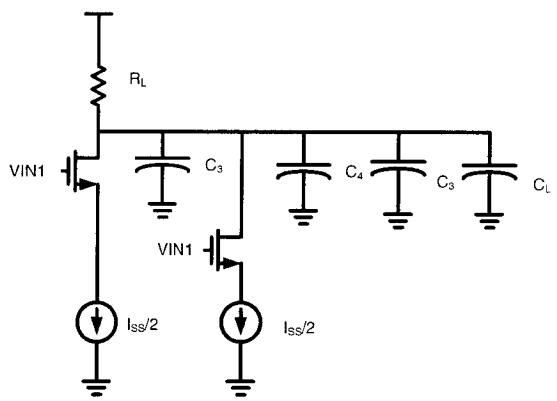


Figure 4.8. Small-Signal Model for the Criss-Crossed Symmetric XOR Gate

When *VIN1* goes high, one current source sees capacitance C_3 and C_L, whereas the other source sees C_3, C_4 and C_L. The transient output voltage during a transition is given by

$$V_{CRISS-CROSS}(t) = VDD + I_{SS}R_L(e^{-t/R(2C_3+C_4+C_L)} - 1). \quad (4.18)$$

While Figure 4.8 shows the transition of $VIN1 = 0 \rightarrow 1$, the same model applies for a change on any of the input signals. Since the transient equations are the same for all transitions, the time at which mid-swing occurs should be the same for all input patterns. Therefore, the theoretical output jitter is equal to 0.

Even though the circuit removes the jitter contribution during transitions, the glitching issue has not been addressed. The gate must be designed carefully to minimize glitching so that jitter characteristics are not affected [66].

4.2.2. 4-Branch Symmetric XOR

In order to maintain glitch magnitude to a minimum and to reduce jitter contribution, a 4-branch symmetric gate is proposed (Figure 4.9) [10]. This gate can be seen as being composed of four branches, each representing a minterm of the XOR or of the XNOR function. The size of each transistor can be roughly half of the values used in Figure 4.1 because there are two parallel paths driving every branch. This configuration allows for the input drivers to be loaded equally since they are all connected to the same number of transistors. In addition, the criss-crossed connections allow for a symmetric drive.

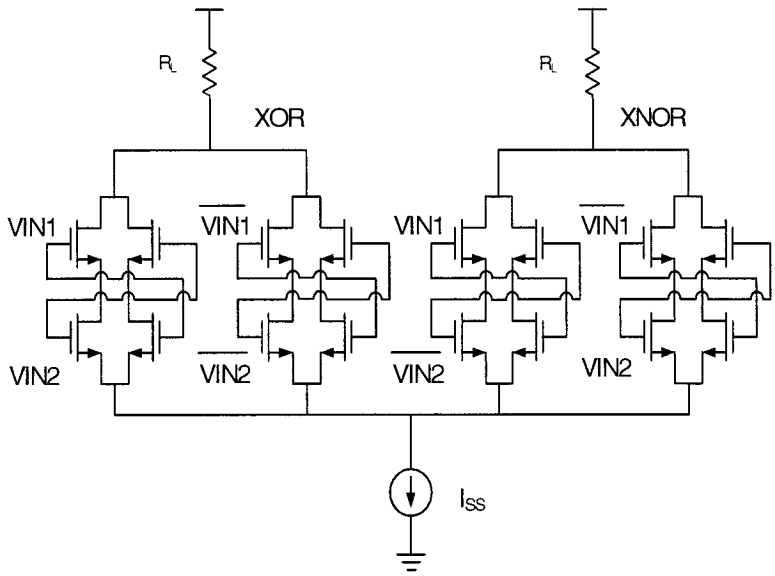


Figure 4.9. 4-Branch Symmetric XOR Gate

For the analysis of this design, it is useful to only consider the branch that is being turned "ON" (Figure 4.10). The branch is modeled as an RC circuit to simplify the analysis and to keep it consistent with previous gate analyses.

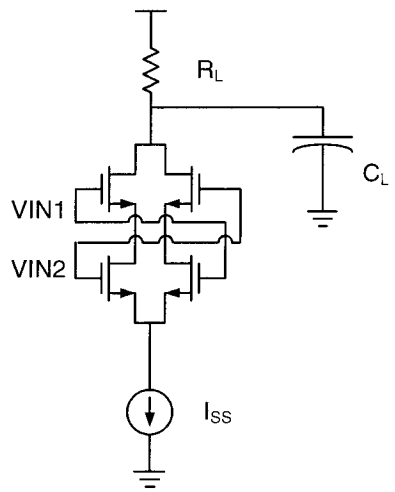


Figure 4.10. Single Branch of the 4-Branch Symmetric XOR Gate

Figure 4.11 shows a model of the XOR gate branch when input *VIN1* is turned "ON". C_5 and C_6 are lumped capacitances that include the various parasitic capacitances associated to the top and bottom transistors respectively. Their values are about half of C_2.

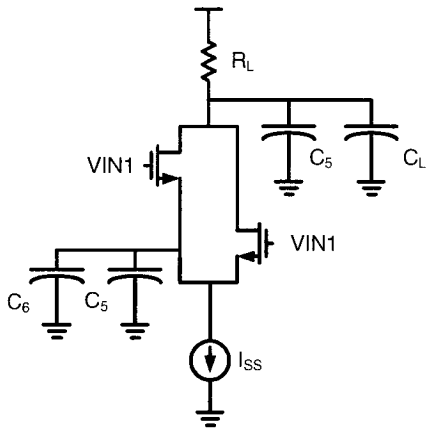


Figure 4.11. Symmetric XOR Model

It should be noted that, although Figure 4.11 shows input *VIN1*, it can be replaced by any other input without loss of generality. Therefore, irrespective of which input is switching, the transient output voltage of the model is given by

$$V_{4SYM}(t) = VDD + I_{SS}R_L(e^{-t/R_L(2C_5+C_6+C_L)} - 1) \tag{4.19}$$

Since all falling transitions follow the same transient equation, it means that their mid-swing times are all the same. A similar analysis performed on rising transitions also yielded a unique equation for all possible input patterns. Therefore, it can be concluded that the deterministic jitter of this circuit is theoretically equal to 0.

Another benefit of this topology is that it helps reduce glitching that occurs when both inputs change at the same time. Recall that glitching occurs when the current flow switches from one branch to another because it is not done instantaneously. The intermediate nodes need to be set to a certain level before steady current flow can be established. When this symmetric topology operates at high speed, it will be shown in the next subsection that, at least one of two intermediate node voltages will have the required level for current flow at all times. With the correct intermediate node voltage, current switching is faster and glitching is reduced considerably.

4.2.3. 4-Branch Symmetric XOR with Reset

This section proposes a symmetric configuration that resets the internal branches of the previously described XOR gate to reduce its jitter even further. In the analysis performed in the previous subsection, it was shown that jitter resulting from the 4-branch symmetric XOR is 0 if we assumed that the intermediate nodes between the criss-crossed transistors are always initially set to the same potential. This, however, is not always the case. The voltage at the intermediate nodes is dependent on the preceding input signals. Let us re-examine the circuit in Figure 4.10. When the input is equal to $VIN1 - VIN2 = 00$, all the transistors are switched off and the intermediate nodes retain their previous values. If the previous input was $VIN1 - VIN2 = 11$, both intermediate nodes would have a relatively low voltage. With low intermediate voltages, when the input is "HIGH", the transistors have the required V_{GS} and V_{DS} to operate as current sources. For

the case where $VIN1 - VIN2 = 01$, one of the intermediate nodes would have a higher voltage and the other would have a lower voltage. In this case, only one of the branches has the required potential to operate as a current source. The other branch would need to discharge its intermediate node before it is able to draw the required current.

There are two conclusions to draw from the previous discussion. The first conclusion is that, when the input goes "HIGH", at least one of the intermediate nodes will be driven by transistors that have the required voltages to act as a current source at all times. The second conclusion is that, since the initial voltages on the intermediate nodes are dependent on the previous input, the symmetric XOR may still generate some jitter. This is depicted in Figure 4.12.

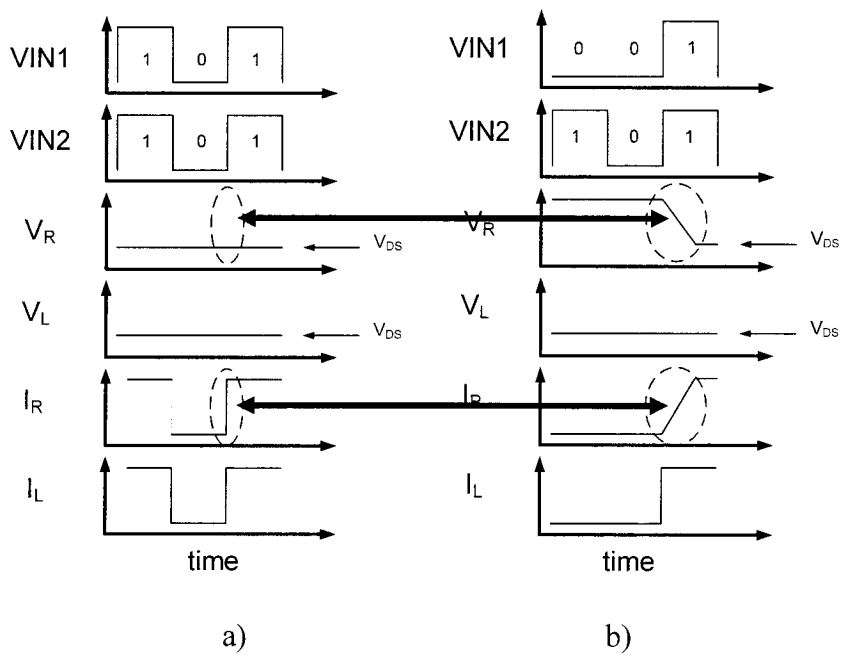


Figure 4.12. Intermediate Nodes of the 4-Branch Symmetric XOR Gate with Reset when Inputs are a) $VIN1 - VIN2 = 11 \rightarrow 00 \rightarrow 11$ and b) $VIN1 - VIN2 = 01 \rightarrow 00 \rightarrow 11$

In Figure 4.12, V_L represents the intermediate node voltage on the left branch and I_L represents its current. Similarly, V_R and I_R respectively represent the current and the intermediate node voltage on the right branch. When the inputs follow the pattern $VIN1-VIN2=11\rightarrow 00\rightarrow 11$ as depicted in Figure 4.12 a), the intermediate voltages are set low before being disconnected. When the input returns to $VIN1-VIN2=11$, the current flow will begin more rapidly because the transistors are already properly biased. In Figure 4.12 b), when the input switches from $VIN1-VIN2=01\rightarrow 00\rightarrow 11$, only one of the intermediate node voltages is set low, whereas the other one has a higher voltage. When inputs become $VIN1-VIN2=11$, the transition will be slightly slower than the previous case because current flow does not start as fast on one of the branches.

In order to remove this remaining source of jitter, the intermediate nodes can be reset to a known value whenever the branch is not driven. A circuit implementing this functionality is shown in Figure 4.13.

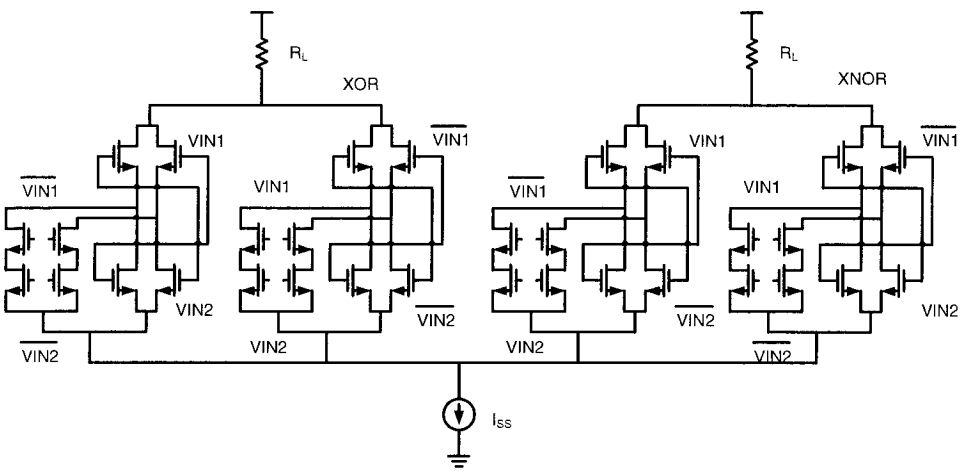


Figure 4.13. 4-Branch Symmetric XOR Gate with Reset

A branch is not driven when its four transistors are not conducting. When this occurs, extra transistors are used to set the intermediate node voltage to that of the tail current source. Through this process, the time required to discharge this node is reduced when the branch is selected. One of the drawbacks of this approach is the addition of parasitic capacitance on the intermediate nodes, which could degrade the maximum achievable speed.

4.3. Simulation Results

To validate the theoretical model and to show more concrete evidence of the circuits' performance, the different XOR gates have been designed and simulated in 0.18 μm CMOS technology. To simulate under more realistic conditions, the XOR gates are driven by MCML buffers and are also loaded with equal size buffers. The different gates used in this experiment are listed and described in Table 4.2. The *reference* column assigns a letter to each gate for faster identification in subsequent figures and tables. Their descriptions are presented in the *Design Note* column of Table 4.2. The different topologies are arranged so that references a) and b) are based on MCML gates whereas references c) to e) are based on symmetric configurations.

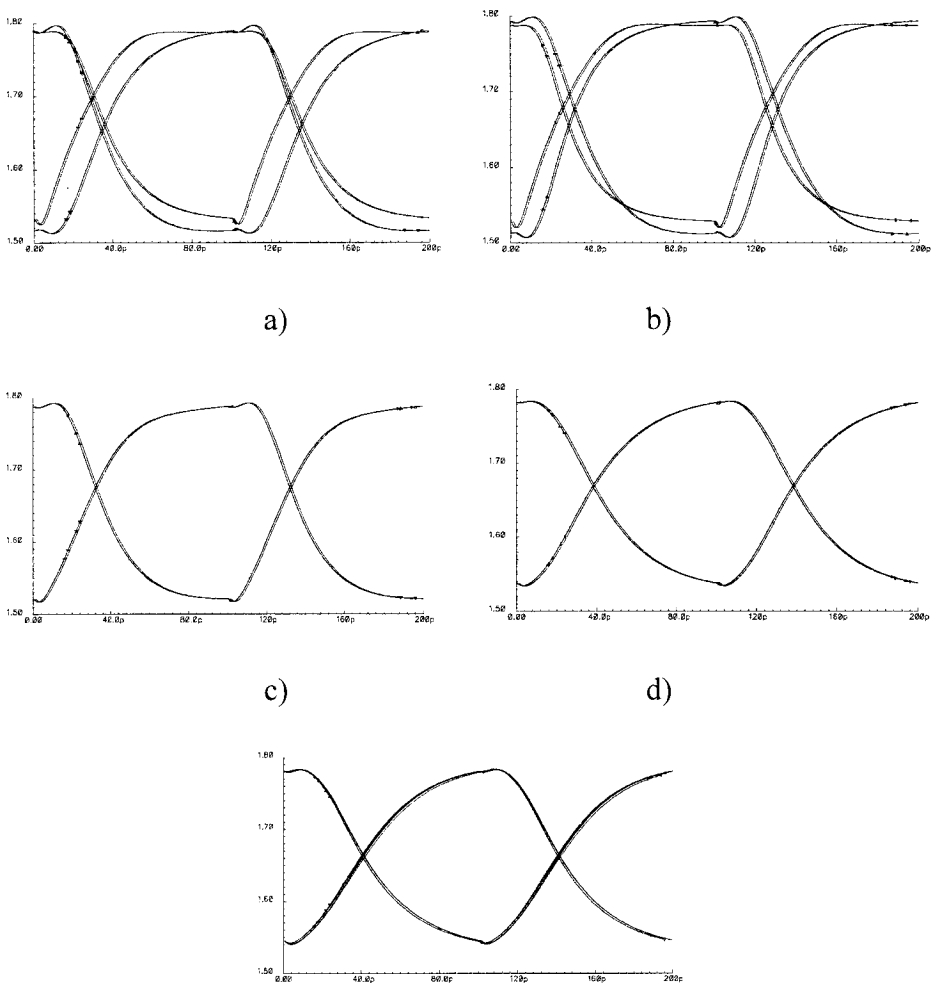
Table 4.2. Description of Simulated XOR Gates

Reference	XOR Topology	Design Note
a)	MCML Same top/bottom	Top and bottom transistors have the same dimensions (Figure 4.1)
b)	MCML Controlled Imbalanced Drivers	Driver circuits have different drive strength and XOR sizes are adjusted accordingly. (Figure 4.1)
c)	Symmetric Criss-Cross	Top and bottom transistors have the same size on both criss-crossed gates (Figure 4.7)
d)	Symmetric 4-Branch	All transistors are half the W/L ratio of the ones in a) (Figure 4.9)
e)	Symmetric 4-Branch with Reset	XOR gate is the same as d). Added transistors are all minimum size. (Figure 4.13)

Two particular input sequences are used to test the performance of the XOR gates: a periodic input and a random input. The first test involves stimulating the design with two square waves operating at the same frequency and with a phase offset of ($\pi/2$). This test helps characterize the circuits in presence of a periodic input as in the case of a phase detector in a PLL.

With the input pattern previously described, it is expected that the output of the XOR gate be a periodic signal with 50% duty-cycle oscillating at twice the frequency of the input signals. The input signals have been chosen so that the duration of the output bit would be 100 ps. The simulated output signals are folded and overlapped to form the eye diagrams shown in Figure 4.14. The figure shows that, for periodic input applications, the conventional XOR MCML topologies exhibit horizontal eye closure. This is primarily due to uneven duty-cycles caused by gate asymmetry. The outputs from the symmetric topologies have consistent transition times and transition moments. Results taken from the simulated eye diagrams are presented in Table 4.3. The horizontal eye opening

percentage is calculated using the expected bit duration of 100 ps as a reference. On the other hand, the vertical eye opening percentage uses the DC voltage swing as a reference.



a) b) c) d) e)

Figure 4.14. Voltage vs. Time Eye Diagram for a (π/2) Phase Offset Periodic Input (See Table 4.2 for Definitions of a), b), c), d) and e))

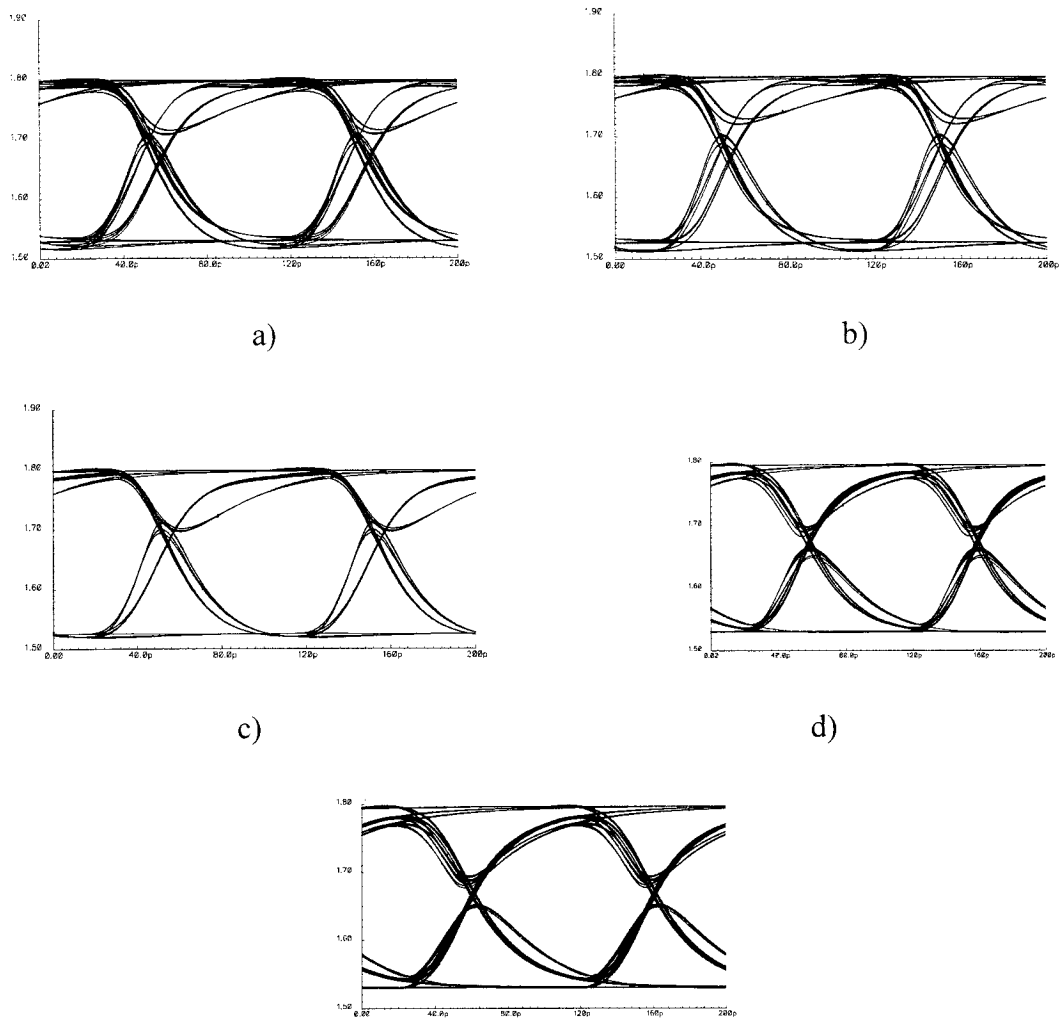
Table 4.3. Eye Diagram for (π/2) Phase Offset Input (100ps output bit time)

	Horizontal Eye Opening (%)	Vertical Eye Opening (%)
a)	85.6	87.5
b)	92.9	90.7
c)	98.8	96.6
d)	98.9	94.2
e)	97.9	94.8

It should be noted that XOR gate asymmetry is not due primarily to bandwidth limitation, but rather to the asymmetry of the gate. Even when bandwidth enhancement techniques are used [5], the output asymmetry is still present.

The second type of input signal used to characterize the circuits is a random bit pattern. Instead of relying on a pseudo-random bit sequence (PRBS) generator circuit, an ideal AHDL model was used. The reason for using this ideal model is to limit the jitter introduced from external sources. The output waveforms are manipulated to form the eye diagrams of Figure 4.15. This allows the deterministic jitter to be measured. Simulations are run with the XOR gates being exposed to data rates of 10 Gb/s.

By examining Figure 4.15 a), it is possible to see the different jitter components described in Section 4.1: the load and drive jitters that cause uncertainty during transitions and the glitch-related jitter. In Figure 4.15 a) and b), load and drive jitters are responsible for the different transition times and the different moments at which they occur.



a) b)

c) d)

e)

Figure 4.15. Voltage vs. Time Eye Diagram for a Random Input (See Table 4.2 for Definitions of a), b), c), d) and e))

When symmetric solutions are used, transitions are more uniform. Although the criss-crossed symmetric gate has the most uniform transitions of all simulated gates, it does not solve the glitch issue which is responsible for its large jitter contribution. This is shown in plot c) of Figure 4.15, where the magnitude of the glitches is larger than mid-swing. On

the other hand, the two solutions derived from the 4-branch topology are not affected to the same extent by the glitches. These results confirm that the glitch model proposed in a previous section provides a good description of the behavior. It also shows that the design of XOR gates based on the conventional MCML gate needs to be done carefully to reduce the magnitude of the glitches.

The dimensions of the eye diagrams are presented in Table 4.4. Results show that there is a slight decrease in the vertical eye opening. Since this chapter is concerned with jitter reduction, such small decreases in vertical eye opening are acceptable.

Table 4.4. Eye Diagram for Random Input

	Horizontal Eye Opening (%)	Vertical Eye Opening (%)
a)	85.4	78.6
b)	88.2	79.2
c)	89.9	83.3
d)	95.6	79.5
e)	96.4	74.6

Although the first order mathematical models show that all the proposed solutions could eliminate the output jitter, this was not observed during simulations. The discrepancy is due to certain effects that are not taken into account by the model, such as transistor channel resistance and transistor operating region. In addition, at such high bit rates, the effects of inter-symbol interference (ISI) also contribute to the jitter.

Results in Table 4.4 show that the circuit with the worst eye diagram is the classic MCML XOR gate with transistors of equal dimensions. All symmetric configurations seem to have improved characteristics, although the benefits of the criss-crossed topology

are not noticed due to the large glitches. The 4-branch symmetric XOR gate with reset is the gate with the best performance.

To broaden the scope of this analysis, the jitter performance of the different XOR gates has also been examined with different bit durations ranging from 100ps to 500ps. The results are plotted in Figure 4.16.

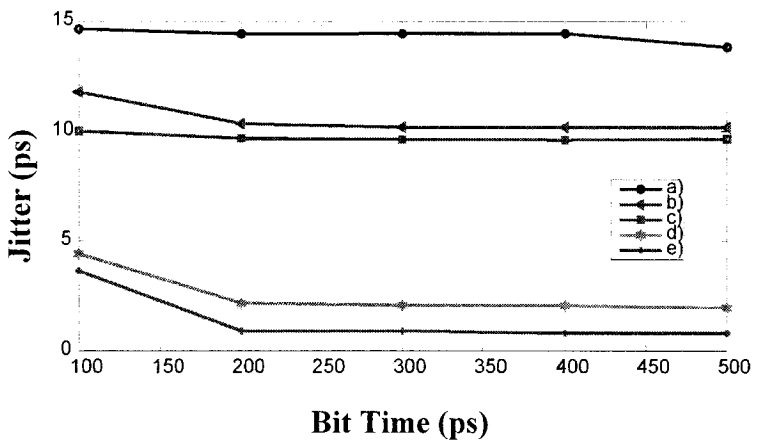


Figure 4.16. Jitter as a Function of Bit Period

Figure 4.16 shows that, depending on the chosen solution, reduction in output jitter ranges from about 26% to near 95% at slower bit rates. At 10 Gb/s, jitter improves by up to 75%. The symmetric gate topologies demonstrate better performance than the optimized MCML XOR even though all models predict 0 output jitter. This can be attributed to the fact that jitter compensation is a process that is dependent on the accuracy of the model while the symmetric gate compensation is based on matching, which works even on characteristics that are not present in the model.

4.4. Experimental Results

The circuit-level contribution of this chapter comes from the 4-branch symmetric XOR gate. To demonstrate its functionality, the gate has been fabricated in 0.18μm CMOS technology and packaged in a CQFP. Since the goal is to check for functionality rather than measuring its high speed performance, the inputs and outputs are connected to regular analog pads. The relevant portion of the micrograph is shown in Figure 4.17.

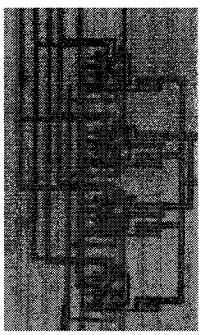


Figure 4.17. Micrograph of the 4-Branch Symmetric XOR Gate

The input signals to the chip are generated by the Hewlett Packard 81130A pattern generator and the experimental setup is shown in Figure 4.18.

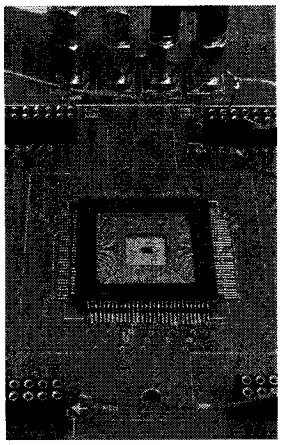


Figure 4.18. Experimental Setup for the 4-Branch Symmetric XOR Gate

The pattern generator has been programmed to generate a 2^{16}-1 PRBS for one of the inputs while leaving the other input steady. The reason for this setup is that only four high speed connections are available on the board and two of them are required to display the output signal. The signals coming from the signal generator provide a stimulus with a speed of 2 MHz. Figure 4.19 shows the output when the chip is stimulated with the random input. The thick lines are due to an "infinite persistence" switch that was set on the oscilloscope which overlaps the signals. Under these conditions, no convincing jitter measurements have been obtained due the large amount of ISI and signal reflection coming from the setup.

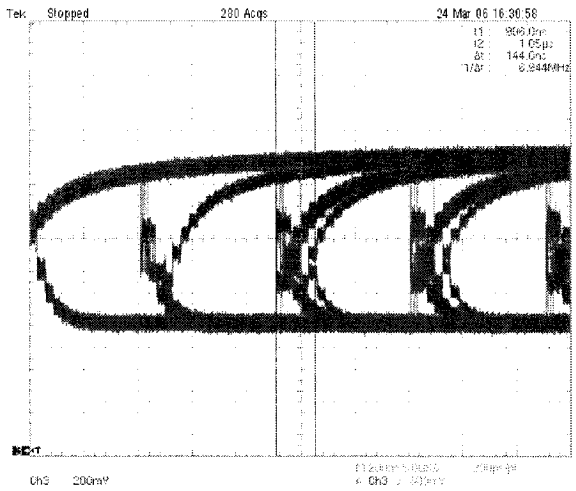


Figure 4.19. Chip Output in Response to $\pi/2$ Phase Offset Square Waves

The verification of gate symmetry in the proposed XOR is done using DC analysis. To perform this test, one differential input signal is held stable while the other input is swept over its operating range. This task is performed for every possible combination of inputs as shown in Table 4.5.

Table 4.5. Test Cases for the Symmetric XOR Gate

	$VIN1/\overline{VIN1}$	$VIN2/\overline{VIN2}$
Case 1	0/1	Sweep
Case 2	1/0	Sweep
Case 3	Sweep	0/1
Case 4	Sweep	1/0

The input voltage sweep is done using 17 discrete points. The output results for all four test cases are plotted and overlapped in Figure 4.20. The figure shows that the DC input-output relation is nearly identical for all input combinations. This shows that the

structure of the XOR gate is indeed symmetric. The maximum difference between the measured output signals at any given point is less than 10mV. This difference can be attributed to process variations. It can be shown that a variation of 10% on any single transistor size can cause such departure from perfect symmetry in the 4-branch symmetric XOR. For the purpose of comparison, DC simulations have been performed on the MCML XOR and results show output voltage differences of about 40mV.

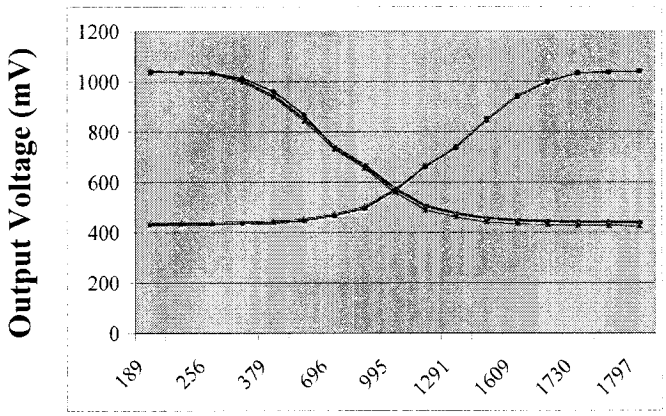


Figure 4.20. DC Measurements of the 4-Branch Symmetric XOR Gate

4.5. CONCLUSIONS

This chapter presented different strategies to design high speed and low jitter XOR gates. We started by developing a simple and intuitive model from which the transient characteristics of the MCML XOR gate have been analyzed. The model accounts for issues such as uneven load jitter, uneven drive jitter and glitches. The analysis of this

model shows that properly sizing the transistors of an XOR gate can theoretically remove the output jitter. The drawback of this approach is susceptibility to process variations and large glitches. For high speed and low-jitter applications, this chapter presented three solutions: the criss-crossed configuration, the 4-branch symmetric XOR and the 4-branch with reset. Depending on the chosen solution, reduction in jitter ranges from 26% to 95% at lower frequencies. Unlike the results expected from the mathematical model, the simulated output jitter is not 0. This is due to the fact that certain characteristics are not modeled. At high frequencies, the effects of ISI become a factor and the improvements are a little lower. According to simulations, with 100 ps input data bit time, it was possible to reduce the jitter by 75% using symmetric topologies.

The results show that jitter reduction using the symmetric XOR topologies yield better performance than circuits based on the MCML XOR. This is true because the symmetric gates are based on load and drive matching, and even non-modeled characteristics are matched in the process. Although the criss-crossed symmetric configuration yielded the smallest transition area, the glitches incurred actually increases the total jitter. Solutions based on the 4-branch symmetric XOR gate exhibit the least amount of glitching, reduce jitter and could potentially simplify the design process.

To demonstrate the functionality of the 4-branch symmetric XOR gate, a chip has been fabricated in 0.18μm CMOS technology. Bench tests show that the XOR gate operates properly and is indeed symmetric.

Chapter 5

Design of a High-Speed Differential Frequency-to-Voltage Converter and its Applications

The process of converting frequency to voltage is an important task in applications such as FLLs, PLLs and temperature stabilized ring oscillators [9]. Its main function is to convert an oscillating signal frequency into a corresponding voltage that can be processed conveniently. While there are multiple ways of converting a frequency into a corresponding voltage, two of the most important ones are counter-based and integrator-based converters [39]. The method based on counters uses synchronous digital circuitry with a very high speed clock to count the number of clock cycles that elapse within one period of the measured signal. This number can then be converted to a voltage or can be processed digitally. The counter-based conversion process is shown in Figure 5.1. It has the advantage of being simple and precise, provided the clock signal has a frequency much higher than the input signal.

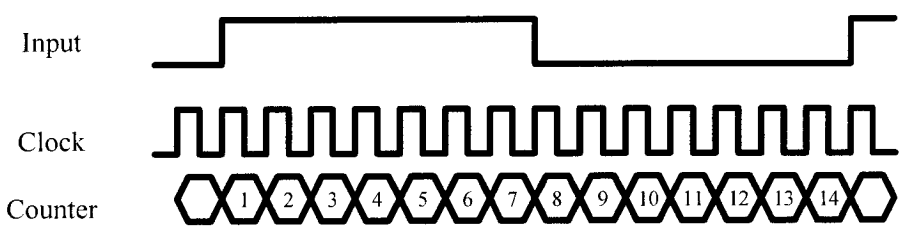


Figure 5.1. Counter-Based Conversion

The other prevalent method uses the integration operation to convert frequency to voltage (Figure 5.2). To describe its functionality, let us assume that the incoming signal is a 50% duty cycle square wave with a magnitude of $\pm A$. The output of the integrator would be a ramp function with a slope of $\pm Ak$ and a peak value of $Ak(\frac{T}{2})$, where T is the period of the incoming signal and k is the gain of the integrator. If the integrated signal is sampled at its peak, the result would be an output value that is proportional to the input period and hence, inversely proportional to the input frequency. The advantage of using this method is that it does not require a high speed clock that would limit the maximum input frequency that can be processed. However, this circuit has intrinsic accuracy limitations caused by sampling/switching noise and non-linearities in the integrator.

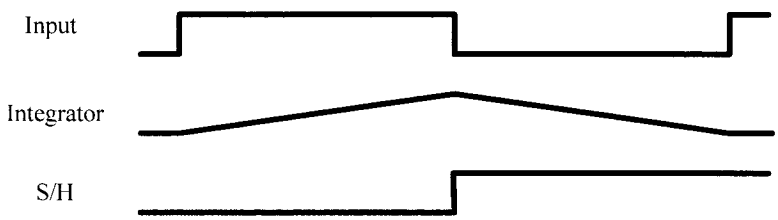


Figure 5.2. Integrator-Based Conversion

For high speed applications, integrator-based FVC are preferred. In 2001, Djemouai et al. proposed an algorithm and circuit implementation for an integrating FVC [20]. While this circuit operates reliably at speeds of several hundred MHz, it shows deficiencies when used in multi-GHz applications. As discussed in Chapter 2, the main

reasons are that rail-to-rail voltage swings and pulse-controlled sequencing of events limit the maximum operating speed.

5.1. Single-Phase Architecture

To overcome the deficiencies of the previous solutions, this thesis proposes a new single-phase algorithm for FVCs. This algorithm is a modification of the one proposed by Djemouai et al. It consists of removing one of the phases to transform the converter into a two phase system as shown in Figure 5.3. With the proposed conversion algorithm, phases *S1* and *S2* from Djemouai's algorithm are combined.

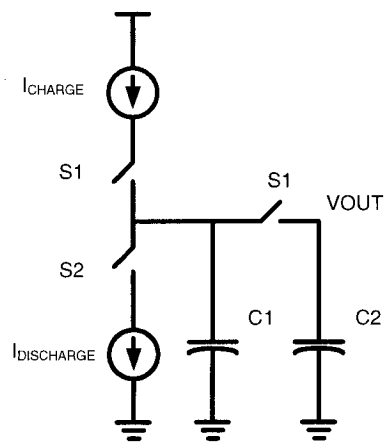


Figure 5.3. Ideal Model of an Implementation of the Single Phase Algorithm

The operation of the system can be analyzed in three parts as shown in Figure 5.4. During the first phase, both capacitors are charged to a potential that is proportional to the duration of the phase. During the second phase, only capacitor C1 is discharged. As the

cycle starts over again, the charges are distributed between C1 and C2 and the integration process begins once again from that point.

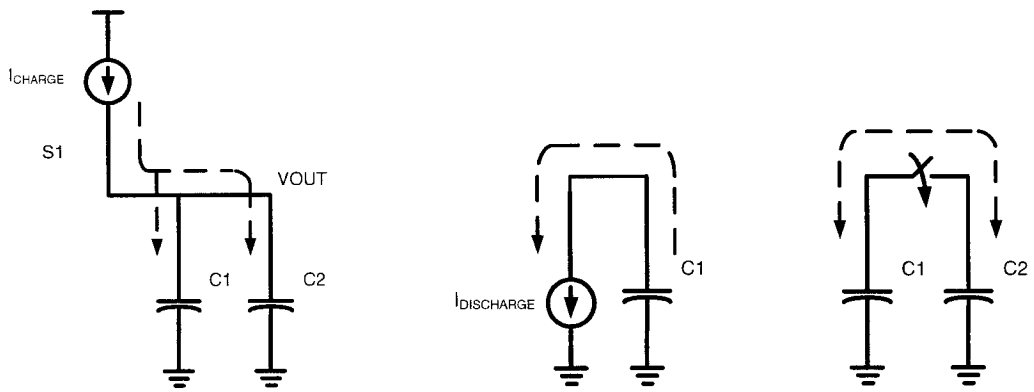


Figure 5.4. Steps in Single Phase Operation

Since there is charge-sharing occurring at every cycle, it may not seem evident that the output would converge to a known value after a given time. An analysis of the system is therefore required. Let C_{TOTAL}=C1+C2. Assuming a stable input frequency, ideal current sources and an initial potential of 0V, the output voltage at the end of the first phase (Figure 5.4 a)) is equal to $\frac{I_{CHARGE} \cdot t}{C_{TOTAL}}$. During the second phase (Figure 5.4 b)), only capacitor C1 is discharged. When the cycle starts over again, the charges in capacitor C2 are shared between C1 and C2, which would make the output voltage drop. The value of this drop depends on C1, C2 and the switch resistance. If the switch is ideal, charge-sharing occurs instantaneously and the output voltage becomes $(\frac{I_{CHARGE} \cdot t}{C_{TOTAL}})(\frac{C2}{C_{TOTAL}})$

(Figure 5.4 c)). This indicates that charge-sharing can be reduced by making C2 larger than C1.

An ideal waveform representation of the algorithm behavior is shown in Figure 5.5.

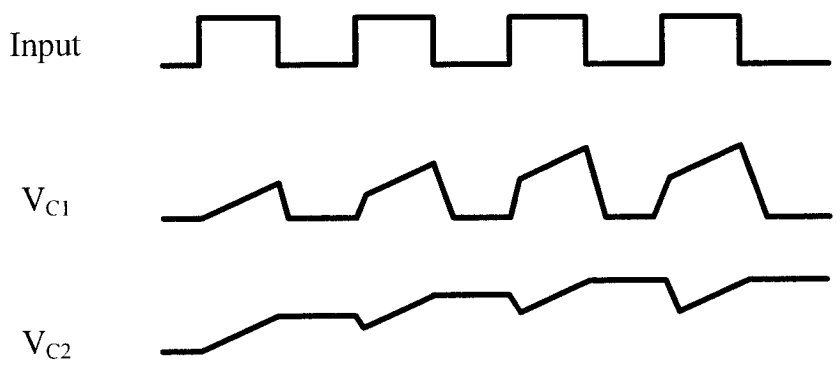


Figure 5.5. Ideal Simulation Waveform of Single Phase Algorithm

If the switch is not considered ideal, the value of the switch resistance will determine the time required for the distribution of the charges. For simplicity, let us assume that the switch is ideal and that the output voltage is given by the previous equation.

After charge-sharing has occurred, another integration phase begins. Since the effects of charge-sharing and integration are additive, output voltage at the end of this phase tends towards $(\frac{I_{CHARGE} \cdot t}{C_{TOTAL}})(\frac{C2}{C_{TOTAL}}) + (\frac{I_{CHARGE} \cdot t}{C_{TOTAL}})$. At the beginning of the following cycle, the charges would be distributed between the capacitors reducing the voltage to $(\frac{I_{CHARGE} \cdot t}{C_{TOTAL}})(\frac{C2}{C_{TOTAL}})^2 + (\frac{I_{CHARGE} \cdot t}{C_{TOTAL}})(\frac{C2}{C_{TOTAL}})$. After going through the iterative analysis a few times, it becomes obvious that the output voltage at the end of the integration phase of iteration N is given by

$$V_{FVC} = I_{CHARGE} t\left(\frac{C2^{N-1}}{C_{TOTAL}{}^{N}} + \frac{C2^{N-2}}{C_{TOTAL}{}^{N-1}} + \ldots \frac{1}{C_{TOTAL}}\right) \tag{5.1}$$

It should be noted that this first-order model only approximates the behavior of the circuit. Eq. (5.1) can be rewritten as a series:

$$V_{FVC} = \frac{I_{CHARGE} \cdot t}{C2} \sum_{i=1}^{N} \left(\frac{C2}{C_{TOTAL}}\right)^{i} \tag{5.2}$$

This equation closely resembles that of a geometric series, which has the following form:

$$y = \sum_{i=0}^{N} x^{i} \tag{5.3}$$

The main difference between Eq. (5.2) and Eq. (5.3) is that the series in Eq. (5.2) starts at i=1 instead of i=0. It is therefore possible to replace Eq. (5.2) by a geometric series that has a missing first element.

As N→∞, when x < 1, the sum of the geometric series becomes equal to $\frac{1}{1-x}$. Since C2 will always be smaller than C_{TOTAL}, Eq. (5.2) can be rewritten as:

$$V_{FVC} = \frac{I_{CHARGE} t}{C2}\left[\frac{1}{1-(C2/C_{TOTAL})} - 1\right] \tag{5.4}$$

Note that the term between brackets is equal to the summation of the infinite geometric series that has been decremented by 1. This accounts for the missing i=0 term in Eq. (5.2). From Eq. (5.4), with some algebraic manipulation, the following can be obtained:

$$V_{FVC} = \frac{I_{CHARGE} \cdot t}{C1} \tag{5.5}$$

Eq. (5.5) states that the output voltage of the FVC at steady-state is independent of C2. This seems to indicate that C2 can be made arbitrarily large in order to minimize charge-sharing effects. However, the statement is true only when N is infinitely large. For finite values of N, the equation is not the same. Transient behavior of the circuit can be found by replacing the infinite geometric series by a finite one. Knowing that the sum of a geometric series with N elements is equal to $\frac{x^{N+1}-1}{x-1}$, Eq. (5.2) can be written as:

$$V_{FVC} = \frac{I_{CHARGE}t}{C2}\left(\frac{(C2/C_{TOTAL})^{N+1}-1}{(C2/C_{TOTAL})-1}-1\right) \tag{5.6}$$

Once again, since element i=0 is missing, the result is decremented by 1. Eq. (5.6) shows the behavior of the system when values of N are small. As the value of N increases, the output voltage will asymptotically tend towards the steady-state value of $\frac{I_{CHARGE}\cdot t}{C1}$.

When designing a system, it may not always be convenient to wait until the output voltage has reached steady-state before using it. In those cases, it is useful to know when the output is reasonably close to the final value so that it can be used. This information can be found as follows. Let p represent the percentage of error of the current voltage with respect to the final output voltage. To find the number of cycles required to reduce the output error to p, Eq. (5.6) can be made equal to a fraction of the infinite geometric series summation:

$$V_{FVC} = (\frac{I_{CHARGE}\cdot t}{C2})[\frac{(C2/C_{TOTAL})^{N+1}-1}{(C2/C_{TOTAL})-1}-1] = (1-p)(\frac{I_{CHARGE}\cdot t}{C1}) \tag{5.7}$$

When this equation is solved for N, the result is

$$N = \frac{\ln(p)}{\ln(C2/C_{TOTAL})} \tag{5.8}$$

This equation shows that the convergence time will be long when the required error fraction is small and/or when C2 is large. While C2 can be made arbitrarily large for optimal steady-state behavior, a large C2 can be detrimental to transient performance. In the extreme case where C2 approaches C_{TOTAL}, the denominator would be close to 0 and the number of cycles required to reach steady-state would tend to infinity. Therefore, the equation shows that there is a trade-off to be made between the convergence time and the magnitude of output glitches.

To visualize the dynamic behavior of an ideal implementation of the algorithm, a MATLAB model has been created. Figure 5.6 shows the simulation results for the FVC with different values for capacitor C2. In the first case, where C1=C2 (dotted line), the output voltage converges rapidly but contains large glitches. In the second case, where C2=5*C1 (solid line), the convergence process is slower, but the glitches are smaller. This simulation illustrates the trade-off between convergence time and output glitches. In both cases, at steady-state, the output voltage is the same.

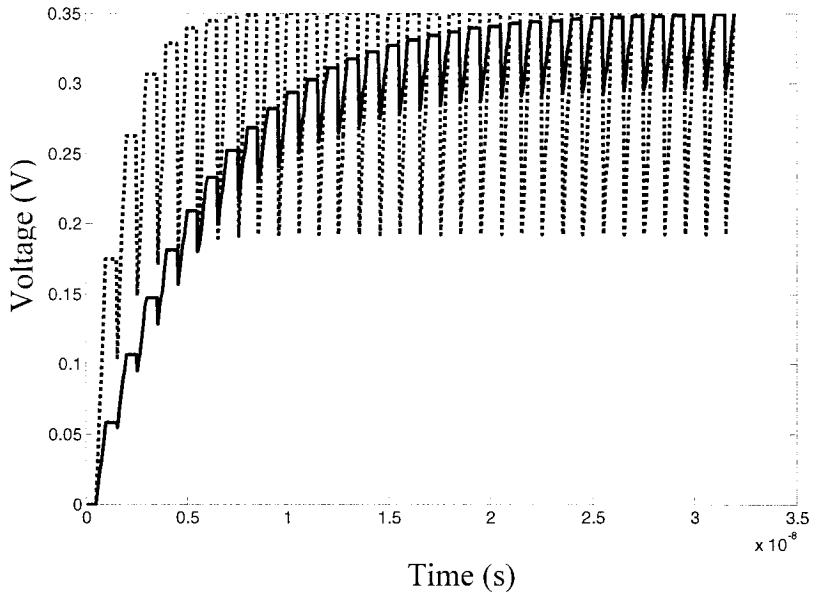


Figure 5.6. MATLAB Simulation with Two Different Values of C2

The purpose of an FVC is to convert a frequency into a corresponding voltage. Different frequencies should be converted to different voltages. However, because of the glitches present in the output signal, the voltage changes momentarily at those instances. This voltage instability may make it difficult to distinguish between two frequencies that are too close. Let F_A and F_B be frequencies that are nearly equal and let V_A and V_B be their respective outputs before glitching. Assume, without loss of generality, that $F_A > F_B$. Since the FVC output is inversely proportional to the frequency, V_B will be larger than V_A. The best resolution obtainable using this FVC is defined as being the minimum frequency difference F_A-F_B, such that the V_B - V_{GLITCH} > V_A. To determine this resolution, we proceed as follows. The magnitude of a glitch is equal to the difference between the steady-state voltage and the voltage after charge-sharing. After simplification, the equation for the magnitude of the glitch is given by the following:

$$V_{GLITCH} = \left(\frac{I_{CHARGE}t}{C_{TOTAL}}\right) \tag{5.9}$$

Knowing that the final output voltage is given by Eq. (5.5), it follows that a change in input period of Δt will cause a change in output equal to

$$\Delta V = \frac{I_{CHARGE}\Delta t}{C1}. \tag{5.10}$$

By solving for Δt, the resulting equation shows how a glitch in the output signal can be wrongly interpreted as a change in input frequency:

$$\Delta t = \frac{\Delta V \cdot C1}{I_{CHARGE}} \tag{5.11}$$

Replacing ΔV with the value of V_{GLITCH}, the following can be obtained:

$$\Delta t = \frac{C1}{C_{TOTAL}} \cdot t \tag{5.12}$$

Eq. (5.12) gives the maximum resolution that can be obtained using the proposed FVC. It shows that the maximum resolution of the FVC is dependent on the chosen design parameters. It also indicates that a small value of C1 with respect to C2 helps increase the maximum resolution of the FVC.

5.2. Circuit-Level Implementation

As stated previously, the key factors that limit the performance of the previously reported FVCs are the use of three phases and the use of full-swing signals. The algorithmic solution proposed in the previous section solves one of the problems by using a 2-phase approach that can be controlled by a single differential clock. This is known as

true single phase clocking and is a known good strategy to enable high speed operation. To further increase speed, a second good strategy is to reduce the voltage swings. The topologies shown in Figures 5.7 and 5.8 use both design strategies. Figure 5.7 shows the circuit used for input signal integration. This circuit replaces the two switches to the left of Figure 5.3.

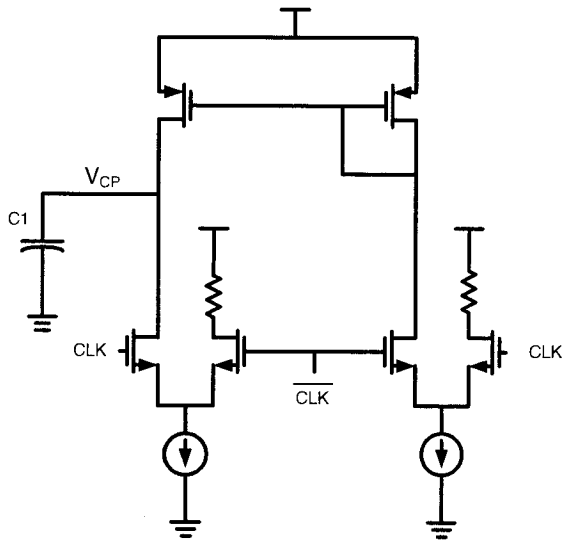


Figure 5.7. Charging/Discharging Switch

This current-switch configuration is well-known in applications such as charge pumps in PLLs [58]. The differential pairs operate as current switches when the differential input is larger than ΔV given by Eq. (2.1), which is typically less than VDD. The use of this structure removes the need to have full-swing signals.

For the remaining charge-transfer switch of Figure 5.3, a simple NMOS can be used. Since the control signal is differential, a second transistor is added in order to balance the

load seen by the driver. In addition, two dummy switches are included to reduce the effects of charge injection. The resulting circuit is shown in Figure 5.8.

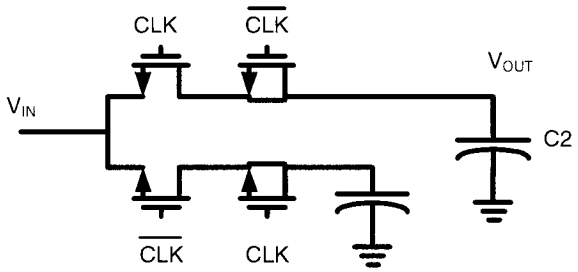


Figure 5.8. Charge Transfer Switch

Although NMOS switches are typically used in full-swing applications, they can also be used in certain cases where the swing is lower.

The conditions that must be respected for the switch to operate properly are the following:

1) When the control signal is high, the switch must be conducting
2) When signal is low, the switch should be shut off.

Let us assume that the control signal switches between VDD and V_{LOW}. Also, assume, without loss of generality, that the input signal is applied to the source of the transistor. When the input signal is equal to VDD, the switch should be "ON". To make the transistor conduct, V_{GS} must be at least equal to V_{TH} and therefore, the input voltage at the source cannot be larger than VDD-V_{TH}. Similarly, when the control signal becomes equal to V_{LOW}, the switch should turn "OFF". For this to occur, the V_{GS} of the transistor

must be less than V_{TH} and therefore, the input voltage cannot be lower than V_{LOW}-V_{TH}. The valid range of input voltages is shown in Figure 5.9.

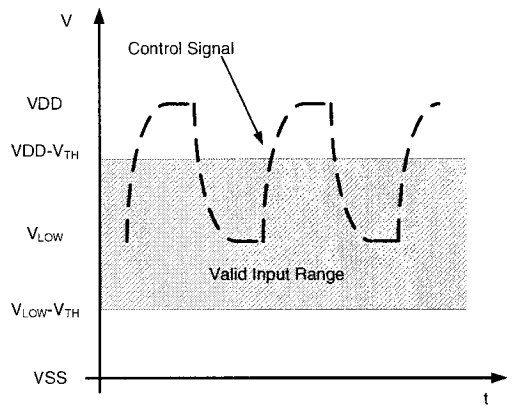


Figure 5.9. Valid Input Voltages

Under these conditions, the NMOS transistor can work as a switch without requiring full voltage swing.

In order to reduce charge-sharing, the integrator is implemented backwards: it starts at VDD and integrates downwards. Thus, at the beginning of every cycle, the input voltage is higher than VDD-V_{TH}, which prevents the charge transfer switch from conducting. Since the switch does not conduct, charge sharing cannot occur at that point. This charge distribution only happens when the input voltage is below VDD-V_{TH} and the NMOS switch is turned "ON". At that point, the voltage difference at its nodes is lower and hence, the magnitude of charge sharing is reduced.

Figure 5.10 shows a typical circuit simulation of a frequency to voltage conversion working at 5 GHz. The signal looks thick due to glitches caused by charge injection. In

Figure 5.11, the steady-state voltage output of signals with three different periods is shown (199 ps, 200 ps and 201 ps). It also shows that the implemented circuit, in this case, is even able to discriminate between period differences of 1 ps.

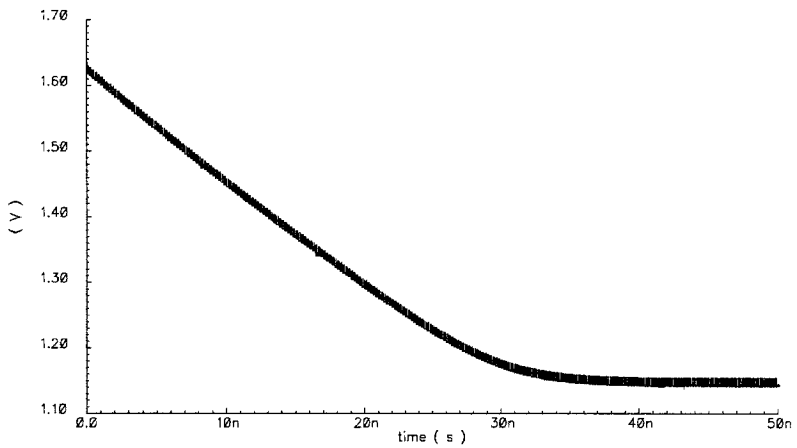


Figure 5.10. Frequency-to-Voltage Conversion at 5 GHz

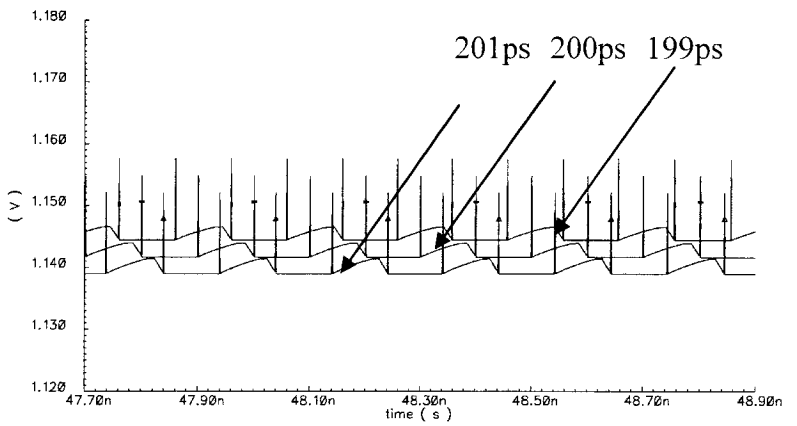


Figure 5.11. Outputs when Input Signal Periods are 199, 200 and 201 ps

5.3. Sample Application: Pulse Width Control Loop

In recent years, a growing number of applications have been using dual clock edges to synchronize operations [14][63]. Compared to the single active edge concept, the use of both clock edges allows the speed of the clock to be reduced by half, which can potentially lead to reduced power consumption. In synchronous digital design, this can be accomplished by using dual edge-triggered flip-flops [50]. The dual-edge concept can also be used in high speed applications, such as double data rate (DDR) random access memories (RAM) [14] and half-rate CDR circuits [63]. Systems that use both edges of a clock typically require a reference clock with 50% duty cycle for optimal operation. Any deviation from this value reduces timing margins and could lead to lower performance.

Under ideal conditions, clock signals coming from MCML oscillators generally have 50% duty cycle. In practice, however, process variations, such as differential pair mismatch and resistive load mismatch, can cause the signal to be distorted. Asymmetry in the layout can also be the cause of uneven characteristics. The resulting distortions include variations in signal-crossing levels and uneven duty cycles. Consequently, designers are faced with tighter constraints to avoid potentially damaging results such as metastability or data recovery error.

In order to ensure a 50% duty cycle signal, some designers resort to pulse width control loops (PWCLs) [40][47][73]. While many solutions exist for single ended signals in applications operating at moderate speeds, not many solutions are currently known for high speed differential applications. This section proposes a novel circuit to resolve the

problem of duty cycle distortion in high speed differential applications such as 10 Gb/s half-rate CDR circuits. The proposed architecture is based on the FVCs proposed in the previous section to allow duty cycle correction on differential multi-GHz clock signals.

The conventional architecture of a PWCL is shown in Figure 5.12. It consists of a feedback control loop that adjusts the duty cycle of the input signal to match that of a reference signal. The reference signal is generated by the ring oscillator located at the bottom of the figure. Since the duty cycle of the output signal depends on that of the reference signal, the ring oscillator should be designed with a duty cycle of 50%. The top part of Figure 5.12 shows the path through which the input signal propagates. It consists of a common-source amplifier, sometimes known as a pseudo-inverter, and an odd number of inverter stages. The common-source amplifier operates as an inverter whose transition time can be adjusted. This transition time adjustment helps change the duty cycle of the signal and ultimately, restores it to 50%. The output of the pseudo-inverter is then reconditioned by a series of inverters.

Both the top and the bottom portions of the PWCL in Figure 5.12 propagate an oscillating signal to their respective charge pumps *CP1* and *CP2*. Each charge pump is connected to a capacitor and forms an integrator whose average output gives a measure of the duty cycle. The measurements provided by both integrators are compared using an integrating amplifier that controls the bias voltage of the pseudo-inverter. Through the control of the pseudo-inverter biasing, the duty cycle of the output clock can be adjusted.

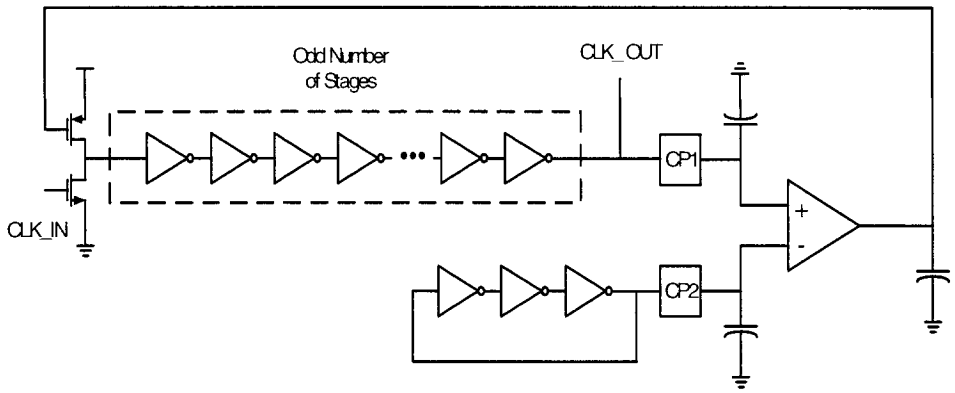


Figure 5.12.Conventional PWCL

One of the problems associated to the conventional PWCL is that the outputs of *CP1* and *CP2* vary over time, and they do so at different frequencies. The output of *CP1* varies at a frequency that depends on the input signal, whereas the output of *CP2* varies at a frequency determined by the ring oscillator. Since the charge pump outputs are changing at different frequencies, the output of the amplifier depends on the phase relationship between the two signals: this can lead to loop instability. To resolve this problem, Lin and Huang [40] propose avoiding the use of an independent reference signal. Instead, they designed a circuit that measures the duration of one phase and compares it to the duration of the other phase. When the durations of both phases are equal, the output has 50% duty cycle and the process is complete. This solution ensures that the signals at the output of both charge pumps change in phase, which solves the loop instability problem.

While this circuit is an improvement over the conventional PWCL, there are still several remaining issues that need to be addressed. The first problem concerns the output of the charge pump. Since the charge pump is controlled by an oscillating signal, its

output will also oscillate. In turn, this can cause the output of the amplifier to fluctuate. Although the amplifier should have common mode rejection, it is important to minimize these input oscillations to reduce noise on the control line.

In addition, this solution is not suitable for the target application because it does not work with differential signaling. Since high-speed CDRs typically use differential signals, a differential solution is required.

5.3.1. Proposed PWCL

A general block diagram of the proposed PWCL is shown in Figure 5.13. It shows how the input signal propagates through the different modules. The first module is called *duty cycle adjust* and its role consists of modifying the shape of the incoming signal according to a control voltage. It can be seen as a variable speed integrator, which charges at one rate and discharges at a different rate. The asymmetry in the charge and discharge times helps create a signal that can eventually be shaped into a 50% duty cycle signal. The output of the *duty cycle adjust* module is then sent to the *waveform shaper*, which consists of a series of high gain buffers. The *waveform shaper* restores the transition times and the DC levels of the clock signal.

The output of the *waveform shaper* and its complement are then fed to the FVCs. In this circuit, the FVCs are used to measure the duration of the phases and provide a stable output of this value. Although the primary function of an FVC is to convert an input frequency into a corresponding voltage, some FVCs can also be used to measure the duty

cycle of a signal. These are grouped into a category called integrator-based FVCs [9] and they are ones used in this application. The outputs of the FVCs are sent to an integrating amplifier, which controls the *duty cycle adjust* module accordingly.

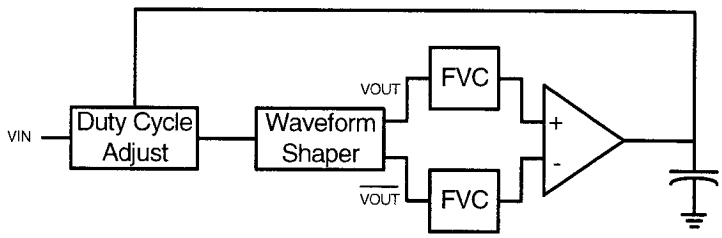


Figure 5.13. Block Diagram of PWCL

To better illustrate the functionality of the proposed PWCL, a sample waveform is shown in Figure 5.14. The top plot shows an input signal that does not have 50% duty cycle. The output of the *duty cycle adjust* module is a triangular wave with different slopes, and changing these slopes is what enables readjusting the duty cycle. This signal is then sent through a series of high gain buffers, collectively known as the *waveform shaper*, to restore the transition times.

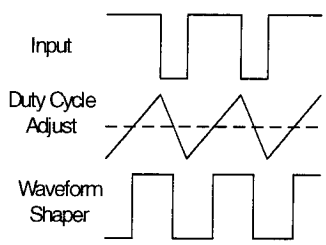


Figure 5.14. Ideal Waveform for PWCL

5.3.1.1. Duty Cycle Adjust

The role of the *duty cycle adjust* module is to change the shape of the incoming waveform. The output is adjusted such that, when it goes through the *waveform shaper*, the outcome is a 50% duty cycle signal. This operation can be accomplished using the circuit shown in Figure 5.15. It is a differential amplifier whose current-source loads can be adjusted independently on each branch. This tuning capacity is illustrated by the dependent current sources I_A and I_B in Figure 5.15.

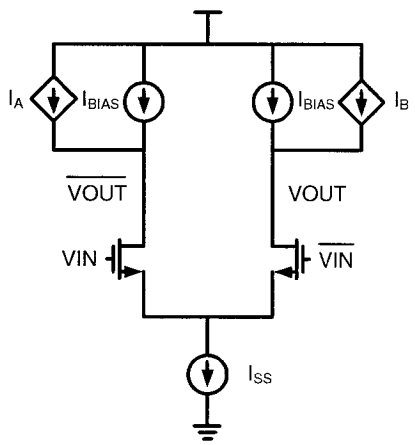


Figure 5.15. Differential Duty Cycle Adjust

To correct the duty cycle of the output signal, the current supplied by the different current sources is adjusted. By varying the current flowing through the branches, the transition times and the DC levels can be changed in order to obtain the desired output signal.

5.3.1.2. FVC

In the PWCL, the FVC is used to measure the duration of a phase. It accomplishes this task by integrating for the duration of the measured phase and then holds this value for the rest of the period. When the duration of the measured phase is long, the output value is large. Similarly, when the duration of the measured phase is short, the output value is small. In these extreme cases, the output values may be larger than VDD-V_{TH} or smaller than V_{LOW}-V_{TH}. In the previous section, it was shown that the FVC cannot operate properly beyond these limits due to the restrictions imposed by the low-swing S/H. Although the previously described limits prevent the FVC from functioning, it does not always cause the PWCL to fail. The reason is that a PWCL does not require the FVCs to be precise nor to be stable. It only requires that the FVCs indicate which phase of a signal has a longer duration. With this information, the amplifier steers the PWCL and brings the loop closer to convergence. Therefore, even with the FVC operating out of its range, the PWCL can still converge.

One situation where the FVC limitation can cause the PWCL to fail is when the charge pump outputs of both FVCs end up above the operating range. In that case, the outputs of the FVCs will likely saturate at VDD-V_{TH} and the PWCL will lock even though duty cycle distortion may still be present. On the other hand, when both charge pump outputs are below the operating range, duty cycle distortions can still be detected. However, as previously described, the S/H switches do not turn "OFF" under these

conditions. Therefore, the inputs to the amplifier will oscillate causing noise on the control line.

Taking this into account, the design of the FVC should be done so that the capacitor voltage at the end of the integration phase of a 50% duty cycle ends up in the middle of the two limiting values. This value is equal to $\frac{VDD+V_{LOW}-2V_{TH}}{2}$.

5.3.2. Simulation Results

To verify the functionality of the proposed PWCL, the circuit has been designed and simulated in 0.18 μm CMOS technology. The FVCs have been designed to have the fastest possible response time in order to facilitate the convergence of the PWCL.

The main purpose of the PWCL is to reshape the clock signal in presence of process variations causing differential signal skew and duty cycles that depart from 50%. To verify the performance of the proposed PWCL, the circuit is stimulated with a distorted 5 GHz signal. A signal with significant distortion is created by propagating a fully symmetrical signal generated by a voltage source through a buffer that has been purposely rendered asymmetric: the resistive loads are mismatched by 10% whereas the W/L ratios of the transistors in the differential pair are mismatched by 15%. In addition, to simulate strongly uneven loads and layout asymmetries, a load of 100fF has been added to only one side of the buffer. Each of these changes increases the level of asymmetry in the signal. The output of the asymmetric buffer is shown in Figure 5.16. It

can be seen in this figure that the two components of the differential signal differ in both magnitude and transition times.

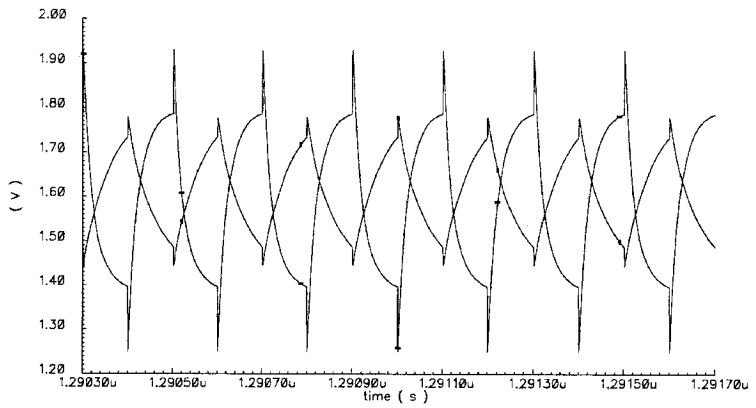


Figure 5.16. Input Signal with Uneven Loading Condition

After the PWCL has converged, the duty cycle of the output clock is fully restored to 50% as shown in Figure 5.17.

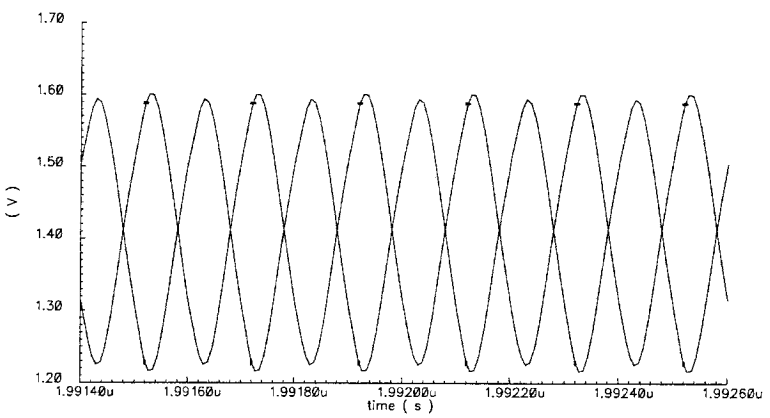


Figure 5.17. Output Signal with Uneven Loading Condition

To determine the maximum clock distortion that can be corrected by this implementation, a series of simulations have been executed. The results show that the implemented circuit reaches its limits when the input signal has about 25% duty cycle (50 ps) with a skew of 49 ps between the differential components. This input signal is shown in Figure 5.18.

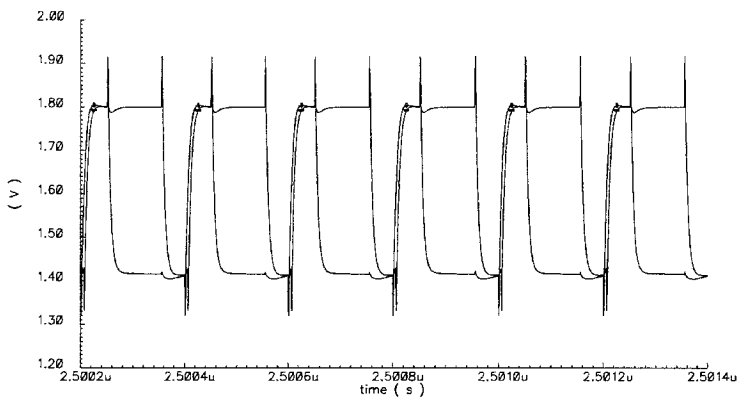


Figure 5.18. Input Signal with 25% Duty Cycle and 49ps Skew

After lock, the output signal is shown in Figure 5.19. The Figure shows that the two DC components of the differential signal are slightly different: this can be adjusted with a few extra stages of buffering. The resulting duty cycle is a little under 48%.

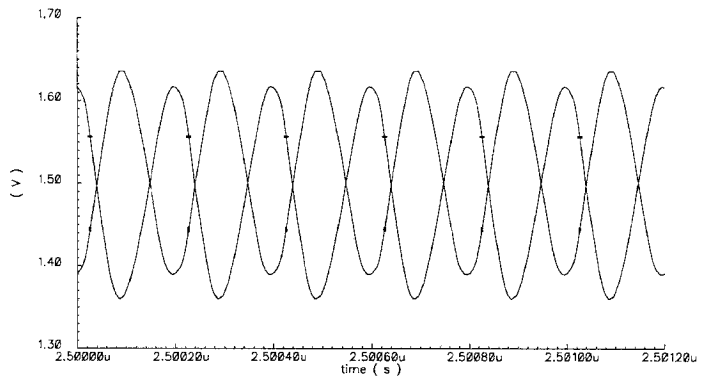


Figure 5.19. Output Signal when Input has 25% Duty Cycle and 49ps Skew

5.4. Frequency Locked Loops: A Second Sample Application

High speed FVCs can be used in a wide variety of applications. Section 5.3 demonstrated the usefulness of such circuit in the design of PWCLs. This section proposes another sample application for the FVC in FLLs. FLLs are used to synchronize the frequency of an internal oscillator to that of an external oscillator. As opposed to PLLs, which need to synchronize both phase and frequency, FLLs do not require any predefined phase relationships. FLLs can be found in a number of applications, including dual-loop PLLs, FM modulators and low-jitter VCOs. In the context of designing multi-GHz transceivers, FLLs can be used as a frequency-acquisition aid that can extend the range of supported frequencies.

A possible architecture for FLLs is shown in Figure 5.20. As stated previously, the role of FLLs is to synchronize the frequency of a local oscillator to that of a reference signal. To accomplish this task, each of these signals is converted to a DC voltage that is

proportional to its frequency. The DC voltages are then compared and the difference is used to adjust the frequency of the local oscillator. The adjustments steer the frequency of the local oscillator closer to that of the reference. The resulting signal is then converted to a DC voltage and fed back to the comparator. This cycle continues until the DC values of both oscillators become equal, at which point the changes are minimal.

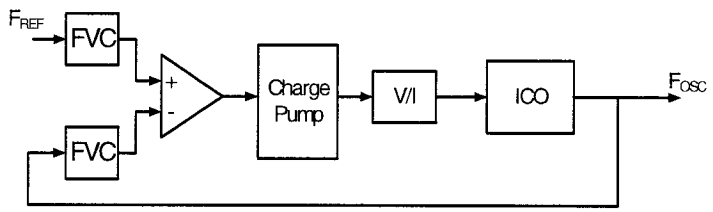


Figure 5.20. Block Diagram of FLL

5.4.1. Analytical Model

The design process for an FLL is similar to that of a PLL in that designers typically start with an analytical model. The analytical model is used to approximate the expected behavior of the loop and to determine the value of certain critical parameters. Before getting to the model of the FLL, it is useful to go over the PLL model and see what the main differences are. The basic linearized model of a PLL is shown in Figure 5.21.

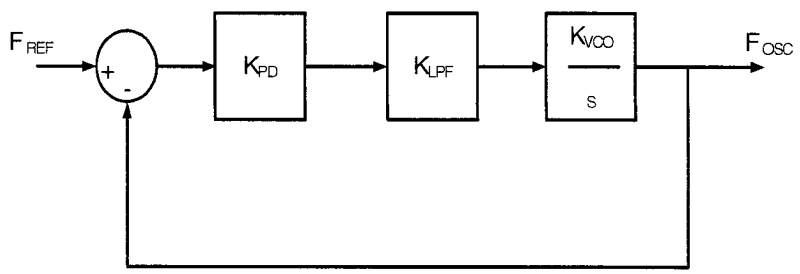


Figure 5.21. Linearized Model of a PLL

In this model, K_{PD}, K_{LPF} and K_{VCO} are the gain and unit conversion factors associated to the phase detector, the low pass filter and the VCO respectively. The first element of a PLL is a phase detector, used to indicate the phase relationship between the reference and generated clocks. The output of the phase detector, which may contain high-frequency noise, is sent to a low-pass filter so that only the DC phase difference is kept. This phase difference signal steers the oscillator to increase or decrease its frequency in order to adjust the phase. This modified signal is then fed back to the phase comparator and the cycle starts over.

It should be noted that the output of the VCO is a frequency whereas the input of the phase detector is a phase. To ensure that the units are consistent, the VCO block also includes a factor of $\frac{1}{s}$ in its transfer function. This shows that the VCO frequency is integrated to produce a phase.

While this model is widely used in the design of PLLs, it cannot simply be applied to the design of FLLs. The main reason is that the input and output signals of the FLL are in terms of frequency, whereas the PLL uses phases. This eliminates the need for the

integrating factor of $\frac{1}{s}$ in the VCO. Another difference between the models is that, while the phase difference can be measured very quickly, a difference in frequency requires more time to detect.

Since the classic PLL model cannot be directly applied to FLLs, a different model is required. Over the years, several authors have proposed FLL models [39]. While these models provide a good starting point, they do not take into account the delay caused by frequency measurement. Since the delay in the feedback path can lead to instability, it is important to model this effect. To account for this delay, this section proposes a new analytical model for FLLs. The proposed block diagram is shown in Figure 5.22.

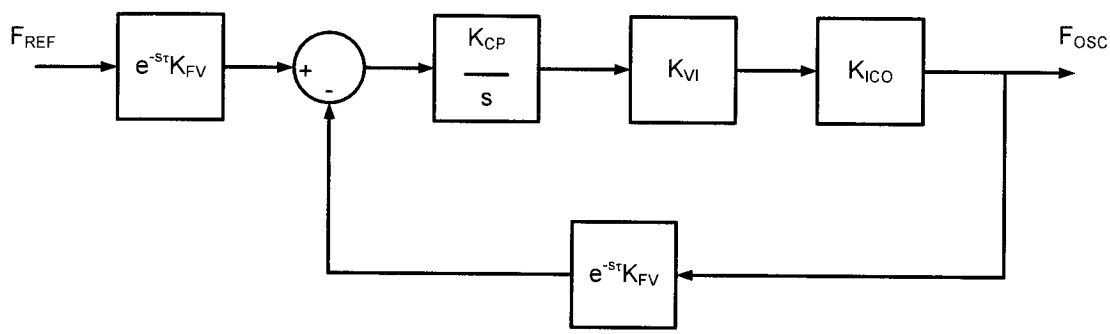


Figure 5.22. Block Diagram of FLL with Transfer Function of Each Block

The FVCs are modeled by the expression $e^{-s\tau} K_{FVC}$, which consists of two distinct parts: the first part is K_{FVC}, which is the gain and unit conversion factor. The second part, $e^{-s\tau}$, is used to model the time delay τ caused by the FVC. As previously explained, this is necessary because frequency measurements cannot be made instantaneously. K_{VI} and K_{ICO} are constants that are used to account for the gain and the unit conversion for the

Voltage-to-Current (V/I) converter and the current-controlled oscillator (ICO). As expected from our previous discussion, a discrepancy is noted between the classic PLL model and the FLL model: in the classic PLL model, the ICO would have been modeled as $\frac{K_{ICO}}{s}$, whereas the FLL model has it as a simple gain K_{ICO}. The charge pump is modeled by a block with a gain of $\frac{K_{CP}}{s}$. In addition to having a gain, it also functions as an integrator, which is modeled as $\frac{1}{s}$.

In this analysis, the reference clock signal is assumed to be constant and therefore, the output of the FVC has reached steady-state. The transfer function in the forward path is found to be:

$$G(s) = \frac{K_{CP} K_{VI} K_{ICO}}{s} \tag{5.13}$$

The feedback path, which only consists of an FVC, has the following transfer function:

$$H(s) = K_{FVC} \cdot e^{-s\tau} \tag{5.14}$$

According to control systems theory, the transfer function of a feedback system is given by $F(s) = \frac{G(s)}{1 + G(s)H(s)}$. Applying this to the FLL transfer function yields:

$$F(s) = \frac{\frac{K_{CP} K_{VI} K_{ICO}}{s}}{1 + \left(\frac{K_{CP} K_{VI} K_{ICO}}{s} \right) \left(K_{FVC} \cdot e^{-s\tau} \right)} \tag{5.15}$$

In order to simplify the analysis, the time delay, $e^{-s\tau}$, can be approximated by a rational function. A good approximation function of time delay can be provided by

Padé's first order relation: $\frac{1-s\tau/2}{1+s\tau/2}$. This allows the transfer function to be analyzed using a conventional approach. Replacing the time delay element by Padé's first order relation, the transfer function becomes:

$$F(s)=\frac{\frac{K_{CP}K_{VI}K_{ICO}}{s}}{1+\left(\frac{K_{CP}K_{VI}K_{ICO}}{s}\right)K_{FV}\left(\frac{1-s\tau/2}{1+s\tau/2}\right)} \tag{5.16}$$

Eq. (5.16) can be rewritten as follows:

$$F(s)=\frac{K_{CP}K_{VI}K_{ICO}(1+s\tau/2)}{s^2\left(\frac{\tau}{2}\right)+s\left(1-\frac{K_{CP}K_{VI}K_{ICO}K_{FV}\tau}{2}\right)+K_{CP}K_{VI}K_{ICO}K_{FV}}. \tag{5.17}$$

Substituting $K_{CP}K_{VI}K_{VCO}$ with K_1 and K_{FVC} with K_2, the following equation is obtained:

$$F(s)=\frac{K_1(1+s\tau/2)}{s^2\left(\frac{\tau}{2}\right)+s\left(1-\frac{K_1K_2\tau}{2}\right)+K_1K_2}. \tag{5.18}$$

The resulting transfer function shows that the system is of 2nd order. The analysis of the system can be done by finding the position of the poles. These are found by calculating the roots of the denominator in Eq. (5.18):

$$s=\frac{-\left(1-\frac{K_1K_2\tau}{2}\right)\pm\sqrt{\left(1-\frac{K_1K_2\tau}{2}\right)^2-4\left(\frac{\tau}{2}\right)K_1K_2}}{2\left(\frac{\tau}{2}\right)}. \tag{5.19}$$

The desired response of the system is obtained by carefully selecting the values of K_1, K_2 and τ. Recall that the behavior of a system mainly depends on the position of the poles in the complex plane. Poles that are located in the right-hand side of the s-plane indicate that the system is unstable. When poles are on the y-axis (imaginary), the system is considered marginally stable and produces an oscillatory response to any given input.

Overdamped and critically damped responses are due to poles being located on the negative x-axis (real and negative). Otherwise, when the poles have both real and imaginary components, the system is underdamped.

Underdamped systems tend to react quickly, but are subject to overshoots. While overdamped systems are not subject to these undesirable responses, the response is typically slow. In a critically damped system, the response is the fastest without causing any overshoot. To design this critically damped system, the two poles of the transfer function need to be real and identical. This is accomplished by making the part below the square root equal to 0. The resulting second order equation is:

$$K_1K_2\tau^2 - 12K_1K_2\tau + 4 = 0 \tag{5.20}$$

By substituting $X = K_1K_2\tau$, the following is obtained:

$$X^2 - 12X + 4 = 0 \tag{5.21}$$

This equation has two possible solutions, which are $6 \pm 4\sqrt{2}$. The two solutions represent the two different points in the complex plane where poles are equal: one point is in the right-hand side of the s-plane and one is one on its left-hand side. The poles on the right-hand side are unstable and the corresponding solution is therefore discarded. The solution indicates that the system is critically damped if $K_1K_2\tau = 6 - 4\sqrt{2} \cong 0.343$. The relation indicates that, for a given system. If the delay τ is increased, the gains need to be reduced in the same proportions if the damping characteristics of the system are to be preserved. This shows that the delay in the FVC contributes as much to the stability of the loop as do the gains K_1 and K_2.

The root locus plot of the FLL is shown in Figure 5.23. This plot traces the gains K_1 and K_2 as they are varied from 0 to $+\infty$. The root locus leaves the real axis at two different points given by $\frac{2(1-\sqrt{2})}{\tau}$ and $\frac{2(1+\sqrt{2})}{\tau}$. The points at which the root locus plot crosses the imaginary axis is given by $\pm\frac{2j}{\tau}$.

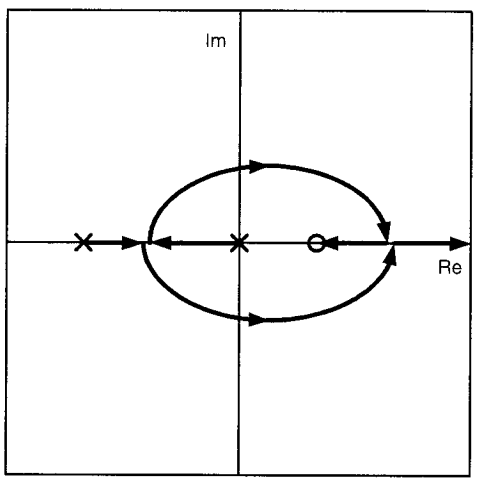


Figure 5.23. Root Locus Plot of FLL

The theoretical analysis provides more insight into the behavior of FLLs. It shows how the different parameters are related and how they affect the overall response of the circuit. These results then can be used as basic guidelines in the design of the FLL building blocks.

5.4.2. Building Blocks

The FLL proposed in the previous section (Fig 5.20) has been designed and fabricated in TSMC 0.18 μm CMOS technology. The circuit on the test chip consists of a ring oscillator that tracks the frequency of a 5 GHz reference ring oscillator through the circuitry of the FLL. The different building blocks are described in the following subsections.

5.4.2.1. ICO

The ring oscillators used in this chip consist of 4 stages and have a frequency that is controlled by a current input. The delay cells are made up of MCML buffers with a standard PMOS load. These delay elements are biased from the top and from the bottom of the cell with the same current. This way, the charging and discharging process should be symmetric and helps maintain a constant output swing. The biasing circuit, along with one stage of the oscillator, is shown in Figure 5.24.

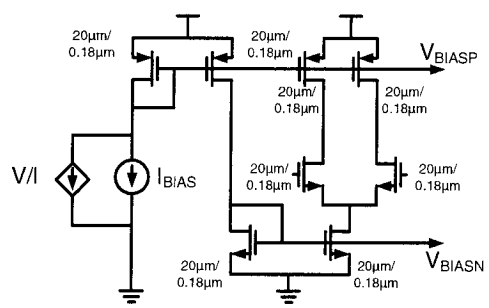


Figure 5.24. Biasing Scheme and Single Delay Element

5.4.2.2. Comparator

The comparator used in the FLL is implemented with a 2-stage fully differential operational amplifier topology, as shown in Figure 5.25. To help with loop stability, the DC gain of the amplifier has been limited to 25. The operational amplifier has a -3 dB frequency of little over 800 MHz.

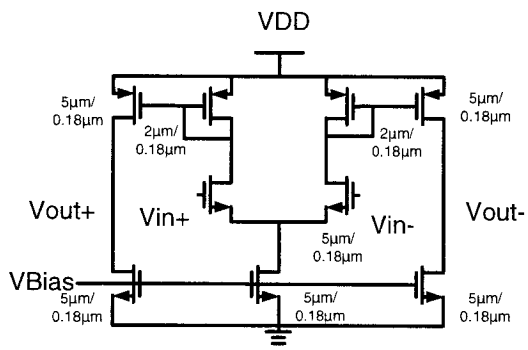


Figure 5.25. Fully Differential Two Stage Amplifier

5.4.2.3. V/I

The signal controlling the frequency of the local oscillator is a voltage coming from the charge pump. The ICO, on the other hand, requires a current to adjust its frequency. To convert from a voltage to a corresponding current, the V/I converter shown in Figure 5.26 is implemented. The V_{CTRL} signal is used to control the current that flows through the drain of the output transistor. The maximum value for I_{OUT} is limited by the current mirror. Although the conversion is not linear, it offers a large input voltage range and is

monotonously increasing. These characteristics are satisfactory for a closed-loop system, such as an FLL.

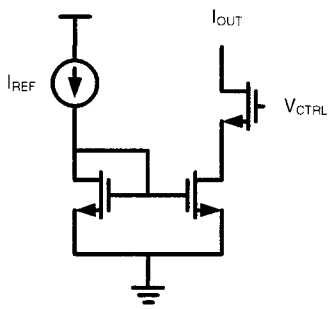


Figure 5.26. V/I Circuit

5.4.2.4. Divide-by-64

In order to test a circuit for proper functionality, the most straight-forward method is to send the signals off-chip. However, when dealing with signals oscillating at 5 GHz, this task is not trivial. To facilitate the process of testing the FLL, the output speed of 5 GHz is reduced to 78.125 MHz through a divide-by-64 circuit. This relatively slow output signal reduces the complexity of the test setup and low-frequency oscilloscopes can be used to view the output. The divide-by-64 circuit is implemented using a cascade of six digital divide-by-2 circuits.

5.4.3. Top-Level Simulations After Extracted Layout

The implemented top-level circuit uses 4-stage ring oscillators for both the reference signal and for the tracking oscillator. The reference signal has an oscillation frequency that can be adjusted through an external pin, whereas the tracking oscillator's frequency is adjusted through the loop.

To demonstrate the full functionality of the system, a sample simulation was run from the layout, after extraction of parasitic capacitances. The advantage of such simulation is the increased visibility and the access to internal parts of the circuit. Figure 5.27 shows the output of the two oscillators at the beginning of the simulation. It shows that the frequencies are different. At the end of the simulation, the frequency of the local oscillator matches that of the reference clock. This is shown in Figure 5.28.

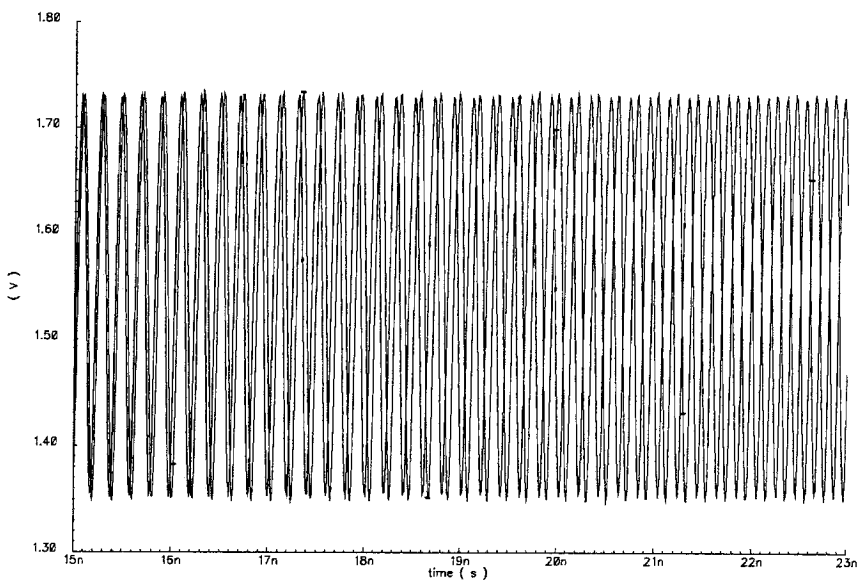


Figure 5.27. Full Speed Oscillators before Lock

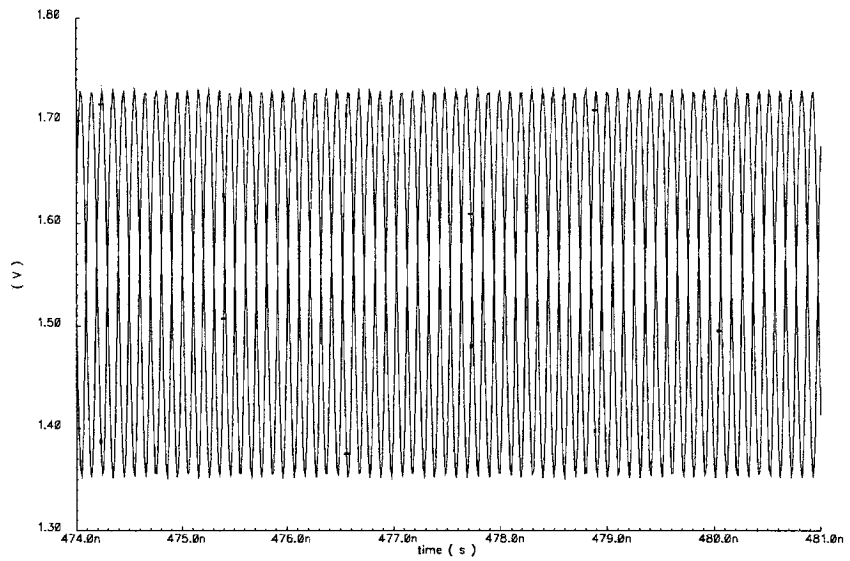


Figure 5.28. Full Speed Oscillators after Lock

5.4.4. Experimental Results

The FLL chip has been fabricated, placed in a CQFP package and mounted on a printed circuit board (PCB). The total chip size, including bonding pads, is about 1 mm^2 and the power consumption of the FLL is about 77.4mW at 1.8V. The micrograph of the chip is shown in Figure 5.29.

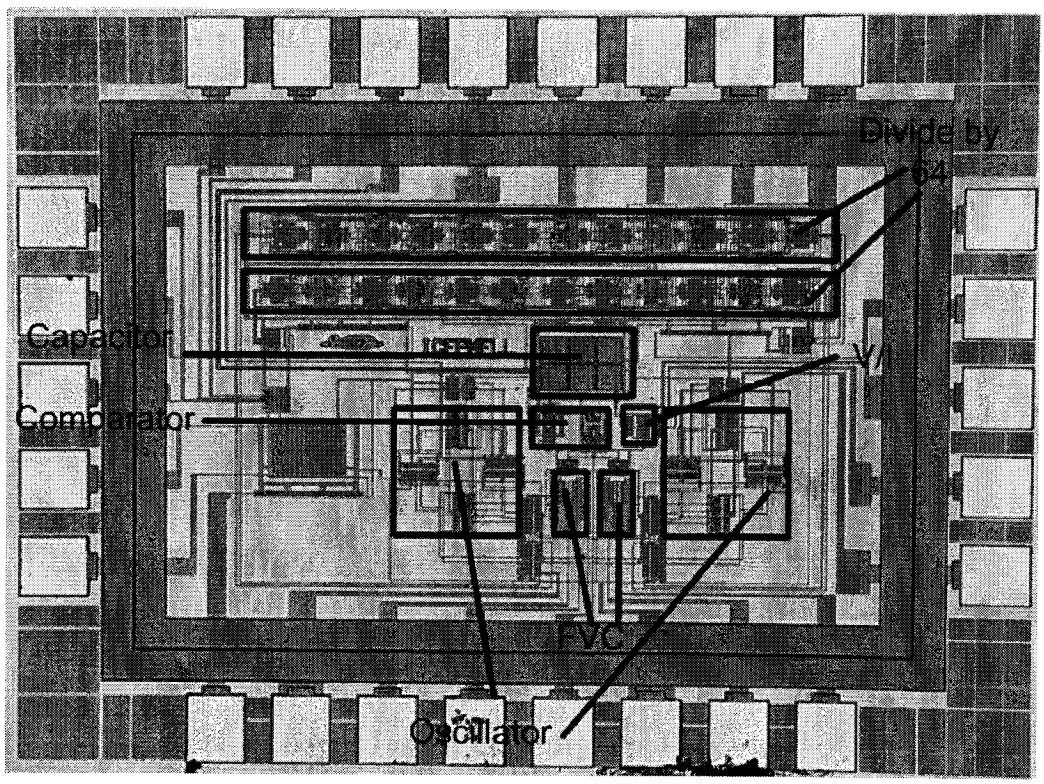


Figure 5.29. Chip Micrograph of the FLL

The experimental setup used to test the chip is shown in Figure 5.30. The oscillation frequency of the reference clock is adjusted through an external potentiometer. The reference and the local oscillator signals are divided by 64 before being sent to the output pads.

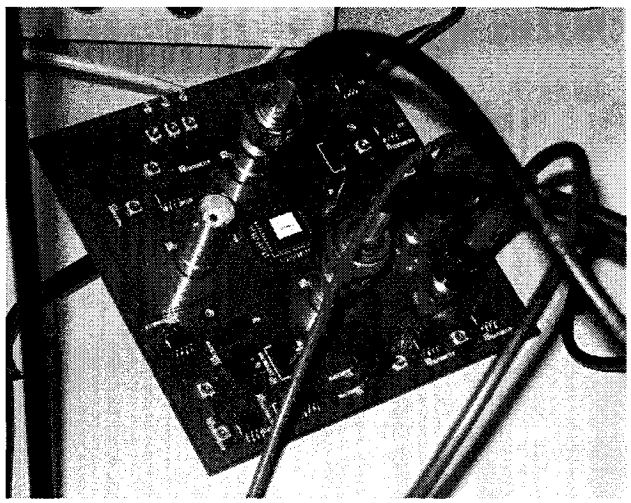


Figure 5.30. Experimental Setup for FLL

Although the expected frequency of operation was 5 GHz, the implemented circuit was able to operate reliably up to 3.65 GHz. A more in-depth investigation showed that the difference in operating speeds is due to process variations. Since the FLL chip also carries the process and temperature measurement circuit presented in Chapter 3, it was possible to observe that the chip was made in a slow-slow process.

Figure 5.31 shows the output signals before the oscillators are locked. The time-division used for the measurement on the oscilloscope is 10 ns per square. Since the oscillators are divided by 64, the signals that are observed on the oscilloscope are about 57 MHz. The figure clearly shows that there is a frequency difference between the two signals before locking. After synchronization, the two oscillating signals with almost same frequency are shown in Figure 5.32.

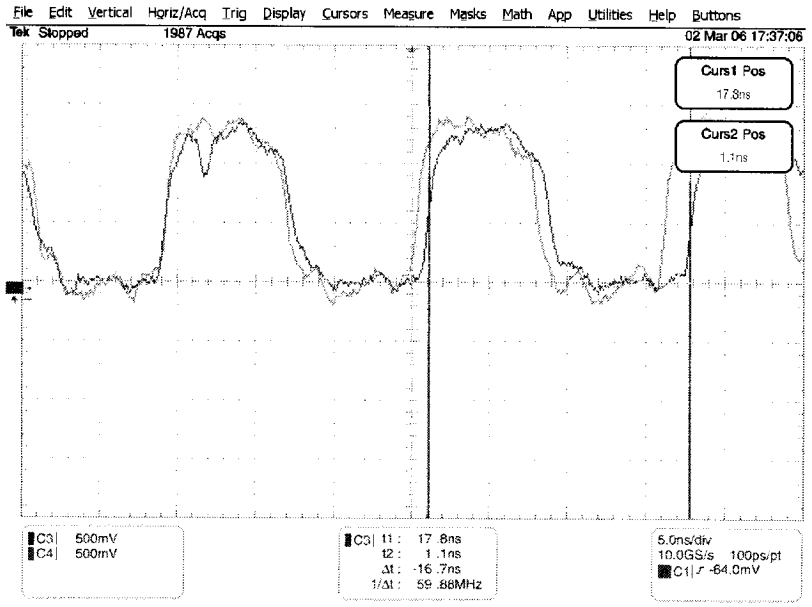


Figure 5.31. Output Signals before Lock

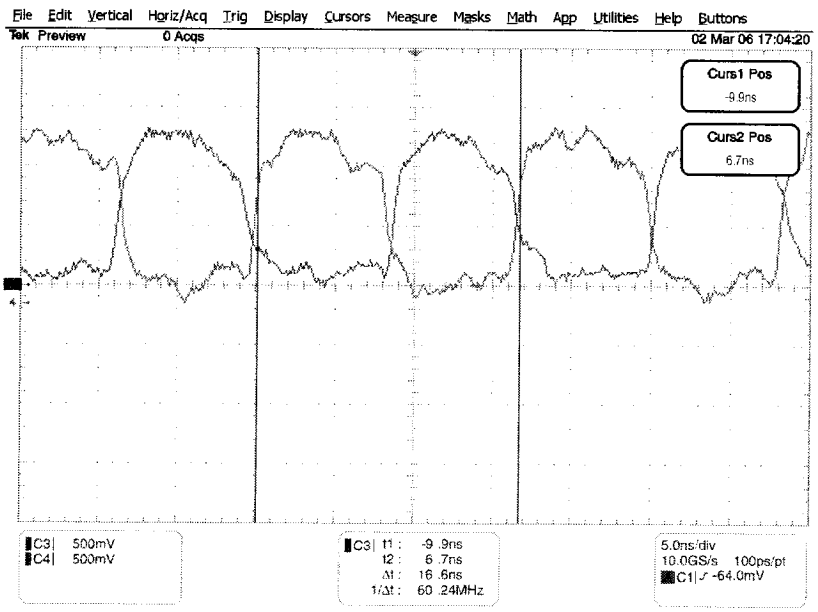


Figure 5.32. Output Signals after Lock

5.5. Conclusions

This chapter proposed the use of a single phase algorithm and low-swing circuit techniques that allow implementing high-speed FVCs. It overcomes the problems of a previous structure by using these techniques and was possible to operate at speeds of up to 5 GHz in 0.18 μm CMOS technology.

In addition to showing that the FVC functions at the circuit level, several analytical equations have been derived. These equations help characterize the FVC and also help the designer choose critical parameters.

To show the usefulness of a high-speed FVC, two sample applications have been demonstrated. The first application is a PWCL architecture that can restore a 50% duty cycle differential signal from a distorted one. When designed and simulated in 0.18μm CMOS technology, it can restore a 5 GHz signal that has 25% duty cycle and whose differential components are skewed by up to 49ps. Even though the output signal is not extremely well-behaved for this worst case, it is more than adequate for use in MCML gates and it produces signals with duty cycles that are very close to 50%. While the output in this worst case produces uneven DC levels, this can be corrected by extra buffering stages.

The second application for the FVC is in the design of FLLs. It was first shown that existing FLL models are deficient in that they neglect the time delay associated to frequency measurement. Since delay in the feedback path is important when considering loop stability, a new model was proposed. The proposed model includes time delay in the analysis by using Padé's approximation. This provides a more complete FLL model.

The FLL was designed and fabricated in TSMC's 0.18μm CMOS technology. The design uses two FVCs to convert the frequency of a local oscillator and that of a reference clock into corresponding voltages. Extracted layout simulations show that the FLL can lock onto a 5 GHz signal whereas the fabricated chip can operate reliably at 3.65 GHz. It was found that the discrepancy between frequencies is due to process variations.

Chapter 6

Conclusion

In this thesis, several novel ideas were presented to help improve the design of high speed transceivers. Contributions were made at the algorithmic level, the architectural level as well as circuit level.

The use of active inductors in the design of MCML gates was proposed in Chapter 3. This technique, known as active shunt-peaking, increases the bandwidth of circuits using it. It allows for MCML gates to operate at higher frequencies than gates designed using conventional methods. The small signal model of the active shunt-peaked buffer was analyzed to provide more insight into the design process and optimal design parameters were calculated. As a result, the transition time of these gates is improved by as much as 17%. Even though this improvement is not as high as those achievable using planar spiral inductors, this solution requires much less circuit area and does not require a separate design environment. The main drawback to active inductors is their sensitivity to process variations. Since their performance is highly dependent on fabrication technology, process variations can cause the behavior to change. To alleviate this concern, a process measurement and compensation technique has been proposed. The measurement scheme has been implemented on a chip and test results confirm that the circuit is functionally correct.

Another contribution of this thesis relates to the quality of the signals. In Chapter 4, it was shown that typical MCML gates are asymmetric and that this characteristic is detrimental to the performance of CDRs. In particular, the XOR gate was analyzed because it is one of the most commonly used logic gates in CDR phase detectors. A simple and intuitive jitter model of the XOR gate was developed. It shows that transistor sizing may help remove the jitter from MCML XOR gates. However, the required size for low jitter and the required size for high speed are usually not the same. This indicates that there is a trade-off to be made between jitter and speed. To resolve this conflict, several solutions have been proposed. These solutions allow for reduction of 26% to 95% in jitter at low speeds and up to 75% at speeds of 10 Gb/s. To verify the functionality and the symmetry of the proposed solution, a chip has been fabricated and tested. The results show that the proposed gate is in fact symmetric and works well.

Chapter 5 dealt with the issue of frequency to voltage conversion. The chapter introduced a new true single phase algorithm to convert frequency to voltage. It also presented novel techniques that allow circuits to operate with low voltage swings. The resulting FVC was designed and simulated and results show that it could operate at 5 GHz. As a sample application, the FVC was used in the design of a PWCL that can restore 50% duty cycle to a clock signals. This 50% duty cycle is especially important in high speed transceivers that use the increasingly popular half-rate concept. The implemented FVC was able to restore duty cycle of signals operating at 5 GHz. A second application of the FVC is in FLLs. A new analytical model for FLLs has been developed and provides a more complete description of the system. Using the proposed FVC, a 5

GHz FLL was designed and simulated. A chip was also fabricated and tested. Results from this chip show that the FLL can operate reliably at frequencies of up to 3.65 GHz. Further investigation showed that the difference between expected performance and measured performance was probably due to process variations.

The design techniques proposed throughout this thesis help increase the bandwidth, improve the quality of the recovered signals and proposes new approaches to designing multi-GHz transceivers. It is the author's belief that these techniques can help design high speed circuits with less silicon area, better signal quality and with less design effort.

6.1. Future Work

The work presented in this thesis paves the way for many possible research topics. This section provides a brief overview of some of the possibilities.

Intra-Die Process and Temperature Variations: Inter-die process variations have long been a concern of circuit designers. In recent years, intra-die variations have become more of an issue. Having demonstrated the possibility of measuring process and temperature variations on a chip, it would be interesting to examine the distribution of process variations within a die. Perhaps, with a small number of strategically placed process measurements, it is possible to extrapolate the characteristics of the whole chip. This technique would allow fabricated chips to have less uncertainty in their performance.

Standard Cells Shunt-Peaking: This thesis showed that active shunt-peaking can be applied to the design of MCML gates. A possible extension to this topic could be its application to the design of standard cells. If possible, it would improve the overall design speed and allow for the speed of standard cell ASICs to approach that of custom ASICs.

Full Rate Circuit Measurements: The experimental results presented in this thesis mainly come from measurements made at slow speed. The reason was to facilitate the process of designing and testing the circuit. Now that the concepts have been demonstrated, it would be important to characterize these circuits using high speed measurements and directly examine the output signals.

Integration of the Proposed Techniques: The techniques presented in this thesis have been demonstrated individually and have shown improvements over conventional methods. The next step is to integrate the different contributions and to design a multi-GHz CDR.

REFERENCES

[1] S. B. ANAND, B. RAZAVI, *A CMOS Clock Recovery Circuit for 2.5 Gb/s NRZ Data*, IEEE Journal of Solid-State Circuits, Vol. 36, No.3, March 2001. pp. 432-439.

[2] A. BALATSOS, D. LEWIS, *Low Skew Clock Generator with Dynamic Impedance and Delay Matching*, International Solid-State Circuits Conference 1999.

[3] J. BAKER, H. W. LI, D. E. BOYCE, CMOS: Circuit Design, Layout and Simulation,1st Ed., Wiley-IEEE, 1997.

[4] H. BRAUNISCH, R. NAIR, *On the Techniques of Clock Extraction and Oversampling*, Hot Interconnects, 2001

[5] H. T. BUI, Y. SAVARIA, *Shunt-peaking in MCML Gates And Its Application In The Design Of A 20 Gb/S Half-Rate Phase Detector,* International Symposium on Circuits and Systems, Vol. 4, May 2004, pp. 369-372.

[6] H. T. BUI, Y. SAVARIA, *Design and Analysis of XOR Gates for High-Speed and Low-Jitter Applications,* World Multi-Conference on Systemics, Cybernetics and informatics, Vol. 6, July 2005, pp. 60-65.

[7] H. T. BUI, Y. SAVARIA, *Shunt-Peaking of MCML Gates Using Active Inductors*, North East Workshop on Circuits and Systems, Montreal, Canada, June 2004, pp. 361-364.

[8] H. T. BUI, Y. SAVARIA, *High-Speed Differential Frequency-to-Voltage Converter*, Northeast Workshop on Circuits and Systems, June 2005, pp. 76-79.

[9] H. T. BUI, Y. SAVARIA, *10 GHz PLL Using Active Shunt-Peaked MCML Gates and Improved Frequency Acquisition XOR Phase Detector in 0.18 μm CMOS*, International Workshop on System on Chips, Banff, Canada, July 2004, pp. 115-118.

[10] H. T. BUI, Y. SAVARIA, *A Generic Method for Embedded Measurement and Compensation of Process and Temperature Variations in SoCs,* International Workshop on System-on-Chip for Real-Time Applications, July 2005, pp. 557-562.

[11] H. T. BUI, Y. SAVARIA, *High Speed Differential Pulse-Width Control Loop Based on Frequency-to-Voltage Converters*, Submitted to the International Symposium on Circuits and Systems 2006.

[12] H. T. BUI, Y. WANG, Y. JIANG, *Design And Analysis Of Low-Power 10-Transistor Full Adders Using Novel XOR-XNOR Gates,* Transactions on Circuits and Systems II: Analog And Digital Signal Processing, Vol. 49, No. 1, January 2002, p 25-30.

[13] J. CAO, M. GREEN, A. MOMTAZ, K. VAKILIAN, D. CHUNG, J. KEH-CHEE, M. CARESOSA, X. WANG, T. WEE-GUAN, C. YIJUN; L. FUJIMORI, A. HAIRAPETIAN, *OC-192 Transmitter and Receiver in Standard 0.18um CMOS*, IEEE Journal of Solid-State Circuits, Vol. 37, No. 12, December 2002. pp. 1768-1780.

[14] F. CARRARA, P. FILORAMO, G. PALMISANO, *High-Dynamic-Range Variable Gain Amplifier With Temperature Compensation And Linear-In-Decibel Gain Control*, Electronics Letters, Vol. 40 , Issue 6 , March 2004, pp. 363-364.

[15] U.-R. CHO, T.-H. KIM, Y.-J. YOON, J.-C. LEE, D.-G. BAE, N.-S. KIM, K.-Y. KIM, Y.-J. SON, J.-S. YANG, K.-I. SOHN, S.-T. KIM, I.-Y. LEE, K.-J. LEE, T.-G.

KANG, S.-C. KIM, K.-S. AHN, H.-G. BYUN,, *A 1.2-V 1.5-Gb/s 72-Mb DDR3 SRAM,* Journal Of Solid-State Circuits, Vol. 38, No. 11, Nov. 2003, pp. 1943-1951.

[16] E. CRAIN, M. PERROTT, *A Numerical Design Approach For High Speed, Differential, Resistor-Loaded, CMOS Amplifiers*, International Symposium on Circuits and Systems, Vol.5, May 2004, pp. 508-511.

[17] N. DA DALT, C. SANDNER, *A Subpicosecond Jitter PLL for Clock Generation in 0.12-μm Digital CMOS*, IEEE Journal of Solid-State Circuits, Vol. 38, No. 7, July 2003, pp. 1275–1278.

[18] W. J. DALLY, *Digital Transmitter with Equalization*, US Patent No. US 6,266,379 B1, July 24 2001.

[19] D. DATTA , R. GANGOPADHYAY, *Performance Analysis Of The Delay And Exclusive-OR Type Clock Recovery Circuit In An APD-Based Optical Receiver*, IEE Proceedings on Optoelectronics, Vol. 138, No. 1, February 1991, pp.21-32.

[20] A. DJEMOUAI, M. A. SAWAN M. SLAMANI, *New Frequency-Locked Loop Based on CMOS Frequency-to-Voltage Converter: Design and Implementation,* IEEE Transactions on Circuits and Systems—II: Analog and Digital Signal Processing, Vol. 48, No. 5, May 2001, pp. 441-449.

[21] R. C. D. DULK, *Digital PLL Lock-Detection Circuit*, Electronic Letters, Vol. 24, No. 14, July 1988, pp. 880-882.

[22] W. F. ELLERSICK, *Data Converters for High-Speed CMOS Links*, Ph.D. Dissertation, Stanford University, USA, August 2001.

[23] Y. FOUZAR, Y. SAVARIA, M. SAWAN., *A CMOS phase-locked loop with an auto-calibrated VCO,* International Symposium on Circuits and Systems, Vol. 3, May 2002, pp. 177-180

[24] M. M. GREEN, U. SINGH, Design Of Cmos CML Circuits For High-Speed Broadband Communications International Symposium on Circuits and Systems, Vol. 2, May 2003, pp. 204-207.

[25] H. M. GREENHOUSE, *Design of Planar Rectangular Microelectronic Inductors*, IEEE Transactions on Parts, Hybrids and Packaging, Vol. PHP-10, No.2, June 1974.

[26] A. HATI, M. GHOSH AND B.C. SARKAR, *Phase Detector for Data-Clock Recovery Circuit,* Electronics Letters 14th February 2002, Vol. 38 No. 4, pp. 161-163.

[27] J. HAUENSCHILD, H.-M. REIN, L SCHMIDT, K. WORNER, *Versatile Silicon Bipolar XOR Gate for Signal Processing up to 8Gbit/s*, Electronics Letters 18th January 1990 Vol. 26, No. 2, pp. 114-115.

[28] P. HEYDARI, R. MOHAVAVELU, *Design of Ultra-High Speed CMOS CML Buffers and Latches*, International Symposium on Circuits and Systems, Vol. 4, May 2003, pp. 169-172.

[29] C. R. HOGGE, *A Self-Correcting Clock Recovery Circuit*, Journal of Lightwave Technology, Vol LT-3, No. 6, December 1985, pp. 1312-1314.

[30] K. INIEWSKI, S. MAGIEROWSKI; M. SYRZYCKI, *Phase Locked Loop Gain Shaping for Gigahertz Operation*, International Symposium on Circuits and Systems, Vol. 4, May 2004, pp. 157-60.

[31] H. O. JOHANSSON, C. SVENSSON, *Time Resolution of NMOS Sampling Switches Used on Low-Swing Signals*, IEEE Journal of Solid-State Circuits, Vol. 33, No. 2, February 1998, pp. 237-245.

[32] K. KANDA K. NOSE, H. KAWAGUCHI, T. SAKURAI,. *Design Impact of Positive Temperature Dependence on Drain Current in Sub-1-V CMOS VLSIs*, IEEE Journal of Solid-State Circuits, Vol. 36, No. 10, October 2001, pp. 1559-1564.

[33] D. KEHRER, H. D. WOHLMUTH, H. KNAPP, M. WURZER, A. L. SCHOLTZ, *40-Gb/s 2:1 multiplexer and 1:2 demultiplexer in 120-nm standard CMOS*, IEEE Journal of Solid-State Circuits, Vol. 38, No. 11, November 2003, pp. 1830-1837.

[34] K. KISHINE, K. FUJIMOTO, S. KUSANAGI, H. ICHINO, *PLL Design Technique by a Loop-Trajectory Analysis Taking Decision-Circuit Phase Margin into Account for Over-10-Gb/s Clock and Data Recovery Circuits,* IEEE Journal of Solid-State Circuits, Vol. 39, No. 5, May 2004, pp. 740-750.

[35] S. R. KYTHAKYAPUZHA, Modeling Of Spiral Inductors And Transformers, Master's Thesis, Kansas State University, USA, 2001.

[36] J. LEE B. RAZAVI, A 40-GHz Frequency Divider in 0.18μm CMOS Technology, IEEE Journal Of Solid-State Circuits, Vol. 39, No. 4, April 2004, pp. 594-601.

[37] J. LEE, B. RAZAVI, *A 40-Gb/s Clock and Data Recovery Circuit in 0.18μm CMOS Technology*, IEEE Journal Of Solid-State Circuits, Vol. 38, No. 12, December 2003, pp. 2181-2190.

[38] M.-J. E. LEE, *An Efficient I/O and Clock Recovery Design for Terabit Integrated Circuits*, Ph.D. Dissertation, Stanford University, USA, August 2001.

[39] T. H. LEE, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, 1998.

[40] C. E. LIN, A. S. HOU, *Design of Frequency-to-Voltage Converter Using Successive-Approximation Technique,* Instrumentation and Measurement Technology Conference, Vol. 2, May 2003, pp. 1438-1443.

[41] W.-M. LIN, H.-Y. HUANG, *A Low-Jitter Mutual-Correlated Pulsewidth Control loop Circuit,* Journal of Solid-State Circuits, Vol. 39, No. 8, August 2004, pp. 1366-1369.

[42] L. C. LIU, B. H. LI, *Fast locking scheme for PLL frequency synthesiser,* Electronic Letters, Vol. 40 No. 15, July 2004, pp. 918–920.

[43] R.-F. LIU, Y.-M. LI, H.-Y. CHEN, *A Fully Symmetrical PFD for Fast Locking Low Jitter PLL*, International Conference on ASIC, Vol. 2 , October 2003, pp. 725-727.

[44] J. G. MANEATIS, M. A. HOROWITZ, *Precise Delay Generation Using Coupled Oscillators*, IEEE Journal of Solid-State Circuits, Vol. 28, No. 12, December 1993, pp. 1273-1282.

[45] M. MANSURI, C.-K. K. YANG, *A Low-Power Low-Jitter Adaptive-Bandwidth PLL and Clock Buffer*, IEEE International Solid State Circuits Conference 2003, vol. 1, pp. 430-505.

[46] S. S. MOHAN, The Design, Modeling And Optimization Of On-Chip Inductor And Transformer Circuits Ph.D. Dissertation, Stanford University, December 1999.

[47] S. S. MOHAN, M. HERSHENSON, S. P. BOYD, T. H. LEE, *Bandwidth Extension in CMOS with Optimized On-Chip Inductors*, IEEE Journal of Solid-State Circuits, Vol. 35, No. 3, March 2000, pp. 346–355.

[48] F. MU, MEMBER, C. SVENSSON, *Pulsewidth Control Loop in High-Speed CMOS Clock Buffers,* Journal of Solid State Circuits, Vol. 35, No. 2, February 2000, pp. 134-141.

[49] K. MURATA, T. OTSUJI, T. ENOKI AND Y. UMEDA, *Exclusive OR/NOR IC for > 40Gbit/s Optical Transmission Systems*, Electronics Letters, 16th April 1998, Vol. 34, No. 8, pp. 764-765.

[50] J. MUSICER, *An Analysis of MOS Current Mode Logic for Low-Power and High Performance Digital Logic*, Master's Thesis, University of California in Berkeley, USA, 2000.

[51] N. NEDOVIC, V. G. OKLOBDZIJA, Dual-Edge Triggered Storage Elements and Clocking Strategy for Low-Power Systems, IEEE Transactions On Very Large Scale Integration (VLSI) Systems, Vol. 13, No. 5, May 2005, pp. 577-590.

[52] E. PANAYIRCI, N. EKMEKCIOGLU, *Analysis of a Serial Symbol Timing Recovery Technique Employing Exclusive-OR Circuit*, IEEE Transactions On Communications, Vol. 38, No. 6, June 1990, pp. 915-924.

[53] J. S. PANGANIBAN, *A Ring Oscillator Based Variation Test Chip*, Master's Thesis, Massachusetts Institute Of Technology, June 2002.

[54] B. RAZAVI, *Challenges in the design high-speed clock and data recovery circuits*, IEEE Communications Magazine, Vol. 40, No. 8, August 2002, pp. 94-101.

[55] B. RAZAVI, *Prospects of CMOS Technology for High-Speed Optical Communication Circuits*, IEEE Journal of Solid-State Circuits, Vol. 37, No. 9, September 2002, pp. 1135-1145.

[56] H-M. REIN, *Multi-Gigabit-Per-Second Silicon Bipolar IC's for Future Optical-Fiber Transmission Systems*, IEEE Journal Of Solid-State Circuits, June 1988, Vol.. 23. No. 3, pp. 664-675.

[57] M. RENAUD, Y. SAVARIA, *A Linear Phase Detector For Arbitrary Clock Signals*, International Symposium on Circuits and Systems, May 2002, Vol 4, pp. 775-778.

[58] D. J. RENNIE, *Design and Optimization of Source Coupled Logic in Multi-Gbit/s Clock and Data Recovery Circuits*, Master's Thesis, University of Waterloo, Canada, 2003.

[59] A. REZAYEE, K. MARTIN, *A 9-16Gb/s Clock and Data Recovery Circuit with Three-State Phase Detector and Dual-Path Loop Architecture*, European Solid-State Circuits Conference, September 2003, pp. 684-686.

[60] W. RHEE, *Design Of High- Performance CMOS Charge Pumps In Phase-Locked Loops,* International Symposium on Circuits and Systems, Vol. 2, June 1999, pp. 545-548.

[61] J. ROUTAMA, K. KOLI, K. HALONEN, *A Novel Ring-Oscillator with a Very Small Process and Temperature Variation*, International Symposium on Circuits and Systems, Vol. 1, June 1998, pp. 181-184.

[62] E. SÄCKINGER, *Broadbnd Circuits for Optical Fiber Communication,* John Wiley and Sons, New Jersey, 2005.

[63] E. SÄCKINGER, W. C. FISCHER, *A 3-GHz 32-dB CMOS Limiting Amplifier for SONET OC-48 Receivers*, IEEE Journal of Solid-State Circuits, Vol. 35, No. 12, December 2000, pp. 1884-1888.

[64] J. SAVOJ, B. RAZAVI, *A 10-Gb/s CMOS Clock and Data Recovery Circuit with a Half-Rate Linear Phase Detector*, IEEE Journal Of Solid-State Circuits, Vol. 36, No. 5, May 2001, pp. 761-768.

[65] J. SAVOJ, B. RAZAVI, *High-Speed CMOS Circuits for Optical Receivers*, Kluwer Academic Publishers, 2001.

[66] L. SCHMIDT H-M. REIN, *New High-Speed Bipolar Xor Gate with Absolutely Symmetrical Circuit Configuration*, Electronics Letters, 29th March 1990, Vol. 26, No. 7, pp. 430-431.

[67] S. SIDIROPOULOS, *High Performance Inter-Chip Signalling*, Ph.D. Dissertation, Stanford University, USA, April 1998.

[68] B. G. STREETMAN, *Solid State Electronic Devices*, 4th ed., Prentice Hall, New Jersey, 1995.

[69] C. SVENSSON, A. EDMAN, *10-100 Gb/s Throughput CMOS Techniques*, Symposium on VLSI Circuits Digest of Technical Papers, 1999.

[70] J. TAKASOH, T. YOSHIMURA, H. KONDOH, N. HIGASHISAKA, *A 12.5Gbps Half-rate CMOS CDR Circuit For l0Gbps Network Applications*, Symposium on VLSl Circuits Digest of Technical Papers, June 2004, pp. 268-271.

[71] A. TANABE, M. UMETANI, I. FUJIWARA, T. OGURA, K. KATAOKA, M. OKIHARA, H. SAKURABA, T. ENDOH, F. MASUOKA, *0.18 μm CMOS 10-Gb/s*

Multiplexer/Demultiplexer ICs Using Current Mode Logic with Tolerance to Threshold Voltage Fluctuation, IEEE Journal of Solid-State Circuits, Vol. 36, No. 6, June 2001, pp. 988-996.

[72] A. TANABE, Y. NAKAHARA, A. FURUKAWA, T. MOGAMI, *A Redundant Multivalued Logic for a 10-Gb/s CMOS Demultiplexer IC*, IEEE Journal of Solid-State Circuits, Vol. 38, No. 1, January 2003, pp. 107-113.

[73] A. THANACHAYANONT, *CMOS Transistor-Only Active Inductor For IF/RF Applications*, IEEE International Conference on Industrial Technology, Vol. 2, December 2002, pp. 1209-1212.

[74] K. VICHIENCHOM, W. LIU, *Analysis of Phase Noise due to Bang-Bang Phase Detector in PLL-Based Clock and Data Recovery Circuits*, Circuits and Systems, 2003. International Symposium on Circuits and Systems, Vol. 1, May 2003, pp. 617-620.

[75] J. S. WANG AND P. H. YANG, *Low-voltage CMOS Pulsewidth Control Loop Using Push-Pull Charge Pump,* Electronics Letters 29th Mar. 2001 Vol. 37 No. 7 pp. 409-411.

[76] S. WILLIAMS, H. THOMPSON, M. HUFFORD, E. NAVIASKY, *An Improved CMOS Ring Oscillator PLL with Less than 4ps RMS Accumulated Jitter*, Custom Integrated Circuits Conference, October 2004, pp. 151-154.

[77] A. WORAPISHET, M. TAMSIRIANUNT, *An NMOS Inductive Loading Technique for Extended Operating Frequency CMOS Ring Oscillators*, Midwest Symposium on Circuits and Systems, Vol. 1, August 2002, pp. 116-119.

[78] W.-C. WU, C.-C. HUANG, C.-H. CHANG, N.-H. TSENG, *Low-Power CMOS PLL for Clock Generator*, International Symposium on Circuits and Systems, Vol. 1, May 2003, pp. 633-636.

[79] M. YAMASHINA, H. YAMADA, *An MOS Current Mode Logic (MCML) Circuit for Low-Power Sub-GHz Processors*, IEICE Transactions on Electronics, Vol. E75-C, No. 10, October 1992, pp. 1188-1195.

[80] C.-K. K. YANG, *Design of High Speed Serial Links in CMOS*, Ph.D Dissertation, Stanford Unversity, USA, December 1998.