ADAPTIVE TECHNIQUES FOR ANALOG AND MIXED SIGNAL INTEGRATED CIRCUITS

DISSERTATION

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

By
Ayman Adel Fayed, M.S.

The Ohio State University
2004

Dissertation Committee:
Professor Mohammed Ismail, Adviser
Associate Professor George J. Valco
Associate Professor Joanne DeGroat

Approved by

Adviser
Graduate Program in Electrical Engineering
Modern electronic systems and circuits' designers are facing the increasingly difficult task of designing systems that have to handle very tight sets of specifications, or even multiple sets of specifications, required for adequate performance. Given the numerous variables encountered by the designers during the design process including—but not limited to—the variations in the fabrication process, temperature, transmission media, it becomes very challenging to implement systems that can meet very tight specifications, and yet keep high levels of integration and preserve cost effectiveness. Adaptive circuits and systems are a commonly used technique to combat the variations encountered during the design process. Essentially, an adaptive circuit is a circuit that has the ability to detect any drift in its own characteristics (drift due to temperature or a fabrication process parameter change for example) as well as any change in the characteristics of other circuits that interface with it (a transmission media for instance), and then tune itself back to its desired state accordingly.

In this thesis, a categorization of the variables encountered during the design of electronic systems will be presented. A description of the idea of adaptive systems will be presented along with a discussion of the different circuit elements required for any adaptive system to operate successfully. Two different adaptive techniques are going to be proposed in this research. The first technique is a digital adaptive algorithm for tuning on-chip
resistors that are subject to wide errors due to the fabrication process and temperature variations. The proposed algorithm improves the accuracy of on-chip resistors to be within ±5% instead of the typical ±25% guaranteed by the process, and can potentially achieve more accuracy if more chip area is used. The second technique is an analog adaptive equalizer used to compensate for Intersymbol Interference caused by the limited bandwidth of the transmission media. The Equalizer uses a new voltage-controlled transconductor proposed by the author to achieve wider, highly linear adaptive range, and is used to enable high speed data transmission across a UTP CAT-5 Ethernet cable up to 100 meters in length.
Dedicated to my parents,
my wife,
and my lovely baby girl Layla
ACKNOWLEDGMENTS

I would like to thank my adviser, Mohammed Ismail, for his great support and encouragement through out the course of my studies and research. He hasn't been only a technical and professional mentor, but a personal friend as well. He has been always available when I needed him, and I'm very grateful for his commitment.

I would also like to thank all my co-workers and management at the connectivity solutions department at Texas Instruments Incorporated who with their technical help made this work possible. I'm grateful to Richard Griffith for his great technical insights, Krunali Patel, Andy Tsong, Jurgen Bareither, David K. Johnson for their great encouragement and support during my masters and Ph.D. work.

I would also like to thank my friends Mohammad Al-Shyoukh and Mohammad Kamal for pointing out formatting errors in my thesis as well as for the helpful technical discussions. Special thanks for my friends Rasha Hitata, Sherief Gharraph, Abdulhamid Akel, and Wassim Shaar who handled my graduation paper work at OSU.
VITA

October 14, 1975 ........................................... Born – Portsaid, Egypt

July, 1998 .......................................................... B.S. Electrical and Communications Engineering, Cairo University

1998 - 2001 .......................................................... Teaching Assistant, Dept. of Electrical Engineering, The Ohio State University


PUBLICATIONS


PATENTS


FIELDS OF STUDY

Major Field: Electrical Engineering
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xii</td>
</tr>
</tbody>
</table>

## Chapters

1. **Introduction**                                   1
   1.1 Innovation Areas                                6
   1.2 Thesis Outline                                  8
   List of References                                  12
2. **On-Chip Resistors and Capacitors**               13
   2.1 Passive Resistors                               14
   2.1.1 Accuracy of Passive Resistors                 20
   2.1.2 Matching Properties of Passive Resistors      27
   2.2 Passive Capacitors                              32
   2.2.1 Accuracy of Passive Capacitors                46
   2.2.2 Matching Properties of Passive Capacitors     48
   List of References                                  52
3. **Tuning Architectures**                           54
3.1 Automatic Tuning ........................................................................... 56
  3.1.1 References .................................................................................. 62
    3.1.1.1 Voltage References ............................................................. 63
    3.1.1.2 Current References ............................................................. 66
    3.1.1.3 Time References ................................................................. 68
  3.1.2 Electronically Tunable Elements ............................................. 69
    3.1.2.1 Tunable Resistance .............................................................. 69
    3.1.2.2 Tunable Capacitors ............................................................ 81
    3.1.2.3 Tunable Transconductors ................................................. 82
  3.2 Post Fabrication Tuning ............................................................... 88

List of References ............................................................................. 90

4. A Voltage-Controlled Transconductor ........................................... 92
  4.1 Introduction .................................................................................... 93
  4.2 The Proposed Transconductor .................................................... 98
  4.3 Simulations Results ...................................................................... 102
  4.4 Conclusion ................................................................................... 110

List of References ............................................................................. 111

5. A Digital Adaptive Technique for On-Chip Resistors Tuning .......... 112
  5.1 The Proposed Technique ............................................................ 114
  5.2 Practical Aspects .......................................................................... 126
  5.3 Design Examples .......................................................................... 130

List of References ............................................................................. 136

6. Equalization .................................................................................... 138
  6.1 Intersymbol Interference ............................................................... 139
  6.2 Eye Diagrams ............................................................................... 149
  6.3 Equalization Architectures .......................................................... 153
    6.3.1 The Feed-Forward Architecture ......................................... 155
    6.3.2 The Feed-Back Architecture ............................................... 160
6.3.3 The Mixed Feed-Forward Feed-Back Architecture 163
6.4 Equalization Implementation Techniques 165

List of References 167

7. An Analog Adaptive Equalizer for Wire Line Transceivers Applications 169
7.1 Motivation 171
7.2 Transmission Line Modeling 172
7.3 The Equalization Architecture 176
7.3.1 The Forward Equalizer 180
7.3.2 The Slicer 184
7.3.3 The Reference Signal Generator 188
7.3.4 The Band-Pass Filter 190
7.3.5 The Differential-Difference Squarer 192
7.3.6 The Integrator 194
7.4 Simulation Results 196
7.5 Measurements Results 207

List of References 216

8. Conclusion 218
8.1 Future Research 220

Bibliography 222


LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Resistors values for an accuracy of ±8.75%</td>
<td>131</td>
</tr>
<tr>
<td>5.2</td>
<td>Resistors values including the switches</td>
<td>131</td>
</tr>
<tr>
<td>5.3</td>
<td>Measured resistance values</td>
<td>132</td>
</tr>
<tr>
<td>5.4</td>
<td>Resistors values for an accuracy of ±5%</td>
<td>134</td>
</tr>
<tr>
<td>5.5</td>
<td>Resistors values including the switches</td>
<td>134</td>
</tr>
<tr>
<td>7.1</td>
<td>The equalizer's differential output eye horizontal opening</td>
<td>208</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>(a) Resistor's dimensions</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>(b) Layout of an integrated resistor</td>
<td></td>
</tr>
<tr>
<td>2.2</td>
<td>(a) Multiple fingers resistor</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>(b) Layout of a serpentine structure</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>The diffusion process</td>
<td>17</td>
</tr>
<tr>
<td>2.4</td>
<td>A well resistor with a guard ring to reduce noise coupling from the substrate</td>
<td>19</td>
</tr>
<tr>
<td>2.5</td>
<td>(a) Etching and lateral diffusion errors in well resistors</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>(b) Etching errors in polysilicon resistors</td>
<td></td>
</tr>
<tr>
<td>2.6</td>
<td>(a) Boundary dependent Etching</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>(b) Etching with dummies present</td>
<td></td>
</tr>
<tr>
<td>2.7</td>
<td>(a) Resistor layout with 45° corners and dummy strips</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>(b) Resistor layout with rounded corners and dummy strips</td>
<td></td>
</tr>
<tr>
<td>2.8</td>
<td>Two matched resistors layout</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>(a) good layout</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) poor layout</td>
<td></td>
</tr>
<tr>
<td>2.9</td>
<td>Layout of two matched interdigitized</td>
<td>30</td>
</tr>
<tr>
<td>2.10</td>
<td>(a) Cross coupled layout</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>(b) Layout of two matched common centroid interdigitized resistors</td>
<td></td>
</tr>
<tr>
<td>2.11</td>
<td>A MOS Structure</td>
<td>33</td>
</tr>
<tr>
<td>2.12</td>
<td>C-V characteristics for a MOS Structure</td>
<td>36</td>
</tr>
</tbody>
</table>
2.13 Using transistors as capacitors
(a) In a P-substrate technology
(b) In an N-substrate technology ........................................... 38

2.14 A Well capacitor on P-substrate technology .......................... 40

2.15 A comparison between the C-V characteristics of a PMOS
transistor used as a capacitor and a well capacitor .................. 40

2.16 (a) Parasitic capacitance in a well capacitor
(b) Using a guard ring for noise protection ............................. 41

2.17 A Poly-poly capacitor ..................................................... 42

2.18 Effect of scaling on lateral flux capacitors
(a) Before scaling: lower specific capacitance and higher side area
(b) After scaling: Higher specific capacitance and lower side area .... 45

2.19 Vertical and lateral flux capacitors
(a) Conventional vertical flux capacitor.
(b) Lateral flux capacitor incorporated within
    a vertical flux capacitor ................................................. 45

2.20 An interdigitized lateral flux capacitor .............................. 46

2.21 Layout of two matched capacitor ..................................... 49

2.22 (a) Layout of an integer multiple of a unit capacitor
(b) Layout of a non integer multiple of a unit capacitor .......... 51

3.1 Direct tuning architecture
(a) continuous tuning
(b) discontinuous tuning .................................................. 58

3.2 Indirect tuning architecture .............................................. 59

3.3 Digital based tuning:
(a) with a DAC
(b) with direct digital control ............................................ 62

3.4 Schematic diagram of a simple band-gap circuit .................. 64
3.5 A reference currents generator ........................................... 67

3.6 Linearization techniques based on:
(a) a single transistor
(b) two transistors with common gates
(c) two transistors with common sources
(d) two transistors with common drains .................................... 72

3.7 Linearization techniques based on:
(a) the four-quadrant cell with common drains
(b) the four-quadrant cell with common sources
(c) two transistors with common drains and common sources .......... 78

3.8 A tunable integrator .............................................................. 80

3.9 A triode-mode based transconductor ...................................... 85

3.10 (a) saturation-mode based transconductor
(b) The COMFET structure ...................................................... 87

4.1 Schematic diagram of a linear mode based transconductor ........... 94

4.2 Schematic diagram of the proposed transconductor ..................... 99

4.3 Output currents for the proposed circuit and the original circuit in figure 4.1 at 100mV control voltage ............................................ 104

4.4 Output currents for the proposed circuit and the original circuit in figure 4.1 at 1V control voltage ............................................. 104

4.5 $G_{m+}$ and $G_{m-}$ for the proposed circuit and the original circuit in figure 4.1 at 100 mV control voltage ........................................ 105

4.6 $G_{m+}$ and $G_{m-}$ for the proposed circuit and the original circuit in figure 4.1 at 400 mV control voltage ................................. 105

4.7 $G_{m+}$ and $G_{m-}$ for the proposed circuit and the original circuit in figure 4.1 at 1 V control voltage ........................................ 105
4.8 Transconductance error for the proposed circuit and
the original circuit in figure 1 at 100 mV control voltage 106
4.9 Transconductance error for the proposed circuit and
the original circuit in figure 1 at 400 mV control voltage 107
4.10 Transconductance error for the proposed circuit and
the original circuit in figure 1 at 1 V control voltage 107
4.11 Transconductance differential error for the proposed circuit
and the original circuit in figure 4.1 at 100 mV control voltage 108
4.12 Transconductance differential error for the proposed circuit
and the original circuit in figure 4.1 at 400 mV control voltage 108
4.13 Transconductance differential error for the proposed circuit
and the original circuit in figure 4.1 at 1 V control voltage 109
4.14 The transconductance versus control voltage for the proposed
circuit and the original circuit in figure 4.1 109
5.1 The quantization concept 115
5.2 The resistor block 116
5.3 The proposed tuning architecture 121
5.4 The reference generator 124
5.5 The control matrix 125
6.1 The multi-path effect in wireless networks 140
6.2 The response of a band-limited channel to a single pulse 142
6.3 Response of channel 1 to the first input pattern 143
6.4 Response of channel 2 to the first input pattern 144
6.5 Response of channel 3 to the first input pattern 144
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.6</td>
<td>Response of channel 1 to the second input pattern</td>
<td>145</td>
</tr>
<tr>
<td>6.7</td>
<td>Response of channel 2 to the second input pattern</td>
<td>145</td>
</tr>
<tr>
<td>6.8</td>
<td>Response of channel 3 to the second input pattern</td>
<td>146</td>
</tr>
<tr>
<td>6.9</td>
<td>Response of channel 1 to the third input pattern</td>
<td>146</td>
</tr>
<tr>
<td>6.10</td>
<td>Response of channel 2 to the third input pattern</td>
<td>147</td>
</tr>
<tr>
<td>6.11</td>
<td>Response of channel 3 to the third input pattern</td>
<td>147</td>
</tr>
<tr>
<td>6.12</td>
<td>The Eye diagram of a random data stream across channel 1</td>
<td>150</td>
</tr>
<tr>
<td>6.13</td>
<td>The Eye diagram of a random data stream across channel 2</td>
<td>151</td>
</tr>
<tr>
<td>6.14</td>
<td>The Eye diagram of a random data stream across channel 3</td>
<td>151</td>
</tr>
<tr>
<td>6.15</td>
<td>The effect of enhancing the high frequency components of the received signal on the effective bandwidth of the transmission channel</td>
<td>154</td>
</tr>
<tr>
<td>6.16</td>
<td>The feed-forward equalization architecture</td>
<td>156</td>
</tr>
<tr>
<td>6.17</td>
<td>The frequency response of an all-pass forward equalizer</td>
<td>158</td>
</tr>
<tr>
<td>6.18</td>
<td>The adaptive feed-forward equalization architecture</td>
<td>159</td>
</tr>
<tr>
<td>6.19</td>
<td>The feed-back equalization architecture</td>
<td>160</td>
</tr>
<tr>
<td>6.20</td>
<td>The adaptive feed-back equalization architecture</td>
<td>162</td>
</tr>
<tr>
<td>6.21</td>
<td>The mixed feed-forward feed-back equalization architecture</td>
<td>164</td>
</tr>
<tr>
<td>6.22</td>
<td>The adaptive mixed feed-forward feed-back equalization architecture</td>
<td>164</td>
</tr>
<tr>
<td>7.1</td>
<td>An RLC transmission line model for 100m UTP CAT 5 cable</td>
<td>174</td>
</tr>
<tr>
<td>7.2</td>
<td>The transmission channel including termination networks</td>
<td>174</td>
</tr>
</tbody>
</table>
7.3 The frequency response of 100m and 0m transmission channel 175
7.4 The eye diagram of the 100m transmission channel input 175
7.5 The eye diagram of the 100m transmission channel output 176
7.6 The adaptive equalizer block diagram 177
7.7 The first-order G_m-C high-pass filter with common mode feedback 181
7.8 The frequency response of the boosting filter for different control voltages 183
7.9 Schematic diagram of the three-stage slicer 185
7.10 The frequency response of the slicer's first stage 188
7.11 The reference signal generator schematic diagram 189
7.12 The schematic diagram of the band-pass filter 191
7.13 The frequency response of the band-pass filter 191
7.14 The schematic diagram of the differential difference squarer 193
7.15 The DC response of the differential difference squarer 193
7.16 The schematic diagram of the integrator 195
7.17 The frequency response of the integrator 195
7.18 The frequency response of the forward equalizer, the 100m transmission channel, and the combined frequency response of the channel and the equalizer 197
7.19 The frequency response of the forward equalizer, the 0m transmission channel, and the combined frequency response of the channel and the equalizer 197
7.20 The transient response of the 100m channel, the forward equalizer, and the slicer ........................................ 199
7.21 The output differential eye diagram of the 100m channel ................................................................. 199
7.22 The output differential eye diagram of the forward equalizer for a 100m ................................................. 201
7.23 The output differential eye diagram of the slicer for a 100m channel ....................................................... 201
7.24 The transient response of the error and control signals for a 100m channel ............................................. 202
7.25 The transient response of the 0m channel, the forward equalizer, and the slicer ........................................ 204
7.26 The output differential eye diagram of the 0m channel .............................................................................. 204
7.27 The output differential eye diagram of the forward equalizer for a 0m channel ........................................... 205
7.28 The output differential eye diagram of the slicer for a 0m channel ............................................................ 205
7.29 The transient response of the error and control signals for a 0m channel .................................................. 206
7.30 The layout of the equalizer ......................................................................................................................... 209
7.31 The chip micrograph showing the location of the equalizer on the chip ..................................................... 210
7.32 (a) The input differential eye diagram of the equalizer for a 12 inches cable (b) The output differential eye diagram of the equalizer for a 12 inches cable ................................................................. 211
7.33 (a) The input differential eye diagram of the equalizer for a 10 meters cable
(b) The output differential eye diagram of the equalizer for a 10 meters cable 212

7.34 (a) The input differential eye diagram of the equalizer for a 30 meters cable
(b) The output differential eye diagram of the equalizer for a 30 meters cable 213

7.35 (a) The input differential eye diagram of the equalizer for a 50 meters cable
(b) The output differential eye diagram of the equalizer for a 50 meters cable 214

7.36 (a) The input differential eye diagram of the equalizer for a 100 meters cable
(b) The output differential eye diagram of the equalizer for a 100 meters cable 215
CHAPTER 1

INTRODUCTION

Modern electronic systems are facing many challenges due to the numerous variables and restrictions implied by the nature of the application, the quality of available fabrication technologies, cost effectiveness, and so on. While those variables and restrictions make it an increasingly difficult task for circuits and systems designers, they still -given those variables- have to meet the very tight specifications required for adequate performance of modern systems. The variations in a typical analog and mixed signal system could be categorized into three main categories. The first category of variations is the circuit level category, i.e. variations encountered during the design of each individual circuit block in a system. Those variations include process variations, supply voltage variations, and temperature variations. Process variations are caused by the limited control over the conditions and the characteristics of the fabrication process. This limited control causes variations in the length, width, and threshold voltage of the transistors as well as the absolute values of on-chip resistors and capacitors. In addition to process variations, the system has to meet the required set of specifications for a range of supply voltages and temperatures as well. The typical supply voltage range is ±10% of the nominal value, while the typical temperature
range is from -40 to 125 °C. Taking all those variations into account, it becomes a challenging task to design circuits that can accurately meet the tight specifications set on a system with all those variations present. For instance, in a digital CMOS process, the design of continuous-time filters with on-chip resistors and capacitors becomes impractical for many applications since the values of those resistors and capacitors could vary as much as ±25% from their intended nominal value. This makes it very difficult to control the positions of poles and zeros in a given filter. Another example is the implementation of integrated transmission line termination resistors for digital high speed wire line communication systems. In this case, the variations in those resistors from the nominal value will cause mismatches with the transmission line characteristic impedance, which will cause reflections and hence the loss of signal power and integrity [1-4]. Generally speaking, the design of circuits that need accurate absolute values of resistors, capacitors, voltages, or transistors becomes a difficult task. This includes filters, oscillators, delay elements, transmission lines terminations, and so on. Even in circuits that wouldn't need resistors and capacitors, the variations in the threshold voltage of transistors as well as the length and the width could still be a challenging problem. For example, signal drivers in wire line systems usually have restrictions on the rise and fall times as well as duty cycle distortion. The system specifications usually put an upper and a lower limit on those parameters (usually given in an eye diagram format). If the range of rise and fall times allowed by the specifications is tight, it becomes very difficult to meet the specifications required given the variations in the absolute parameters of the transistors, the supply voltage, temperature, as well as the inherent mismatch between a PMOS and an NMOS transistor.
Different techniques have been developed in the literature to avoid the dependency on the absolute values of devices parameters in general. One technique is based on developing circuits for implementing the required system (filter, oscillator, etc.) such that the targeted parameter (oscillation frequency of an oscillator, cut-off frequency of a filter, .....etc) of the system would depend on a ratio between similar on-chip elements (resistors, capacitors, or transistors) rather than their individual absolute values (Switched-Capacitor techniques is an example) [5]. As will be discussed later in detail, the ratios between similar on-chip elements have a much better accuracy than their individual absolute values. Unfortunately though, it is not always possible to implement the required system with the parameter of interest as a function of a ratio between similar elements. In these cases, depending on the absolute values of on-chip elements is inevitable.

The second category of variations is the system level category. The variations in this category are related to the specifications and the type of the system, i.e. variations that cause the specifications of the system to be changeable. As an example, there are more than one mobile phone standard (GSM, TDMA, etc..). TDMA systems are widely used in the USA, while the GSM standard is widely used in the rest of the world. The need for a single transceiver that can automatically handle the two standards has been increasing recently. This means that there will two completely different sets of specifications that the transceiver has to handle. The impact of having two different sets of specifications on the circuit level is very significant. There are two options: designers can either use two separate circuit blocks with each block handling one set of specifications and switch between them, or one circuit block that can handle both sets of specifications. Obviously, the first option is impractical since it will consume almost double the space even though it might be a logical starting...
point. The second option will introduce another design challenge added to the one already present because of the variations in the first category. For example, designers have to find a way to design a filter that can automatically switch between two different specifications on the bandwidth and the center frequency. This applies also to oscillators, low noise amplifiers, analog to digital converters, and so on. Therefore, designers do not only have to compensate for process, power supply, and temperature variations to comply with one set of specifications, they need to do that to comply with different sets of specifications as well.

The third category of variations is the network level category, i.e. variations in the transmission media. There are two different types of networks: wired networks, and wireless networks. Usually the two biggest problems introduced to digital communication systems through the transmission media is inter symbol interference (ISI) and cross channel interference. ISI is generally caused by the limited bandwidth of the transmission media. The limited bandwidth causes each transmitted symbol to be spread in time beyond its allocated time slot, consequently, interfering with other symbols in the same data stream. In wireless networks, and in addition to limited bandwidth, ISI could also happen due to the multi-path effect, in which an indirect delayed signal that usually results from multiple reflections interferes with the direct signal received by the antenna causing the current received symbol to be contaminated with previous symbols in the same channel. In wired networks, ISI is dependent on the cable length and characteristics. For example, in Ethernet networks applications, the cable length could be any where between zero meters and hundred meters. For each length, ISI is different due to the different frequency response of different cable length. For both wired and wireless systems, equalizers are employed to fix the ISI problem, but the designers are challenged by the fact that they don't only need to equalize for a fixed
channel, they need to equalize for a channel that can change characteristics in an unpredictable way.

Considering the three categories of variations discussed above, it is very obvious that the common thread between all of them is the problem of designing circuits that can handle those variations, preferably in an automatic way. Whether the variations are in the elements composing the circuits (as in the first category), or in the specifications the circuits have to meet (as in the second category), or in the transmission media characteristics (as in the third category), the common solution is adaptation. Adaptation is a very broad technique that can be applied to so many fields in the area of analog and mixed signal circuit design. The first step in any adaptation scheme is to sense the parameter of interest, followed by the second step, which is comparing the sensed parameter to a constant reference resulting in an error signal that indicates the amount of variation in the parameter of interest. The third step involves further processing of the error signal according to a specific algorithm, which results in a control signal that is then could be used to either directly control the parameter of interest in order to minimize or perhaps eliminate the variation represented by the error signal, or to control other different elements in the system to compensate for those variations and keep the performance within the required specifications. Circuits and systems that employ adaptation techniques are called adaptive circuits and systems. Adaptive equalization, automatic impedance control, time constant control, offset cancellation techniques (in comparators), and multi-mode systems are just few examples of systems that apply adaptation techniques in one way or another.
1.1. Innovation Areas

Adaptation techniques are based on sensing the variation in a specific parameter in the system and then quantify this variation. Based on the variation amount, the rest of the system adapts itself to compensate for this variation. Taking that into consideration, using adaptation techniques requires two levels of innovation, i.e. the circuits' level and the system level. On the circuits' level, innovation includes the design of circuits that enables adaptation, while on the systems level, innovation includes the different algorithms or architectures used to perform the adaptation process. As mentioned in the introduction, in order to adapt for a variation in a specific parameter, the parameter has to be sensed. This will require the design of a circuit that can track and sense the wide variations in that parameter. Therefore, the sensing circuit has to have a very wide input and output range and it also has to be insensitive as much as possible to process and system variations in order to be used efficiently in the sensing process. Sensing circuits might be amplifiers, OPAMPS, transconductors, filters and so forth. The output of the sensing circuit has to be then compared to a reference signal, which implies that comparison techniques have to be developed. Those comparison techniques have to be able to handle wide input ranges and might have to have a highly linear performance for adequate comparison. In the analog domain, the comparison could be simply done using a differential amplifier, while in the digital domain, analog to digital converters may have to be used. Any comparison technique needs some sort of a reference signal, and in order for the comparison result to be meaningful, this reference has to be insensitive to all the variations and the parameters that the system is trying to detect and compensate for. Designing a reference could be a very
difficult and challenging task, which leads to another level of innovation as will be discussed later in details. The result of the comparison process is an error signal that gets processed further to generate a control signal that is used to tune the parameter of interest back to its desired value. This tuning process introduces another challenge, i.e. the design of circuits that have variable –or in other words- programmable characteristics that can be controlled with a control signal (a voltage or a current signal). This includes the design of variable gain amplifiers, voltage controlled transconductors, programmable filters, programmable resistors, programmable time constants, and so on. Those circuits have to have a large programmable range (perhaps linear as well) in order to adapt for large variations in the parameter of interest. Taking into account the sensing process, the comparison process, and the programmable behavior of circuits, using adaptation techniques opens a wide range of possibilities for the circuits' level innovation.

The system level innovation includes two elements: the algorithm used for processing the error signal to generate the control signal used to tune the parameter of interest, and the architecture of the adaptation process in general. The algorithms used for processing the error signal are very important since they determine the total accuracy, speed, and stability of the adaptation process. Heavy mathematical analysis has to be performed in order to develop the best and most efficient algorithm for a given application. There are generally three possible techniques that could be used in the adaptation process, i.e. analog techniques (continuous or discrete), digital techniques, or mixed signal techniques. The error signal that results from the comparison process could be an analog signal (continuous or discrete) or a digital word. Based on the nature of the error signal, the further processing needed on it in order to generate the control signal could be done either in the digital or the
analog domain. Once the control signal is generated and based on its nature, the programmable circuits in the system could be either analog or digitally programmable blocks. Therefore, the adaptation process could be done purely in the analog or the digital domain, or it can employ both techniques simultaneously. Digital techniques are very powerful since they enable very complex algorithms to be done using digital signal processors (DSP), but they usually require analog to digital converters (ADC) and digital to analog converters (DAC). Particularly, if the variations range of the parameter of interest is much larger than the desired accuracy, the design of the ADC becomes very difficult since high resolution is required. Analog techniques on the other hand don't have the resolution limitation discussed above, but mathematical processing required on the error signal becomes relatively complicated in the analog domain. Mixed signal architectures are usually used to achieve the advantages of both analog and digital techniques. No matter what technique or architecture is used for the adaptation process, stability, convergence time, and speed have to be considered when developing the adaptation architecture and algorithm.

1.2. Thesis Outline

As mentioned before, three categories of variations can be encountered in analog and mixed signal circuits and systems design. This thesis concentrates on adaptation techniques for variations in the first and third categories only, i.e. variations due to process, temperature, supply voltage, and transmission media. The second category variations are left for future research. As discussed before, the absolute value of the different parameters of
on-chip elements (resistor, capacitors, and transistors) usually suffer from wide variations due to process and temperature variations (first category variations). In order to understand and quantify those variations, chapter 2 starts with an overview of the available types of on-chip resistors and capacitors. For each type, a discussion of the sources of errors in their absolute values due to variations in the fabrication process as well as temperature will be presented. Chapter 2 will also give an overview of design and layout techniques to minimize those errors. An important aspect of on-chip resistors and capacitors is their matching properties. Chapter 2 will also give an overview of those matching properties with a discussion of the different layout techniques to improve and achieve accurate matching.

Once the sources of errors in the absolute values of on-chip elements are discussed in chapter 2, chapter 3 goes over the different techniques and possibilities to tune for those errors. A discussion of the different on-chip automatic adaptive tuning techniques that could be used to control the accuracy of the absolute value of any parameter in the system including the absolute values of on-chip resistors and capacitors will be presented. An overview of the different elements used in the tuning process including references and electronically tunable elements will be discussed. Post fabrication tuning options will also be briefly described.

An important element of any automatic tuning process is the presence of an electronically programmable circuit that can be used to tune the parameter of interest to the desirable accuracy. Chapter 4 is dedicated for the discussion of a new voltage controlled transconductor proposed by the author that could be used in conjunction with any automatic tuning process. The proposed transconductor is specifically used in the implementation of an analog adaptive equalizer for a high speed digital wire line
communication system presented in chapter 7. The proposed transconductor achieves a highly linear wide input range as well as a wide programmable range.

Once the different tuning architectures and elements are presented in chapter 3 and 4, chapter 5 goes on with presenting a new digital tuning algorithm for on-chip resistors proposed by the author. The chapter concentrates on the theoretical and mathematical background behind the algorithm along with a discussion of the advantages of the proposed tuning technique over other widely used techniques. The significance of this part of the research that it introduces a simple, yet powerful methodology for implementing accurate on-chip resistors that can enable the integration of some applications on chip that were not possibly done except by using external components. For instance, the proposed technique was successfully used for integrating accurate on-chip transmission line termination resistors for a high speed digital communication transceiver (480Mbps). The proposed tuning technique can also be used in any RC based application (filters, oscillators, etc.), where accurate resistors are needed.

Chapter 6 concentrates on the third category variations, i.e. variations in the transmission media. The chapter starts by a discussion of the problem of Intersymbol Interference (ISI) in digital wireless and wire line communication systems, followed by a discussion of the available techniques to minimize the impact of ISI on the quality of the data reception. In particular, equalization theory and the different algorithms used in the equalization process are going to be discussed. This includes fixed and adaptive equalization and the different architectures that could be used (analog, digital, and mixed signal) to implement equalization algorithms along with the advantages and disadvantages of each technique.
Chapter 7 uses purely analog adaptive equalization architecture to implement a 125Mbps transceiver that operates over a variable length of an Untwisted-Shielded-Pair (UTP) Category 5 (CAT 5) Ethernet cable. The length of the cable varies from a foot up to 100 meters. A discussion of the requirements that every circuit block used in the system needs to meet will be presented, along with simulations and measurements results. In chapter 8, a conclusion of the thesis will be presented by summarizing all the information that has been discussed. A proposal of future research topics and extensions to the current research will also be discussed.
LIST OF REFERENCES


CHAPTER 2

ON-CHIP RESISTORS AND CAPACITORS

Resistors and capacitors are both very important basic building blocks in any analog IC design. Numerous applications in the area of analog signal processing make use of these two elements, either separately, or combined. Those applications include – but not limited to- oscillators, filters, delay elements, matching networks, amplifiers, and so on. Using off-chip resistors and capacitors have the advantage of being very accurate, but are limited by the number of pins that could be dedicated to them. This is in addition to the higher cost associated with using external components. The ability to implement resistors and capacitors on the chip along with the circuits that use them enables high levels of integration. CMOS processes offer different ways to implement resistors and capacitors on-chip. This chapter will give an overview of the available techniques to implement on-chip resistors and capacitors. The different sources of errors in their absolute values due to variations in the fabrication process as well as temperature will be reviewed. An overview of design techniques to minimize those errors along with the matching properties of on-chip resistors and capacitors will be presented. Different layout techniques to improve and achieve accurate matching will be also discussed.
2.1. Passive Resistors

Passive resistors are implemented using the classic concept of the uniform resistivity and the geometry of a conducting material. The concept of passive here is presented versus the concept of active, which means implementing the resistor function using an active element (a transistor). The universal equation that relates the uniform resistivity of a material and its geometry (assuming rectangular geometry) to its total resistance is presented in equation 2.1:

\[ R = \rho \frac{L}{A} = \rho \frac{L}{W \times t} \]  

(2.1)

where \( \rho (\Omega \cdot \mu m) \) is the resistivity of the material, \( L (\mu m) \) is the length in the direction of the current, \( W (\mu m) \) is the width perpendicular to the direction of the current, and \( t (\mu m) \) is the depth of the material [1]. Figure 2.1a shows a diagram of the different dimensions and the layout of an integrated resistor. As equation 2.1 shows, the total resistance could be determined by \( L \) and \( W \) given \( \rho \) and \( t \). A modification for equation 2.1 is given in equation 2.2:

\[ R = \rho_s \frac{L}{W} = n \times \rho_s \]  

(2.2)

where \( \rho_s (\Omega/\square) \) is the sheet resistance of the material per square, and \( n \) is the number of squares formed by \( L \) and \( W \). Taking the resistance of the ohmic contacts into account, the total resistance of the integrated resistor showed in figure 2.1b will be:

\[ R_{tot} = 2R_{Cont} + n \times \rho_s \]  

(2.3)

Where \( R_{Cont} \) is the contact resistance.
Counting the number of squares is straightforward if the resistor is laid out as a single slap of the conductive material as shown in figure 2.1b or as a multiple discontinuous slaps as in figure 2.2a. Very often though, when a large resistor is needed, it is more efficient from an area perspective to layout the resistor in a serpentine structure as shown in figure 2.2b [2]. At the corner squares, the current will be changing direction, therefore the simple formula in equation 2.2 will not be accurate in estimating the resistance of the corner square. Roughly, corner squares are counted as a half square. Methods for counting corner squares could be found in [1].

The layers available in CMOS integrated circuits for implementing passive resistors are diffusion (n⁺, p⁺, well) and polysilicon [2]. Each type has its own characteristics, behavior with temperature change, and reasons for inaccuracy. As mentioned before, the inaccuracy in the integrated resistance value could be as high as ±25%. Diffusion is one of the primary methods used to introduce different types of impurities (Boron, Phosphorus, etc.) to the semi-conductor material. Using diffusion, the type of the
majority current carriers in silicon, as well as its resistivity can be controlled. The diffusion process starts with introducing a high concentration of the impurity material on the surface of the substrate at high temperature (typically varying from 900 to 1200 °C depending on the type of impurity) for a specific period of time. The impurity material will then diffuse to a certain depth inside the substrate causing its electrical characteristics to change. Figure 2.3 shows the diffusion process of a donor material into a P-substrate. The donor material diffused into the P-substrate converts the p-type silicon to n-type silicon and therefore changes the majority current carries which essentially changes the resistivity of the silicon. The point where the type of the silicon changes from n-type to p-type is called the junction depth, which is a function of the diffusion process (concentration of the impurities introduced, the diffusion time and temperature) [1].

As shown in equation 2.2 the total resistance of the diffused material could be calculated by knowing the sheet resistance and the number of squares. The sheet resistance
could be easily calculated if the resistivity of the conducting material is uniform across the depth of the material, which is not the case in the diffusion process. In this case the sheet resistance can be calculated using the following equation:

\[
\rho_s = \frac{1}{\int_{X_j}^\infty q\mu N(X)dx}
\]

(2.4)

where \( X_j \) is the junction depth, \( \mu \) is the majority carrier mobility, and \( N(X) \) is the net impurity concentration as a function of depth. It can be shown that for a given diffusion profile, the sheet resistance \( \rho_s \) is uniquely related to the surface concentration of the diffused material as well as the background concentration of the substrate [1]. Depending on the type and the concentration of the diffused impurity, the sheet resistance of the diffused resistor could be in the order of \( K\Omega/\square \) like in the case of well resistors (n or p type), or it
could be in the order of tens of $\Omega/\square$ like in the case of source and drain diffusion resistors (n$^+$ or p$^+$).

The polysilicon layer is usually used to form the gates of the transistors. It is made up of small crystalline regions of silicon. Therefore, in the strictest sense polysilicon is not amorphous silicon, and it is not crystalline silicon such as the wafer [3]. Amorphous silicon is made up of randomly organized silicon atoms, while crystalline silicon is formed by silicon atoms organized in an orderly fashion. The use of polysilicon gates is a key advance in modern CMOS technology since it allows the source and drain regions to be self-aligned to the gate, thus eliminating parasitics from overlay errors between the gate and the source and drain diffusion. The polysilicon layer used for the gates however has to be highly doped, otherwise the depletion of the gate itself will result in a series capacitance with the gate oxide capacitance, which in turn leads to a reduced inversion layer charge density and degradation of the MOSFET transconductance [4]. Therefore, polysilicon has a very low resistivity in the order of 20 to 30 $\Omega/\square$, and it can be used to implement relatively low resistors. A special type of polysilicon resistors is known as silicide-block poly resistor. This type of polysilicon resistors requires an extra mask during the fabrication process to block the high doping step performed on the polysilicon used for the gates of the transistors from affecting the silicide-block polysilicon. Therefore the doping of the silicide-block resistors is kept around a critical value that will minimize the variations of the sheet resistance of that layer with temperature. This issue will be discussed later in detail. The critical sheet resistance of the silicide-block polysilicon at which a minimum variation with temperature could be achieved is around 300 $\Omega/\square$. 

18
Diffusion resistors –specifically well diffusion- have the advantage of having a higher sheet resistance, therefore implementing the required resistance will consume a relatively smaller area than polysilicon resistors. It suffers though from a poor noise and speed performance due to its direct contact with the substrate. This direct contact with the substrate will inject noise from the other circuits on the same substrate into the body of the resistor. The parasitic capacitance between the resistor and the substrate will also be significant. One way to reduce the noise coupling from the substrate is the use of a grounded guard ring around the well diffusion to absorb the noise from the substrate [2]. Figure 2.4 shows the guard ring. It is also worth mentioning that well resistors have a slight voltage dependent behavior. The reason behind this behavior is that the depletion region depth in the well resistance is a function of the reverse bias between the substrate and the common mode voltage of the plus and minus terminals of the resistor. Since the extension of the depletion region into the well reduces the effective depth of the well, it causes the sheet resistance to slightly change with voltage, therefore causing some nonlinearity in the resistor performance.

Figure 2.4 A well resistor with a guard ring to reduce noise coupling from the substrate.
Polysilicon resistors have the advantage of being isolated from the substrate, which enhances the noise and speed performance of the resistor due to the lower parasitic capacitance with the substrate. It also does not have a voltage dependent behavior as diffusion resistors do. Silicide-block polysilicon resistors specifically have the advantage of minimal variation with temperature, but they require an extra processing step during fabrication. Generally polysilicon resistors have the disadvantage of having low sheet resistance, which implies a higher area consumption to implement the required resistance.

2.1.1. Accuracy of Passive Resistors

In the development of any process, it is very important to evaluate how well the process can control the parameters of different devices. For that purpose, test structures of the device of interest are built and then measured for different wafers and different lots. This evaluation process will result in a statistical distribution for the parameters of interest across different process conditions. The test structures used have to be simple to minimize measurement errors, yet they have to be a good and accurate representative of the different parameters under test. The task of building those test structures might be complicated especially if too many parameters need to be characterized. In the case of resistors, typically van der pauw structures are used to measure the sheet resistance $\rho_s$, a pair of resistors with similar lengths to determine the encroachment, and a Kelvin contact resistance structure to determine the combined contact and spreading resistance introduced by a contact. There
have been also other proposed structures in the literature that can accurately determine the accuracy and matching properties of on-chip resistors across process variations [3].

As equation 2.2 suggests, the error in the value of \( R \) is a function of the error in \( \rho_s \), \( L \), and \( W \). The First source of error is \( \rho_s \), which is mainly a function of the fabrication process, and temperature variations during the circuit operation. During fabrication, the uncertainty in the diffusion time, temperature, and doping concentration determines the amount of error in the absolute value of \( \rho_s \). This error varies from one process to another depending on how well the different process steps are controlled, usually a good characterization of different resistor structures using the process could give an accurate estimation of the amount of error expected. On the other hand, after the fabrication process is done and a certain value of \( \rho_s \) is achieved to within a certain accuracy, during the operation of the circuit using the resistor, the temperature of the resistor material will change up and down depending on the current flow in the resistor. It is a well known fact that the resistivity of any conducting material will vary with temperature, therefore certain error will be added to the absolute value of the resistor due to the variation in \( \rho_s \) with temperature [1]. The variation with temperature is more difficult to handle since it varies with time and with the circuit operation, unlike the errors introduced by the fabrication process which once they happen they stay the same. Generally speaking, the circuit designer has to make sure that his circuit can tolerate those variations, otherwise a tuning scheme has to be continuously monitoring the value of the resistor and correcting for the errors. A measure of the variation in the resistor value with temperature is defined by the temperature coefficient [4]:
where $R$ is the resistance value at nominal temperature, and $T$ is the temperature in Kelvin. $TC$ is usually measure in ppm (parts per million).

Generally speaking, the sign of the temperature coefficient (+ or -) is a function of the current carriers concentration in room temperature. When the temperature increases, more energy becomes available to the electrons (or holes) to be liberated from the nucleus and contribute to the current flow, hence reducing the effective total resistivity of the resistor (negative temperature coefficient). On the other hand though, increasing the temperature leads to more vibrations in the atoms of the conductor's material, which consequently reduces the mean free path the current carrier can travel before it hits the lattice and loses its energy, i.e. phonon-lattice scattering. This reduction in the mean free path of the current carriers increases the resistivity of the resistor (positive temperature coefficient). Since there are two contradicting factors, the effective temperature coefficient will depend on which factor is more dominant. Generally, if the current carriers' concentration at room temperature is already high, the temperature coefficient tends to be positive, while if the concentration is low, the temperature coefficient tends to be negative. Since the doping level determines the concentration of the current carriers, then by changing the doping level, the temperature coefficient could be positive, negative, or zero.

Diffusion resistors generally (well, n+, and p+) are usually doped enough that they tend to have a positive temperature coefficient. Well resistors have a temperature coefficient ranging from 1500 to 2500 ppm. Polysilicon resistors on the other hand have two different types. The first type is the one that is used for the gate of the MOS transistor. This type is highly doped and has a very low resistivity in the order of 20 to 30 $\Omega/\square$. Since it is highly
doped (higher concentration of carriers), this type has a positive temperature coefficient ranging from 1000 to 2000 ppm. The second type is the silicide-block polysilicon resistor, where the doping of the silicide-block resistors is kept around a critical value that will give almost zero, or very low temperature coefficient. The critical sheet resistance of the silicide-block polysilicon at which a very low temperature coefficient could be achieved is around $300 \, \Omega/\square$. This type of resistors usually has a temperature coefficient ranging from -10 to 10 ppm.

The second source of error in the resistor value is the error in $W$, and $L$. This is mainly caused by the etching process during the photo step that defines the geometry of the resistor. Figure 2.5 shows the etching process for a well resistor and a ploy resistor. As shown in figure 2.5a, the actual width (or length) of the well area -where the diffusion will take place- will be wider than the drawn width on the photo mask, therefore the effective resistance of the well area could be higher or lower than expected depending on the error $\Delta W$ and $\Delta L$. The same concept also applies to polysilicon resistors as shown in figure 2.5b with the difference that the actual $W$ or $L$ will always be less than the drawn ones. A unique

![Figure 2.5](image-url)

\textbf{Figure 2.5} (a) Etching and lateral diffusion errors in well resistors.
(b) Etching errors in polysilicon resistors.
phenomenon that happens only in diffusion resistors is the lateral diffusion process. As shown in figure 2.5a, the diffused material will also tend to diffuse laterally. Therefore the effective width (or length) of the well resistor will also be higher than the original drawn value. This phenomenon will add to $\Delta W$ and $\Delta L$ as well. If we take into account the errors in $\rho_s$, $W$, and $L$, then the total resistance value will therefore be:

$$R = \left(1 + \frac{\Delta \rho_s}{\rho_s}\right) \left(1 + \frac{\Delta L}{L}\right) \rho_s \frac{L}{W} = EF \cdot R_{\text{ideal}}$$  \hspace{1cm} (2.6)$$

where $R_{\text{ideal}} = \rho_s \frac{L}{W}$ is the required resistance value, and the error factor

$$EF = \left(1 + \frac{\Delta \rho_s}{\rho_s}\right) \left(1 + \frac{\Delta L}{L}\right)$$

is the error due to the variations in the different parameters.

The percentage error in the resistor value will then be:

$$\text{Error}(\%) = [EF - 1] \times 100$$  \hspace{1cm} (2.7)$$

As shown by equation 2.7, in order to minimize the error in the resistance value, the error factor $EF$ has to be as close as possible to 1. Generally $\frac{\Delta \rho_s}{\rho_s}$ is a function of the process and there is nothing that could be done in the layout of the resistor to minimize it. On the other hand, $\Delta L$ and $\Delta W$ are solely dependent on the etching process not the absolute value of $L$ and $W$, therefore in order to minimize $\frac{\Delta L}{L}$ and $\frac{\Delta W}{W}$ it is a good practice to design the resistor with a large $L$ and $W$. For example if the required resistance is three squares, it is better to design the resistor with $W = 2 \mu m$ and $L = 6 \mu m$ instead of $W = 1 \mu m$ and
\[ L = 3 \mu \text{m} \]. The obvious drawback though is the larger area the resistor will occupy. Another phenomenon that is worth mentioning is that the etching process is generally boundary dependent [2]. That means that the amount of active etching in the body of the resistor will be a function of the shape of the boundary of the resistor and also of what is on the side of the resistor during the etching process. Generally, active etching is more efficient if there was a free space around the boundary of the resistor. Figure 2.6a shows an example of how the error \( \Delta W \) (or \( \Delta L \)) could be different from the left and the right of the resistor due to different boundary condition. In addition to the fact that \( \Delta W \) (or \( \Delta L \)) will be higher if there was a free space on the boundary, the non uniformity in \( \Delta W \) (or \( \Delta L \)) across the structure of the resistor will cause more uncertainty in the value of the resistor as well. For this reason it is a good practice in precise applications to add dummy resistors around the main resistor as shown in figure 2.6b to guarantee less active etching and also a uniform boundary condition, hence a lower and uniform \( \Delta W \) and \( \Delta L \). Another problem that is worth mentioning is related to the layout of the resistor in figure 2.2b. Since the etching process is boundary dependent, there will be a higher active etching that happens at the edges of the corner.

![Diagram](a) Boundary dependent Etching, (b) Etching with dummies present.

25
square of the resistor shown in figure 2.2b [2]. This higher etching will cause the edges of the corner square to be rounded and their shape to be poorly controlled. The poor control over the shape of the corner square will cause uncertainty in the contribution of the corner square to the total resistance. For precise applications, the sharp edges should be designed to be rounded or with 45 degrees as shown in figure 2.7. This way more control over the shape of the corner square could be achieved and therefore more accuracy in the value of the resistor.

Figure 2.7 (a) Resistor layout with 45° corners and dummy strips. (b) Resistor layout with rounded corners and dummy strips.
2.1.2. Matching Properties of Passive Resistor

As discussed in the previous section, the uncertainty in the absolute value of integrated resistors could be as high as ±25% from the designed value. This loose accuracy might be inadequate for many applications. Matching between resistors on the other hand could achieve much more accuracy than the absolute value of each resistor. In fact, with careful layout, matching between resistors could be kept within less than 0.1% error for most processes [2]. Generally speaking, the reason for the ability of the process to achieve good matching is that the variations in the process parameters from one point to another on the die will be very small if the two points are very close to each other. Therefore, if two resistors are located very close to each other, then it is very likely that any change in the process parameters in the area these two resistors occupy will be the same for both. It is worth mentioning that the ability of the fabrication process to match two elements applies not only to resistors, but also for transistors and capacitors. This fact has driven the development of many circuit techniques that depend solely on matching rather than absolute values, switch capacitor circuits is a good example.

There are many guide lines for design and layout to achieve good matching between integrated resistors. Those guide lines apply to transistors and capacitors as well. Generally, if the two resistors are close to each other, then the process parameters for both of them will be very close, but this is not enough to achieve the best matching. The layout has to also guarantee that the two resistors will be affected in the same way by any change in the process parameters even if the change is similar for both. There are different levels of layout complexity in order to achieve good matching between two resistors. The first level is simply
laying out the two resistors in a very close proximity to each other and keeping the same orientation for both. Figure 2.8 shows a good and a poor layout. Keeping the same orientation will likely cause the two resistors to be affected in same way by any process variation, and if the two resistors are close enough, then the process parameters will be the same for both of them, and they will track each other in value even though their absolute value is significantly changing with any change in the process. Note that presence of the dummy elements will guarantee a uniform etching rate across the two resistors, which will also help matching.

Generally, even if the application requires only good matching between two resistors, it is still a good practice to try to reduce the effect of $\Delta W$ and $\Delta L$ on the absolute value of each individual resistor since a lower $\Delta W$ and $\Delta L$ means also lower mismatch between the two resistors. Therefore, it is not recommended to use minimum dimensions allowed by the process for either $L$ or $W$. IF the area allows, then at least double the minimum dimension

![Figure 2.8](image.png) Two matched resistors layout (a) good layout, (b) poor layout.
allowed by the process should be used for either $L$ or $W$. One still has to be careful though because increasing the sizes of $L$ and $W$ will improve the matching because of smaller effect of $\Delta W$ and $\Delta L$ but only up to a limit. Increasing the sizes of the resistors beyond that limit will introduce another source of error due to $\Delta \rho_S$ since the larger the area the structure occupies the larger the difference in the process parameters from one end of the structure to the other end, which obviously reduces the matching quality between the two resistors in the structure.

Sometimes, it is not possible to design each resistor as a single slap of material as shown in figure 2.8 specially if the resistor value is too high (which requires very long structures) or too low (which requires very wide structures). In that case laying out the resistor as a single slap of material will suffer from long range process variations and therefore poor matching. In these cases laying out the resistors in a serpentine structure as shown in figure 2.2b is very common. The serpentine layout will suffer from little $\Delta \rho_S$ in both the X and Y dimensions, which usually gives a lower overall error in the value of the resistor than suffering from a large $\Delta \rho_S$ in only one dimension. Matching between two serpentine resistors will not be very good if simply the two resistors are laid out next to each other. This will lead to the second level of complexity in the layout of two matched resistors, which is the interdigitized approach. In this approach, each resistor is designed using multiple fingers instead of a serpentine structure, then the two resistors are laid out as shown in figure 2.9. In the interdigitized structure, the centroid of both resistors is kept very close to each other, and therefore better matching is achieved, usually within 0.5% [2]. As shown in figure 2.9, at each given location on the X dimension there will be a finger from one
resistor and another finger from the other resistor at very close proximity and both of them will suffer from the same variations in the process, therefore minimizing the mismatch due to the longer X dimension. As was mentioned before, the wider the structure becomes, the more $\Delta \rho_s$ will be across it. The interdigitized approach fights that problem by guaranteeing that at each given location on the X dimension there will be a finger from one resistor and another finger from the other, but it does not take into account the variations in the Y dimension, which will be a function of X. Usually this is not a problem if the dimension of each resistor finger in the Y direction is short enough to neglect the $\Delta \rho_s$ in the Y direction. If this is not the case, then the only solution is to divide each resistor into more number of fingers with shorter dimension in the Y direction, but this essentially means increasing the width of the whole interdigitized structure in the X direction due to the larger number of fingers, which will again defeat the purpose of matching. This leads us to the third level of

Figure 2.9 Layout of two matched interdigitized resistors.
complexity in the layout of the two matched resistors, which is the common centroid (sometimes is called a cross coupled) approach shown in figure 2.10. In this approach, each resistor is divided into two elements and then laid out as shown in figure 2.10a. Starting from the center of the common centroid structure, the variations in the process parameters in both the X and Y direction will always see an element of each resistor, therefore the two resistors will suffer from process variations equally in the X and the Y dimension. If the number of fingers that compose each resistor element in figure 2.10a is designed to be an even number, then yet another level of accurate matching could be achieved by interdigitizing each resistor element as shown in figure 2.10b.

Figure 2.10 (a) Cross coupled layout.
(b) Layout of two matched common centroid interdigitized resistors.
2.2. Passive Capacitors

Parallel plate capacitors are the most common form of capacitors. Two plates of a highly conductive material with a dielectric in between are simply all what you need to implement a capacitor. The value of the capacitor can be determined using the following equation:

\[ C = \frac{\varepsilon_r \varepsilon_0 \cdot WL}{t} = C_s \cdot A \quad (2.8) \]

where \( \varepsilon_0 \varepsilon_r (\text{fF/\text{m}}) \) is the permittivity of the dielectric material (\( 3.45 \times 10^{-6} \text{ fF/ \text{um} } \) for silicon dioxide), \( t (\text{\text{um}}) \) is the dielectric thickness, and \( WL (\text{\text{um}^2}) \) is the area of the capacitor plate [2]. The specific capacitance \( C_s (\text{fF/\text{um}^2}) \) is the capacitance per unit area or per square. The CMOS technology naturally enables the implementation of capacitors. Two of the highly conductive layers available in the process could be used to implement the lower and the top plates of the capacitor with the silicon dioxide layer in between acting as the dielectric. The conductive materials available in CMOS processes are the diffusion layers (p+, n+), the polysilicon layer, and the metal layers. Before we discuss the different capacitors available in CMOS processes, we will first discuss the fundamentals of Metal-Oxide-Semiconductor (MOS) capacitor since evaluating the capacitor value is a little more complicated than the formula presented in equation 2.8.

A simple MOS structure is shown in figure 2.11. Here the semiconductor or the well is chosen to be a P-type silicon, but the analysis is essentially the same for an N-type silicon. For simplicity the well is assumed to be at zero potential, while the metal potential is \( V_g \). When \( V_g \) is negative, more holes will be attracted to the well silicon-dioxide interface,
which effectively increases the conductivity of the well right beneath the oxide. This condition is called "accumulation" condition since the majority current carriers (holes in this case) are being accumulated at the well-oxide interface due to the negative voltage applied on the metal. Since the conductivity of the well right beneath the oxide is significantly high during the accumulation condition, any charge stored in the capacitor will practically be stored in a very thin layer right beneath the oxide, which means that the total capacitance of the structure will essentially be the oxide capacitance $C_{ox}$ and can be determined using equation 2.8. When $V_g$ is equal to zero, no accumulation of the majority carriers happens right beneath the oxide, therefore the charge stored in the capacitor will be distributed across a certain depth in the well, which effectively adds a series capacitor to $C_{ox}$. This condition is called the flat band condition. The flat band capacitance will be a little smaller than $C_{ox}$ and can be determined by [5]:

![Figure 2.11 A MOS Structure.](image)
\[
\frac{1}{C} = \frac{1}{C_{\text{ox}}} + \sqrt{\frac{KT}{\varepsilon_s q^2 N_a}}
\]

(2.9)

where \( \varepsilon_s \) is the silicon permittivity \((1.04 \times 10^{-12} \text{ F/cm})\), \( q \) is the charge of the electron \((1.6 \times 10^{-19} \text{ C})\), and \( N_a \) is the doping concentration.

When \( V_g \) starts to slightly turn positive, the holes will start to be repelled away from the well-oxide interface due to the positive voltage applied on the metal, which will create a depletion region beneath the oxide, therefore this condition is called "depletion" condition. In the depletion condition the charge stored in the capacitor will be distributed across the depth of the depletion region created. This distributed storage of charge can be modeled as an extra capacitor in series with the oxide capacitor \( C_{\text{ox}} \), which will cause the total capacitance of the MOS structure to decrease. The total value of the capacitance in the depletion condition could be shown to be [5]:

\[
C = \frac{C_{\text{ox}}}{\sqrt{1 + \left(2C_{\text{ox}}^2 V_g / \varepsilon_s q N_a\right)}}
\]

(2.10)

As \( V_g \) goes more and more positive, an inversion layer of minority carriers (electrons in this case) starts to form at the well-oxide interface, which effectively again increases the conductivity of the well layer right beneath the oxide. This condition is called the "inversion" condition. In this condition, any charge stored in the capacitor will practically be stored in the thin layer right beneath the oxide, which brings the total capacitance back to \( C_{\text{ox}} \) again as it was during the accumulation condition. It is worth mentioning though that if the variations in \( V_g \) is fast (100 Hz or more), the total capacitance during the inversion condition will not rebound to \( C_{\text{ox}} \), in fact it will keep decreasing until it hits a \( C_{\text{min}} \) given by [5]:

34
\[
\frac{1}{C_{\text{min}}} = \frac{1}{C_{\text{ox}}} + \sqrt{\frac{4K\ln\left(\frac{N_d}{n_i}\right)}{\varepsilon_a q^2 N_a}}
\]  
(2.11)

where \( K \) is Boltzmann's constant \((1.38 \times 10^{-23} \text{ } J/\text{K})\), \( T \) is temperature in Kelvin, and \( n_i \) is the intrinsic carrier density in silicon. The fundamental reason behind that behavior is that the minority carriers in the inversion layer need some time to completely charge or discharge the thin layer right beneath the oxide in response to the variations in \( V_g \). If the variations in \( V_g \) is relatively fast, then the minority carriers will be effectively charging and discharging a wide depth in the well instead of a very thin layer underneath the oxide. Hence an effective series capacitance will form with \( C_{\text{ox}} \), which will keep the total value of the capacitance at the lower value shown in equation 2.11. Figure 2.12 shows the capacitance variation of a MOS structure with P-type well versus \( V_g \) for both low and high frequencies. In an N-type MOS structure, the same capacitance-voltage behavior will happen with an inverted polarity of the metal-to-well voltage. Analyzing figure 2.12, we can see that generally in order to get the highest capacitance from a MOS capacitor, \( V_g \) has to bias the capacitor into either the accumulation or the inversion region and to avoid the depletion condition. However, if the metal voltage \( V_g \) is a high frequency signal then the only reliable region for different frequencies is the accumulation region. The only problem with the accumulation region though that in order to get the full \( C_{\text{ox}} \), the capacitor has to be biased with around 1V more negative than the flat band condition since the flat band capacitance is a little less than \( C_{\text{ox}} \) as shown in equation 2.9. A work around this problem could be found by analyzing equations 2.9 and 2.10. The more the doping \((N_d)\) of the well is, the closer the flat band capacitance and the depletion capacitance are to \( C_{\text{ox}} \). Therefore, by increasing \( N_d \), the dip
in the C-V curve shown in figure 2.12 between the accumulation and the inversion regions will decrease, which effectively reduces the restriction on how negative (or positive) the metal-to-well voltage has to be in order to get the total capacitance to \( C_{ox} \) in both the accumulation and inversion regions. This is especially very important to maximize the constant-capacitance voltage range. It is worth mentioning that even though biasing the MOS capacitor in the depletion region is generally not desirable, but has a very interesting application as a voltage controlled capacitor (a varactor). So it could be used to tune the frequency of an oscillator or the center frequency of a filter, and so forth. After reviewing the fundamentals of MOS capacitors, a discussion of the different options for implementing capacitors in CMOS technology will now be presented.
The simplest capacitor used in CMOS processes is the transistor itself where both the source and the drain are connected together to the bulk of the transistor and they represent the bottom plate, while the gate represents the top plate. In a P-substrate technology, an NMOS transistor can be only used as capacitor to ground since the substrate—which is common for the rest of the devices—has to be connected to ground, while a PMOS transistor could be used as a floating capacitor as shown in figure 2.13a. In a N-substrate technology, a PMOS transistor can be only used as capacitor to supply since the substrate—which is common for the rest of the devices—has to be connected to the supply, while an NMOS transistor could be used as a floating capacitor as shown in figure 2.13b. Note that the transistor connected as a capacitor is essentially a MOS structure like the one we discussed earlier. The major disadvantage for using the transistor as a capacitor is the voltage dependent behavior of the value of the capacitor as discussed earlier in the MOS structures. In a P-substrate technology, using an NMOS transistor as a capacitor to ground is only useful for low frequency applications given that the voltage on the gate of the transistor is higher than the threshold voltage of the transistor to ensure that it is biased in the inversion condition, hence giving the highest constant capacitance possible. Therefore, if the gate-to-bulk potential difference across the transistor is a small signal (less than the threshold voltage of the transistor), then using the NMOS transistor will not be desirable because of the voltage-dependent capacitance in the depletion region. Note that the NMOS transistor can not be biased in the accumulation region since the bulk of the transistor has to always be connected to the lowest voltage in the circuit. A PMOS transistor could also be used as a floating capacitor, if the gate-to-bulk voltage is positive, then the capacitor will be biased at the accumulation condition, hence the capacitance will be equal to $C_{ox}$. If the gate-to-bulk
voltage is negative (more than the threshold voltage), the capacitor will be biased at the inversion condition and the capacitance will be equal to $C_{ox}$ only at low frequencies. As mentioned before, the reliable region for all frequencies in MOS structures is the accumulation region, therefore if the application using the capacitor can guarantee that the potential difference across the capacitor will always be positive, then using a PMOS transistor with the gate as the plus side of the capacitor and the bulk as the minus side will generally have a better frequency performance than an NMOS transistor. Note that the previous discussion is also valid for an N-type substrate technology with the exception that the PMOS and NMOS transistors will exchange roles.

As explained before in the case of MOS structures, the only problem with the accumulation condition is that at the flat band (zero potential difference across the capacitor), the capacitance is a little less than $C_{ox}$ and it starts to increase with the potential

Figure 2.13 Using transistors as capacitors (a) In a P-substrate technology, (b) In an N-substrate technology.
difference until it gets to the full $C_{ox}$ value at around 1V, which is a voltage dependent behavior for the voltage range from 0 to 1V. In order to reduce the voltage dependent behavior for that range of potential difference, another device was introduced specially to implement capacitors in CMOS technology, which is the well capacitor. The well capacitor is the most commonly used capacitor in CMOS technology. It is simply a MOS structure, where the top plate is implemented using the polysilicon layer, and the bottom plate is implemented using an n$^+$ well in a P-substrate technology, or a p$^+$ well in an N-substrate technology. Figure 2.14 shows a well capacitor built on a P-substrate technology. You can also look at the well capacitor in this case as a PMOS transistor without a source and a drain, and with an n$^+$ bulk instead of a lightly doped n-type bulk. Therefore, the same things said about using the PMOS transistor as a capacitor will also apply for the well capacitor with the exception that the highly doped well in the well capacitor case will have the effect of shifting the C-V curve more to the left. This effectively improves the constant-capacitance voltage range. Figure 2.15 shows a comparison between the C-V curves at low frequencies of a PMOS transistor used as a capacitor and an N-type well capacitor in a P-substrate technology. Note that in the well capacitor case the threshold voltage is sufficiently high (due to the higher doping of the well) that for practical negative voltages the well capacitor will always be in the depletion condition.

In modern technologies the specific capacitance for well capacitors ranges from $0.5 \text{fF/}\mu\text{m}^2$ to $4.7 \text{fF/}\mu\text{m}^2$. Therefore in order to implement a $1\text{pF}$ capacitor, an area ranging from $2000\mu\text{m}^2$ to $213\mu\text{m}^2$ is needed. The two common problems with transistors connected as capacitors and well-capacitors as well are their noise performance and parasitic capacitances. Since both the bottom and top plates are relatively very close to the substrate,
**Figure 2.14** A Well capacitor on P-substrate technology.

**Figure 2.15** A comparison between the C-V characteristics of a PMOS transistor used as a capacitor and a well capacitor.
the resulting capacitor will be actually represented by the structure shown in figure 2.16a, where the parasitic capacitance between the polysilicon (the top plate) and the substrate is $C_{p1}$, and the parasitic capacitance between the well (the bottom plate) is $C_{p2}$ [2]. These two parasitic capacitors will affect the absolute accuracy of the intended value of the capacitor. Moreover, noise from the substrate can be injected to the main capacitor through the parasitic capacitors, which will degrade the noise performance of the main capacitor. One way to avoid the noise injection is the use of a guard ring around the capacitor to absorb the noise from the substrate as shown in figure 2.16b. In some processes, there are two available polysilicon layers that could also be used to implement capacitors. In that case, the bottom plate is implemented using the first polysilicon layer and the top plate is implemented using the second polysilicon layer. A poly-poly capacitor is shown in figure 2.17.

![Figure 2.16](image_url)

**Figure 2.16** (a) Parasitic capacitance in a well capacitor, (b) Using a guard ring for noise protection.
Generally a poly-poly capacitor will have a little less parasitic capacitance to the substrate since the top plate is farther away from the substrate. Also in order to reduce the noise injected from the substrate, a shielding well biased at a quite voltage could be used to isolate the capacitor from the substrate. Therefore, poly-poly capacitors have generally a better noise performance than transistor and well capacitors. Another major advantage of poly-poly capacitors over well capacitors is that they have relatively less voltage-dependent and frequency-dependent behavior. The fundamental reason behind that is since both the top and the bottom plates are implemented with the highly doped polysilicon layer, accumulation, depletion, and inversion have much less effect on the capacitance value. In fact the value of the capacitor could be found using the simple formula of the parallel plate capacitor shown in equation 2.8. However, poly-poly capacitors require a second polysilicon layer which will add to the cost of production due to the extra mask needed. Most digital CMOS processes doe not have a second polysilicon layer.

Figure 2.17 A Poly-poly capacitor.
Another type of capacitors is the metal capacitor (sometimes called vertical flux metal capacitor). Those capacitors use two metal layers to implement the top and the bottom plates of the capacitor. Since the metal layers are isolated from the substrate with thick field oxide, the parasitic capacitance from the top and bottom plates to the substrate are very small. Therefore this type of capacitors has good noise performance. Since the metal layers are highly conductive, metal capacitors don't suffer from voltage-dependent and frequency-dependent behavior as in MOS capacitors. Since metal layers are isolated using thick field oxide, the specific capacitance of metal capacitors is very low. Therefore, it is almost impractical—from an area perspective—to use them to implement any significant capacitance.

In order to improve the specific capacitance of vertical metal capacitors there has been a recent growth in the use of metal-insulator-metal capacitors or MIM capacitors. MIM capacitors use a thin insulator between two metal plates instead of the thick field oxide and therefore achieve higher specific capacitance [6]. MIM capacitors also have the same advantages of poly-poly capacitors, but also share the same disadvantage of needing extra masks and processing steps that increase the cost of production. They are also not available in standard digital CMOS technologies.

Despite the performance advantages of poly-poly and vertical metal capacitors, and aside from the generally lower specific capacitance, they have the major disadvantage of not scaling with technology scaling. Generally, the oxide thickness between different layers of metals and polysilicon stay almost the same even with the lateral scaling of the transistors [7]. This simply means that the specific capacitance of those capacitors stay the same and therefore occupy the same area, while the rest of the devices shrink in size, which increases the relative area of those capacitors with scaling. In order to solve this problem, another
type of capacitors was introduced, which is the lateral flux capacitor (sometimes is called a fringing capacitor). In this type of capacitors both plates are implemented using two metal strips of the same metal layer instead of two different layers. The capacitance is achieved through the fringing field between the two metal strips, and therefore in that case the separation between the two metal strips represent \( t \) in equation 2.8 and the area of the capacitor is the side area of the metal strip. Since the separation between metal lines on the same metal level scale down with the process, the specific capacitance of lateral flux capacitors increases, which means that in order to get the same total capacitance after scaling, the side area of the metal strips have to also scale down with process scaling. Figure 2.18 show a lateral flux capacitor and how the capacitor scale with technology. Note that the different shading is just to differentiate between the two plates of the capacitor.

In addition to scaling with technology, lateral flux capacitors can also be incorporated with vertical capacitors to achieve a significantly higher specific capacitance [7]. As shown in figure 2.19, dividing a conventional vertical flux capacitor to multiple cross connected sections has the effect of increasing the total capacitance due to the lateral flux capacitors added to the structure, while keeping the total area almost the same. Effectively, this means higher specific capacitance. Note that the top and bottom plates of the equivalent capacitor are distributed between two metal layers. Different shadings are used in figure 2.19 to emphasize this fact. The same idea could also be extended not only to two metal layers, but to even multiple metal layers to increase the specific capacitance even further.

Since lateral flux capacitors depend on the side area of the metal strips, then increasing the periphery of the structure for the same area will increase the effective specific capacitance even further. There are many ways to increase the periphery, one way is shown
Figure 2.18 Effect of scaling on lateral flux capacitors
(a) Before scaling: lower specific capacitance and higher side area.
(b) After scaling: Higher specific capacitance and lower side area.

Figure 2.19 Vertical and lateral flux capacitors
(a) Conventional vertical flux capacitor.
(b) Lateral flux capacitor incorporated within a vertical flux capacitor.
in figure 2.20 where multiple metal fingers are interleaved together in an interdigitized approach [8,9]. This structure is done in multiple metal layers and then all the resulting capacitors are connected in parallel. This technique could result in a specific capacitance that ranges from $0.25 \text{fF/} \mu\text{m}^2$ to $0.4 \text{fF/} \mu\text{m}^2$. However, using multiple metal layers has the drawback of not permitting any routing above the capacitors.

### 2.2.1. Accuracy of Passive Capacitors

As equation 2.8 suggests, the error in the value of $C$ is a function of the error in $t$, $L$, and $W$. The error in $t$ or the oxide thickness is mainly a function of the accuracy of the fabrication process and there is nothing in the design or layout process of capacitors that could be done to minimize its impact. The error in $W$ and $L$ on the other hand is due to the etching process as discussed earlier in the resistor's case. Taking the error in $W$ and $L$
into account, and assuming that $\Delta W = \Delta L = 2x$, the actual area of the capacitor is going to be [2]:

$$ A_{\text{actual}} = (W - 2x)(L - 2x) \approx WL - 2(W + L)x = A - Px = A \left(1 - \frac{P}{A}x\right) $$

(2.12)

where $A$ is the drawn area of the capacitor, and $P$ is the perimeter. The relative error in the area of the capacitor becomes:

$$ \frac{A_{\text{actual}} - A}{A} = -\frac{P}{A}x = -2 \left(\frac{x}{W} + \frac{x}{L}\right) $$

(2.13)

As shown in equation 2.13, in order to reduce the effect of the error $x$, $W$ and $L$ has to be larger. Unlike the resistor case, it is not possible to increase $W$ and $L$ without increasing the value of the capacitor. Therefore, if a lower value of capacitance is needed, then the dimensions of the capacitor have to be lower, hence the error will be higher and has to be tolerated. The higher the value of the capacitor is, the higher the accuracy of the capacitor. Adding dummies around the capacitors though helps to reduce the errors in $W$ and $L$ due to the etching process as discussed before in resistors case.

Well capacitors vary with process in the range of $\pm 10\%$, which is generally less than the variation range in the cases of resistors. The reason behind that is that the error in the resistor value is also a direct function of the error in the doping in addition to the error in $W$ and $L$, while doping in case of well capacitors has a little effect on the capacitance since the well is already highly doped. This minimizes the effect of doping errors in the case of well capacitors, while those errors are significant in resistors.

Metal capacitors — especially interdigitized lateral flux capacitors — on the other hand vary widely with process, roughly around $\pm 25\%$. This is essentially due to the fact that flux capacitors are composed of many fingers interleaved together, each finger has its own $W$, 

47
$L$, and $t$ errors. In addition to that and in order to increase the specific capacitance, multiple flux capacitors on multiple metal layers are usually connected together in parallel, and each metal layer introduces its own error as well. All those errors can add up to a significant total error in the value of the capacitor.

Generally capacitors do not vary significantly with temperature since the errors in $W$, $L$, and $t$ are not functions of temperature, and carrier concentration in both the top and bottom conductive plates of the capacitors have a little effect on the capacitance value. Therefore flux capacitors temperature coefficient is as low as -24 to -12 ppm, while well capacitors have a temperature coefficient varying from 120 ppm to 180 ppm. Well capacitors have a little higher temperature coefficient since the carrier concentration in the well changes with temperature, however the impact on the capacitance is still considered very low as compared to resistors.

2.2.2. Matching Properties of Passive Capacitors

As discussed in the matching properties of integrated resistors, capacitors could also match very well given a careful layout. This fact has driven the development of many circuit techniques that depend solely on matching between capacitors rather than their absolute values, switch capacitor circuits is a good example. The same guide lines for design and layout of integrated resistors to achieve good matching, apply as well in the case of integrated capacitors. Therefore using dummies to minimize the effect of etching and to keep the etching profile uniform across the whole structure is very important to achieve
good matching. As mentioned in the resistors case, using rounded or 45 degrees edges for the corners of the capacitor helps in minimizing the etching errors.

All the different levels of layout complexity in order to achieve good matching between two resistors also apply in the case of capacitors. Figure 2.21 shows two matched capacitors laid out in common centroid fashion. Note how the contacts between the polysilicon top plates and the metal layer are done on top of the thin oxide, which is permissible in some processes. In some other processes, those contacts have to be done on top of the field oxide as shown in figures 2.14 and 2.16b. Very often, a specific ratio between two capacitors is needed rather than two matched capacitors. Since the ratio between two capacitors is essentially equal to the ratio between the respective areas, then equation 2.12 suggests that if the two capacitors had the same perimeter to area ratio, the effect of the

![Figure 2.21 Layout of two matched capacitor.](image)

49
etching error $x$ on matching could be eliminated. Hence, it is very crucial to keep $\frac{P}{A}$ the same for the two capacitors. One way to do that is to use a unit capacitor to build both capacitors as multiple parallel unit capacitors. Figure 2.22a shows two matched capacitors with a ratio 1:2 built of a unit capacitor. This method is useful if the ratio needed is a ratio between integers. If the ratio between the two capacitors is not a ratio of integers, then a unit capacitor is used to implement the integer part of the ratio, and a non unit capacitor with the same $\frac{P}{A}$ is used to implement the fraction part of the ratio. Figure 2.22b shows the layout of two matched capacitors with a ratio 4:3.2.
Figure 2.22 (a) Layout of an integer multiple of a unit capacitor. (b) Layout of a non integer multiple of a unit capacitor.
LIST OF REFERENCES


CHAPTER 3

TUNING ARCHITECTURES

Resistors and capacitors are used in many applications in the area of analog signal processing. They are fundamental elements in oscillators, filters, delay elements, termination networks, amplifiers, and so on. Usually, if simple designs are used to implement any of the above applications, the main characteristics (oscillation frequency of an oscillator, cut-off frequency of a filter, .....etc) of any of them will highly depend on the absolute values of the resistors and capacitors used. The accuracy of the absolute values of the resistors and capacitors used in any design is always a problem that floats to the surface, especially if the targeted application requires high accuracy for adequate performance. Off-chip discrete resistors and capacitors can achieve high accuracy (1% for resistors and 5% for capacitors), but they have to be used externally. This implies a higher cost and also raises other issues, like the number of pins that can be used to connect those external elements to the internal circuit. For example, if too many filters are implemented and each one is using a different resistor, it becomes impractical to assign a pin for each one. Also the parasitic elements represented by bond-wire capacitance and inductance that are introduced in the path to the external resistor or capacitor could affect the performance of the circuit causing the loss of
accuracy intended originally by using external elements. The problems introduced by those parasitic elements become severe at high speed systems.

On-chip resistors and capacitors are a cheaper alternative to external ones and also easier to integrate into the system, but as discussed in the previous chapter, they suffer from the relatively poor control—especially in digital CMOS Processes—over their absolute values. Most modern processes can guarantee the accuracy of on-chip resistors and capacitors to within ±25%. Many techniques were developed to solve the accuracy problem of the absolute values of on-chip resistors and capacitors mentioned above. One solution is based on developing circuit techniques for implementing the required system (filter, oscillator, ...etc.) such that the targeted parameter (oscillation frequency of an oscillator, cut-off frequency of a filter, .....etc) of the system would depend on a ratio between similar on-chip elements (resistors or capacitors) rather than their individual absolute values (Switched-Capacitor techniques is an example) [1]. As discussed in the previous chapter, the ratios between similar on-chip elements have a much better accuracy than their individual absolute values. Unfortunately though, it is not always possible to implement the required system with the parameter of interest as a function of a ratio between similar elements. In these cases, depending on the absolute values of the resistors and capacitors becomes inevitable, and a tuning technique has to be used.

Tuning techniques can be categorized into two main categories. The first category is on-chip automatic tuning (self tuning), and the second category is post fabrication tuning. In this chapter, an overview of the different tuning techniques used by both categories along with the pros and cons will be presented. A discussion of the different elements used in the tuning process including references and electronically tunable elements will be presented.
3.1. Automatic Tuning

Automatic tuning is a term that refers to a system that has the ability to tune a set of its parameters to an accurate specific set of values automatically by itself without the need of any external intervention, thus it is also called a self tuned system. While the set of parameters in the system might be affected by many different factors, the rest of discussion will assume that those parameters depend on the absolute values of on-chip resistors, capacitors, or both. Usually there are two methodologies for performing the automatic tuning of a certain parameter in the system, either tuning the values of the on-chip resistors and capacitors used in the system, which automatically means tuning the parameter of interest that depends on those resistors and capacitors to the same accuracy, or tuning the parameter of interest itself to compensate for the errors in the on-chip resistors and capacitors used. Obviously if the parameter of interest is the resistance or the capacitance, then the second methodology is not applicable. The second methodology implies that the parameter of interest has to be not only a function of the resistors and capacitors, but it also has to be a function of another control parameter that can be changed to compensate for the errors in the resistors and capacitors, in other words, the parameter of interest has to be electronically tunable. For example, the location of a pole in a filter's transfer function could be made as a function of the gain of an amplifier in addition to its dependency on the absolute value of a resistor and a capacitor. By designing the amplifier such that its gain is controlled by a voltage signal, the location of the pole can be adjusted.
The automatic tuning process has three fundamental steps. The first step is measuring the parameter of interest. The second step is the error signal generation, which results from comparing the measured value of the parameter of interest to an accurate reference in order to quantify how off the value of the parameter is from its desired value. The third step is generating a tuning signal by using the error signal in conjunction with a negative feedback loop. The tuning signal is then used to tweak the parameter of interest to match the reference. The automatic tuning process could be done using three different architectures depending on the application. The first architecture is the direct continuous architecture in which the tuning process is continuously happening during the normal operation of the tuned circuit, and the tuning loop does not need to use the input of the tunable circuit. Figure 3.1a shows a conceptual block diagram of the direct continuous tuning architecture. Many times though, the tuning loop requires the use of the input of the tunable circuit during the tuning process, therefore the input of the tunable circuit has to be disconnected from the input signal during the tuning period. This leads to the second architecture, which is the direct discontinuous architecture. In this architecture, the tunable circuit has to be taken off line during the tuning process until a control signal value has been reached, then this value is held and the tunable circuit is returned back on line. Figure 3.1b shows a conceptual block diagram of the direct discontinuous architecture. Note that if the application does not tolerate taking the tunable circuit off line, then during the tuning period a copy of the tunable circuit with a previously held control signal can be used as shown in figure 3.1b. Direct architectures (continuous and discontinuous) have the advantage that the same circuit that is used by the input signal is the same circuit that is being tuned. Therefore, the highest level of tuning accuracy could be achieved using these two architectures.
In some cases when the input of the tunable circuit needs to be used by the tuning loop but can't be disconnected from the input signal or taken off line as in direct discontinuous architectures, a master copy of the tunable circuit is usually used to perform the tuning, and the resulted control signal is then used to tune a slave copy of the tunable circuit, where the slave copy is the one that is always connected to the input. Since the tuning process is done on a copy of the tunable circuit (the master), this architecture is referred to as indirect tuning architecture. Indirect tuning architectures could also be continuous, where the tuning process is continuously running, or they could be discontinuous, where the tuning is performed and the control signal is held at the resulted value. Figure 3.2 shows a conceptual block diagram of the indirect tuning architecture. Note that the input to the master copy of the tunable circuit could be connected to the input of

Figure 3.1 Direct tuning architecture (a) continuous tuning, (b) discontinuous tuning.
the slave copy, or it could be connected to an excitation signal. Usually the excitation signal is used in the cases where in order to measure the parameter of interest, a special signal has to be applied to the input (the excitation signal) of the circuit. The excitation signal could be simply an accurate periodic clock signal generated from a crystal oscillator, which is very common in the tuning of time references. The main disadvantage of indirect architectures though is that the tuning accuracy will be limited by the mismatches between the master and slave copies of the tunable circuit.

The previous categorization of tuning techniques was based on when the tuning is done (continuous or discontinuous), and on what the tuning is performed (direct and indirect tuning), but the tuning process could be also categorized based on the comparison nature. The nature of the reference usually dictates the nature of the comparison circuit. As will be discussed later in details, references could be either a voltage or a current level (magnitude), or a frequency (phase) of an oscillation. Therefore, the tuning process could be
either a Magnitude Locked Loop (MLL) based tuning, or a Phase Locked Loop (PLL) based tuning. In the Magnitude Locked Loop (MLL) based tuning, the reference is a constant voltage or current level, and the information about the measured parameter of interest is in another voltage or current level. By comparing the magnitude of the two levels, i.e. the measured parameter of interest and the reference, an error signal can be generated and then manipulated to generate the control signal (the tuning signal). Therefore, the comparison circuit shown in figures 3.1 and 3.2 will have to be a magnitude comparator. MLL based comparison is generally susceptible to DC offsets in the comparison circuit, which is generally done through the aid of a differential amplifier that subtracts the reference level from the measured parameter of interest level and gives an analog error signal, or simply through the aid of a comparator, which gives a 1 or 0 result of the comparison. Those DC offsets cause a systematic error in the resulted tuned parameter. Note that the MLL based tuning relies on the ability to measure the parameter of interest and convert this measurement to a current or a voltage level, which –even though possible- in some cases is not a straight forward process if the nature of the parameter of interest is time related (an oscillation frequency for example). The Phase Locked Loop (PLL) based tuning on the other hand, relies on a time reference instead of a level reference, i.e. a reference phase or frequency. In that case the information about the measured parameter of interest is in the phase or frequency of a signal instead of its level. The phase of this signal is then compared to the phase of the reference using a phase detector to generate the error signal. In this case the comparison circuit shown in figures 3.1 and 3.2 will have to be a phase detector.
The tuning process could also be categorized based on the nature of the error and tuning signal generation (the feedback circuit) to an analog based tuning and digital based tuning. In the analog based tuning, the comparison process that generates the error signal and the further processing on the error signal that generate the tuning signal are purely done in the analog domain. Therefore, the tuning signal can take continuous values in the tuning range. In the digital based tuning, the measured parameter of interest and the reference signal are both converted to digital signals (either jointly or independently) using an Analog to Digital Converter (ADC), and then all the comparisons and the further processing is done in the digital domain. The resulted digital tuning signal can then be converted to an analog signal using a Digital to Analog Converter (DAC), where it's then used to control the parameter of interest. In that case, the parameter of interest can only have finite discrete values in the tuning range, thus, the accuracy of the tuning will be a function of the resolution of the ADC and the DAC used. Figure 3.3a shows a conceptual block diagram of a digital based tuning architecture. In some cases though, if the parameter of interest can be controlled directly with a digital signal, i.e. the parameter of interest is digitally programmable, then the DAC can be eliminated altogether from the loop. An example of that is shown in figure 3.3b, where the parameter of interest is a function of a bank of resistors and capacitors in which individual resistors and capacitors can be turned on or off using a direct digital signal.

Generally as shown by figures 3.1, 3.2 and 3.3, there is an absolute necessity of a reference, an electronically tunable element, and a mean to measure the parameter of interest in order for any tuning architecture to operate. In the next few sections, voltage and time references will be discussed followed by an overview of some available electronically tunable elements, which include tunable resistors, tunable capacitors, and tunable transconductors.
Unfortunately there are not many absolute accurate quantities that can be used in IC design. This fact is the main drive for developing circuits that solely depend on a ratio between two similar elements rather than the absolute characteristics of each individual element. While this technique is very successful, still some times there is no way around depending on an absolute quantity. Generally, any quantity that has dimensions or units is always a suspect of being inaccurate unless it is equal to a universal constant that does not depend on the fabrication process, supply voltage, or temperature. Accurate references usually used in IC designs are voltage references, current references, and time references. Voltage references can be generated on-chip without the need of any external elements, while current references can be generated using an on-chip voltage reference in conjunction with an external resistor. Time references are usually generated using an external crystal oscillator in conjunction with an on-chip PLL (Phase Locked Loop).

Figure 3.3 Digital based tuning: (a) with a DAC, and (b) with direct digital control.

3.1.1. References
3.1.1.1. Voltage References

A good voltage reference has to be independent from the supply voltage, fabrication process, and temperature. There are numerous ways available in literature to generate a reference voltage [2], this section will concentrate on the circuits that generates the most accurate voltage reference, which are band-gap referenced circuits. The voltage drop across a forward biased diode can be represented by the following equation [3]:

\[ V_d = nV_T \ln \frac{I}{I_s} \]  \hspace{1cm} (3.1)

where \( n \) is the emission constant (varies from 1 to 2), \( V_T \) is the thermal voltage, and \( I_s \) is the leakage current (typically equals to \( 10^{-15} \) A). The thermal voltage is represented by:

\[ V_T = \frac{KT}{q} \]  \hspace{1cm} (3.2)

where \( K \) is Boltzmann’s constant (1.38\times10^{-23} \text{ J/K}) , \( T \) is the temperature in Kelvin, and \( q \) is the electrons charge (1.6\times10^{-19} \text{ C}). The thermal voltage \( V_T \) is equal to 26mV at room temperature \( (T = 300^\circ K) \) and varies by 0.085mV/°C with temperature. Even though it might look from equation 3.1 that \( V_d \) should have a positive temperature coefficient due to its proportionality with the thermal voltage, but it actually has a negative temperature coefficient. The reason behind that is that the leakage current \( I_s \) is a strong function of the band-gap energy of silicon \( E_g \) (eV) which can be represented by [3]:

\[ I_s \propto e^{-E_g/KT} \]  \hspace{1cm} (3.3)

\[ E_g = 1.16 - \left( 702 \times 10^{-6} \right) \left( \frac{T^2}{T+1108} \right) \]  \hspace{1cm} (3.4)
Analyzing equations 3.3 and 3.4, since the band-gap energy $E_g$ of silicon has a negative temperature coefficient, the leakage current $I_s$ will have a positive temperature coefficient. The effect of the positive temperature coefficient of $I_s$ on $V_d$ is larger than the effect of the positive temperature coefficient of the thermal voltage $V_T$, therefore $V_d$ will have a negative temperature coefficient. The variation in $V_d$ with temperature is around $-2\text{mV}^/\degree\text{C}$ [2]. Because of the negative temperature coefficient of the forward bias voltage of the diode, it can not be used as a temperature stable voltage reference. Band gap circuits though, compensate for the negative temperature coefficient of the forward bias voltage of the diode by adding to it a voltage that has a positive temperature coefficient or proportional to absolute temperature (PTAT). Equation 3.1 suggests that if the leakage current $I_s$ was eliminated from the equation, then the resulted voltage will be proportional to the thermal voltage $V_T$. To achieve that, if the same current is flowing into two diodes that have different areas, then the difference between the forward bias voltages of the two diodes will be proportional to the thermal voltage $V_T$.

Band-gap circuits use the parasitic pnp BJT transistor formed by the P$^+$ implant (the emitter), the n-well (the base), and the p-type substrate (the collector) available in any n-well process in conjunction with CMOS transistors to generate a temperature stable reference voltage. Figure 3.4 shows a simple band-gap circuit [2]. In this circuit, the parasitic transistors D$_2$ and D$_3$ have $N$ times the area of D$_1$, while the resistor in series with D$_3$ is $L$ times the resistor in series with D$_2$. The CMOS current mirrors keep the currents in D$_1$ and D$_2$ the same, therefore the current $I$ in figure 3.4 is given by:
Note how the generated current is proportional to absolute temperature (PTAT). The reference voltage can then be written as:

\[ V_{\text{ref}} = ILR + V_{d3} \]  

(3.6)

where \( V_{d3} \) is the voltage drop across the BJT \( D_3 \). Replacing \( I \) in equation 3.6 with equation 3.5, the reference voltage becomes:

\[ V_{\text{ref}} = (ln \, N)I_T + V_{d3} \]  

(3.7)

As mentioned before, the forward bias voltage of the diode will have a negative temperature coefficient, while the voltage resulting across the resistors is a PTAT voltage, therefore both parts of equation 3.7 compensate for each other. The variation in the reference voltage \( V_{\text{ref}} \) with temperature is defined as:

\[ \text{Figure 3.4 Schematic diagram of a simple band-gap circuit.} \]

65
\[
\frac{\partial V_{\text{ref}}}{\partial T} = (\ln \ln N) \frac{\partial V_T}{\partial T} + \frac{\partial V_d}{\partial T}
\] 

(3.8)

In order to have a temperature stable reference voltage, equation 3.4 has to be equal to zero.

The variation in the diode voltage with respect to temperature \( \frac{\partial V_d}{\partial T} \) is equal to \(-2\text{mV/}^\circ\text{C}\), while the variation in the thermal voltage \( \frac{\partial V_T}{\partial T} \) is \(0.085\text{mV/}^\circ\text{C}\) at the room temperature \((T = 300^\circ\text{K})\) [2]. Therefore in order for equation 3.4 to be equal to zero, the following equality has to hold:

\[
\ln \ln N = \frac{2}{0.085} = 23.5
\]

(3.9)

By choosing the appropriate value for \( L \), an accurate temperature stable reference voltage could be created. Since \( L \) is a ratio between two similar on-chip resistors, it could be designed to be very accurate. Note also that the CMOS transistors are used as current mirrors and their absolute characteristics are of no concern. Using op-amps, multiple copies of the reference voltage could be created. Band-gap references can achieve an accuracy of \(\pm3\%\) without any trimming of the on-chip resistors used. The accuracy could be increased to \(\pm1\%\) by trimming the on-chip resistors used.

### 3.1.1.2. Current References

The reason for having an on-chip accurate voltage reference without any external components is that the thermal voltage \( V_T \) and the band-gap voltage are universal constants that have the units of voltage. Unfortunately, there is no universal constant that has the units
of current. There are many available techniques to generate current references [2], the most accurate one is using an accurate voltage reference across an accurate resistance in order to generate an accurate current reference. While band-gap circuits can generate an accurate voltage on-chip, on-chip resistors can vary up to ±25\%, therefore an external resistor (can have accuracy up to ±1\%) has to be used in conjunction with a band-gap circuit and an op-amp to generate an accurate current reference. Figure 3.5 shows a simple reference current generator. The op-amp copies the band-gap voltage reference to an external resistor, and the current generated is then mirrored to be used in multiple locations in the IC. The obvious disadvantage of such a technique is the need of an external resistor.

Figure 3.5 A reference currents generator.
3.1.1.3. Time References

Time references are not as easy to generate on-chip as in voltage references case. The reason behind that again is that there is no an on-chip universal constant that has the units of time. Even though there exist some techniques to generate an on-chip accurate time constants based on RC products [4], but the most popular techniques rely heavily on an external crystal oscillator that is used in conjunction with a PLL to generate multiple frequency references on-chip. A piezoelectric crystal, such as quartz, exhibits electromechanical-resonance characteristics that are very stable with time an temperature and highly selective (having very high quality factors). The extremely stable resonance characteristics and the high quality factors of quartz crystals result in oscillators with very accurate and stable frequencies. Crystals are available with resonance frequencies in the range of few KHz to hundreds of MHz. Temperature coefficients of the frequency of oscillation ranges from 1 to 2 ppm per °C. However, crystal oscillators, being mechanical resonators, are fixed frequency circuits [5]. Therefore, there is always a need of an on-chip PLL to generate multiple frequencies, or multiple phases of the same frequencies. Those different frequencies and phases serve as a very accurate time references.
3.1.2. Electronically Tunable Elements

As discussed before, in order to tune for a certain parameter, this parameter has to be electronically tunable using either a voltage, or current signal. Therefore, the parameter of interest has to be a function of a variable resistance, a variable capacitance, a variable gain of an amplifier, a variable transconductance, and so on. Those variable elements have to be either voltage or current controlled. In the next three sub-sections, examples of electronically tuned resistors, capacitors, and transconductors will be presented. In the next two chapters, a voltage controlled transconductor, and an electronically tunable offset comparator proposed by the author will be discussed.

3.1.2.1. Tunable Resistance

One of the most important electronically tunable elements is the resistivity of the transistor itself. A transistor current-voltage relationship can generally be represented by [6]:

\[
I_D = (W/L) \int_{V_c=V_s}^{V_c=V_D} f(V_G, V_C) dV_C
\]

(3.10)

with

\[
f(V_G, V_C) = -\mu Q_C + \mu (KT/q)(dQ_C/dV_C)
\]

(3.11)

where \(W\) and \(L\) are the dimensions of the transistors, \(V_s, V_G,\) and \(V_D\) are the source, gate, and drain voltages respectively, \(V_C\) is the channel voltage, \(\mu\) is the current carriers mobility,
$\frac{KT}{q}$ is the thermal voltage, and $Q_c$ is the channel charge. By first order approximation, equations 3.10 and 3.11 are valid for all operation regions of the transistor including strong inversion, weak inversion, saturation, and non-saturation regions [7]. The second part of equation 3.11 represents the weak inversion channel charge, while the first part represents the strong inversion channel charge. The function $f$ can be a very complicated function in $V_G$ and $V_C$ if phenomena like mobility reduction and body effect are included. In order to simplify the computation without losing much accuracy, we will assume that the mobility is constant ($\mu = \mu_o$) and that the transistor is operating in strong inversion, therefore the channel charge could be represented by [8]:

$$Q_c = -[V_G - V_{TB}(V_C)]C_{ox} \quad (3.12)$$

where $V_{TB} = \text{threshold voltage as a function of the channel voltage } V_C$, and $C_{ox}$ is the gate oxide capacitance. Note that all voltages are referenced to the substrate potential. The variation in the threshold voltage with the channel potential, i.e. the body effect, can be approximated by the linear function [7]:

$$V_{TB}(V_C) = V_{TO} + \alpha V_C \quad (3.13)$$

where $V_{TO}$ is the threshold voltage at $V_C = 0$, and $\alpha$ is a process dependent constant. In modern CMOS processes, $\alpha$ varies from 1.05 to 1.35. If the body effect is to be ignored, then $\alpha$ has to be assumed unity. Using equations 3.12 and 3.13, equation 3.10 will yield:

$$I_D = \mu_o C_{ox} \left( \frac{W}{L} \right) \int_{V_C = V_G}^{V_C = V_D} (V_G - V_{TO} - \alpha V_C) dV_C \quad (3.14)$$
The transistor is said to be in the triode region of operation as long as \( V_D \) is less than the channel saturation voltage \( V_{C_{sat}} = (V_G - V_{TO})/\alpha \). In that case, by integrating equation 3.14, the drain current is represented by:

\[
I_D = \left( K / 2\alpha \right) \left[ (V_G - V_{TO} - \alpha V_S)^2 - (V_G - V_{TO} - \alpha V_D)^2 \right] 
\]

(3.15)

where \( K = \mu_a C_{ox} (W / L) \). When the drain voltage is higher than \( V_{C_{sat}} \), then the channel pinches at a lower voltage than \( V_D \) and the transistor is said to be in the saturation region. In that case the drain current is represented by:

\[
I_D = \mu_a C_{ox} (W / L) \int_{V_c=V_{C_{sat}}}^{V_c=V_S} (V_G - V_{TO} - \alpha V_c) dV_c 
\]

(3.16)

\[
I_D = \left( K / 2\alpha \right) (V_G - V_{TO} - \alpha V_S)^2 
\]

(3.17)

Note that the current equation in the saturation mode of operation is simply the same as the triode mode with the second part of equation 3.15 omitted.

As equations 3.15 and 3.17 suggest, the current-voltage relationship of a transistor could be tuned using the gate voltage \( V_G \), therefore it can be used as a tunable resistance or a tunable voltage to current converter, i.e. transconductor. The only problem with that is the non-linear behavior of the current-voltage equations presented in 3.15 and 3.17. In order to use the transistor as a tunable resistor (or a voltage to current converter), some non-linearity cancellation techniques has to be employed. There are many available non-linearity cancellation techniques in the literature based on both the triode and saturation modes of the transistor operation [1,7]. Generally, non-linearity cancellation techniques relies on manipulating the terminal voltages and currents of single or multiple transistors to cancel
non-linear factors in equations 3.15 and 3.17. The simplest non-linearity cancellation technique is shown in figure 3.6a, in which a single transistor operating in the triode region is used with manipulated source and drain voltages to achieve a tunable linear resistance. If the drain and source voltages are assumed to be:
where \( V_B \) is a constant voltage, then equation 3.15 reduces to:

\[
R = \frac{2V_y}{I_D} = \frac{1}{K(V_G - V_{T0} - \alpha V_B)}
\]  

(3.19)

The resulted resistance is a strong function of temperature and process due to the appearance of \( K \), \( V_{T0} \), and \( \alpha \) in equation 3.19, however, while those factors affect the absolute value of the resistance, but they have not effect on the linearity of the resistance and the inaccuracies in the absolute value can be tuned out by changing \( V_G \).

Using a single transistor and manipulating terminal voltages to cancel the non-linearity in the current-voltage relationship does not give enough degrees of freedom on the terminal voltages. In fact, as shown in the previous example, the drain and source voltages have to have a constant common mode, which is not always the case. Using multiple transistors to cancel the non-linearity gives more degrees of freedom to the terminal voltages. Figure 3.6b shows two matched transistors sharing the same gate voltage \( V_G \), while the drain and source voltages are separated. Using equation 3.15, the difference between the currents flowing into \( M_1 \) and \( M_2 \) will be:

\[
I_1 - I_2 = \frac{K}{2\alpha} \left[ (V_G - V_{T0} - \alpha V_{S1})^2 - (V_G - V_{T0} - \alpha V_{S2})^2 \right] 
\]  

which can be further reduced to:

\[
I_1 - I_2 = K \left[ (V_{S2} - V_{S1}) \left( V_G - V_{T0} - \frac{\alpha}{2} (V_{S1} + V_{S2}) \right) - (V_{D2} - V_{D1}) \left( V_G - V_{T0} - \frac{\alpha}{2} (V_{D1} + V_{D2}) \right) \right] 
\]  

(3.21)
Note that equations 3.20 and 3.21 assume that the two transistors are operating in the triode region, but they can also be used in saturation mode by omitting the bottom parts of the equations, which evaluates to zero in saturation. As shown by equation 3.21, in order to achieve linearity between the differential current and a terminal voltage, there are two options: either eliminating the bottom part of equation 3.21 by forcing the drain voltages of both transistors to be equal, i.e. $V_{D_1} = V_{D_2} = V_D$, or eliminating the top part of equation 3.21 by forcing the source voltages of both transistors to be equal, i.e. $V_{S_1} = V_{S_2} = V_S$. In the first case, equation 3.21 will be reduced to:

$$I_1 - I_2 = K \left[ (V_{S_2} - V_{S_1}) \left( V_G - V_{T_0} - \frac{\alpha}{2} (V_{S_1} + V_{S_2}) \right) \right]$$

(3.22)

Furthermore, if the source voltages were manipulated such that $V_{S_1} = V_B - V_X$ and $V_{S_2} = V_B + V_X$, then equation 3.22 becomes:

$$R = \frac{2V_X}{I_1 - I_2} = \frac{1}{K \left( V_G - V_{T_0} - \alpha V_B \right)}$$

(3.23)

which is again similar to equation 3.19. However, this method has the advantage of being valid for both the triode and saturation modes of the two transistors since the bottom part of equation 3.21 that was eliminated by forcing the drain voltages of both transistors to be equal is equal to zero in the saturation mode anyway, which will make equations 3.22 and 3.23 valid in that mode as well. The second option (forcing the source voltages to be equal) will reduce equation 3.21 to:

$$I_1 - I_2 = K \left[ (V_{D_1} - V_{D_2}) \left( V_G - V_{T_0} - \frac{\alpha}{2} (V_{D_1} + V_{D_2}) \right) \right]$$

(3.24)
furthermore, if the drain voltages were manipulated such that $V_{D1} = V_b + V_x$ and $V_{D2} = V_b - V_x$, then equation 3.22 becomes:

$$R = \frac{2V_x}{I_1 - I_2} = \frac{1}{K(V_G - V_{TO} - \alpha V_b)}$$  \hspace{1cm} (3.25)

which is again similar to equation 3.19, but unlike equation 3.23, it is only valid when the two transistors are operating in the triode region.

Equations 3.22 and 3.24 were based on sharing the gate voltage between the two transistors in figure 3.6b. Another way to achieve non-linearity cancellation is shown in figure 3.6c and 3.6d where the gate voltages are now separated, while sharing either the same source or the same drain terminals between the two transistors. If the two transistors are sharing the same source as shown in figure 3.6c, and assuming that both transistors are in the triode mode, then the differential current becomes:

$$I_1 - I_2 = \frac{K}{2\alpha} \left[ \frac{(V_{G1} - V_{G2})(V_{G1} + V_{G2} - 2V_{TO} - 2\alpha V_S)}{(V_{G1} - V_{G2} - \alpha(V_{D1} - V_{D2}))(V_{G1} + V_{G2} - 2V_{TO} - \alpha(V_{D1} + V_{D2}))} \right]$$  \hspace{1cm} (3.26)

If the drains are forced to be at the same voltage, i.e. $V_{D1} = V_{D2} = V_D$, then equation 3.26 reduces to:

$$R = \frac{V_D - V_S}{I_1 - I_2} = \frac{1}{K(V_{G1} - V_{G2})}$$  \hspace{1cm} (3.27)

The advantage of equation 3.27 over equation 3.19 is that the value of the resistance is not dependent on the threshold voltage of the transistor, which improves the accuracy of the resulting resistance.
When the two transistor structure is sharing the same drain as shown in figure 3.6d instead of the same source like figure 3.6c, and assuming triode mode operation, the differential current become:

\[
I_1 - I_2 = \frac{K}{2\alpha} \left[ \left( V_{G1} - V_{G2} - \alpha(V_{S1} - V_{S2}) \right) \left( V_{G1} + V_{G2} - 2V_{TO} - \alpha(V_{S1} + V_{S2}) \right) - \right]
\]

(3.28)

If the sources are forced to be at the same voltage, i.e. \( V_{S1} = V_{S2} = V_S \), then equation 3.28 reduces to:

\[
R = \frac{V_D - V_S}{I_1 - I_2} = \frac{1}{K(V_{G1} - V_{G2})}
\]

(3.29)

which is the same resistance resulted in equation 3.27.

The structure in figure 3.6c, which results in a resistance value shown by equation 3.27 is only valid if both transistors are operating in the triode mode. In order to expand the operation to the saturation mode as well, the circuit is figure 3.7a could be used. In that case the differential current could be represented by:

\[
I_{O1} - I_{O2} = (I_1 - I_4) - (I_2 - I_3)
\]

(3.30)

with

\[
I_1 - I_4 = K \left[ (V_{S34} - V_{S12}) \left( V_{G1} - V_{TO} - \frac{\alpha}{2} (V_{S12} + V_{S34}) \right) \right]
\]

(3.31)

\[
I_2 - I_3 = K \left[ (V_{S34} - V_{S12}) \left( V_{G2} - V_{TO} - \frac{\alpha}{2} (V_{S12} + V_{S34}) \right) \right]
\]

(3.32)

Substituting 3.31 and 3.32 in 3.30, the resulting resistance becomes:

\[
R = \frac{V_{S34} - V_{S12}}{I_{O1} - I_{O2}} = \frac{1}{K(V_{G1} - V_{G2})}
\]

(3.33)
Since equations 3.31 and 3.32 are valid in both the triode and saturation modes (given that all the transistors have the same drain voltage), the resulted resistance value is valid if the transistor pairs $M_1,M_4$ and $M_2,M_3$ are both in the saturation mode, or in the triode mode, or one pair is in the saturation mode and the other is in the triode mode. The circuit presented in figure 3.7a is also called four-quadrant multiplier, and it has the best non-linearity cancellation. In fact it was shown that this circuit is practically insensitive to distributed effects in the transistor channel, and that it has inherent compensation to such effects [9,10].

A similar structure to the four-quadrant cell shown in figure 3.7a could be also used with the source terminals kept the same for all the transistors instead of the drain terminal. This structure is shown in figure 3.7b. In this case the differential current assuming triode region operation could be represented by:

$$I_{O1} - I_{O2} = (I_1 - I_4) - (I_2 - I_3)$$ \hspace{1cm} (3.34)

with

$$I_1 - I_4 = \frac{K}{2} \left[ (V_{D12} - V_{D34})(V_{G1} - V_{TO} - \alpha(V_{D12} + V_{D34})) \right]$$ \hspace{1cm} (3.35)

$$I_2 - I_3 = \frac{K}{2} \left[ (V_{D12} - V_{D34})(V_{G2} - V_{TO} - \alpha(V_{D12} + V_{D34})) \right]$$ \hspace{1cm} (3.36)

Substituting 3.35 and 3.36 in 3.34, the resulting resistance becomes:

$$R = \frac{V_{D12} - V_{D34}}{I_{O1} - I_{O2}} = \frac{1}{K(V_{G1} - V_{G2})}$$ \hspace{1cm} (3.37)

Note that equation 3.37 is still only valid if the all the transistors are operating in the triode region. In fact, if all the transistors are in the saturation region, the differential current becomes zero. Thus, the structure in figure 3.7a is more valuable than the one in figure 3.6b.
Figure 3.7 Linearization techniques based on (a) the four-quadrant cell with common drains, (b) the four-quadrant cell with common sources, and (c) two transistors with common drains and common sources.
Another technique based on using the structure shown in figure 3.6d is shown in figure 3.7c, where both transistors share the same source and drain terminals and different gate voltages. In this case, assuming triode mode operation the total current can be represented by:

\[ I = I_1 + I_2 \]  \hspace{1cm} (3.38)

with

\[ I_1 = K \left( V_D - V_S \right) \left( V_{G1} - V_{TO} - \frac{\alpha}{2} \left( V_D + V_S \right) \right) \]  \hspace{1cm} (3.39)

\[ I_2 = K \left( V_D - V_S \right) \left( V_{G2} - V_{TO} - \frac{\alpha}{2} \left( V_D + V_S \right) \right) \]  \hspace{1cm} (3.40)

If the gate voltages are forced to be:

\[ V_{G1} = V_C + V_D \] and \[ V_{G2} = V_C + V_S \]  \hspace{1cm} (3.41)

where \( V_C \) is a control voltage, and by neglecting the body effect (\( \alpha = 1 \)), then equation 3.38 reduces to:

\[ R = \frac{V_D - V_S}{I} = \frac{1}{2K(V_C - V_{TO})} \]  \hspace{1cm} (3.42)

As discussed in this section, there are multiple ways to implement an electronically tunable resistance using linearization techniques of transistors whether in the triode mode, or in the saturation mode. While using transistors has the advantage of being electronically tunable, and the consumption of relatively smaller area than passive resistors, but it also suffers from major disadvantages. The first disadvantage is the need of a non linearity cancellation technique, which increases the complexity of the circuit. The second disadvantage is the that usually there is a limit on the usable voltage range of the resulted
resistance due to the requirement that all the transistors should be at least on, in addition to the fact that some tunable resistors require a triode-only or a saturation-only mode of operation, which limits the usable range even further. The third disadvantage is the poor noise performance of the resulted resistor due to the fact that transistors are in direct contact with the substrate. Therefore, their noise behavior will be worse than a poly resistor for example.

A popular application for linearization techniques of the resistance of the transistor is the implementation of voltage controlled integrators that are used extensively in the area of MOSFET-C filters where classical analog filters, e.g. active RC filters, are being converted to an all MOS implementation with the aid of the linearization techniques of MOSFET transistors [7]. Figure 3.8 shows an example of such an integrator using the four-quadrant structure shown in figure 3.7a, in this case the integrator time constant can be electronically

![A tunable integrator.](image)

Figure 3.8 A tunable integrator.
tunable using $V_{c_1}$ and $V_{c_2}$. In fact, using equation 3.33, the differential output voltage can be written as:

$$V_{o+} - V_{o-} = -\frac{K(V_{c_1} - V_{c_2})}{C} \int_{-\infty}^{t} (V_{i+} - V_{i-}) dt$$

(3.43)

### 3.1.2.2. Tunable Capacitors

As discussed in the previous section, having an electronically tunable resistance is a key element to perform the tuning process. Having electronically tunable capacitors on the other hand, can greatly facilitate the tuning process of some specific blocks like oscillators and filters. There are not as many ways to implement electronically tunable capacitors as in the case of resistors, and usually electronically tunable capacitors have a relatively limited tuning range (around 15%), which limits their usage in most of the cases to the fine tuning of an already course-tuned parameter. There are two main phenomena that are used to implement an electronically tunable capacitor (also called a varactor). The first phenomenon is the voltage dependent capacitance behavior of a MOS capacitor as discussed in detail in the previous chapter (accumulation, depletion, inversion capacitance), and there is also the voltage dependent capacitance of the depletion region of a lightly-doped reverse-biased PN junction. The unit depletion capacitance of a reverse-biased PN junction can be shown to be [3]:

81
\[ C_d = \frac{\varepsilon_{si}}{W_d} \]  

(3.44)

where \( \varepsilon_{si} \) is the silicon dielectric constant, and \( W_d \) is the depletion region thickness. The depletion region thickness can also be shown to be [3]:

\[ W_d = \sqrt{\frac{2\varepsilon_{si}(N_a + N_d)V_d}{qN_aN_d}} \]  

(3.45)

where \( V_d \) is the reverse voltage across the PN junction, \( N_a \) and \( N_d \) are the acceptor and donor doping density in the P and N sides of PN junction respectively, and \( q \) is the electron charge. As shown by equation 3.45 the width of the depletion region can be controlled by the reverse voltage across the junction, which consequently makes the capacitance in equation 3.45 an electronically tunable capacitance.

In addition to the relatively poor noise performance of the MOS or the PN depletion capacitance due to their close proximity to the substrate, they also lack a degree of freedom since the control voltage that controls the capacitance has to be one of the terminal voltages of the capacitor itself, which substantially limits the configurations they can be used with since one of their terminals has to be always AC grounded.

### 3.1.2.3. Tunable Transconductors

A transconductor is simply a circuit block that converts an input voltage to a current signal with a linear transformation factor referred to as the transconductance \( G_m \) -which should not be confused with the transconductance of the transistor \( g_m \) - for a specific range
of the input voltage. Transconductors can have a single ended input and output, a
differential input and a single ended output, or a differential input and a differential output.
Figure 3.9 shows a symbol for a fully differential transconductor. Transconductors have a
wide range of applications in the area of analog signal processing. Those applications include
active filters, amplifiers, equalizers, and numerous other applications [1,7]. Programmable
transconductors particularly are very important in the area of automatic tuning since they
enable the tweaking of the circuit behavior to compensate for process and temperature
variations. Most programmable transconductors are controlled with a voltage signal that
changes the transconductance value $G_m$. There have been numerous techniques to
implement programmable transconductors introduced in the literature [11-14]. Generally,
there are two types of transconductors. The first type is based on the triode mode of
operation of the MOSFET transistor, while the second type is based on the saturation mode
of operation. Triode mode based transconductors give a better linearity performance, i.e.
constant $G_m$, and they have a better single ended performance, i.e. each output current
could have a good linear behavior. On the other hand, transconductors based on the
saturation mode of operation of the MOSFET transistor have the advantage of better speed
performance for the same power consumption, but they have moderate linearity
performance when compared to the triode based transconductors since they rely on the
square-law model of the transistor which is usually not very accurate, specially in the short
channel processes. In addition, usually the difference between the two output currents of the
transconductor is linear, while each single ended current output suffers from significant non-
linearity, which is a disadvantage that triode based transconductors don’t suffer from [1].
In order to evaluate the performance of a transconductor, several factors have to be taken into consideration. The first factor is the input range that will give a specific linear performance i.e. constant $G_m$. The linear range of operation is very important for determining the level of distortion for a given input voltage range. The second factor that is important in the case of fully differential output transconductors is the symmetry of the two differential outputs, i.e. the quality of the fully differential nature of the transconductor. In other words, if the transconductance of the positive output is $G_{m+} = \frac{I_{O+}}{V_{id}}$ and for the negative output is $G_{m-} = \frac{I_{O-}}{V_{id}}$, then how close $G_{m+}$ and $G_{m-}$ are to each other. Where $I_{O+}$, $I_{O-}$, and $V_{id}$ are the positive output current, the negative output current, and the input differential voltage respectively. This is particularly important for determining the level of common mode distortion the transconductor introduces at the output. The third factor that has to be considered is the control voltage range, i.e. the tuning range of the transconductor given a certain input voltage range and distortion levels at the output. Wide control range enables a more robust and controlled performance as well as a wider tuning range.

An example of a transconductor based on the triode mode of operation of the MOSFET transistor is shown in figure 3.9 [15]. In this circuit all the transistors are operating in the saturation mode except for $M_9$, which is operating in the triode mode of operation. Since the current in both $M_1$ and $M_2$ is constant and assuming the two transistors are identical, using equation 3.17, the source voltages of these two transistors are:
Since $M_9$ is operating in the triode mode, then substituting equations 3.46, and 3.47 into 3.15, the current flowing in $M_9$ could be written as:

$$I_9 = 2 \left( V_c - V_{CM} + \frac{2I_{1}}{\alpha K} \right) V_{id}$$

(3.48)

where $V_{id}$ is the input differential voltage $V_{i+} - V_{i-}$, and $V_{CM}$ is common mode voltage $V_{i+} + V_{i-} / 2$. Since $I_{O+} = I_{O-} = I_9$, then the transconductance could be written as:
As shown in equation 3.49, the transconductance can be electronically tunable using $V_C$. However, the circuit in figure 3.9 suffers from some disadvantages. The main disadvantage is the distortion that will be introduced to the outputs due to the presence of the common mode voltage in the transconductance formula. This is particularly a problem if the input signal has a time-variable common mode voltage. On the other hand, the circuit has the advantage of having multiple output terminals by simply taping more current mirrors from $M_3$ and $M_4$, which could save a lot of die area and power consumption.

An example of a transconductor based on the saturation mode of operation of the MOSFET transistor is shown in figure 3.10a [16,17,18]. The basic element of this transconductor is shown in figure 3.10b and is referred to as the COMFET structure, where both transistors are operating in the saturation mode. For the COMFET structure, using equation 3.17 and assuming $M_1$ and $M_2$ are identical and ignoring the body effect ($\alpha = 1$), the differential current could be written as:

$$I_{d1} - I_{d2} = (K/2)(V_{in} - 2V_X + V_C)(V_{in} - V_C - 2V_{TO})$$  \hspace{1cm} (3.50)

In the transconductor shown in figure 3.11b, and assuming that $M_1$, $M_2$, $M_3$, $M_4$, $M_5$, and $M_6$ are all identical, then $V_X$ and $V_Y$ could be written as:

$$V_X = \frac{V_1}{2} \quad \text{and} \quad V_Y = \frac{V_2}{2}$$  \hspace{1cm} (3.51)

Using equations 3.50 and 3.51, the differential current in the transconductor, and the transconductance could be written as:

$$G_m = \frac{I_{Oc}}{V_{id}} = 2 \left( V_C - V_{CM} + \sqrt{2\alpha I_x \frac{K}{K}} \right)$$  \hspace{1cm} (3.49)
As shown in equation 3.53, the transconductance can be electronically tunable using $V_C$.

However, the circuit in figure 3.10b suffers from some disadvantages. The first disadvantage is that –like most saturation based transconductors- only the differential current is linear with the input voltage not each individual output current. The second disadvantage that is the distortion that will be introduced to the output current due to the body effect of the transistors $M_1, M_2, M_4$, and $M_5$. The third disadvantage is that the input linear range is limited by the condition needed to turn on both $M_1$ and $M_3$. 

\[
I_1 - I_2 = (I_{d1} - I_{d2}) - (I_{d3} - I_{d4}) = \left(\frac{K}{2}\right)(V_1 - V_2) V_C
\]  

(3.52)

\[
G_m = \frac{I_1 - I_2}{V_1 - V_2} = \left(\frac{K}{2}\right)V_C
\]  

(3.53)
3.2. Post Fabrication Tuning

An alternative approach to the automatic tuning process is the post fabrication tuning. It is based on tuning the absolute values of the on-chip elements themselves after fabrication. This method is particularly useful when the parameter of interest in the system can not be controlled by a voltage or a current signal, and depends solely on the accuracy of the absolute value of the on-chip element. The post fabrication tuning is based on developing an on-chip programmable block that is equivalent to a resistor or a capacitor and can be controlled with a digital signal. For example, a resistor could be implemented using multiple resistors in parallel with each resistor being controlled by a switch that could be either turned on or off using a digital control signal. By measuring the targeted parameter (that depends on the absolute value of the on-chip resistor or capacitor) and comparing it with its ideal required value, an error signal could be produced. Using a tuning algorithm, the equivalent absolute value of the on-chip elements can then be tuned to the required accuracy by setting the digital control signal. The tuning algorithm in this case is done externally, where a software program on the production tester is used to measure the error signal and determine accordingly the control signal. The control signal could then be permanently hardwired on the chip using laser trimming, poly fuses, or an on-chip EPROM.

Laser trimming is based on using a laser beam to cut an on-chip metal line. By leaving or cutting the metal line, a digital 1 or a 0 could be permanently programmed inside the chip. The poly fuses are just poly lines that could be blown electrically using a current signal, again a digital 1 or a 0 could be permanently programmed inside the chip depending on whether the fuse is blown or not. The EPROM is an electrically programmable read only
memory implemented on the chip, a 1 or a 0 could be permanently written on that memory. It is worth mentioning that all the three previous methods could be done only one time during the production test phase, where the tester software implements the tuning algorithm and copies the required control signal to the chip using one of the above methods. Therefore, the control signal can not be changed later and is permanently hardwired.

As discussed before, the on-chip automatic tuning relies on an on-chip circuit that measures the error signal and sets the control signal accordingly. Therefore, it has the advantage of being much cheaper than the post fabrication tuning for three major reasons. First, it saves tester time (since the tester will not be used to implement the tuning algorithm), which consequently reduces the production cost. Second, it does not need any special processing steps as in laser trimming, poly fuses, or EPROM since the control signal is produced internally and does not need to be hardwired on the chip, consequently reducing production cost. Third, the on-chip circuit that implements the tuning algorithm can be continuously used during operation to tune for changes due to temperature variations. This advantage could not be achieved using laser trimming, poly fuses, and EPROM since they provide a one time permanent programmability.
LIST OF REFERENCES


CHAPTER 4

A VOLTAGE-CONTROLLED TRANSCONDUCTOR

Transconductors have a wide range of applications in the area of analog signal processing. Those applications include active filters, amplifiers, equalizers, and numerous other applications [1, 2]. Programmable transconductors particularly are very important in the area of automatic tuning since they enable the tweaking of the circuit behavior to compensate for process and temperature variations. Most programmable transconductors are controlled with a voltage signal that changes the transconductance value $G_m$ for a specific range of the input voltage. In this chapter, a low voltage (1.8 V), highly linear voltage controlled transconductor with a wide differential input range and a wide control voltage range will be presented [3]. A simple technique is used to cancel the second order nonlinearities caused by MOSFET transistors in the triode mode of operation to significantly improve the linearity of the transconductor as well as the differential input range. The proposed implementation gives a significant improvement in the control voltage range and also makes the circuit truly fully differential. A technique used in the literature will be used as a starting point for the proposed transconductor and a comparison between them will be discussed to highlight the improvements in the proposed implementation.
4.1. Introduction

As mentioned in chapter 3, a transconductor is simply a circuit block that converts an input voltage to a current signal with a linear transformation factor referred to as the transconductance $G_m$. The input voltage range that gives a specific linear performance i.e. constant $G_m$, the symmetry of the two differential outputs, i.e. the quality of the fully differential nature of the transconductor, and the control voltage range, i.e. the tuning range of the transconductor given a certain input voltage range and distortion levels at the output are all very important factors that determine the quality of the transconductor. A transconductor that is based on the triode mode of operation of the MOSFET transistor is shown in figure 4.1 [2]. As mentioned in chapter 3, triode mode based transconductors give better linearity performance than saturation mode based transconductors. For the circuit shown in figure 4.1 and assuming the ideal triode mode current equation [2]:

\[
I_1 = K \left[ \left( V_{CM} + \frac{V_{id}}{2} - V_T \right) V_C - \frac{V_C^2}{2} \right] 
\]

(4.1)

\[
I_2 = K \left[ \left( V_{CM} - \frac{V_{id}}{2} - V_T \right) V_C - \frac{V_C^2}{2} \right] 
\]

(4.2)

\[
I_{CM} = K \left[ \left( V_{CM} - V_T \right) V_C - \frac{V_C^2}{2} \right] 
\]

(4.3)

\[
I_{O+} = I_{CM} - I_2 = \frac{K}{2} V_{id} V_C 
\]

(4.4)

\[
I_{O-} = I_1 - I_{CM} = \frac{K}{2} V_{id} V_C 
\]

(4.5)
Where $V_{id}$ is the differential input voltage, $V_{CM}$ is the common mode voltage of the input, and $V_C$ is the control voltage. Since equations 4.1 through 4.7 assume that both $M_1$ and $M_2$ are operating in the triode mode the following condition has to be satisfied:

$$V_C < V_{CM} - \frac{V_{id}}{2} - V_T$$

(4.8)

Equations 4.1 through 4.7 suggest that ideally the transconductor is achieving the desired function with a perfect linearity performance. However, in order to evaluate the linearity
performance of the transconductor, a more accurate model of the triode mode operation of
the transistor that models second order effects has to be used. Equations 4.1 and 4.2 are
assuming that in the triode mode of operation the current will be linearly proportional to the
gate source voltage (the ideal MOSFET current formula), and that is not very accurate
especially for short channel transistors. In order to model the nonlinearity with respect to the
gate source voltage for a MOSFET, the current equation for the linear mode of operation
could be rewritten as:

\[ I_D = K \left[ \left( V_{GS} + \beta_V V_{GS}^2 + \beta_2 V_{GS}^3 + \ldots \right) - V_T \right] V_{DS} - \frac{V_{DS}^2}{2} \]  (4.9)

where \( \beta \) is a process dependent parameter. Since the cubic and higher order terms are very
small, they can be ignored and the square term only can be used. Equation 4.8 will then be
reduced to:

\[ I_D = K \left[ \left( V_{GS} + \beta_V V_{GS}^2 \right) - V_T \right] V_{DS} - \frac{V_{DS}^2}{2} \]  (4.10)

Using equation 4.10, equations 4.1 through 4.7 for the circuit in figure 4.1 can be rewritten
as:

\[ I_1 = K \left[ \left( V_{CM} + \frac{V_{id}}{2} \right) + \beta_1 \left( V_{CM} + \frac{V_{id}}{2} \right) - V_T \right] V_C - \frac{V_C^2}{2} \]  (4.11)

\[ I_2 = K \left[ \left( V_{CM} - \frac{V_{id}}{2} \right) + \beta_1 \left( V_{CM} - \frac{V_{id}}{2} \right) - V_T \right] V_C - \frac{V_C^2}{2} \]  (4.12)

\[ I_{CM} = K \left( V_{CM} + \beta V_{CM}^2 - V_T \right) V_C - \frac{V_C^2}{2} \]  (4.13)

\[ I_{O+} = I_{CM} - I_2 = K \left[ \left( \frac{V_{id}}{2} + \beta_1 \left( 2V_{CM} - \frac{V_{id}}{2} \right) \right) V_C \right] \]  (4.14)
\[ I_{O-} = I_1 - I_{CM} = K \left[ \frac{V_{id}}{2} + \beta_i \left( 2V_{CM} + \frac{V_{id}}{2} \right) \frac{V_{id}}{2} \right] \; V_C \] (4.15)

\[ G_{m+} = \frac{\partial}{\partial V_{id}} I_{O+} = K \left[ \frac{1}{2} + 2\beta_i V_{CM} - \frac{\beta_i}{4} V_{id} \right] V_C \] (4.16)

\[ G_{m-} = \frac{\partial}{\partial V_{id}} I_{O-} = K \left[ \frac{1}{2} + 2\beta_i V_{CM} + \frac{\beta_i}{4} V_{id} \right] V_C \] (4.17)

Equations 4.10 through 4.17 could now be used to give a more accurate evaluation of limitations in the performance of the circuit. The first limitation in the performance of the circuit shown in figure 4.1 is the linearity of the output currents with respect to the differential input voltage. As equations 4.16 and 4.17 show, the value of both \( G_{m+} \) and \( G_{m-} \) is a function of the differential input voltage \( V_{id} \), which introduces nonlinearity to the output current and consequently increases the total harmonic distortion in the output signal. Hence, if a specific linearity performance is required, the input voltage range has to be limited to a small value. This fact further decreases the valuable input voltage range beyond the condition specified in equation 4.8. The second limitation is that \( G_{m+} \) and \( G_{m-} \), are not equal, which means that the design is not truly fully differential, which will introduce common mode distortion in the output. The third limitation is the control voltage range of the circuit. Since the circuit in figure 4.1 assumes that \( M_1 \) and \( M_2 \) are operating in the triode mode of operation, the condition shown in equation 4.8 has to be satisfied. Hence, if a wide input range is needed, the control voltage range has to decrease and vice versa, which is a disadvantage at low supply voltages. In some situations though, even with losing linearity due to exceeding the control voltage limit shown in equation 4.8, it is desirable to still be able to control the transconductance beyond the limit shown in equation 4.8. In fact, in order to
evaluate the effect on the transconductance if the condition stated in 4.8 is not met, and assuming positive differential voltage, the current equation for $M_2$ has to be changed to the saturation mode of operation formula:

$$I_2 = \frac{K}{2} \left( V_{CM} - \frac{V_{id}}{2} - V_T \right)^2$$

(4.18)

hence equations 4.16 and 4.17 will be changed to:

$$G_{m+} = \frac{\partial}{\partial V_{id}} I_{O+} = K \left( V_{CM} - \frac{V_{id}}{2} - V_T \right)$$

(4.19)

$$G_{m-} = \frac{\partial}{\partial V_{id}} I_{O-} = K \left( \frac{1}{2} + 2 \beta_1 V_{CM} + \frac{\beta_2 V_{id}}{4} - V_T \right) V_C$$

(4.20)

As equations 4.19 and 4.20 show, the transconductance of the positive output saturates and does no longer depend on $V_C$ -which means it can't be controlled any further- while the transconductance of the negative output—even though suffers from non linearity because of the presence of $V_{id}$—but it can still be controlled further with $V_C$. Since the transconductance of one output saturates while the transconductance of the other output does not when condition 4.8 is not met, the circuit completely loses its truly fully differential nature. This is in addition to the fact that distortion will be introduced due to the dependence of $G_{m-}$ on $V_{id}$ as shown in equation 4.20. The fourth limitation in the circuit in figure 4.1 is that it needs a third op-amp to generate the $I_{CM}$ signal. This is in addition to the extra circuitry needed to extract the common mode voltage of the input signal, which increases the size and power consumption of the transconductor. In the next section a modification to the circuit shown in figure 4.1 will be proposed. The introduced
modification extends the control voltage range and also significantly improves the linearity of the circuit. The modification also makes the transconductor truly differential.

4.2. The Proposed Transconductor

The proposed circuit modifies the circuit shown in figure 4.1 in a way to significantly improve the linearity of the transconductor i.e. making both \( G_{m+} \) and \( G_{m-} \) independent from \( V_{id} \). The modification also makes the design truly fully differential by making \( G_{m+} \) and \( G_{m-} \) equal for all input and control voltage range. The proposed circuit also significantly increases the control voltage range by making both \( G_{m+} \) and \( G_{m-} \) proportional to \( V_C \) even when the transistors \( M_1 \) and \( M_2 \) enter the saturation region. The proposed circuit also eliminates the need for a third op-amp or a common mode extraction circuit, which will save power and area. The transconductor is shown in figure 4.2. The positive input stage is comprised of \( M_1, M_5, \) and \( M_9 \), while the negative input stage is comprised of \( M_2, M_6, \) and \( M_{10} \). Both the positive and negative input stages are identical. \( M_3, M_7, \) and \( M_{11} \) are identical to \( M_1, M_5, \) and \( M_9 \) respectively, while \( M_4, M_8, \) and \( M_{12} \) are identical to \( M_2, M_6, \) and \( M_{10} \) respectively. Since transistors \( M_1 \) and \( M_5 \) are connected in a cascode configuration as well as transistors \( M_3 \) and \( M_7 \), and assuming that \( M_5 \) and \( M_7 \) are operating in the saturation mode (a condition that could be met by properly sizing \( M_9, M_{10}, M_{11}, \) and \( M_{12} \)), the output impedance at the drains of both \( M_5 \) and \( M_7 \) will be
substantially high. This means that the drain voltages of $M_5$ and $M_7$ will have a small effect on the values of the currents $I_3$ and $I_4$ respectively. Since $M_5$ and $M_7$ are identical and they have the same gate voltage, then the currents $I_3$ and $I_4$ will be identical if the source voltages of $M_5$ and $M_7$ are identical. Furthermore, since $M_1$ and $M_3$ are identical and they have the same gate to source voltage $V_{GS}$, and since they both operate in the triode mode, then the currents $I_3$ and $I_4$ will be identical if the drain voltages of $M_1$ and $M_3$ are identical. If we assume that $I_4$ is less than $I_3$, then the source voltage of $M_7$ (the drain of $M_3$) will have to be higher than the source voltage of $M_5$ (the drain of $M_1$), but having a higher drain voltage on $M_3$ means that $I_4$ will be higher than $I_3$, which contradicts the

Figure 4.2 Schematic diagram of the proposed transconductor.
assumption. Furthermore, if we assume that $I_3$ is less than $I_4$, then the source voltage of $M_5$ (the drain of $M_1$) will have to be higher than the source voltage of $M_7$ (the drain of $M_3$), but having a higher drain voltage on $M_1$ means that $I_3$ will be higher than $I_4$, which again contradicts the assumption. Considering the arguments mentioned above, the only stable condition for the circuit will have to be that the currents $I_3$ and $I_4$ flowing in transistors $M_1$ and $M_3$ respectively are essentially equal, and that the source voltages of $M_5$ and $M_7$ (the drain voltages of $M_1$ and $M_3$ respectively) are also essentially identical. In other words, a virtual short circuit between the drains of transistors $M_1$ and $M_3$ is established. Since the op-amp sets the drain of $M_1$ to $V_C$, the drain of $M_3$ will also be set to $V_C$. This guarantees that the control voltage $V_C$ is controlling the drains of both $M_1$ and $M_3$, and that the currents $I_3$ and $I_4$ are identical without the need for an extra op-amp to force $V_C$ on the drain of the transistor $M_3$. The same analysis could be applied to the negative input stage as well.

In order to demonstrate how the proposed circuit improves the linearity and input range of the transconductor, using equation 4.9, the following set of equations are derived:

$$I_1 = I_2 = K \left[ \left( V_{CM} - \frac{V_{id}}{2} \right) + \beta_i \left( V_{CM} - \frac{V_{id}}{2} \right)^2 - V_T \right] V_C - \frac{V_C^2}{2} \right]$$ (4.21)

$$I_3 = I_4 = K \left[ \left( V_{CM} + \frac{V_{id}}{2} \right) + \beta_i \left( V_{CM} + \frac{V_{id}}{2} \right)^2 - V_T \right] V_C - \frac{V_C^2}{2} \right]$$ (4.22)

$$I_{O+} = I_4 - I_1 = K \left[ (1 + 2 \beta_i V_{CM}) V_{id} V_C \right]$$ (4.23)

$$I_{O-} = I_3 - I_2 = K \left[ (1 + 2 \beta_i V_{CM}) V_{id} V_C \right]$$ (4.24)
Comparing equations 4.23 through 4.26 to equations 4.14 through 4.17, it is clear that the proposed modification forced both the positive and negative current outputs, as well as the transconductances to be equal. This equality makes the design truly fully differential and reduces the common mode distortion as well as the offsets introduced when $I_{O+}$ and $I_{O-}$ are not equal. Equations 4.25 and 4.26 show that the transconductance of both current outputs of the proposed circuit is also linear and constant with respect to the differential voltage input (as opposed to equations 4.16 and 4.17). This improvement significantly reduces the total harmonic distortion at the output currents. Hence, if a certain non-linearity error in the transconductance is acceptable, the proposed circuit gives a much wider differential input range for the same amount of error than the original circuit in figure 4.1. In order to also see how the proposed circuit extends the control voltage range as well, if $M_2$ and $M_4$ enters the saturation mode of operation (equation 4.8 is not satisfied), the following set of equations could be derived:

\[
I_1 = I_2 = \frac{K}{2} \left[ \left( V_{CM} - \frac{V_{id}}{2} - V_T \right)^2 \right] \tag{4.27}
\]

\[
G_{m+} = \frac{\partial}{\partial V_{id}} I_{O+} = \frac{K}{2} \left[ \left( V_{CM} - \frac{V_{id}}{2} - V_T \right) \right] + \frac{1}{2} + \beta \left( V_{CM} + \frac{V_{id}}{2} \right) V_C \tag{4.28}
\]

\[
G_{m-} = \frac{\partial}{\partial V_{id}} I_{O-} = \frac{K}{2} \left[ \left( V_{CM} - \frac{V_{id}}{2} - V_T \right) \right] + \frac{1}{2} + \beta \left( V_{CM} + \frac{V_{id}}{2} \right) V_C \tag{4.29}
\]
Comparing equations 4.28 and 4.29 to equations 4.19 and 4.20, equations 4.28 and 4.29 show that for higher values of $V_C$ when equation 4.8 is not satisfied and some of the transistors go into the saturation mode, the transconductances –although dependent on $V_{id}$- still can be controlled by $V_C$ in a linear fashion. This is very beneficial especially at low supply voltages (1.8 V) since it extends the range of control $V_C$ have over the transconductance. Equations 4.28 and 4.29 also show that both $G_{m+}$ and $G_{m-}$ are equal, hence keeping the truly fully differential nature of the circuit even when the transistors enter into the saturation mode of operation.

### 4.3. Simulation Results

Simulations results show the improvements that the proposed circuit in figure 4.2 has over the original circuit shown in figure 4.1. Figures 4.3 and 4.4 show the output currents of the transconductor shown in figure 4.1 and the proposed transconductor in figure 4.2. Notice the linearity difference in the currents. Figures 4.5, 4.6, and 4.7 show the transconductances of both circuits at different control voltages. Note the significant improvement in the transconductance variation with input voltage. The proposed circuit is giving a much less variation in the transconductance value with respect to the input differential voltage, an advantage that will significantly reduce the total harmonic distortion. To assess the improvement in performance, figures 4.8, 4.9, and 4.10 show the percentage error in the transconductance value relative to the transconductance at 0 input differential voltage for both circuits. In other words:
As those figures show, if a 3 percent error is acceptable in the transconductance value and a control voltage of 100 mV is used, the circuit in figure 4.1 offers only 120 mV differential input range, while the proposed circuit in figure 4.2 offers 793 mV differential input range. At a control voltage of 400 mV, the circuit in figure 4.1 offers a 220 mV differential input range, while the proposed circuit in figure 4.2 offers 481 mV. At a control voltage of 1 V, the circuit in figure 4.1 offers a 214 mV differential input range, while the proposed circuit in figure 4.2 offers 513 mV. As those numbers show, there is a significant improvement in the differential input range, or in other words, the output currents are much more linear with respect to the differential input voltage. Another advantage of the proposed circuit is the truly fully differential nature of the circuit. Figures 4.11, 4.12, and 4.13 show the differential error in the transconductance value (the difference between $G_{m+}$ and $G_{m-}$ relative to the transconductance at 0 differential input). In other words:

$$\text{Differential Transconductance Error (\%)} = \frac{G_{m+}(V_{id}) - G_{m-}(V_{id})}{G_m(0)} \times 100$$  \hspace{1cm} (4.31)

As shown in those figures, the original circuit (shown in figure 4.1) offers a differential input range of 19.7 mV at 100 mV control voltage for a 1 percent error, while the proposed circuit offers 710 mV differential input range, which is a very significant improvement. Another advantage of the proposed circuit is the improvement in the control voltage range. Figure 4.14 shows the transconductance versus the control voltage for both circuits. The new circuit has a higher linear range of control voltage. Also figure 4.14 shows that for the proposed circuit, the transconductance value saturates at a higher value, which agrees with equations 4.28 and 4.29.
Figure 4.3 Output currents for the proposed circuit and the original circuit in figure 4.1 at 100mV control voltage.

Figure 4.4 Output currents for the proposed circuit and the original circuit in figure 4.1 at 1V control voltage.
Figure 4.5 \( G_{m+} \) and \( G_{m-} \) for the proposed circuit and the original circuit in figure 4.1 at 100 mV control voltage.

Figure 4.6 \( G_{m+} \) and \( G_{m-} \) for the proposed circuit and the original circuit in figure 4.1 at 400 mV control voltage.
Figure 4.7 $G_{m+}$ and $G_{m-}$ for the proposed circuit and the original circuit in figure 4.1 at 1 V control voltage.

Figure 4.8 Transconductance error for the proposed circuit and the original circuit in figure 1 at 100 mV control voltage.
Figure 4.9 Transconductance error for the proposed circuit and the original circuit in figure 1 at 400 mV control voltage.

Figure 4.10 Transconductance error for the proposed circuit and the original circuit in figure 1 at 1 V control voltage.
Figure 4.11 Transconductance differential error for the proposed circuit and the original circuit in figure 4.1 at 100 mV control voltage.

Figure 4.12 Transconductance differential error for the proposed circuit and the original circuit in figure 4.1 at 400 mV control voltage.
Figure 4.13 Transconductance differential error for the proposed circuit and the original circuit in figure 4.1 at 1 V control voltage.

Figure 4.14 The transconductance versus control voltage for the proposed circuit and the original circuit in figure 4.1.
4.4. Conclusion

In this chapter, a simple technique to improve the linearity of a voltage controlled transconductor was introduced. The high linearity of the circuit enables a high differential input range as well as a high control voltage range with a low supply voltage (1.8 V). The proposed transconductor achieves a truly fully differential nature for the output currents with $G_{m+}$, $G_{m-}$ almost identical. The technique achieves this high performance by eliminating second order effects in linear mode based transconductors. The technique also eliminates the use of an extra op-amp and a common mode extraction circuitry, which save power and area without compromising the high performance. The proposed transconductor was filed to the United States Patent Office, and a patent was granted on it on April 20, 2004 under the number 6724258. It is going to be used in implementing an all-pass adaptive $G_m$-C filter used in an analog adaptive equalizer that will be presented in chapter 7.
LIST OF REFERENCES


CHAPTER 5

A DIGITAL ADAPTIVE TECHNIQUE FOR ON-CHIP RESISTORS TUNING

A resistor is a very fundamental building block in any analog IC design. Multiple applications in the area of analog signal processing use some sort of a resistor to achieve the performance required. As mentioned in the previous chapters, using off-chip resistors—even though achieve higher accuracy—puts a limit on the level of achievable integration, as well as the increase in cost associated with them. Using on-chip resistors is an attractive alternative due to the cheaper cost and the higher integration levels that could be achieved. Yet again, the poor control over their absolute values puts another limit on the performance of the whole system. An important application for accurate on-chip resistors is in the area of wired communication networks, where cables are usually modeled as transmission lines. In this case, accurate termination resistors have to be used to minimize reflections and keep signal power and integrity [1-4]. Typically an accuracy of ±10% across process and temperature is adequate for high speed transmission lines terminations, yet higher accuracy should be targeted to improve the performance. As discussed in chapter 3, tuning on-chip resistors
(automatic, or post fabrication tuning) achieves a good accuracy and still keeps a higher level of integration with an acceptable cost possible. Chapter 3 introduced different tuning architectures that can be used to tune for any on-chip parameter. A key element to achieve the tuning is to have an electronically tunable element, which in the case of resistors, could simply be a combination of transistors configured in a specific way to cancel any non-linearity in the transistor behavior [5-11]. Even though using active devices with non-linearity cancellation techniques might achieve an acceptable tunable and linear performance for some applications, yet they still generally suffer from range, linearity, and noise problems. The range problem evolves from the fact that active transistors need a specific minimum voltage to be turned on, as well as any other range restriction to keep them in a specific operation mode (triode, or saturation), while the linearity problem evolves from the fact that non-linearity cancellation techniques are relying on the square-law representation of the transistor, which has an accuracy limit specially for short channel transistors [12]. On the other hand, since active elements are in direct contact with the substrate, they are relatively more vulnerable to noise coupling from the substrate, which limit the noise performance of the implemented resistor.

This chapter introduces a digital tuning algorithm for implementing accurate on-chip resistances using passive elements. The technique is based on an indirect version of the digital-based tuning architecture discussed in chapter 3 (figure 3.3b) with a digitally programmable resistor matrix [13]. The technique has the advantage of using a single external resistor to tune all the required on-chip resistors automatically even if they are of different values. It could achieve –if needed- an arbitrary tight control over on-chip resistors that will be limited only by the accuracy of the external resistor (usually has ±1% accuracy),
current mirrors mismatches, and area. The technique does not require any accurate voltage references and it implements the required resistors using on-chip passive resistors, which reduces the noise introduced to the system and improves linearity and range.

5.1. The Proposed Technique

The proposed architecture uses a single tightly controlled external resistor to control the on-chip resistors within the required accuracy. The technique needs only one external resistor even if different on-chip resistors with different values need to be tuned in the same time. Any modern mixed-signal IC design will essentially have an external resistor and a band-gap reference circuit, where both are used to generate the required biasing currents and voltages needed by the analog parts of the circuit. The proposed technique though does not need the band-gap reference or any kind of an accurate voltage reference.

Before the technique is introduced, a discussion of the concept of "variation range quantization" shown in figure 5.1 will be presented. Let's assume that an on-chip resistor is designed to have the value $R_{req}$ in nominal conditions. Given process and temperature variations, this value will drift. Let's define the maximum relative percentage the on-chip resistor value can drift from its nominal value $R_{req}$ to be $\pm D_{Max} \%$. $D_{Max}$ is a characteristic of each specific process, physical dimensions of the resistor, and temperature. It is usually available through process characterization. Now, let's define the maximum relative percentage variation we can allow for the on-chip resistor to drift from its nominal value $R_{req}$ to be $\pm C_{Max} \%$. As shown in figure 5.1, the whole $\pm D_{Max}$ range can be divided into
\[(n+1)\text{ different regions, where each region is } C_{Max}\text{ wide except for region } \left(\frac{n}{2}\right)\text{ which is } 2C_{Max}\text{ wide. Region } \left(\frac{n}{2}\right)\text{ is essentially where the tuning algorithm needs to place the variation in the value of the on-chip resistance to be within. This division leads to the following equation that determines } n:\]

\[n = \left(\frac{D_{Max}}{C_{Max}} - 1\right) \times 2 \quad (5.1)\]

Therefore, what the tuning algorithm should achieve is to always tune the variation in the resistor value to the region \(\left(\frac{n}{2}\right)\) if the variation in the resistor value fell within any of the different \((n+1)\) regions in figure 5.1. Hence, what is needed is to actually have a resistor
block that is composed of \((n+1)\) parallel on-chip resistors as shown in figure 5.2 instead of a single resistor, where each resistor is controlled with a switch that can be turned on or off using a digital control signal. In figure 5.2, \(R_0\) is the core resistor that has a value close to \(R_{req}\), while the rest of the resistors are just modulating resistors to adjust the total resistance of the resistance block to be within \(R_{req} \pm C_{Max} \%\). This chapter develops –in addition to the concept- a mathematical way for determining the value of each resistor in the resistor block shown in figure 5.2 that will make the whole technique very simple, yet very practical.

**Figure 5.2** The resistor block.
To start off, if each variation region in figure 5.1 is assigned to one of the parallel resistors in the resistor block that could be switched on or off, we can find adequate values for those resistors to guarantee that the total resistance of the resistor block is within region $\left(\frac{n}{2}\right)$. Therefore, the resistor block needs to be designed such that in nominal condition $R_1$ through $R_n$ are switched on, while $\frac{R_n}{2}$ through $R_n$ are switched off. Note that $R_0$ is always connected, which is an advantage that will be discussed later in details. According to the variation in the values of those resistors, the tuning algorithm should either switch on or switch off more resistors in parallel to $R_0$ in order to always place the total resistance of the resistor block to be in region $\left(\frac{n}{2}\right)$. The previous discussion implies that:

$$R_0 \parallel R_1 \parallel \ldots \parallel \frac{R_n}{2} = R_{req}$$

(5.2)

Note that the resistors names represent their nominal values. In order to develop mathematical equations that determine the nominal values of the $(n+1)$ resistors, couple of assumptions will be made. First, we will assume that all the resistors in the resistor block vary in the same way and also with the same relative magnitude. As discussed in chapter 2, this assumption is well justified since all of the resistor are going to be placed in a very close proximity to each other and also have close values. It is also a well known fact that on-chip resistors maintain their ratio very well, which implies that their relative magnitude change is the same, in fact on-chip resistors usually match within better than 0.1% [14]. Provided that the resistors are laid out carefully, and are designed to have wide enough dimensions to minimize under etching effects, i.e. $\frac{\Delta W}{W}$ and $\frac{\Delta L}{L}$, this assumption is very safe. The second
assumption is that the switch resistance is negligible relative to the values of the resistors, which is an assumption that is not immediately obvious, specially when the required resistance value is relatively small, but as will be showed later in this chapter, the proposed tuning algorithm makes this assumption valid, which is one of the important advantage of the proposed technique, i.e. the non-linearity of the switch resistance has a very small effect on the performance of the resistor block.

In order to develop an intuitive methodology for calculating the values of the resistors in the resistor block, let's start by assuming that the on-chip resistors in the resistor block changed from their nominal value by a percentage value between \((-D_{\text{Max}})\)% and \((-D_{\text{Max}} + C_{\text{Max}})\)% , which is equivalent to region (0) in figure 5.1. Note that \((-D_{\text{Max}} + C_{\text{Max}}) = -\frac{n}{2} C_{\text{Max}}\). Since this is the worst case negative variation (the smallest value of all the resistors), all the resistors \(R_1\) through \(R_{\frac{n}{2}}\) should be switched off leaving only \(R_0\) connected. Note that using this strategy, \(R_0\) does not need a switch, which is a very beneficial fact as will be explained later on. Since in that case only \(R_0\) is present then it has to satisfy the following two inequalities to cover its variations:

\[
R_0 \left(100 - D_{\text{Max}}\right) > \left(100 - C_{\text{Max}}\right) R_{\text{req}} \quad (5.3)
\]

\[
R_0 \left(100 + C_{\text{Max}}\right) < \left(100 - \frac{n}{2} C_{\text{Max}}\right) R_{\text{req}} \quad (5.4)
\]

Inequality 5.3 guaranties that if the variation is at the bottom of region (0) (which is \((-D_{\text{Max}})\)%), then the nominal value of \(R_0\) will be high enough to place the total resistance of the resistor block (including the variation) at a value higher than the bottom edge of
region \( \left( \frac{n}{2} \right) \). Inequality 5.4 guaranties that if the variation is at the top of region \( (0) \) (which is \((- D_{\text{Max}} + C_{\text{Max}})\%\)), then the nominal value of \( R_0 \) will be low enough not to place the total resistance of the resistor block (including the variation) at a value higher than the top edge of region \( \left( \frac{n}{2} \right) \). Both inequalities guarantee that for any variation within \((- D_{\text{Max}})\%\) and \((- D_{\text{Max}} + C_{\text{Max}})\%), the total resistance of the resistor block (including the variation) will be always within \( \pm C_{\text{Max}}\% \) from the required nominal value \( R_{\text{req}} \). Combining inequalities 5.3 and 5.4 into a single inequality:

\[
\frac{100 - C_{\text{Max}}}{100 - D_{\text{Max}}} R_{\text{req}} < R_0 < \frac{100 + C_{\text{Max}}}{100 - \frac{n}{2} C_{\text{Max}}} R_{\text{req}}
\] (5.5)

Inequality 5.5 determines the range the nominal value of \( R_0 \) can take. Now assuming that the on-chip resistors in the resistor block changed from their nominal value by a percentage value between \((- D_{\text{Max}} + C_{\text{Max}})\%\) and \((- D_{\text{Max}} + 2C_{\text{Max}})\%), and assuming that in this case both \( R_0 \) and \( R_1 \) will be switched on, then following the same logic used to determine \( R_0 \), the following inequality can be used to determine \( R_1 \):

\[
\frac{100 - C_{\text{Max}}}{100 - \frac{n}{2} C_{\text{Max}}} R_{\text{req}} < R_1 \ // R_0 < \frac{100 + C_{\text{Max}}}{100 - \frac{n - 2}{2} C_{\text{Max}}} R_{\text{req}}
\] (5.6)

solving inequality 5.6 for \( R_1 \):

\[
\frac{R_0 \left(100 - C_{\text{Max}}\right) R_{\text{req}}}{100 \left(R_0 - R_{\text{req}}\right) + C_{\text{Max}} \left(R_{\text{req}} - \frac{n}{2} R_0\right)} < R_1 < \frac{R_0 \left(100 + C_{\text{Max}}\right) R_{\text{req}}}{100 \left(R_0 - R_{\text{req}}\right) - C_{\text{Max}} \left(R_{\text{req}} + \frac{n - 2}{2} R_0\right)}
\] (5.7)
Again, inequality 5.7 defines a range of values the nominal value of \( R_i \) can take as a function of \( R_0 \), while \( R_0 \) could be determined using inequality 5.5.

If the same methodology (adding more resistors in parallel) is followed to calculate the rest of the resistors values, the following set of recursive inequalities can be determined:

For \( i < \frac{n}{2} \):

\[
\frac{(100 - C_{\text{Max}})R_{\text{req}}RT_i}{100 \left( RT_i - R_{\text{req}} \right) + C_{\text{Max}} \left( R_{\text{req}} - \left( \frac{n}{2} + i - 1 \right)RT_i \right)} < R_i \tag{5.8}
\]

For \( i > \frac{n}{2} \):

\[
\frac{(100 + C_{\text{Max}})R_{\text{req}}RT_i}{100 \left( RT_i - R_{\text{req}} \right) - C_{\text{Max}} \left( R_{\text{req}} + \left( \frac{n}{2} - i - 1 \right)RT_i \right)} < R_i \tag{5.9}
\]

\[
\frac{1}{R_n} = \frac{1}{R_{\text{req}}} - \frac{1}{RT_n} \tag{5.10}
\]

where \( i \) is an index that starts from 0 to \( n \) in order to calculate the resistors from \( R_0 \) up to \( R_n \), and \( RT_i \) is the total parallel combination of \( R_0 \) up to \( R_{i+1} \), or mathematically:

\[
\frac{1}{RT_i} = \sum_{k=0}^{\frac{i-1}{2}} \frac{1}{R_k} \quad \text{for } i > 1 \quad \text{and} \quad RT_0 = \infty \quad \text{for } i > 0 \tag{5.11}
\]
Note that for the special case when \( i = 0 \) and \( RT_0 = \infty \), inequality 5.8 will be reduced to equation 5.10. Note also that equation 5.10 is another form of equation 5.2.

After the mathematical equations that determine the nominal values of the resistors in the resistor block have been developed, a technique to detect and quantify the drift in the value of the on-chip resistors from their nominal values has to also be developed. When that detection is possible, the digital control signals needed to control the switches in the resistor block can be determined. In order to achieve that, the circuit in figure 5.3 is proposed. The op-amp copies a reference voltage \( V_B \) to an external resistor \( R_{ext} \). The current generated is then mirrored using a high quality current mirror and injected to an on-chip resistor \( R_{int} \).

Figure 5.3 The proposed tuning architecture.
Note that the simple current mirror showed in figure 5.3 is only an example, a high quality current mirror designed to minimize the effects of transistor mismatches on the mirroring ratio \( K \) and also has a high output impedance should be used. Usually a cascode or a low voltage cascode current mirror is sufficient. The voltage across \( R_{int} \) \((V_{Br})\) will be:

\[
\frac{V_{Br}}{V_B} = \frac{R_{int}}{K R_{ext}}
\]  

(5.12)

Note again that the names of the resistors represent their nominal values. Equation (5.12) shows that the relative percentage change in \( R_{int} \) from its nominal value will cause the same relative change in \( V_{Br} \) from \( V_B \). Therefore, if \( R_{int} \) is implemented in a close proximity to the resistor block, and then \( V_{Br} \) is compared to \( V_B \), the relative variation in the values of the on-chip resistors in the resistor block due to process and temperature variations can be detected. It is worth mentioning that since \( V_{Br} \) is being compared to \( V_B \) to find the relative change, the absolute value of \( V_B \) or its accuracy is of no concern. In fact, it could simply be just a potential divider from the supply and it doesn't matter if it changes with process or temperature. The factor \( K \) as well as \( R_{int} \) could simply be anything desired, but since \( V_{Br} \) is being compared to \( V_B \), it is much easier just to set the ratio in equation 5.12 to be nominally equal to unity. This implies that the nominal value of \( R_{int} \) will be \( K \) times \( R_{ext} \). This gives more flexibility in choosing the value of \( R_{int} \), for example picking \( R_{int} \) to be double \( R_{ext} \) will reduce the power consumption by a factor of two on the expense of doubling the area needed to implement \( R_{int} \).
As shown by the concept of quantization shown earlier, what determines the state of the switches controlling the resistors in the resistor block is really the region where the relative change in the values of the on-chip resistors from their nominal value happens to be within. In order to determine the variation region, reference voltages that correspond to the borders of those regions have to be generated. In order to develop the mathematical equation that determines those reference voltages, the following steps will be followed:

If \( \frac{V_{Br}}{1 - \left(1 - \frac{n}{2 \times 100} C_{Max}\right)V_B} \), then the resistor \( R_i \) will definitely be switched on. Let's call the control signal that controls the switch to be \( V_{C1} \) and let's assume that it's active high. This leads to the logical equation:

\[
V_{C1} = \text{high if } V_{Br} > V_{r1} \text{ and } V_{C1} = \text{low if } V_{Br} < V_{r1}
\]  

(5.13)

where \( V_{r1} = \left(1 - \frac{n}{2 \times 100} C_{Max}\right)V_B \). Using the same steps, the following set of recursive equations could be found:

\[
V_{ri} = 1 - \left(1 - \frac{\frac{i}{2} + 1 - i}{100}\right) C_{Max} V_B \quad \text{for } i \leq \frac{n}{2} 
\]  

(5.14)

\[
V_{ri} = 1 - \left(1 - \frac{n - i}{2 \times 100}\right) C_{Max} V_B \quad \text{for } i > \frac{n}{2} 
\]  

(5.15)

\[
V_{Ci} = \text{high if } V_{Br} > V_{ri} \text{ and } V_{Ci} = \text{low if } V_{Br} < V_{ri} \quad \text{for } i \leq \frac{n}{2}
\]  

(5.16)

\[
V_{Ci} = \text{low if } V_{Br} < V_{ri} \text{ and } V_{Ci} = \text{high if } V_{Br} > V_{ri} \quad \text{for } i > \frac{n}{2}
\]  

(5.17)
where \( i \) is an index that starts from 1 to \( n \). Note how the logic level of \( V_{Gi} \) is inverted for \( i > \frac{n}{2} \) during nominal conditions, i.e. when \( V_{n/2} < V_{Br} < V_{n/2+1} \). Using equations 5.14 to 5.17, the reference generator and the control matrix shown in figures 5.4 and 5.5 respectively can be implemented. In figure 5.4, the op-amp forces \( V_B \) on the series on-chip resistors \( R_{r0} \) to \( R_{r\frac{n}{2}} \). Using equations 5.14 and 5.15 to develop the resistors values in the potential divider in figure 5.4, the following recursive equation can be found:

\[
R_n = \left( 1 - \frac{n-1}{100} C_{Max} \right) R_0 - \sum_{k=0}^{k=i-1} R_{rk}
\]  

(18)

**Figure 5.4** The reference generator.
where \( i \) is an index that starts from 0 to \( n \), and \( R_r = \sum_{k=0}^{k=n} R_{rk} \). \( R_r \) along with \( V_p \) determines the amount of current flowing in the potential divider, and its value is optional to the designer. Note that when \( i = 0 \), \( \sum_{k=0}^{k=n} R_{rk} = 0 \). The control matrix shown in figure 5.5 is just a stack of \( n \) comparators that implement equations 5.16 and 5.17, and accordingly control the switches in the resistor block. This concludes the implementation of the proposed tuning technique. The next section will discuss some practical aspects and advantages of the proposed technique.

\[
\begin{align*}
    V_{(n/2)} & \quad V_{C1} \\
    V_{r2} & \quad V_{C2} \\
    V_{r1} & \quad V_{C(n/2)} \\
    V_{Br} & \quad \text{Control Word} \\
    V_{r(n/2+1)} & \quad V_{Cn} \\
    V_{rn} & \quad V_{C(n/2+2)} \quad V_{C(n/2+1)}
\end{align*}
\]

**Figure 5.5** The control matrix.
5.2. Practical Aspects

In this section some practical advantages of the proposed technique as well as some design notes will be pointed out. The first practical advantage of proposed technique is the recursive nature of equations 5.8 to 5.10 and 5.14 to 5.18. This recursive nature makes the design process very simple. An excel sheet or a matlab code could simply be developed to calculate those equations. The required resistance value $R_{req}$, the maximum variation $D_{Max}$, the required tolerance $C_{Max}$, and $R_t$ could be introduced to the excel sheet or the matlab code as inputs, and all the resistors in the resistor block as well as the reference generator could be calculated.

The second practical advantage is that equations 5.8 to 5.10 give a range of values that each resistor in the resistor block can take rather than a specific accurate value. This makes the design very flexible to inaccuracies and very easy to implement. Normally, the value of each resistor in the resistor block should be in the middle of the range specified by equations 5.8 to 5.10 in order to achieve the maximum margin possible.

The third practical advantage is that the modulating resistors $R_1$ through $R_n$ are much higher than the core resistor $R_0$. For example, if $R_{req} = 50\Omega$, $D_{Max} = 25\%$, and $C_{Max} = 5\%$ then $R_0$ will be about $65\Omega$, while $R_1$ up to $R_n$ will be in the kilo ohms range, which is much higher than $R_0$. This should come of no surprise since $R_1$ up to $R_n$ are just used to tune the total resistance of the resistor block, while $R_0$ is the resistor that will bear most of the current flowing in the block. Since only $R_1$ up to $R_n$ have switches and not $R_0$, the design of the switches become really easy due to the resistors relatively high values.
Consequently, the switches do not have to be designed with large sizes in order to make their resistance relatively negligible (switches usually need to have a relatively negligible resistance in order not to affect the linearity of the total resistance). Not having to design the switches to be large will also reduce the capacitive loading of the switches, which will widen the frequency range the resistor block could be used for. This advantage becomes very clear taking into account that some techniques use a single resistor instead of a resistor block, and to achieve the tuning, a bank of resistors is used with each resistor having a slightly drifted value from $R_{req}$. Depending on process variations, only one resistor out of the bank of resistors will be used. This means that the switch has to be negligible relative to values close to $R_{req}$, which will require significantly larger transistors as opposed to the proposed technique. Note that if the switch does not have a relatively negligible resistance, it will degrade the linearity of the total resistance and will also increase the variation in the total resistance with temperature. This clarifies the advantage of the proposed technique and also justifies the second assumption made during the development of the recursive equations 5.8 to 5.10.

The fourth practical advantage is that the proposed technique does not require any accurate voltage references. Essentially, $V_b$ is arbitrary and does not need to be accurate. This is due to the fact that the reference voltages are generated using the same voltage $V_b$ that is used to generate the current from $R_{ext}$. This is specifically an advantage for systems that do not have a band-gap circuit available on-chip.

The fifth practical advantage of the technique is that it's using the same components that are usually used to generate accurate biasing currents for the analog parts in the system without disturbing it. The structure shown in figure 5.3 could be actually used to generate
accurate biasing currents to the rest of the circuit by taping off more current mirrors. In this case \( V_g \) will be the output of a band-gap circuit, which will have to be used anyway to generate the biasing currents whether the proposed technique is used or not. In some analog parts, the biasing currents do not have to be very accurate, and instead of using the band-gap voltage across an external resistor, it is used across an on-chip resistor. In that case, the structure in figure 5.4 could be used to generate those biasing currents as well by setting \( R_{int} \) in equation 5.18 to the desired value. Whether the biasing currents are generated using the band-gap voltage across an external resistor or an on-chip resistor, the proposed architecture can be easily accommodated. Hence, the power consumption of the proposed technique could be considered to be only the power consumed in \( R_{int} \) and the control matrix. The previous few paragraphs discussed some practical advantages of the proposed tuning architecture. In the next few paragraphs, some design notes that the designer has to keep in mind during the design process will be considered.

The first note is that if the application specifies a certain required tolerance on the on-chip resistance, then \( C_{Max} \) has to be chosen to be a little less than the required tolerance. This is to account for the variation in the value of \( R_{ext} \) (usually varies within 1\%), any mismatch errors in the current mirror used to inject the current into \( R_{int} \), and also the offset introduced by the op-amp in figure 5.3. Usually a 2 or 3 percent margin in \( C_{Max} \) is sufficient.

The second note is that it is not desirable to have to tune for temperature change since it is time dependent. Having to tune for temperature variations will consequently require the tuning circuit to be active at all times, which increases the power consumption. Even though the proposed technique could achieve that, but it is wiser to avoid it. Therefore, under nominal process conditions and across all temperature range, the designer
has to make sure that each resistor in the resistor block does not vary more than \( \pm C_{Max} \% \) from the nominal value. This could be achieved by using silicide-block poly resistors to implement the resistors in the resistor block. Note that silicide-block poly resistors have a very small temperature coefficient. Having a negligible switch resistance helps a lot too in minimizing variations with temperature. If there is no need to tune for temperature variations, then the tuning circuit could be activated only one time during power up and then deactivated for the rest of the operation time without worrying about temperature effects.

The third note is that if the required on-chip resistance is small (50\( \Omega \) for example), then \( R_{int} \) will have to be much higher in order to save power. Otherwise if \( R_{int} \) is chosen to be 50\( \Omega \), the current needed to generate \( V_{Br} \) will be significantly higher. The designer has to be careful to design \( R_{int} \) and all the resistors in the resistor block to have wide enough dimensions and are laid out in a close proximity to guarantee that all of them are affected by the same relative magnitude with process variation even though their values are different. This might require a little extra space but it is very important. If different resistor blocks with different values are required then one \( R_{int} \) could be used provided that all the resistor blocks are laid out in a close proximity. If the resistor blocks have to be scattered to different locations on the chip, then there are two options. The first option is to dedicate an \( R_{int} \) resistor for each individual resistor block that is laid out close it, and then a state machine along with the control matrix could be used to do the comparison with the reference voltages for each resistor block successively. The second option is to use a single \( R_{int} \) resistor for all the resistor blocks, and simply reduce \( C_{Max} \) by another 1 percent to account for the mismatch between the scattered resistor blocks. The first option could be
complicated and requires a state machine, while the second option is easier to implement. If
the second option is chosen though, it is desirable to implement $R_{int}$ at the center of the
chip in order to give an average estimation of the errors in the resistors in the rest of the
chip.

5.3. Design Examples

In this section, two design examples along with simulation and measurements results
will be presented. The first design example is the implementation of termination resistors for
a single port high speed transceiver. The transceiver operates at 480Mbps and it requires a
pair of termination resistors of nominally $45\, \Omega$ each. The transceiver specifications require a
tolerance of no more than $\pm 10\%$ in the value of the termination resistors. For this
transceiver, $C_{Max}$ was chosen to be 8.67% in order to give a 1.33% margin to account for
any mismatches in the current mirrors. Process characterization showed that the maximum
change from the nominal value for on-chip silicide-block poly resistors is $\pm 26\%$ from the
nominal value including variations with temperature as well, which implied that $D_{Max}$ should
be 25%, but for extra margin $D_{Max}$ was chosen to be 26%. Using equation 5.1, the number
of parallel resistors needed to achieve this accuracy was found to be $n = 4$, therefore, and by
using inequalities 5.8 to 5.10, the range of values each resistor in the resistor block was
calculated. The resistance range for each resistor is shown in table 5.1. As discussed before,
since the average value of the range allowed for each resistor gives the maximum margin, the
average value was simply used as the nominal value for each resistor. Note how the
<table>
<thead>
<tr>
<th>Resistor</th>
<th>Minimum (Ω)</th>
<th>Maximum (Ω)</th>
<th>Average (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>56</td>
<td>59</td>
<td>57.5</td>
</tr>
<tr>
<td>R1</td>
<td>374</td>
<td>806</td>
<td>590</td>
</tr>
<tr>
<td>R2</td>
<td>323</td>
<td>323</td>
<td>323</td>
</tr>
<tr>
<td>R3</td>
<td>238</td>
<td>564</td>
<td>401</td>
</tr>
<tr>
<td>R4</td>
<td>262</td>
<td>953</td>
<td>607.5</td>
</tr>
</tbody>
</table>

Table 5.1 Resistors values for an accuracy of ±8.75%.

<table>
<thead>
<tr>
<th>Resistor</th>
<th>Silicide-Block Resistance (Ω)</th>
<th>Switch Resistance (Ω)</th>
<th>Total Resistance (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>57.5</td>
<td>N/A</td>
<td>57.5</td>
</tr>
<tr>
<td>R1</td>
<td>531</td>
<td>59</td>
<td>590</td>
</tr>
<tr>
<td>R2</td>
<td>290.7</td>
<td>32.3</td>
<td>323</td>
</tr>
<tr>
<td>R3</td>
<td>360.9</td>
<td>40.1</td>
<td>401</td>
</tr>
<tr>
<td>R4</td>
<td>546.75</td>
<td>60.75</td>
<td>607.5</td>
</tr>
</tbody>
</table>

Table 5.2 Resistors values including the switches.

Modulating resistors $R_1$ through $R_4$ are order of magnitude higher than the required total resistance ($45\, \Omega$). Since the values in table 5.1 show the total resistance, the resistance of the switches has to be accounted for. Hence, the total resistance is divided between the switch and an actual silicide-block poly resistor, where the switch resistance was designed to be 10% of the total resistance. Table 5.2 shows the resistance of each silicide-block poly resistor along with the switch resistance. The 10% switch resistance limited the linearity error due to the switches to be within ±0.017%. In the implementation of this transceiver, accurate biasing currents were needed anyway, hence, a band-gap circuit and an external resistor $R_{\text{ext}}$ were readily available. Therefore, the output of the band-gap circuit was used as the source of $V_B$, while $R_{\text{ext}}$ was chosen to be 6.3K$\Omega$. The mirroring factor $K$ was simply chosen to be...
unity, which makes the nominal value of \( R_{\text{int}} \) 6.3K\( \Omega \) as well. Each resistor in the resistor block as well as \( R_{\text{int}} \) was checked under nominal conditions and under temperature range from -40 to 125 °C to make sure that the relative magnitude change in the value of each resistor with temperature variations is less than ±8.67%. That was easy to achieve due to the low temperature coefficient of silicide-block poly resistors. The reference generator block resistors were calculated using equation 5.18 with \( R_r \) chosen to be 6.3K\( \Omega \). \( R_{r0} \) was found to be 5.208K\( \Omega \), while the rest of the resistors were found to be all equal to 546\( \Omega \). Simulations were done using different process corners, temperature variations (-40 to 125 °C), and different supply voltages (3.3V ±10%). Statistical simulations were also performed on the whole structure. Statistical simulations take into account the errors due to mismatches between the resistors, current mirror mismatches, the offset introduced by the op-amps and the comparators in the control matrix, and process variations. All the previous simulations were done taking into account a ±1% error in \( R_{\text{ext}} \). Simulation results showed that the total resistance of the resistor block is well controlled within 41.25\( \Omega \) and 48.82\( \Omega \), with a nominal value of 44.8\( \Omega \), which is about ±8.5% from 45\( \Omega \). As mentioned before, those results are including all different variables that the circuit could encounter. This design example was

<table>
<thead>
<tr>
<th>Process</th>
<th>-3( \sigma )</th>
<th>Mean</th>
<th>+3( \sigma )</th>
<th>Error from 45( \Omega )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fast Process</td>
<td>42.55( \Omega )</td>
<td>45.62( \Omega )</td>
<td>48.69( \Omega )</td>
<td>-5.4% to +8.2%</td>
</tr>
<tr>
<td>Nominal Process</td>
<td>42.98( \Omega )</td>
<td>45.88( \Omega )</td>
<td>48.78( \Omega )</td>
<td>-4.5% to +8.4%</td>
</tr>
<tr>
<td>Slow Process</td>
<td>43.18( \Omega )</td>
<td>45.78( \Omega )</td>
<td>48.41( \Omega )</td>
<td>-4.1% to +7.5%</td>
</tr>
</tbody>
</table>

| Table 5.3 | Measured resistance values. | 132 |
implemented on a standard 0.18µm technology, and measurements were performed on the
design for different process corners. The resistance was measured at different voltage levels
up to 70% of the 3.3V supply voltage (∼2.3V), and the maximum error from the resistance
value measured at the middle of this range (35% of the 3.3V supply ∼1.1V) was found to be
within ±0.017%, which demonstrate the high linearity performance of the proposed
algorithm. Table 5.3 shows the measurement results, which align very well with the
simulation results.

The second design example is the implementation of termination resistors for an 8
ports high speed transceiver. The transceiver operates at 1.65Gbps and it requires a pair of
termination resistors of nominally 50Ω each for each port. The transceiver specifications
require a tolerance of no more than ±10% in the value of the termination resistors. The 8
ports that the transceiver has are not in a close proximity, in fact, they were scattered to
different locations on the chip. Therefore, every pair of termination resistors has to be tuned
individually. As discussed before, this requires that each termination resistors pair has to
have its own \( R_{\text{int}} \) as well as its own control matrix, which will consume area and power. An
alternative approach was chosen, which is reducing the value of \( C_{\text{Max}} \), i.e. tune to a tighter
range in order to give more margin to account for the scattered locations of termination
resistors pairs, and use the same \( R_{\text{int}} \) and control matrix for all 8 termination resistors pairs.
In order to achieve that, \( C_{\text{Max}} \) was chosen to be 5% in order to give a 5% margin to account
for any mismatches due the current mirrors as well the scattered locations of the
terminations. Process characterization showed that the maximum change from the nominal
value for on-chip silicide-block poly resistors is ±25% from the nominal value including
variations with temperature as well, which implied that \( D_{\text{Max}} \) should be 25%. As in the
<table>
<thead>
<tr>
<th>Resistor</th>
<th>Minimum (Ω)</th>
<th>Maximum (Ω)</th>
<th>Average (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>63.33</td>
<td>65.63</td>
<td>64.48</td>
</tr>
<tr>
<td>R1</td>
<td>750.06</td>
<td>1467.16</td>
<td>1108.61</td>
</tr>
<tr>
<td>R2</td>
<td>673.94</td>
<td>1366.23</td>
<td>1020.08</td>
</tr>
<tr>
<td>R3</td>
<td>642.62</td>
<td>1420.44</td>
<td>1031.53</td>
</tr>
<tr>
<td>R4</td>
<td>610.00</td>
<td>610.00</td>
<td>610.00</td>
</tr>
<tr>
<td>R5</td>
<td>475.00</td>
<td>1050.00</td>
<td>762.50</td>
</tr>
<tr>
<td>R6</td>
<td>541.59</td>
<td>1685.53</td>
<td>1113.56</td>
</tr>
<tr>
<td>R7</td>
<td>499.74</td>
<td>1544.06</td>
<td>1021.90</td>
</tr>
<tr>
<td>R8</td>
<td>481.91</td>
<td>1609.12</td>
<td>1045.52</td>
</tr>
</tbody>
</table>

Table 5.4  Resistor values for an accuracy of ±5%.

<table>
<thead>
<tr>
<th>Resistor</th>
<th>Silicide-Block Resistance (Ω)</th>
<th>Switch Resistance (Ω)</th>
<th>Total Resistance (Ω)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R0</td>
<td>65</td>
<td>NA</td>
<td>65</td>
</tr>
<tr>
<td>R1</td>
<td>997.2</td>
<td>110.8</td>
<td>1108</td>
</tr>
<tr>
<td>R2</td>
<td>918</td>
<td>102</td>
<td>1020</td>
</tr>
<tr>
<td>R3</td>
<td>928.8</td>
<td>103.2</td>
<td>1032</td>
</tr>
<tr>
<td>R4</td>
<td>549</td>
<td>61</td>
<td>610</td>
</tr>
<tr>
<td>R5</td>
<td>686.7</td>
<td>76.3</td>
<td>763</td>
</tr>
<tr>
<td>R6</td>
<td>1002.6</td>
<td>111.4</td>
<td>1114</td>
</tr>
<tr>
<td>R7</td>
<td>919.8</td>
<td>102.2</td>
<td>1022</td>
</tr>
<tr>
<td>R8</td>
<td>941.4</td>
<td>104.6</td>
<td>1046</td>
</tr>
</tbody>
</table>

Table 5.5  Resistor values including the switches.

previous example, using equation 5.1, the number of parallel resistors in each resistor block needed to achieve this accuracy was found to be $n = 8$, therefore, and by using inequalities 5.8 to 5.10, the range of values each resistor in the resistor block was calculated. The resistance range for each resistor is shown in table 5.4. As in the previous example, since the average value of the range allowed for each resistor gives the maximum margin, the average value was simply used as the nominal value for each resistor. Again, the division between
silicide-block resistors and the switches was chosen to be 10%. Table 5.5 shows the values of the resistors and the switches. As in the previous example, a band-gap circuit and an external resistor $R_{\text{ext}}$ were readily available. Therefore, the output of the band-gap circuit was used as the source of $V_B$, while $R_{\text{ext}}$ was chosen to be $6.3\,\text{k}\Omega$. The mirroring factor $K$ was simply chosen to be unity, which makes the nominal value of $R_{\text{int}}$ $6.3\,\text{k}\Omega$ as well. Each resistor in the resistor block as well as $R_{\text{int}}$ was checked under nominal conditions and the temperature range from -40 to 125 °C to make sure that the relative magnitude change in the value of each resistor with temperature variations is less than ±5%. Again, that was easy to achieve due to the low temperature coefficient of silicide-block poly resistors. The reference generator block resistors were calculated using equation 5.18 with $R_{\text{r}}$ chosen to be $6.3\,\text{k}\Omega$. $R_{\text{r0}}$ was found to be $5.04\,\text{k}\Omega$, while the rest of the resistors were found to be all equal to $315\,\Omega$. The design was implemented on a 0.13μm technology. Simulations were done using different process corners, temperature variations (-40 to 125 °C), and different supply voltages (3.3V ±10%). Statistical simulations were also performed on the whole structure. Statistical simulations take into account the errors due to mismatches between the resistors, current mirror mismatches, the offset introduced by the op-amps and the comparators in the control matrix, and process variations. All the previous simulations were done taking into account a ±1% error in $R_{\text{ext}}$. Simulation results showed that the total resistance of all resistor blocks (16 total) is well controlled within $46.76\,\Omega$ and $53.26\,\Omega$, which is about ±6.5% from $50\,\Omega$. Note that the ±6.5% includes only the short range mismatches between the resistor blocks (assuming they are laid out in close proximity). Taking into account that the resistor blocks are scattered across the chip, a 1% more loss in accuracy is roughly estimated.
LIST OF REFERENCES


CHAPTER 6

EQUALIZATION

The main topic of the previous chapters was the first category of variations and the available adaptive techniques to combat those variations. As discussed in chapter 1, the first category of variations is the circuit level category, i.e. variations encountered during the design of each individual circuit block in a system. Those variations are essentially the variations in the absolute values of the basic devices characteristics (resistors, capacitors, and transistors) due to process variations, supply voltage variations, and temperature variations. This chapter concentrates on the third category of variations, which is the network level category, i.e. variations in the transmission media. As mentioned in chapter 1, one of the major problems with any transmission media used for digital communications is Intersymbol Interference, ISI for short. This chapter will give an overview of the causes of ISI and its effect on the reception quality of digital data. The available techniques for combating ISI, particularly equalization, including fixed and adaptive equalization will be discussed. A comparison between analog and digital implementations of equalization techniques will be presented along with the advantages and disadvantages of both methodologies. Chapter 7 will present the implementation of an analog adaptive equalizer for wire line transceivers.
6.1. Intersymbol Interference

ISI has been traditionally a serious limitation on the data rates that can be sent across a specific communication channel [1-4]. ISI generally refers to the interference that occurs between the current received bit and other previously received bits in the same data stream due to the addition of portions of the previously received bits to the bit that is being received. In the simplest form, the front end digital receiver is just a comparator that compares the received amplitude to a certain threshold to decide if the received bit is a 1 or a 0, therefore if the portions added from the previous bits to the current bit are enough to change the current bit from a 1 to a 0 or vice versa, then ISI could significantly affect the bit error rate of the whole receiver. On the other hand, if the portions added from the previous bits to the current bit are not enough to mislead the receiver, then theoretically it should be tolerable. However, this is not true all the time due to the limitations on the design of the front end receiver itself as will be discussed later. In fact, ISI could still be a potential problem even if it is not enough to change the current received bit. In order to understand the nature of ISI, its causes have to be clearly explained. As mentioned before, there has to be a way for portions of previously received bits to get to the current received bit. In wireless networks, the multi-path effect shown in figure 6.1 is a straightforward cause of ISI. Basically, multiple indirect delayed versions of the transmitted signal (that usually results from multiple reflections) interferes with the direct signal received by the antenna causing the current received symbol to be contaminated with previous symbols, which is essentially ISI. Luckily though, since those delayed versions of the transmitted signal are a result of reflections, they travel for a longer distance than the direct received signal before they get to
the receiver's antenna. Therefore, they are usually attenuated. If the current received symbol is $X_{r_i}$, then it can be represented by:

$$X_{r_i} = \alpha_i X_{r_i} + \alpha_{i-1} X_{r_{i-1}} + \alpha_{i-2} X_{r_{i-2}} + \ldots$$  \hspace{1cm} (6.1)

where $X_{r_i}$ is the transmitted symbol, and $\alpha_i$ is the associated attenuation factor due to the channel path. Equation 6.1 can be rewritten more generally using the Z domain as:

$$X_r(Z) = \left(\alpha + \alpha_1 Z^{-1} + \alpha_2 Z^{-2} + \alpha_3 Z^{-3} + \ldots\right) X_T(Z)$$  \hspace{1cm} (6.2)

Where $X_r(Z)$ and $X_T(Z)$ are the received and transmitted signals respectively. An important thing to notice from equation 6.2 is that the transfer function between the transmitted and the received signals is a low pass function [3].

In wire line communications, there is really no multi-path effect since there is only one path for the signal between the transmitter and the receiver, yet ISI could also happen due to a completely different phenomenon, which dispersion. In order to elaborate on this,
let's assume that the wire line channel could be modeled as a simple RC network, which essentially represent a low-pass filter. Furthermore, let's assume that the channel has a much smaller bandwidth than the digital signal that is being sent across it and that the digital pulses are simple gate pulses. That is, a 1 is represented by a pulse with an amplitude $a$ and a duration of $T_b$, while a 0 is represented by a pulse with an amplitude $-a$ and the same duration. Since an RC filter is a linear filter, its response for a train of pulses could be studied by adding up its response to each individual pulse. Figure 2 shows the response of a band-limited channel to a single pulse. As shown in figure 2, the response of the channel to such a pulse involves two effects. First, the maximum amplitude the output pulse can reach is less than the original amplitude of the input pulse. Second, the output pulse becomes dispersed and its duration extends beyond $T_b$. The rising edge of the output pulse could be written as:

$$V_O(t) = a \left( 1 - e^{-t/RC} \right) \quad \text{for } t < T_b$$

Using equation 6.3, the maximum voltage in the output pulse could be found by replacing $t$ with $T_b$, that is:

$$V_O(T_b) = a \left( 1 - e^{-T_b/RC} \right)$$

On the other hand, using equation 6.4, the falling edge of the output pulse could be written as:

$$V_O(t) = a \left( 1 - e^{-T_b/RC} \right) e^{-t/RC} \quad \text{for } t > T_b$$

As equations 6.3 to 6.5 show, the output pulse is being spread out in time beyond its allocated time period $T_b$, therefore, when another pulse is sent out across the channel, the residue from the previous pulse represented by equation 6.5 is going to contaminate the new received pulse, and potentially mislead the front end receiver [5].
The severity of ISI introduced by the low-pass nature of the communication channel is a function of two factors. The first factor is the bandwidth of the channel, while the second factor is the density of transitions in the data stream being sent across the channel, i.e. the bit pattern. Therefore, ISI is usually categorized as a deterministic jitter since it could be related to the bit pattern. In order to clarify the effect of both factors, the following experiment could simply be performed. Let's assume that the data rate of the transmitted data is 500 Mbps, i.e. the bit duration is 2 ns. Let's also assume that a 1 is a gate pulse with amplitude 1V, while a 0 is a gate pulse of amplitude -1V. Furthermore, let's assume we have 3 different channels, channel 1 has a bandwidth of 250 MHz, channel 2 has a bandwidth of 125 MHz, and channel 3 has a bandwidth of 62.5 MHz. Let's also assume the following three bit patterns: The first pattern is an alternating 1-0 pattern, i.e. a 1 followed by a 0 that repeats periodically, the second pattern is 1-1-1-0 that also repeats periodically, and the third pattern is 1-1-1-1-1-1-0 that also repeats periodically. Figures 6.3-6.5 show the output

![Figure 6.2 The response of a band-limited channel to a single pulse.](image-url)
response of the 3 channels for the first bit pattern, while figures 6.6-6.8 show the output response of the 3 channels to the second bit pattern, while figures 6.9-6.11 show the output response of the 3 channels to the third bit pattern. First off, let’s compare figures 6.3, 6.4, and 6.5, which represent the response of all three channels to the first bit pattern. As shown in figure, the effect of the channel bandwidth is mainly on the amplitude of the received bits, while the duration is not affected. This should come of no surprise since if the input signal is periodic, then the output signal will have to be periodic with the same period as well, and since the 1 and the 0 bits have the same duration, then they will appear as such at the output too. Therefore, for an alternating 1-0 pattern, the effect of ISI appears only as attenuation in the amplitude of the received bits, but not on their time duration. This is an important result

![Figure 6.3](image-url)  
**Figure 6.3** Response of channel 1 to the first input pattern.
Figure 6.4 Response of channel 2 to the first input pattern.

Figure 6.5 Response of channel 3 to the first input pattern.
**Figure 6.6** Response of channel 1 to the second input pattern.

**Figure 6.7** Response of channel 2 to the second input pattern.
Figure 6.8 Response of channel 3 to the second input pattern.

Figure 6.9 Response of channel 1 to the third input pattern.
Figure 6.10 Response of channel 2 to the third input pattern.

Figure 6.11 Response of channel 3 to the third input pattern.
since it can be used to differentiate between ISI and different kinds of jitter at the output of a specific channel. Specifically, it is a common practice to send an alternating 1-0 pattern across a communication channel in order to exclude ISI from the jitter measurements. Second, by comparing figures 6.4, 6.7, and 6.10, which represent the response of channel 2 to all three different patterns, it is obvious that the 0 bit has a different amplitude and duration even though it is going through the same channel. It is also obvious that the longer the identical series of 1s before the 0 is, the worse the effect of ISI on the amplitude and duration of the 0 bit is. This also should come of no surprise since the longer the input stays at constant amplitude, the more the channel charges to that amplitude, which will make it more difficult to discharge when the input switches. Therefore, ISI starts to affect the duration of the received bits only if the input pattern is not symmetrically alternating as in the first bit pattern. Third, by comparing figures 6.6, 6.7, and 6.8, as well as 6.9, 6.10, and 6.11, which represent the response of the 3 channels to both the second and third bit patterns, it is obvious that the lower the channel bandwidth is, the worse the effect of ISI on the received bits amplitude and duration, given that the input bit pattern is not a symmetric alternating pattern. This is also expected since the slower the channel is, the more dispersed the output pulses get, and consequently the more they interfere with each other. In fact, as shown in figure 6.11, the effect of ISI on the output pattern is so severe that the 0 bit does not even reach the threshold of the front end receiver, which will essentially be received as 1 instead of a 0.

To summarize the previous discussion, there are two conclusions. The first conclusion is that as the channel bandwidth gets lower, the effect of ISI becomes higher on both the amplitude and duration of the received bits, with the exception of symmetric
alternating patterns, where the effect of ISI appears only on the amplitude of the received bits but not their duration. The second conclusion is that as the series of identical bits before a complementary bit is sent gets longer, the effect of ISI on the amplitude and duration of the received complementary bit becomes higher. One way to fight that is bit stuffing, which is used in NRZ transmission schemes, where a complementary stuffing bit is inserted when the data stream includes a series of identical 1s or 0s that are longer than a specific length. Bit stuffing is very important in facilitating the clock recovery from NRZ data streams, as well as to minimize the effect of ISI.

6.2. Eye Diagrams

As discussed in the previous section, ISI is a function of the bit pattern being sent across the channel. If the input pattern is very long and random, it becomes a very difficult task to find out the effect of ISI on both the amplitude and duration of the received bits. A good tool to assess ISI (and jitter in general) in the received data stream is eye diagrams [3]. Eye diagrams are created by sending a continuous random data stream and then capturing the output waveform. The captured waveform is then divided to sections with duration equal to the ideal bit period, and then those sections are overlaid on top of each other to create an eye diagram. The two important parameters in an eye diagram are the vertical and the horizontal openings of the eye. The vertical opening represents the minimum amplitude the received bits can have. It usually defines the sensitivity of the front end receiver, i.e. the minimum amplitude the receiver has to be able to detect, which by definition is the
minimum gain of the receiver. The horizontal opening on the other hand represents the minimum duration the received bits could have. Hence, it usually defines the speed of the receiver, i.e. the minimum bit width the receiver has to be able to detect. In order to illustrate the horizontal and vertical openings of an eye, figures 6.12-6.14 show the three eye diagrams that result from sending a long random 500Mbps data stream across channels 1, 2, and 3 defined earlier in the previous section. The ideal horizontal eye opening such a data stream should be 2ns, while the ideal vertical opening should be 2V. As shown in the figure, the first eye diagram that represents the response of channel 1 has the widest vertical and horizontal openings since it has the highest bandwidth. As the bandwidth gets lower, both the vertical and horizontal eye openings gets smaller. At the extreme case of channel 3 shown in figure 6.14, the eye gets completely closed, which essentially means that ISI was

![Eye diagram of a random data stream across channel 1.](image)

**Figure 6.12** The Eye diagram of a random data stream across channel 1.
Figure 6.13 The Eye diagram of a random data stream across channel 2.

Figure 6.14 The Eye diagram of a random data stream across channel 3.
sever enough that some bits could not even cross the threshold of the front end receiver and therefore will be received incorrectly.

In theory, as long as there is a vertical and horizontal eye openings, then it is possible to detect the received bits correctly since ISI was not sever enough to prevent the bits from reaching the threshold of the front end receiver. If there is no eye opening at all as shown in figure 6.6c, then obviously the received data stream can not be received correctly without further processing. Unfortunately though, this is only true to a limit. The reason behind that is that ISI affects both the vertical and horizontal openings simultaneously, which essentially means that when ISI becomes more severe, the front end receiver has to achieve higher gain and higher speed at the same time. Since it a well known fact that the gain and the bandwidth of a comparator (a digital receiver) are two contradictory parameters, i.e. for the same power consumption, higher gain means lower bandwidth and vice versa [6-9]. Therefore, for very small eye openings –even though there is an opening- it becomes very difficult to design a front end receiver with acceptable power consumption. At high data rates, this problem becomes even more dominant.

Whether the eye is completely closed or the vertical and horizontal openings are too narrow, further processing is needed to widen the eye opening in the received signal, i.e. compensate or eliminate the effect of ISI. The technique used to combat the effect of ISI is referred to as equalization since it equalizes the effect of the transmission channel to enable the receive path to achieve higher bit error rates. In the following section an overview of equalization techniques and architectures will be presented.
6.3. **Equalization Architectures**

Whether ISI is introduced through multi-path effects or through dispersion due to the limited bandwidth of the transmission media, the transfer function between the transmitted and received data streams could be modeled as a bandwidth limited low pass transfer function [3]. In other words, for a given data stream, ISI results from the unequal attenuation and delay the different frequency components of the transmitted signal suffer from due to the low pass nature of the transmission channel. In fact, the relatively higher attenuation that the high frequency components of the transmitted signal suffer from when compared to the lower frequency components is the main reason for the degradation in the horizontal eye opening of the received signal since those higher frequency components are the ones that represents the fast rising and falling edges of the transmitted signal. Therefore, when they get attenuated, the received signal losses those fast rising and falling edges. Given the explanation above, it becomes obvious that the straight forward cure for ISI is the design of the front end receiver such that it relatively enhances the high frequency components of the received signal in order to compensate for the higher attenuation they suffered from due to the low pass nature of the transmission channel. In other words, the front end receiver needs to have a high pass transfer function with a unity DC gain to compensate for the low pass nature of the channel [1]. Essentially, what this enhancement does is that it increases the effective bandwidth of the transmission channel. In order to illustrate the idea, figure 6.15 shows the transfer function of a low pass transmission channel, the transfer function of a high pass front end receiver, and their combined frequency response. As shown in figure, the effective bandwidth of the transmission channel with the high pass filter is almost
doubled. This process of enhancing the high frequency components of the transmitted signal is what is referred to as equalization. Figure 6.15 though leaves us with an obvious requirement, which is the pre-knowledge of the transmission channel in order to design the front end receiver with the correct frequency response. Otherwise, over enhancement of the higher frequency components will cause the addition of more ISI instead on fixing it. Those kinds of equalizers are usually called fixed equalizers since their frequency response is fixed and optimized for a specific transmission channel. The pre-knowledge of the transmission media is high unlikely though. For example, wireless communication channels depend on the different paths of the signal, weather conditions, obstacles or buildings in the way, etc...

**Figure 6.15** The effect of enhancing the high frequency components of the received signal on the effective bandwidth of the transmission channel.
Wired transmission channels also have similar problems, i.e. Cable length, connectors' types, temperature.....etc. This makes it very difficult to predict the frequency response of the transmission media, and therefore using a fixed equalizer might not be adequate. For the above reasons adaptive equalizers are widely employed, in which the frequency response of the front end receiver is automatically adapted to the transmission channel. There are many different equalization architectures that could be implemented in the analog domain (continuous or discrete time), or in the digital domain. In the next few sections, an overview of the different equalization architectures will be presented.

6.3.1. The Feed-Forward Architecture

The first equalization architecture is the feed-forward architecture shown in figure 6.16. This architecture is simply based on the fact that if the total transfer function between the transmitter and the receiver has enough bandwidth, then the transmitted waveform will retain its shape and go through without any ISI [3]. As mentioned earlier, in order to achieve that, the front end receiver has to have some sort of a high pass nature with a unity DC gain to compensate for the low pass nature of the transmission channel, hence restoring the transmitted signal shape and pass it to the digital receiver (the slicer). Therefore, the feed-forward architecture shown in figure 6.16 uses one high pass filter as its front end receiver, and it is usually called the Forward Equalizer (FE). An important aspect of this feed-forward architecture is that it only tries to remove the effect of ISI by restoring the original shape of the received waveform by enhancing the speed of the transitions (high-pass response).
Therefore, it does not use any knowledge of the previously received bits in the data stream to remove ISI, i.e. there is no feedback. This is the main reason for calling the forward equalizer the ISI precursor remover [10].

The feed-forward architecture has the major advantage of restoring the shape of the transmitted waveform, or in other words, it widens up the horizontal eye opening of the received signal. As discussed in the previous section, this essentially makes the design of the digital receiver (the slicer) much easier. The other advantage of the feed-forward architecture is that it does not require the knowledge of any previously received bits to determine the current received bit due to its open loop nature, which essentially enables the de-multiplexing of the data at the input of the receiver to allow faster data rates [11]. For example, if the received data is de-multiplexed into $n$ parallel data lines with each line running at only $\frac{1}{n}$ the data rate, then by using $n$ parallel forward equalizers (one for each line) and then multiplexing the outputs, the data stream could be equalized but with each equalizer running only at $\frac{1}{n}$ the data rate. This essentially enables higher data rates to be
equalized without requiring each equalizer to be as fast as the data rate. It is important to note here that other equalization architectures that rely on the knowledge of previously received bits do not have this feature since de-multiplexing the data will essentially alter the data structure. In the next section, a more detailed discussion of the lack of the de-multiplexing feature in feed-back architectures will be presented.

The major disadvantage on the other hand of the feed-forward architecture is its relatively poor noise performance [10]. This stems from the fact that the high pass nature of the forward equalizer not only restores the high frequency components of the transmitted signal, it also enhances high frequency noise as well. This enhanced high frequency noise can significantly degrade the bit error rate and potentially defeat the whole purpose of removing ISI. Therefore, the feed-forward architecture is not usually an adequate choice in highly noisy environments. Even though this is the case, there exist some techniques in the literature that tries to solve the high frequency noise boosting in feed-forward architectures. One technique implements the forward equalizer as an all-pass filter instead of simply a high pass filter. In that case the forward equalizer is usually called an all pass forward equalizer or APFE for short [10, 12-15]. In an all pass filter, the DC gain and the high frequency gain are the same, as opposed to a higher high-frequency gain in the simple high pass filter case. In APFEs, the location of the zeros is chosen to be at a much lower frequencies than the poles in order to enhance only a specific frequency band, while at higher frequencies, when the poles start to have an effect, they stop the enhancement caused by the lower frequency zeros, which effectively, stops the out-of-band high frequency noise enhancement. In fact, ideally the APFE should be designed such that when the input is white noise, the output is also white. Figure 6.17 shows the frequency response of a low-pass transmission channel, an
all-pass forward equalizer, and the combined response of the channel and the equalizer. While APFEs has the advantage of suppressing the high frequency noise enhancement, they add more ISI to the received signal since they simply introduce more poles to the whole transfer function of the transmission channel. So essentially, they improve the total SNR on the expense of adding more ISI.

In the feed-forward architecture shown in figure 6.16, the frequency response of the forward equalizer is optimized for a specific transmission channel. Therefore, it is referred to as a fixed equalizer. In order to equalize for different channels, the adaptive feed-forward architecture shown in figure 6.18 should be employed. In this architecture, the output of the
the forward equalizer is subtracted from the output of the digital receiver (the slicer) -which is assumed to give the right decision- to give an error signal. This error signal is then used to change the frequency response of the forward equalizer to give the minimum error possible.

The theory behind this adaptive architecture is that by assuming that the output of slicer is correct, its output waveform should be identical to the original transmitted waveform. Therefore, once the difference between the output of the slicer and the output of the forward equalizer is minimized, the forward equalizer will be optimized to the transmission channel.

**Figure 6.18** The adaptive feed-forward equalization architecture.
6.3.2. The Feed-Back Architecture

The feed-back architecture is shown in figure 6.19. As mentioned before, ISI is caused by the spread of each individual bit beyond its allocated time, and hence interfering with the later bits being sent. The feed-forward architecture fixes ISI by undoing the dispersion in the received bits by improving the effective bandwidth of the transmission channel between the transmitter and the slicer using the forward equalizer (a high pass filter). The feed-back architecture uses a different approach. Essentially, since the current received bit has been contaminated by portions of the previous bits in the data stream (whether due to dispersion or multi-path effect), then if those previous bits are known, their contribution to the current received bit can be estimated and then negated from the current received bit to remove ISI. As shown in figure 6.19, a filter is used in the feedback path between the

![Figure 6.19 The feed-back equalization architecture.](image_url)
output and the input of the slicer to store and calculate the contributions of the previously received bits—which are assumed to be correct—and then subtract it from the current received bit at the input of the slicer. The filter used in the feedback path is usually referred to as the Decision Feedback Equalizer or DFE for short. Since in this architecture ISI cancellation is done by the knowledge of the previously received bits, the DFE is usually called the postcursor ISI remover[10]. The type of the DFE filter is usually dictated by the nature of the transmission channel since it has to have the same response as the channel in order to produce the same ISI effect in order to negate it from the received waveform. As in the feed-forward architecture, the feed-back architecture shown in figure 6.19 is considered a fixed equalization architecture since the DFE is optimized for a specific channel. Figure 6.20 shows the adaptive version of the feed-back architecture. In this architecture, the adaptive algorithm compares the input and the output waveforms of the slicer and generates an error signal. By minimizing the difference between the two waveforms, the DFE frequency response will adapt to the transmission channel.

The major advantage of the feed-back architecture is that it does not boost the high frequency noise as opposed to the feed-forward architecture [16]. Therefore, it generally has a better performance than the feed-forward architecture in noisy environments. The feed-back architecture is also more effective in equalizing channels with spectral nulls (such as some bad radio channels) and in combating ISI in time variant multi path channels [3]. The disadvantage of the feed-back architecture on the other hand is that although postcursor ISI could be removed, yet the waveform at the input of the slicer was not restored since the high frequency components were not boosted due to the lack of a FE (as opposed to the feed-forward architecture). This makes it more difficult for the slicer to make the right decision.
Another disadvantage of the feed-forward architecture is that in order for the whole system to give the best performance, the total response of the channel has to be raised cosine response, which will require the use of a matched filter at the input [17]. Another disadvantage for the feed-back architecture as opposed to the feed-forward architecture is the data rate limitation. The DFE function is based on calculating the ISI contributed by previously known bits received right before the current bit to remove that contribution from the current bit. This means that the DFE has to finish this calculation and subtraction in a one-bit interval, which means that it has to go as fast as the data rate. This limits the data rates the feed-back architecture can handle. As mentioned in the feed-forward architecture,
this limitation could be overcome by de-multiplexing the received data to $n$ parallel lines and using $n$ parallel forward equalizers operating at $\frac{1}{n}$ the data rate. In the feed-back architectures though, de-multiplexing is not possible. The reason behind that is the theory of operation of the DFE itself. Since the main source of ISI in the current received bit is coming from the immediate previous bits received right before the current bit, the bits received long time before the current bit will have very little or no effect on the current bit. Therefore, if de-multiplexing is used, each parallel DFE in each equalizer will be storing previous bits that are $n$ bits apart from each other and from the current bit due to the de-multiplexing process. This means that those stored bits have very little or no contribution to the ISI in the current bit the equalizer is handling. Hence, the negated values supplied by the DFE will be incorrect and will cause more ISI instead of eliminating it [11].

6.3.3. The Mixed Feed-Forward Feed-back Architecture

This architecture is shown in figure 6.21. As shown in the figure, the architecture uses both the feed-forward technique (FE) and the feed-back technique (DFE) simultaneously. Generally, this architecture is used for the severely distorted channels to give better equalization. The FE removes the precursor ISI while the DFE removes the postcursor ISI. It is worth mentioning that the mixed feed-forward feed-back architecture will suffer from the disadvantages of both the feed-forward and feed-back architecture. The adaptive version of this architecture is shown in figure 6.22 and it works the same way as in the two previous architectures.
**Figure 6.21** The mixed feed-forward feed-back equalization architecture.

**Figure 6.22** The adaptive mixed feed-forward feed-back equalization architecture.
6.4. Equalization Implementation Techniques

In the previous section, the different architectures used for equalization was presented. Given the architecture, there are different possibilities for implementing the architecture from the circuit level perspective. Equalization architectures –specifically adaptive architectures- heavily make use of programmable filters. Those filters could be implemented using purely analog techniques (continuous-time or discrete-time), or digital techniques using a digital signal processors. Analog and digital implementations could be compared on many different levels. From a power consumption perspective, analog implementations consume less power since there is no need for an analog to digital converter (ADC). ADCs relatively consume significant power. In addition to saving the ADC power in analog implementations, analog implementations also do not require a clock, which will save the power used to generate and distribute those clocks among the different circuits. From an area perspective, analog implementations consume less area for the same reasons mentioned above. From a complexity perspective, analog implementations are simpler than the digital counterparts since the blocks used in the system could be implemented in the analog domain with smaller circuits that involve fewer transistors. On the other hand, the digital implementations -which include clocks and much bigger circuits - are usually more complex than the analog counterpart. It is worth mentioning that complexity and difficulty of the design are two different issues. From a difficulty perspective, even though the analog implementations are simpler and smaller than the digital ones, the analog implementations are more difficult to design. For example, to design a filter in the analog domain, process and temperature variations, accuracy of the components used, and
stability and noise issues become much more dominant than in the digital implementations. This makes the design in the analog domain very difficult, yet small and simple. The digital implementations on the other hand, even though are bigger and more complex, but they are straight forward. For example to design a filter in the digital domain, it is a matter of calculating the filter's coefficient, and then automated CAD tools are used to generate the circuits. From a speed point of view, analog implementations do not need to sample any signals. Therefore, the used circuits have to be only fast enough to handle the maximum frequency of the received signal. In the digital implementations though, the received signal has to be sampled and then converted to the digital domain. In order to keep the information in the signal, the sampling rate has to be much higher than the maximum frequency component in the received signal. Hence, any circuits used afterwards have to be as fast as the sampling rate. This puts a lower limit on the data rate a digital implementation can handle as opposed to an analog implementation. From a noise immunity point of view, it is well known that digital implementations are more immune to noise than the analog counterparts. It is worth mentioning though that digital and discrete time implementations add quantization noise to the circuit due to the sampling process.

In this chapter, an overview of the problem of ISI in digital communication systems has been presented. Equalization architectures used to combat ISI have also been reviewed along with the advantages and disadvantages of analog and digital implementations of those architectures. In the next chapter, the implementation of an analog adaptive equalizer for long haul wire line communication systems is going to be presented. The adaptive equalizer is used to equalize for unshielded-twisted-pair category 5 cables (UTP-5 cable) with a length varying from 0.5 to 100 meters at 125 Mbps data rate.
LIST OF REFERENCES


CHAPTER 7

AN ANALOG ADAPTIVE EQUALIZER FOR WIRE LINE TRANSCEIVERS APPLICATIONS

Intersymbol Interference is a major hold-back on the data rate that could be sent across a band limited transmission channel. As discussed in chapter 6, when sending high speed digital signals over a band limited transmission media, the high frequency components of the transmitted signal get attenuated more than the lower frequency components due to the low pass nature of the transmission media. This unequal attenuation of the high frequency components causes the shape of the transmitted pulse to be distorted and spread in time beyond its allocated time frame. It will also cause the rising and falling edges of the received pulse to be very slow. This spread in the transmitted pulse will cause interference between the current received pulse and previously sent pulses causing the receiver to make the wrong decision about the received pulse. Equalization is the technique used to solve the Intersymbol Interference problem. An Equalizer is essentially a system that compensates for the unequal attenuation caused by the transmission media i.e. Enhance the high frequency components of the received signal. An adaptive equalizer is an equalizer that can adapt its frequency response (the high frequency boosting) for transmission channels that
have unpredicted frequency response. In chapter 6, different equalization architectures have been reviewed including fixed and adaptive architectures along with the pros and cons of each of them. Advantages and disadvantages of analog versus digital implementations were also discussed. In this chapter, a purely analog adaptive equalizer based on the feed-forward architecture will be presented. The architecture uses a two stage tunable high-pass analog filter (the boosting filter) as its forward equalizer (FE). The boosting filter poles locations are programmable and could be changed using a control signal. By adjusting this control signal, the frequency response of the boosting filter could be optimized for the frequency response of the transmission channel. The adaptation to the transmission channel is done through comparing the edge rate of the received signal with a standard edge rate to generate an error signal. This error signal indicates the amount of boosting needed from the tunable boosting filter, and is used to generate a control signal that adjusts the frequency response of the filter. The architecture does not require a training signal to perform the adaptation since it is done based on the edge rate of the received data rather than a specific bit pattern. The architecture has been implemented on a standard 180nm digital CMOS process with 1.8V supply voltage, and has been used for a wire line digital transceiver application that operates at 125 Mbps over category 5 (CAT 5) unshielded twisted pair (UTP) cable. The length of the used cable varied from 10 cm to 100 m.
7.1. Motivation

The motivation behind this work emerges from the need of a solid cheap solution for home networking. Even though Ethernet provided a solid, yet cheap solution for data networking through homes, there has been a recent increase in the demand of a home network that can handle video and audio applications in addition to data. The ultimate goal of a home network is be able to get access/control over all audio, video and data equipments in the whole house from any location and to any location in the house through a single cable/connector that can handle all different types of data. The great advantage of Ethernet is its ability to receive and transmit data across the super cheap UTP cable for lengths up to 100m. This makes the Ethernet solution very appealing to home owners due to the cheap cost of wiring the whole house with UTP cables. Yet, Ethernet protocols were not designed to effectively handle audio/video data streams. The reason behind that is that on the hardware level, Ethernet protocols do not handle isochronous data streams. Therefore, it does not guarantee a certain bandwidth to a specific application. For data transmission that could be acceptable, but for real time audio/video, this could cause a problem since real time audio/video streams requires a guaranteed bandwidth. There are different protocols that have been developed to handle isochronous data streams, like IEEE 1394 and USB standards [1-3]. While those standards can handle real time audio/video traffic, they suffer from two major disadvantages. First, they require relatively expensive shielded cables for their data transmission. Therefore, wiring the whole house could be relatively expensive. Second, the limited cable length they can handle. For example, IEEE 1394 and USB transceivers can only transmit and receive over 4m of cable. Since wiring a house will
definitely require longer cables, using IEEE 1394 or USB transceivers is not possible for home networking. The work presented in this chapter was developed to enable high speed transceivers in general to transmit and receive over UTP CAT 5 cables (which is used for Ethernet networks). The presented work helps using standards that are more adequate for real time audio/video applications over the cheap UTP cables used for home networks. In order to achieve that, adaptive equalization was essential since the cable length could be anywhere between practically 0 to 100 meters.

7.2. Transmission Channel Modeling

The first step in any equalization process is the estimation of the transmission channel behavior. This step is very important in determining the amount of equalization needed, i.e. the amount and the frequency range of boosting the forward equalizer need to provide. While adaptive equalization architecture should automatically adapt based on the transmission channel, it is essential to have some sort of worst-case and best-case channel model in order to define the required range of equalization. In wireless channels, this could be a very difficult task since wireless channels are subject to many different factors as mentioned before. In wire line channels on the other hand, the modeling process is relatively much easier. This is due to the fact that most wire line channels (cables) are controlled by specific standards that the manufactures has to abide to. This makes the prediction of the behavior of the channel much simpler. The most commonly used model for wire line channels is the lumped RLC transmission line model [4, 5]. Even though cable models that
are based on S-parameters measurements give much more accurate modeling of the channel, still the RLC models are much easier to use and simulate with acceptable levels of accuracy. For the equalizer implementation, the three section RLC transmission line model shown in figure 7.1 was used to model a 100m of UTP CAT5 cable. This model represents the worst case attenuation that the equalizer has to deal with. In general though, it is well known that when transmitting data across long range cables, the DC component of the transmitted signal has to be blocked to avoid any significant ground shifting from causing any safety hazards. In addition to that, a termination network that matches the characteristic impedance of the cable (100Ω differential) has to be used to avoid any reflections on the line. For that purpose, the whole transmission channel including the cable model shown in figure 7.1 will be as shown in figure 7.2. It's also worth mentioning that for very short cables, the model shown in figure 7.2 can be used -with an acceptable accuracy- by removing the cable model and just leaving the termination networks. From now on, this will be referred to as the 0m cable model, or the best-case transmission channel.

In order to assess the behavior of the transmission channel, figure 7.3 shows the frequency response of the transmission channel shown in figure 7.2 for both 100m and 0m cables. The bandwidth of the 100m channel is at 1.75 MHz -which is almost 100 times less than the targeted data rate of 125 MHz-, while the 0m channel bandwidth is at 280 MHz. In the time domain, eye diagrams are usually the best tool to assess the effect of the limited bandwidth of the transmission channel on the transmitted data. Figures 7.4 and 7.5 show the eye diagrams of a random data stream at the input and the output of a 100m transmission channel. As shown in figures 7.4 and 7.5, the input eye is ideal, while the output eye diagram is completely closed and equalization is necessary to recover the transmitted data correctly.
Figure 7.1 An RLC transmission line model for 100m UTP CAT 5 cable.

Figure 7.2 The transmission channel including termination networks.
Figure 7.3 The frequency response of 100m and 0m transmission channel.

Figure 7.4 The eye diagram of the 100m transmission channel input.
The Equalization Architecture

The adaptive equalizer architecture is shown in figure 7.6 [6]. As mentioned previously, it is based on the feed-forward architecture discussed in chapter 6. The main part of the equalizer is the forward equalizer (the boosting filter), which represents a tunable high-pass filter that can be programmed based on the transmission channel behavior, and is used to boost the high frequency components of the received signal. The automatic tuning loop provides the control voltage to the forward equalizer to control the location of its poles and consequently the amount of the high frequency enhancement. The equalized output of
the forward equalizer then goes to a high speed comparator (the slicer), which resolves the
differential signal coming out of the forward equalizer and converts it to two fully digital
complementary signals with fast rise and fall times. Those two complementary signals are
essentially the received data. In order for the automatic tuning loop to determine if the high
frequency boosting provided by the forward equalizer is enough, it needs to compare the
output of the forward equalizer with a reference signal that has an acceptable rise and fall
times. The result of this comparison generates an error signal that can be further processed

Figure 7.6 The adaptive equalizer block diagram.
to generate a control signal that keeps adjusting the frequency response of the forward equalizer to the optimum high frequency boosting. In order for this comparison to be accurate, the reference signal has to have the same common mode and differential voltage levels as the output of the forward equalizer. This way, the comparison will be only limited to the rise and fall times. Since the two complementary fully digital outputs of the slicer represent the desired output of the whole system, i.e. with fast rise and fall times, they could be used to provide the reference signal needed to perform the comparison explained above. These two signals though are digital rail to rail signals, and therefore they can't be used directly. Instead, they are used along with a reference signal generator block to generate two differential signals with a differential amplitude voltage level that is equal to the differential output level of the forward equalizer only with faster rise and fall times. These two differential signals are used as the reference signal mentioned earlier. In addition to that, the reference signal generator block provides a common mode voltage level (which is the reference signal common mode) that is used in all the common mode feedback circuits in the system to keep the common mode level of all the fully differential circuits (including the forward equalizer) the same as the reference signal. In order to compare the rise and fall times of the reference signal and the output of the forward equalizer, a band-pass filter is used for each signal to pick the high frequency components in both signals. Those high frequency components give an indication on how fast the rise and fall times are, or in other words, it is like measuring the slopes of the signals. Since the parameter of interest is the edge rate rather than its direction (rising or falling edge), the output of each band-pass filter is squared. The reason behind that is that the band-pass filter will give a positive output for a rising edge and a negative output for a falling edge, which will average out to zero and won't
give a correct indication for the edge rate. By squaring the band-pass filter's output, a correct indication of the edge rate could be estimated regardless of the direction of the edge. The squared outputs of the two band-pass filters are then subtracted to generate an error signal. The sign of this error signal indicates whether the forward equalizer is over boosting or under boosting the high frequency components of the received signal. Therefore, if the rising and falling edges of the forward equalizer output are slower than the reference signal, the error signal will have a positive sign, which indicates under boosting. On the other hand, if the rising and falling edges of the forward equalizer output are faster than the reference signal, the error signal will have a negative sign, which indicates over boosting. Optimum boosting will be reached once the error signal converges to zero. The differential difference squarer block shown in figure 7.6 performs the squaring and the subtracting of the outputs of the two band-pass filters to generate that error signal. Since the output signals of the band-pass filters will only have energy during transitions in the input signals and zero otherwise, the differential difference squarer output will also have energy during transitions only. This means that the error signal will have very high frequency components that could potentially cause instability of the whole adaptive loop if it was used directly to control the forward equalizer. In order to remove those high frequency components from the error signal, an integrator is used as a low pass filter to remove the high frequency components, and to also average the error signal over time. The integration process results in a smoothly changing signal that can be used to control the forward equalizer frequency response. The output of the integrator (the control signal) will keep increasing or decreasing as long as the average of the output of the Differential Difference Squarer is positive or negative respectively. Once the optimum control voltage is reached (when the speed of the forward
equalizer output and the reference signal is the same, i.e. zero error signal), the integrator holds this value. This automatic adaptive loop enables the whole system to continuously adapt itself to any changes in the transmission media during normal operation. In this section, the system level operation of the whole equalizer has been discussed. In the following few sections, the circuit design details of each block in the system will be presented.

7.3.1. The Forward Equalizer

The main part of the whole system is the forward equalizer, which will be also referred to as the boosting filter since it boosts the high frequency components of the received signal. The boosting filter is a cascade of two first order filters, where each filter has a single pole and a single zero in its transfer function. The location of the zero in each filter is fixed, while the location of the pole is variable and controlled by a control voltage. By increasing the control voltage, the position of the pole in each filter shifts to higher frequencies allowing more boosting introduced by the zero in the transfer function, while by decreasing the control voltage, the location of the pole in each filter shifts to lower frequencies to reduce the boosting introduced by the zero. In all cases, the location of the pole in each filter is always kept at a higher frequency than the zero in order to allow boosting. The forward equalizer design uses two cascaded fully differential first-order $G_{m}$-C filters. In each filter stage, two identical transconductors, two matched resistor, and two
matched capacitors are used to implement the filter. Since the design is fully differential, a common mode feedback circuit that keeps the common mode level at a constant value is used. Figure 7.7 shows the circuit diagram of the first-order $G_m$-C filter used.

The forward equalizer uses two cascaded stages of this $G_m$-C stage. The transfer function of each filtering stage could be written as:

$$\frac{V_o}{V_i} = \frac{G_m (S \times 2C + g)}{g (S \times C + G_m)}$$  \quad (7.1)$$

where $G_m$ is the transconductance of the transconductor stage, while $g = 1/R$. As equation 7.1 shows, the pole location is controlled by the transconductance value $G_m$, which is
controlled by the control voltage. The zero location in the first stage of the boosting filter is designed to nominally be 3.32MHz with $R = 12\, \text{K}\Omega$ and $C = 4\, \text{pF}$, while in the second stage it is designed to be at 18.95 MHz with $R = 21\, \text{K}\Omega$ and $C = 0.4\, \text{pF}$.

The common mode feedback circuit tracks the common mode voltage of the output of the fully differential $G_m$-C stage and compares it to a set common mode level provided by the reference signal generator. Generally, if the output common mode is equal to the required common mode, the drain current of $M_6$ will be equal to the tail current of the differential pairs. Because $M_7$ and $M_8$ are identical to $M_6$, both of them will source that same current into the two single-ended outputs of the filter, but since there are two current sinks with the same value connected to each single-ended output, the common mode will stay the same. On the other hand, if the common mode voltage of the output was higher than the required common mode, the drain current of $M_6$ will decrease, leading to the reduction of the currents sourced by $M_7$ and $M_8$, which will cause the two current sinks connected to the outputs to pull the common mode voltage down until it gets to the desired value. The same logic applies if the common mode of the output of the filter was lower than the desired common mode.

The voltage controlled transconductor is an essential element for implementing the forward equalizer. Many techniques were reported in the literature for implementing voltage controlled transconductors [7, 8]. In chapter 4, a new proposed voltage controlled transconductor was presented. The boosting filter uses that proposed transconductor to implement the $G_m$-C stages. Figure 7.8 shows the total frequency response of the forward equalizer for different control voltage values. Note that the supply voltage is 1.8V. As shown in the figure, by increasing the control voltage, the high frequency boosting will also increase.
It is worth mentioning that the resistors used in the boosting filter (the forward equalizer) are n-well resistors. The reason of choosing n-well resistors over silicide-block resistors is the positive temperature coefficient of the well resistors. As shown in equation 7.1, the high frequency gain of the boosting filter is equal to $G_M \times R$, and since $G_M$ has a negative temperature coefficient (due to the degradation of the input transistors transconductance with temperature), having positive temperature coefficient resistors tends to compensate for the negative temperature coefficient of $G_M$, and hence gives a stable high frequency gain with temperature.

**Figure 7.8** The frequency response of the boosting filter for different control voltages.
7.3.2. The Slicer

The slicer, which will be referred to as the comparator as well, is the circuit element that resolves the differential output of the forward equalizer. The slicer's bandwidth has to be enough to handle the data rate of 125 Mbps. The slicer serves two purposes in the system. First, it resolves the differential output of the forward equalizer and generates two rail to rail complementary digital signals that serve as the received data. Second, the slicer's output it is used by the reference signal generator to generate the differential reference signal needed by the system to adapt the forward equalizer, as well as the common mode level used by all the common mode feedback circuits in the system. The proposed slicer has three amplification stages. Figure 7.9 shows the schematic diagram of the slicer. The first stage is a current mirror operational transconductance amplifier (OTA) that gives a moderate amplification for the input signal and provides a buffering between the input and the output to prevent kick back effect [7]. Transistors $M_{13}$, and $M_{14}$ have a very important role in the circuit, they prevent $M_3$ and $M_4$ from completely turning off during slew rate limited operation. That way, the circuit will not need to recharge the gates of $M_3$ and $M_4$, which significantly increases the speed of the comparator. The second stage of the slicer is a self biased differential amplifier for each complementary output. This second stage is used to provide more amplification to the signal and to achieve a differential to single-ended conversion of the received signal.

Using the self biased differential amplifier for differential to single-ended conversion has many advantages over using just a simple single-ended inverter, or the classic constant-tail-current differential amplifier [9]. First, as opposed to the classic differential amplifier,
Figure 7.9 Schematic diagram of the three-stage slicer.
the self biased differential amplifier is significantly faster during switching since, due to the feedback, it can supply a relatively high current to speed up the rise and fall times of the transition. In a classic differential amplifier on the other hand, the speed is always limited to the value of the tail current used. Another advantage of the self biased differential amplifier over the classic differential amplifier is that for digital signals, the self biased differential amplifier consumes current only during transitions, while in a classic differential amplifier, the tail current will be always consumed even if there were no transition. The second advantage of the self biased differential amplifier as opposed to a simple single-ended CMOS inverter is its differential input nature. In order to clarify this advantage, let's consider the operation of a single-ended inverter. In the first OTA stage, the cross over voltage of the two differential outputs is essentially controlled by the tail current used and the sizes of the output transistors. If a simple inverter is used for each output to further amplify the signal and convert it to a single-ended output, the difference in the threshold voltage of the inverter and the cross over voltage of the two differential outputs of the OTA will essentially cause a skew between the rise and fall times of the inverter's output, which consequently introduces duty cycle distortion at the inverter's output. The duty cycle distortion added due to the threshold voltage mismatch becomes more significant if the output of the OTA has a slow rise and fall times since the input to the inverter will spend more time around the threshold, which will also make it more susceptible to supply noise. Duty cycle distortion at the output of the slicer can significantly reduce the whole system's jitter tolerance. Even though under nominal operation conditions, the single-ended CMOS inverter's threshold can be designed to match the OTA's output cross-over voltage, still due to the fundamentally different nature of the OTA and the inverter, it is not possible to match the
two thresholds across process and temperature variations. Using the self biased differential amplifier eliminates this problem since its threshold voltage will always be the cross over voltage of its differential input. Therefore its threshold will always match the OTA outputs cross-over voltage, which eliminates any duty cycle distortion introduced due to threshold voltage mismatches.

The third stage of the slicer is two simple single-ended inverters that are used to further increase the gain of the whole slicer. The reason using single ended inverters at this point is acceptable while using them right after the OTA is not is that the outputs of the self biased differential amplifier will have significantly faster rise and fall times than the OTA outputs. This faster rise and fall times minimizes the duty cycle distortion added due to the threshold voltage mismatch between the self biased differential amplifier and the single-ended inverter. Figure 7.10 shows the frequency response of the first stage of the slicer. Since the second and third stages are only for the differential to single-ended conversion, the speed of the whole slicer will be limited by the first stage. As shown in figure 7.10, the bandwidth of the first stage is 21 MHz, while the unity gain frequency is at 1.96 GHz. The DC gain of the slicer is 37 db.

187
The Reference Signal Generator

The slicer after the third stage provides two complementary fully digital rail-to-rail signals that can be used to generate the reference signal needed for the comparison process with the forward equalizer's output. As was mentioned before, this reference signal has to have the same common mode and differential amplitude as the forward equalizer output. In order to achieve that, a potential divider is used to generate two voltage levels as shown in Figure 7.10.

Figure 7.10 The frequency response of the slicer's first stage.

7.3.3. The Reference Signal Generator
The difference between these two levels is equal to the differential output of the forward equalizer. A mid range level is also generated to supply a common mode level to the common mode feedback circuit of the forward equalizer and the band-pass filters to keep all the fully differential signals in the system at the same common mode level. Using the two fully digital signals from the slicer along with PMOS switches, the reference signal and the common mode level could be generated.

**Figure 7.11** The reference signal generator schematic diagram.
7.3.4. The Band-Pass Filter

The purpose of the band-pass filter is to measure the slope of both the reference signal generated by the reference signal generator and the forward equalizer output. The filter used is a second order filter that has one zero and two poles. Since the differentiation process of the input signal is performed due to the presence of a zero in the transfer function, the positions of the poles are not very critical, yet they still have to be there to filter out any high frequency noise. Figure 7.12 shows the schematic diagram of the band-pass filter. The transfer function of the filter could be written as:

\[
\frac{V_{od}}{V_{id}} = \left( \frac{C}{g_{m1}} \right) \left( \frac{S}{1 + S \frac{C}{g_{m1}}} \right) \left( \frac{1 + S \frac{C_{1}}{g_{m2}}}{1 + S \frac{C_{1}}{g_{m2}}} \right) \tag{7.2}
\]

where \( g_{m1} \) is the transconductance of \( M_1 \) and \( M_2 \), while \( g_{m2} \) is the transconductance of \( M_3 \) and \( M_4 \). The center frequency of the filter is designed to nominally be at 140 MHz, while the bandwidth is designed to nominally be at 200 MHz with. The reason behind choosing those numbers is that the fundamental frequency of the data rate is at 125 MHz, and since the band-pass filters is trying the estimate the high frequency component of the data, the center frequency of the filter was chosen to be around the fundamental frequency of the data. It is worth mentioning that since the design is fully differential, a common mode feedback circuit is used to set the common mode level of the band-pass filter's output to the same common mode as the reference signal and the forward equalizer output. Figure 7.13 shows the frequency response of the filter.
Figure 7.12 The schematic diagram of the band-pass filter.

Figure 7.13 The frequency response of the band-pass filter.
7.3.5. The Differential-Difference Squarer

The main function of this block is to square two differential signals and then subtract them. It is used to compare the differential outputs of the two band-pass filters discussed in the previous section to generate the error signal. The operation of the circuit is based on the large signal model of the MOS transistor provided that the common mode voltage of the two differential signals is the same [8]. Figure 7.14 shows the schematic diagram of the differential-difference squarer. The differential signals applied to the two transistor pairs $M_3$, $M_4$ and $M_5$, $M_6$ generate the following two currents:

\[
I_1 = \frac{K_p}{2} \left[ (V_{dd} - V_{1+} - V_{TP})^2 + (V_{dd} - V_{1-} - V_{TP})^2 \right] \quad (7.3)
\]

\[
I_2 = \frac{K_p}{2} \left[ (V_{dd} - V_{2+} - V_{TP})^2 + (V_{dd} - V_{2-} - V_{TP})^2 \right] \quad (7.4)
\]

Since $M_1$ and $M_2$ have the same gate-to-source voltage, and assuming that both of them operate in the saturation region, they will have equal drain currents. Therefore, the difference between $I_1$ and $I_2$ will flow into the resistors. The differential output voltage could then be written as:

\[
V_{O+} - V_{O-} = \frac{K_p}{4} R \left( V_{id1}^2 - V_{id2}^2 \right) \quad (7.5)
\]

where $V_{id1}$ and $V_{id2}$ are the input differential voltages. This differential output serves as the error signal that indicates how the adaptive forward equalizer should behave for a specific transmission channel. Figure 7.15 shows the DC response of the circuit.
Figure 7.14 The schematic diagram of the differential difference squarer.

Figure 7.15 The DC response of the differential difference squarer.
7.3.6. The Integrator

The Integrator is used to filter the high frequency components in the output of the differential difference squarer, i.e. the error signal. It provides a smooth averaged-out version of the error signal that can be used as a control signal for the forward equalizer. This filtering process is essential to avoid any instability in the whole adaptive loop, as well as to guarantee that the adaptive loop will not respond except for a real change in the transmission channel. The integrator has three stages as shown in figure 7.16. The first stage is a simple pre-amplifier to isolate the input and the output of the integrator and to give some amplification to the very small output of the differential difference squarer. The second stage is a single ended OTA with a capacitive load to achieve the integration function. The output of the second stage could be written as:

\[ V_O = \frac{g_m}{S C_1} V_{id} \]  \hspace{1cm} (7.6)

where \( g_m \) is the transconductance of \( M_5 \) and \( M_6 \). The third stage of the integrator is a simple level shifter to shift down the dc level of the output of the integrator to match the control range of the forward equalizer. This shifted-down version of the integrated signal serves as the control voltage of the forward equalizer. Figure 7.17 shows the frequency response of the integrator.

In the previous few sections, a detailed description of the circuit design of each block has been introduced. In the next section, the system level behavior of the equalizer is going to be investigated through the simulation results of the whole system. Lab measurement results of the fabricated system will then be reviewed and a comparison between the simulated and the actual behavior of the system on silicon will be discussed.
Figure 7.16 The schematic diagram of the integrator.

Figure 7.17 The frequency response of the integrator.
7.4. Simulation Results

The presented equalizer was implemented using a standard 180nm digital CMOS process with 1.8V supply voltage. In order to evaluate the performance of the system, two types of simulations were used. First, an AC simulation of the forward equalizer with different transmission channels to evaluate the adaptive behavior of the equalizer. Second, a transient simulation with different transmission channels to evaluate the performance of the rest of the blocks in the system. Two extreme transmission channels were used for the purpose of simulation, the first one is the 100m cable transmission channel, and the second one is the 0m cable transmission channel. Both channels were discussed in detail in section 7.2.

AC simulations were performed on the adaptive forward equalizer and the results are shown in figures 7.18 and 7.19. Figures 7.18 and 7.19 show the frequency response of the transmission channel, the forward equalizer, and the combined frequency response of the channel and the forward equalizer. As shown in figure 7.18, the bandwidth of the 100m channel is only 2 MHz while the data rate is 125 Mbps. Since the bandwidth of the channel is very limited the adaptive forward equalizer provides the maximum high frequency boosting as shown in figure. Therefore, the combined frequency response of the channel and the equalizer (which is the effective bandwidth) will have an improved bandwidth to enable the reception of the 125 Mbps data rate. As shown in figure 7.18, the effective bandwidth is 30 MHz, which is 15 times higher than the original bandwidth of the transmission channel. Figure 7.19 on other hand shows the 0m channel case. In this case the forward equalizer provides almost no high frequency boosting since the transmission
Figure 7.18 The frequency response of the forward equalizer, the 100m transmission channel, and the combined frequency response of the channel and the equalizer.

Figure 7.19 The frequency response of the forward equalizer, the 0m transmission channel, and the combined frequency response of the channel and the equalizer.
channel has enough bandwidth. Therefore, the combined frequency response of the channel and the equalizer (which is the effective bandwidth) will be almost identical to the bandwidth of the forward equalizer alone, which as shown in figure 7.19 is equal to 40 MHz. It is worth mentioning that in the case of wide bandwidth channels (ones that don't need equalization) the effect of the forward equalizer will actually be reducing the effective bandwidth to its own bandwidth, which should be designed to be enough to handle the received data rate adequately.

Transient simulations were performed on the adaptive forward equalizer for both the 100m and the 0m transmission channels. In order to inspect the behavior of the system, a random data stream at 125 Mbps rate (bit width is 8ns) and 500mV differential amplitude is being sent across the transmission channel the outputs of the different blocks will are plotted. For the 100m channel case, figure 20 shows a sample of the input data stream to the channel, the output of the channel (which is also the input to the equalizer), the output of the forward equalizer, and the final digital output of the slicer (which is essentially the received data). As shown in figure 20, the two single-ended outputs (also serve as a differential output) of the transmission channel do not cross over due to the sever ISI introduced by the limited bandwidth of the channel. The single-ended outputs (can also be looked at as a differential output) of the forward equalizer on the other hand cross over normally due to the high frequency boosting introduced by the forward equalizer, which effectively removed the ISI introduced by the channel from the received data. The outputs of the forward equalizer are then introduced to the slicer to generate the digital received data output. As mentioned before, an eye diagram is a useful tool in assessing the performance of the equalizer through inspecting the vertical and horizontal openings. Figure 21 shows
Figure 7.20 The transient response of the 100m channel, the forward equalizer, and the slicer.

Figure 7.21 The output differential eye diagram of the 100m channel.
the differential eye diagram of the output of the 100m transmission channel. As shown in figure, the eye is completely closed which makes it impossible to receive the data without equalization. Figure 22 shows the eye diagram of the differential output of the forward equalizer. Again as shown in figure, the effect of the forward equalizer was to open the received eye with enough vertical and horizontal openings in order to enable the slicer to resolve the received data. The forward equalizer achieves a horizontal opening of 7.2ns, which is only 800ps less than the ideal opening (8ns), while the vertical opening is 275mV (almost half the amplitude of the transmitted data), which is an enough overdrive for the slicer to resolve the data. Figure 23 shows the differential eye diagram of the output of the slicer, which is similar to the output of the forward equalizer (essentially the same data) with the exception of its fully digital rail to rail nature. Note that the vertical opening is 3.6V (double the supply voltage) because the eye is differential. In order to assess the adaptive behavior of the equalizer, figure 24 shows the reference signal, the forward equalizer output, the error signal, and the control voltage (used to adapt the forward equalizer). As shown in figure, at the beginning of the reception, the control voltage starts at a high value (around 1V) to guarantee enough initial boosting, this essentially causes the output of the forward equalizer to have a much faster rise and fall time than the reference signal. Therefore, the error signal will have an average negative value, which will cause the control voltage to drop in value, and consequently reducing the amount of boosting introduced by the forward equalizer. Once the output of the forward equalizer converges to the same rise and fall time as the reference signal, the error signal averages out to 0, and the control voltage starts to hold its value that assures an optimum boosting. In the 100m case, this value of the control voltage is around 236mV, while its convergence time is around 300ns.
Figure 7.22 The output differential eye diagram of the forward equalizer for a 100m channel.

Figure 7.23 The output differential eye diagram of the slicer for a 100m channel.
Transient simulations were also performed using the 0m transmission channel. For this case, figure 25 shows a sample of the input data stream to the channel, the output of the channel, the output of the forward equalizer, and the final digital output of the slicer. In this case, the outputs of the transmission channel cross over normally since the channel has wide bandwidth (279 MHz as shown in figure 7.19), and therefore it hardly introduces any ISI to the data. The outputs of the forward equalizer on the other hand cross over normally but they suffer from some ISI since the bandwidth of the forward equalizer is only 40MHz (as shown in figure 7.19). Yet, 40 MHz is still an enough bandwidth to provide an adequate vertical and horizontal eye opening to the slicer for proper reception at the 125 Mbps data rate. The outputs of the forward equalizer are then introduced to the slicer to generate the

Figure 7.24 The transient response of the error and control signals for a 100m channel.
digital received data output. Figure 26 shows the differential eye diagram of the output of the 0m transmission channel. As expected, the eye is completely open due to the wide bandwidth of the channel, note that the vertical opening represents the peak to peak differential amplitude. Figure 27 shows the eye diagram of the differential output of the forward equalizer. Note that the horizontal and vertical openings have been degraded (6.2ns and 219mV respectively), yet they are still enough for adequate reception by the slicer. Figure 28 shows the differential eye diagram of the output of the slicer, which is similar to the output of the forward equalizer only with a rail to rail nature. In order to assess the adaptive behavior of the equalizer, figure 29 shows the reference signal, the forward equalizer output, the error signal, and the control voltage. Again, at the beginning of the reception, the control voltage starts at a high value (around 1V) to guarantee enough initial boosting, which essentially causes the output of the forward equalizer to have a much faster rise and fall time than the reference signal. Hence, the error signal will have an average negative value, which will cause the control voltage to drop causing the reduction of the boosting introduced by the forward equalizer. Once the error signal averages out to 0, the control voltage converges to its optimum value. In the 0m case, this value of the control voltage is around 35mV, while its convergence time is around 200ns. As expected the control voltage value in the 0m transmission channel is much less than the 100m case since no boosting is required by the forward equalizer.
Figure 7.25 The transient response of the 0m channel, the forward equalizer, and the slicer.

Figure 7.26 The output differential eye diagram of the 0m channel.
Figure 7.27 The output differential eye diagram of the forward equalizer for a 0m channel.

Figure 7.28 The output differential eye diagram of the slicer for a 0m channel.
Transient simulations using a 1.8V supply and nominal process and temperature conditions showed that the rms current consumption of the whole equalizer is around 2.8 mA, which is an extremely low current consumption for such an application. In this section AC and transient simulations of the whole system have been introduced. The adaptive behavior of the equalizer was also investigated using two extreme transmission channels (the 100m and 0m channels). In the next section, lab measurements results will be introduced. The operation of the equalizer with different cable lengths will be measured and assessed through eye diagrams at the input and output of the system.

Figure 7.29 The transient response of the error and control signals for a 0m channel.
7.5. Measurements Results

As explained in the beginning of this chapter, the equalizer was implemented as a part of a larger transceiver/repeater system that was designed to enable two high speed transceivers to communicate across UTP CAT-5 cables that have length up to 100 meters. The transceiver/repeater essentially receives the data from one high speed transceiver (usually through metal traces) and retransmits this data across the CAT-5 cable. On the other side of the cable, another transceiver/repeater receives the data and then retransmits the data to the other high speed transceiver. The data rate used was 125 Mbps (8ns bit time) and the equalizer was used as the receiver that receives the data from the CAT-5 cable and passes it on to the high speed transceiver. The equalizer was implemented on a standard 180nm digital CMOS process with a 1.8V supply voltage. Figure 7.30 shows the layout of the whole equalizer with its different parts highlighted. The layout area of the whole equalizer is 27738 $\mu m^2$. Figure 7.31 shows a chip micrograph of the whole transceiver/repeater system with the equalizer part highlighted.

Lab measurements were performed on the system with different cable lengths. As mentioned previously, eye diagrams is the best way to asses the operation of the equalizer, as well as bit error rate measurements. Differential input and output eye diagrams of the equalizer with different cable lengths were measured and are shown in figures 7.32 to 7.36. Figure 7.32 shows the eyes with a 12 inch cable, figure 7.33 shows the eyes with a 10m cable, figure 7.34 shows the eyes with a 30m cable, figure 7.35 shows the eyes with a 50m cable, and figure 7.36 shows the eyes with a 100m cable. As shown in those figures, the longer the cable gets, the more closed the received eye becomes. Particularly, at the 100m cable case,
the input eye to the equalizer becomes completely closed. Table 7.1 summarizes the output eye's horizontal opening for different cable length. The equalizer was also tested under different process conditions (weak, nominal, and strong conditions), as well as temperature extremes ranging from -40 °C to 125 °C, and with supply voltage varying from 1.6V to 2.0V. The system always achieved at least $10^{-12}$ bit error rate or better. Power consumption measurements were also performed and the rms current consumed by the whole equalizer was found to be between 2.3 mA (at weak process, 1.6V supply, and 125 °C), and 5.15 mA (at strong process, 2.0V supply, and -40 °C).

<table>
<thead>
<tr>
<th>Cable Length</th>
<th>Horizontal Eye Opening (ideal is 8ns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 inches</td>
<td>5.64ns</td>
</tr>
<tr>
<td>10 meters</td>
<td>5.39ns</td>
</tr>
<tr>
<td>30 meters</td>
<td>7.14ns</td>
</tr>
<tr>
<td>50 meters</td>
<td>6.25ns</td>
</tr>
<tr>
<td>100 meters</td>
<td>5.64ns</td>
</tr>
</tbody>
</table>

*Table 7.1* The equalizer's differential output eye horizontal opening.
Figure 7.20 The layout of the equalizer.
Figure 7.21 The chip micrograph showing the location of the equalizer on the chip.
Figure 7.32 (a) The input differential eye diagram of the equalizer for a 12 inches cable. (b) The output differential eye diagram of the equalizer for a 12 inches cable.
Figure 7.33 (a) The input differential eye diagram of the equalizer for a 10 meters cable. (b) The output differential eye diagram of the equalizer for a 10 meters cable.
Figure 7.34 (a) The input differential eye diagram of the equalizer for a 30 meters cable. (b) The output differential eye diagram of the equalizer for a 30 meters cable.
Figure 7.35 (a) The input differential eye diagram of the equalizer for a 50 meters cable. (b) The output differential eye diagram of the equalizer for a 50 meters cable.
Figure 7.36 (a) The input differential eye diagram of the equalizer for a 100 meters cable. (b) The output differential eye diagram of the equalizer for a 100 meters cable.
LIST OF REFERENCES


In this thesis, an introduction to the three categories of variations that could be encountered in analog and mixed signal circuits and systems design has been discussed. The thesis concentrated on adaptation techniques for variations in the first category, i.e. variations due to process, temperature, supply voltage, and variations in the third category, i.e. transmission media. First category variation results in inaccuracies in the absolute value of the different parameters of on-chip elements (resistor, capacitors, and transistors) due to process and temperature variations. Chapter 2 concentrated specifically on on-chip resistors and capacitors with a discussion of the different techniques for implementing those elements and the sources of errors in their absolute values due to variations in the fabrication process as well as temperature effects. Chapter 2 also gave an overview of design and layout techniques to minimize errors in the absolute values of on-chip resistors and capacitors, as well as techniques to improve their matching properties.

Chapter 3 concentrated on the different on-chip automatic tuning architecture techniques that could be used to tune on-chip elements parameters to improve the accuracy of their absolute values. Different elements used in automatic on-chip tuning architectures
have been discussed including voltage, current and time references, and electronically tunable elements (voltage controlled transconductors, resistors, and capacitors). An overview of post fabrication tuning options was also presented.

Chapter 4 introduced a new voltage controlled transconductor proposed by the author that could be used in conjunction with any automatic tuning process. The proposed transconductor used a simple technique to improve the linearity as well as the input and control voltage range by eliminating second order effects in linear mode based transconductors. The high linearity of the circuit enabled a high differential input range as well as a high control voltage range with a low supply voltage (1.8 V). The proposed transconductor also achieved a truly fully differential nature for the output currents. A United States Patent was granted for the design on April 20, 2004 under the number 6724258.

Chapter 5 introduced a new digital tuning algorithm for on-chip resistors proposed by the author. The proposed technique achieves a simple, yet powerful methodology for implementing accurate on-chip resistors that can enable the integration of some applications on chip that were not possibly done except by using external components. The proposed tuning architecture was successfully implemented on a 180nm digital CMOS process for integrating accurate on-chip transmission line termination resistors for a high speed digital communication transceiver.

Chapter 6 discussed the problem of Intersymbol Interference (ISI) in digital wireless and wire line communication systems as a representative of third category variations, i.e. variations in the transmission media. The main causes of ISI (mainly multi-path effects and band limited channels) have been discussed. The chapter also gave an overview of the
available techniques to minimize the impact of ISI on the quality of the data reception. Specifically, fixed and adaptive equalization architectures have been discussed along with the analog, digital, and mixed signal techniques of implementing equalization architecture.

In chapter 7, purely analog adaptive equalization architecture was used to implement a 125Mbps transceiver/repeater that operated using a 1.8V supply. A discussion of the requirements of each block used in the system as well as the circuit design details have been presented, along with SPICE simulation results. The transceiver/repeater operated across an Untwisted-Shielded-Pair (UTP) Category 5 (CAT 5) Ethernet cable that varied in length from one foot up to a 100m. The equalizer was implemented on a standard 180nm digital CMOS process and operated successfully over the desired cable range.

### 8.1. Future Research

In this thesis, there was no consideration for adaptive techniques for the second category of variations, which is left for future research. The second category of variation includes the variations in the system level, i.e. variations that cause the specifications of the system to be changeable. For example, the design of adaptive wireless receivers that can handle more than one mobile phone standard (GSM, TDMA, etc..), or wire line receivers that can handle different networking protocols (Ethernet, IEEE 1394, USB). In these cases, there is a strong need for a single transceiver that can automatically adapt itself to handle different sets of specifications without the redundancy of integrating multiple different receivers with each one handling a set of specifications. Designing a single adaptive
transceiver introduces major design challenges added to the ones already present because of the variations in the first and third category. Therefore, adaptive techniques have to be developed to automatically change the main characteristics of a system block to comply with different sets of specifications. For instance, the design of adaptive filters that can automatically switch between multiple different bandwidths, center and cut-off frequencies becomes a very challenging task. The same argument applies also to oscillators, low noise amplifiers, analog to digital converters, and so on. Those design challenges require significant innovation efforts to enable effective and higher levels of integration.
BIBLIOGRAPHY


224


INDEX

A
Adaptation
architectures-------------------------- 8, 159, 164, 170
techniques ------------------ ii, 5, 6, 8, 138, 218, 220
Amplifiers--- 4, 6, 13, 54, 56, 60, 69, 83, 92, 184, 186, 187, 194, 217, 221, 225, 230
Analog to Digital Converters------ 4, 6, 8, 61, 165, 221

B
Bandwidth---- iii, xvi, 4, 141, 142, 148, 149, 152, 155, 159, 171, 173, 184, 187, 190, 196, 198, 202

C
Cables---- iii, xvi, xviii, 4, 11, 112, 166, 170, 171, 172, 173, 196, 206, 207, 220

D
Digital to Analog Converters------------ xiii, 8, 61
Duty Cycle Distortion ------------------- 2, 186, 187

E
Equalization
adaptive iii, xvi, 9, 10, 11, 110, 138, 154, 166, 169, 170, 172, 176, 220, 228
architectures-------------------ix, x, 152, 155, 159, 163

F
Fabrication
diffusion--xii, 15, 16, 17, 18, 19, 20, 21, 23, 24, 32
doping-------- 18, 21, 22, 34, 35, 39, 42, 47, 81, 82
etching -------------23, 24, 26, 28, 46, 47, 48, 50, 117
junction depth-------------------------- 16, 17
sheet resistance -------------- 14, 16, 17, 18, 19, 20, 23
Filters ---xvi, xv, 2, 3, 4, 6, 10, 13, 36, 54, 55, 56, 80, 81, 83, 91, 92, 110, 136, 141, 153, 155, 157, 159, 160, 162, 164, 168, 170, 176, 178, 180, 181, 182, 183, 189, 190, 192, 194, 221, 224, 225, 227
Flux Capacitors-------------------44, 47, 48

I
Integration----- ii, 10, 13, 112, 113, 179, 194, 219, 221
Intersymbol Interference ---- 4, 10, 138, 139, 140, 142, 144, 148, 149, 151, 152, 155, 157, 159, 160, 161, 163, 166, 198, 202, 219
multi-path ------------ xv, 4, 139, 140, 152, 159, 219

L
Layout
common centroid--------------------- xii, 31, 49
interdigitized------------------- xii, xiii, 29, 30, 46, 47
Linearity-71, 72, 73, 74, 75, 77, 79, 83, 92, 93, 94, 96, 97, 98, 100, 101, 102, 110, 113, 114, 118, 127, 130, 219

M
Matching - 2, 9, 13, 21, 27, 28, 29, 30, 48, 49, 50, 128, 129, 186, 187, 218

N
Noise----xii, xiii, 4, 19, 20, 39, 41, 42, 43, 80, 82, 113, 114, 156, 161, 165, 168, 186, 190, 221, 226

O
On-Chip Elements
capacitors---------------------- 1, 2, 9, 13, 55, 56, 218
metal -------------------------- xiii, 43, 44, 47
poly----------------------------- xiii, 41, 42, 43
well---------------------------- xiii, 39, 42, 47, 48
ratio of, ------------------------- 3, 55
resistors - iii, 1, 2, 9, 10, 13, 21, 55, 56, 66, 67, 112, 113, 114, 116, 117, 118, 119, 121, 122, 123, 124, 218, 219
poly------------------------------- 129, 130, 131, 133, 135
silicide --- 18, 23, 129, 130, 131, 133, 135, 183
well-----------------------------------xii, 17, 19, 183
transistors-----------------------------------2
OPAMPS-------------------------------------6

R
References
current -------------------------------------------- 62, 67
time----------------------------- 59, 60, 61, 62, 68, 219
voltage ------------------- 62, 63, 64, 66, 68, 114, 127

S
Specifications---------------- ii, 1, 2, 3, 5, 130, 133, 220
Standards
Ethernet----------------------------- iii, 4, 11, 171, 220
IEEE 1394---------------------------------- 171, 220
mobile phone ------------------------------------3
USB--------------------------------------171, 220
Substratexi, xiii, 16, 17, 19, 20, 37, 38, 39, 41, 42, 43, 64, 70, 80, 82, 113

T
Temperature ii, iii, 1, 2, 4, 8, 13, 15, 16, 18, 20, 21, 22, 23, 48, 62, 63, 64, 65, 66, 68, 73, 83, 89, 92, 112, 114, 122, 127, 128, 130, 131, 133, 134, 138, 154, 165, 183, 187, 206, 208, 218
Transconductors iii, vi, ix, xiv, 6, 9, 69, 71, 82, 84, 86, 87, 91, 92, 93, 94, 97, 98, 100, 102, 110, 111, 180, 181, 182, 219, 224, 230
Transmission

media--- ii, 4, 5, 8, 10, 138, 152, 154, 169, 180, 218, 219
Transmission Line
modeling---------------------------------------- xvi, 172
terminations2, 10, 12, 112, 130, 133, 136, 219, 222
Tunable
capacitors----------------------------------------ix, 61, 81
resistors--------------------------------------- 61, 71, 80
transconductors-------------------------------ix, 61, 82
Tuning
algorithms - ii, vi, 5, 6, 7, 10, 88, 89, 113, 115, 117, 118, 137, 160, 219, 230
architectures-xiii, xv, 10, 57, 58, 61, 113, 128, 218, 219
automatic --- 9, 55, 56, 57, 83, 88, 89, 92, 176, 177, 218, 219
post fabrication--------------------------55, 88, 89, 113, 219
techniques ---------------------------9, 10, 55, 59, 125

V
Variations
first Category ----------------------------------1
process-------------- 1, 21, 29, 31, 127, 132, 135, 138
second Category--------------------------3
supply------------------------- 1, 138
temperature--iii, 1, 2, 4, 9, 21, 83, 89, 92, 114, 122, 128, 131, 135, 138, 165, 187, 218
third Category ------------------------4

W
Wire Line-- 2, 9, 10, 138, 140, 166, 170, 172, 219, 220
Wireless --------------xv, 4, 10, 139, 153, 172, 219, 220