ABSTRACT


Passive intermodulation distortion can interfere with intended communications signals limiting the capacity and range of a communications system. Many physical mechanisms have been suggested as causes of passive intermodulation distortion. The description of these mechanisms are generally limited to empirical or behavioral models rather than physical descriptions due to the difficulty in isolating passive intermodulation mechanisms. Measurement of passive intermodulation distortion is complicated by the weakly nonlinear behavior of passive components, inhibiting the physical isolation of passive intermodulation producing mechanisms. The dynamic range required to measure the weak nonlinearities of these components can often exceed 100 decibels. A broadband measurement system based on feed-forward cancellation possessing dynamic range in excess of 113 decibels is constructed to overcome passive intermodulation measurement difficulties. Electro-thermal distortion is found to be a dominant passive intermodulation source with a defined non-integer order Laplacian behavior. This behavior results in long-tail transients and a well defined thermal dispersion characteristic in the generated passive intermodulation distortion that cannot easily be explained by integer order differential equations. A fractional calculus description of the phenomena is introduced, accurately modeling both long-tail transients and thermal frequency dispersion. The physics behind electro-thermal distortion is derived analytically for general lumped, lossy microwave components, transmission lines, and antennas. Microwave attenuators, terminations, integrated circuit resistors, transmission lines, and antennas are manufactured to isolate the electro-thermal phenomena. The developed high dynamic range measurement system is used to characterize the thermal dispersion characteristic in the generated passive intermodulation distortion for each manufactured component. Electro-thermal conductivity modulation, dependent only on material parameters, is shown to be a dominant passive intermodulation source in all passive microwave circuits.
©Copyright 2010 by Jonathan Ryan Wilkerson
all rights reserved
Passive Intermodulation Distortion in Radio Frequency Communication Systems

by

Jonathan Ryan Wilkerson

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Electrical Engineering

Raleigh, NC

2010

APPROVED BY:

Dr. Douglas Barlage
Dr. Mohammed Zikry

Dr. Michael B. Steer
Chair of Advisory Committee

Dr. Kevin G. Gard
Co-Chair of Advisory Committee
BIOGRAPHY

Jonathan Ryan Wilkerson was born in Greenville, NC. He received the B.S. degree in both Electrical Engineering and Computer Engineering in 2005, and the M.S. degree in Electrical Engineering in 2006, from North Carolina State University, Raleigh, NC. Since 2005 he has been a graduate Research Assistant with the ERL Laboratory, Electrical and Computer Engineering Department, North Carolina State University. Mr. Wilkerson’s research interests include electro-thermal physics, nonlinear microwave and RF circuits, wave propagation and antennas, thermo-acoustics, electro-acoustics, high dynamic range measurement, and physically based nonlinear modeling.
ACKNOWLEDGMENTS

I would like to express my appreciation for the support and guidance provided by my advisors, Dr. Michael Steer and Dr. Kevin Gard, during my PhD research. I would like to thank Dr. Michael Steer for lending me his considerable knowledge in physics, mathematics, and microwave and RF circuits. He defined my research and guided me through my publications and presentations. I am grateful for access to Dr. Steer’s professional contacts that funded my research, namely ARO. I would like to thank Dr. Kevin Gard, who was an invaluable resource in this work for research direction, measurement methods and ideas, his considerable knowledge of analog and RF circuits, and his assistance in all my publications. I would also like to thank the members of my committee, Dr. Mohammed Zikry and Dr. Doug Barlage, for their comments and discussions on my research.

I would like to thank my graduate student colleagues Gregory Mazzaro, Theodore Robert Harris, Glenwood Garner III, Peter Lam, Chris Saunders, Dr. Frank Hart, Dr. Sonali Luniya, and Zhiping Feng for stimulating conversation and an exchange of knowledge that extended my own research and experience. Through their humor and creativity they provided an enjoyable working environment and the encouragement needed to finish. I extend a special thanks to Robert Harris, Peter Lam, and Dr. Steve Lipa who assisted me greatly in the development of several projects. I also would like to thank Dr. Aaron Walker and Dr. Mark Buff for teaching me to use the laboratory equipment in the Electronics Research Laboratory when I first joined the Ph.D program.

Lastly, I would like to thank my family for providing me encouragement and support, instilling in me a work ethic, and making me believe I could achieve any goal. Their support made this work possible.
# TABLE OF CONTENTS

**LIST OF FIGURES** .............................................................................. ix

**LIST OF TABLES** ............................................................................. xii

1 Introduction .................................................................................. 1
  1.1 Motivations .............................................................................. 2
  1.2 Overview ................................................................................. 2
    1.2.1 Passive Distortion Modeling .............................................. 3
    1.2.2 Passive Intermodulation Measurement ............................ 3
    1.2.3 Objective: Explain and Model the physical source of PIM ... 4
  1.3 Original Contributions ............................................................... 5
    1.3.1 Broadband High Dynamic Range Measurement System .... 5
    1.3.2 Nonlinear Electro-Thermal Theory ................................. 5
    1.3.3 Fractional Electro-Thermal Circuit Model ................. 5
    1.3.4 Foster Expansion Electro-Thermal Circuit Model ....... 6
    1.3.5 Distributed PIM Model .................................................... 6
    1.3.6 Resonant PIM Model .......................................................... 6
    1.3.7 Electromagnetic and Acoustic Anechoic Chamber .......... 6
  1.4 Dissertation Outline ................................................................. 6
  1.5 Published Works .................................................................... 8
    1.5.1 Journals ........................................................................... 8
    1.5.2 Conferences .................................................................... 8
  1.6 Unpublished Works .................................................................. 9

2 Review of Nonlinear Analysis and Measurement Techniques .......... 10
  2.1 Introduction ............................................................................ 11
  2.2 Passive Distortion ................................................................. 12
    2.2.1 Metal-Insulator-Metal and Metal-Metal Contact Nonlinearities . 12
    2.2.2 Ferromagnetic Material Nonlinearity ........................... 20
    2.2.3 Piezoelectric Material Nonlinearity ............................... 21
    2.2.4 Acoustic Nonlinearities ................................................... 23
    2.2.5 Electrical Conductivity Nonlinearity ........................... 24
  2.3 Distortion Analysis Methods .................................................. 25
    2.3.1 Power Series ................................................................... 26
    2.3.2 Volterra Series ................................................................. 26
    2.3.3 Behavioral Models .......................................................... 27
    2.3.4 Physical Models .............................................................. 28
  2.4 Distortion Measurement ........................................................... 29
4.3.2 Foster Approximation ........................................... 113
4.3.3 Summary of Electro-Thermal Modeling .......................... 116
4.4 Case Study: Microwave Terminations ............................. 118
  4.4.1 Thermal Coefficient of Resistance Characterization ......... 118
  4.4.2 Thermal Transient Characterization .......................... 120
  4.4.3 Two-Tone PIM Characterization ............................... 122
  4.4.4 Summary of Electro-Thermal Distortion in Terminations .... 125
4.5 Case Study: Platinum Attenuator ................................ 125
  4.5.1 Thermal Parameter Characterization and Foster Model ........ 126
  4.5.2 Platinum Electro-Thermal Dispersion .......................... 126
  4.5.3 Summary of Electro-Thermal Distortion in Attenuators ....... 129
4.6 Case Study: Integrated Circuit Distortion ....................... 131
  4.6.1 Electro-Thermal Dispersion Measurement ..................... 132
  4.6.2 Wide Thermal Bandwidth Devices ............................. 133
  4.6.3 Narrow Thermal Bandwidth Devices .......................... 140
  4.6.4 Summary of Electro-Thermal Distortion in IC’s .............. 148
4.7 Conclusion ............................................................. 150
5 Distributed Passive Intermodulation Distortion .................... 152
  5.1 Introduction ......................................................... 153
  5.2 Distributed Electro-Thermal Theory ............................... 154
    5.2.1 Heat Conduction on Transmission Lines ..................... 155
    5.2.2 Electro-Thermal PIM of a Finite Element .................... 160
    5.2.3 Distributed PIM Interference ................................. 163
    5.2.4 Summary of Distributed Electro-Thermal PIM Theory ........ 167
  5.3 Nonlinear Conductivity Isolation ................................ 168
    5.3.1 Materials Design for Process Isolation ..................... 168
    5.3.2 Transmission Line Design .................................... 170
    5.3.3 Transmission Line Interface .................................. 172
    5.3.4 Summary of Transmission Line Sample Design ............... 174
  5.4 Case Study: Electro-Thermal Distortion on Transmission Lines ... 175
    5.4.1 Electro-Thermal Dispersion Measurement ..................... 176
    5.4.2 Material and Current Bunching Thermal Effects ............... 178
    5.4.3 PIM Dependency on Line Dimensions .......................... 181
    5.4.4 Summary of Distributed PIM Measurements ................... 185
  5.5 Conclusion ............................................................. 186
6 Passive Intermodulation Distortion in Resonant Structures ........ 189
  6.1 Introduction .......................................................... 190
  6.2 Resonant Structure Electro-Thermal Theory ....................... 191
    6.2.1 Heat Conduction on a Rectangular Patch Antenna ............. 192
    6.2.2 Electro-Thermal PIM of a Finite Element .................... 196
    6.2.3 Summary of Electro-Thermal PIM in Antennas ................ 199
  6.3 Isolation of Physical Electro-Thermal Process ................... 200
6.3.1 Materials Design for Process Isolation ................................................. 200
6.3.2 Transmission Line Interface .............................................................. 204
6.3.3 Summary of Antenna Sample Design .................................................. 206

6.4 Case Study: Microstrip Rectangular Patch Antenna ..................................... 207
6.4.1 Linear Inset Fed Microstrip Rectangular Patch Antenna Design ................. 207
6.4.2 Antenna Electric Fields and Surface Currents ........................................ 210
6.4.3 Antenna Measurement ............................................................................ 216
6.4.4 Summary of Rectangular Patch Antenna ................................................ 225

6.5 Conclusion ................................................................................................. 226

7 Conclusion ..................................................................................................... 228
7.1 Summary of Research and Original Contributions ....................................... 229
7.2 Future Research .......................................................................................... 231

Bibliography ....................................................................................................... 233

Appendices ......................................................................................................... 244

A Electromagnetic and Acoustic Anechoic Chamber ............................................. 245
A.1 Introduction ................................................................................................... 246
A.2 Anechoic Chamber Construction .................................................................. 246
A.3 Anechoic Chamber Acoustic Characterization ............................................. 248
A.4 Anechoic Chamber RF Characterization ...................................................... 250
A.5 Conclusion .................................................................................................... 253

B Matlab Functions ............................................................................................ 254
B.1 PIM in Lossy Lumped Components ............................................................. 255
B.2 PIM on Transmission Lines .......................................................................... 256
B.3 PIM on Antennas ........................................................................................ 260
B.4 IMD Extractor with Drift ............................................................................ 263
B.5 IMD Extractor from Spectrum Shell ............................................................ 264
B.6 Voltage to dBm Converter .......................................................................... 267
B.7 dBm to Voltage Converter .......................................................................... 267
B.8 Cancellation Formula Analysis ................................................................... 267
B.9 Multi-Spectrum File Analysis ...................................................................... 268
B.10 Sum a Vector .............................................................................................. 269
B.11 LimiterModel .............................................................................................. 270
B.12 LimiterGain ................................................................................................ 271
B.13 Function to Find S-parameters ................................................................... 271
B.14 S-parameters Interpolation Routine .......................................................... 272
B.15 Optimized Grunwald-Letnikov Fractional Derivative Routine ................... 272
B.16 Forward Euler Algorithm .......................................................................... 273
B.17 Fractional Simulator ................................................................................. 273
B.18 Riemann-Liouville Fractional Derivative .................................................... 277
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Quantum representation of tunneling process</td>
<td>15</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Intermodulation distortion from tunneling at a micro-contact</td>
<td>17</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Quantum representation of thermionic emission</td>
<td>18</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Constriction resistance at point contacts</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Magnetic hysteresis loop</td>
<td>21</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Polarization hysteresis loop</td>
<td>23</td>
</tr>
<tr>
<td>Figure 2.7</td>
<td>AM-AM and AM-PM measurements</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2.8</td>
<td>AM-AM and AM-PM measurement configuration</td>
<td>31</td>
</tr>
<tr>
<td>Figure 2.9</td>
<td>THD input and output relationships</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.10</td>
<td>Two-tone measurement spectrum</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.11</td>
<td>Two-tone measurement configuration</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.12</td>
<td>Cancellation as a function of phase and amplitude error</td>
<td>36</td>
</tr>
<tr>
<td>Figure 2.13</td>
<td>Feed-forward measurement architecture</td>
<td>37</td>
</tr>
<tr>
<td>Figure 2.14</td>
<td>Two-tone feed-forward measurement spectrum</td>
<td>38</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Cancellation formula performance, simulated</td>
<td>54</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>High dynamic range measurement system</td>
<td>57</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>I/Q matching effects, measurement</td>
<td>59</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Amplifier model</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Amplifier linearity measurement</td>
<td>63</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Amplifier forward wave sweeps, measurement</td>
<td>64</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Amplifier reverse wave sweeps, measurement</td>
<td>65</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>Circulator model</td>
<td>66</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>Isolator S-parameters, measurement</td>
<td>67</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>Isolator forward wave sweeps, measurement</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3.11</td>
<td>Isolator reverse wave sweeps, measurement</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3.12</td>
<td>Isolator two-tone frequency sweeps, measurement</td>
<td>71</td>
</tr>
<tr>
<td>Figure 3.13</td>
<td>Terminator model</td>
<td>72</td>
</tr>
<tr>
<td>Figure 3.14</td>
<td>Attenuator model</td>
<td>73</td>
</tr>
<tr>
<td>Figure 3.15</td>
<td>Transmission line model</td>
<td>74</td>
</tr>
<tr>
<td>Figure 3.16</td>
<td>Limiting amplifier model</td>
<td>77</td>
</tr>
<tr>
<td>Figure 3.17</td>
<td>Limiting amplifier signal</td>
<td>78</td>
</tr>
<tr>
<td>Figure 3.18</td>
<td>Limiting amplifier spectrum, measurement</td>
<td>79</td>
</tr>
<tr>
<td>Figure 3.19</td>
<td>Electromagnetic radiation model</td>
<td>80</td>
</tr>
<tr>
<td>Figure 3.20</td>
<td>Radiative coupling model</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.21</td>
<td>Effects of pre-limiting on cancellation, measurement</td>
<td>82</td>
</tr>
<tr>
<td>Figure 3.22</td>
<td>Limiting amplifier mirroring response, measured</td>
<td>83</td>
</tr>
<tr>
<td>Figure 3.23</td>
<td>Hybrid combiner model</td>
<td>85</td>
</tr>
<tr>
<td>Figure</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.24</td>
<td>Feed-forward measurement system dynamic range, measured</td>
<td>87</td>
</tr>
<tr>
<td>3.25</td>
<td>WCDMA signal cancellation, measured</td>
<td>88</td>
</tr>
<tr>
<td>3.26</td>
<td>WCDMA feed-forward measurement of amplifier</td>
<td>89</td>
</tr>
<tr>
<td>3.27</td>
<td>WCDMA group delay pre-distortion, measurement</td>
<td>90</td>
</tr>
<tr>
<td>4.1</td>
<td>Amplitude modulation in a two-tone signal</td>
<td>99</td>
</tr>
<tr>
<td>4.2</td>
<td>Power envelope of a two-tone signal</td>
<td>99</td>
</tr>
<tr>
<td>4.3</td>
<td>Electro-thermal mixing process</td>
<td>101</td>
</tr>
<tr>
<td>4.4</td>
<td>Long tail electrical response, measured</td>
<td>103</td>
</tr>
<tr>
<td>4.5</td>
<td>Heat conduction in a rectangular chip termination</td>
<td>105</td>
</tr>
<tr>
<td>4.6</td>
<td>Heat conduction in a cylindrical termination</td>
<td>107</td>
</tr>
<tr>
<td>4.7</td>
<td>Single pole electro-thermal model</td>
<td>109</td>
</tr>
<tr>
<td>4.8</td>
<td>Fractional order compact electro-thermal model</td>
<td>111</td>
</tr>
<tr>
<td>4.9</td>
<td>Foster approximation of fractional frequency response</td>
<td>114</td>
</tr>
<tr>
<td>4.10</td>
<td>Approximate fractional order electro-thermal circuit model</td>
<td>117</td>
</tr>
<tr>
<td>4.11</td>
<td>Measurement configuration for thermo-resistance</td>
<td>120</td>
</tr>
<tr>
<td>4.12</td>
<td>Thermo-resistance measurements of several microwave terminations</td>
<td>121</td>
</tr>
<tr>
<td>4.13</td>
<td>Thermal conductivity and capacity measurement configuration</td>
<td>122</td>
</tr>
<tr>
<td>4.14</td>
<td>Termination measurement configuration</td>
<td>123</td>
</tr>
<tr>
<td>4.15</td>
<td>Electro-thermal PIM measurement and model results for terminations</td>
<td>124</td>
</tr>
<tr>
<td>4.16</td>
<td>Single pole and approximate fractional long tail transients</td>
<td>127</td>
</tr>
<tr>
<td>4.17</td>
<td>Approximate electro-thermal model implementation</td>
<td>128</td>
</tr>
<tr>
<td>4.18</td>
<td>Platinum attenuator circuit diagram</td>
<td>129</td>
</tr>
<tr>
<td>4.19</td>
<td>Platinum attenuator PIM measurement</td>
<td>130</td>
</tr>
<tr>
<td>4.20</td>
<td>Platinum attenuator spectrum</td>
<td>130</td>
</tr>
<tr>
<td>4.21</td>
<td>Integrated circuit resistor measurement circuit</td>
<td>133</td>
</tr>
<tr>
<td>4.22</td>
<td>High resistivity polysilicon IC-resistor dimensions</td>
<td>135</td>
</tr>
<tr>
<td>4.23</td>
<td>High resistivity polysilicon IC resistor PIM measurement</td>
<td>136</td>
</tr>
<tr>
<td>4.24</td>
<td>Medium resistivity polysilicon IC resistor dimensions</td>
<td>138</td>
</tr>
<tr>
<td>4.25</td>
<td>Medium resistivity polysilicon IC resistor PIM measurement</td>
<td>139</td>
</tr>
<tr>
<td>4.26</td>
<td>Tantalum nitride IC resistor dimensions</td>
<td>141</td>
</tr>
<tr>
<td>4.27</td>
<td>Tantalum nitride IC resistor PIM measurement</td>
<td>142</td>
</tr>
<tr>
<td>4.28</td>
<td>High resistivity interdigitated polysilicon IC-resistor dimensions</td>
<td>144</td>
</tr>
<tr>
<td>4.29</td>
<td>High resistivity interdigitated polysilicon IC resistor PIM measurement</td>
<td>145</td>
</tr>
<tr>
<td>4.30</td>
<td>Low resistivity silicided gate polysilicon IC-resistor dimensions</td>
<td>147</td>
</tr>
<tr>
<td>4.31</td>
<td>Low resistivity silicide gate polysilicon IC resistor PIM measurement</td>
<td>148</td>
</tr>
<tr>
<td>5.1</td>
<td>Transmission line heat conduction model</td>
<td>156</td>
</tr>
<tr>
<td>5.2</td>
<td>PIM generation along a transmission line</td>
<td>165</td>
</tr>
<tr>
<td>5.3</td>
<td>SEM cross section image of silver on sapphire</td>
<td>170</td>
</tr>
<tr>
<td>5.4</td>
<td>SEM top view image of silver on sapphire</td>
<td>171</td>
</tr>
<tr>
<td>5.5</td>
<td>SEM top view image of silver on sapphire with constriction resistances</td>
<td>172</td>
</tr>
<tr>
<td>5.6</td>
<td>Photolithography mask for silver on sapphire transmission line</td>
<td>173</td>
</tr>
<tr>
<td>5.7</td>
<td>Photolithography mask for silver on quartz transmission line</td>
<td>174</td>
</tr>
</tbody>
</table>
Figure 5.8  Transmission line test system interface material cross section  ......... 175
Figure 5.9  PCB interface to transmission lines ........................................ 176
Figure 5.10 Measurement and model of transmission line PIM ......................... 178
Figure 5.11  Electro-thermal PIM dispersion for silver alone, simulated .............. 180
Figure 5.12  Current density simulation of a transmission line ......................... 182
Figure 5.13  Simulated growth of PIM along the length of a transmission line .... 183
Figure 5.14 Simulated PIM versus line length with varied line width ............... 184
Figure 5.15  Simulated PIM versus line length with varied characteristic impedance 185
Figure 5.16 Simulated PIM thermal dispersion versus substrate thickness .......... 186

Figure 6.1  Rectangular patch antenna heat conduction model .................... 193
Figure 6.2  SEM cross section image of silver on sapphire ......................... 202
Figure 6.3  SEM top view image of silver on sapphire ............................. 203
Figure 6.4  SEM image of silver on sapphire constriction resistances ............. 203
Figure 6.5  Photolithography mask for rectangular patch antenna .................. 204
Figure 6.6  Rectangular patch antenna interface material cross section .......... 205
Figure 6.7  PCB interface to the rectangular patch antenna ......................... 205
Figure 6.8  Simulated and measured antenna S-parameters ......................... 211
Figure 6.9  Simulated silver on sapphire antenna current density ................. 213
Figure 6.10 Simulated current density on a matched feed line .................... 214
Figure 6.11 Simulated silver on sapphire antenna radiated E-field ................. 215
Figure 6.12 Simulated FR4 antenna current density .................................. 217
Figure 6.13 Simulated FR4 radiated E-field ............................................ 218
Figure 6.14 Simulated antenna radiation patterns .................................... 220
Figure 6.15 Antenna PIM transmission measurement configuration ................. 221
Figure 6.16 Antenna PIM reflection measurement configuration .................. 223
Figure 6.17 Measured and modeled PIM on rectangular patch antenna ........... 225

Figure A.1  Anechoic chamber constructed blueprint .............................. 247
Figure A.2  Anechoic chamber acoustic insertion loss ............................... 250
Figure A.3  Anechoic chamber acoustic return loss ................................ 251
Figure A.4  Anechoic chamber electromagnetic insertion loss ..................... 252
Figure A.5  Anechoic chamber electromagnetic return loss ....................... 253

Figure C.1  Cancellation system software initialization flow ....................... 287
Figure C.2  Cancellation system software execution flow .......................... 288
Figure C.3  Cancellation system software control algorithm ....................... 289
Figure C.4  Cancellation system software user interface .......................... 290
LIST OF TABLES

Table 4.1 Process Metallization Thicknesses ................................................. 134
Table 4.2 Simulation Model Parameters ......................................................... 137
Table 4.3 Simulation Model Parameters ......................................................... 140
Table 4.4 Simulation Model Parameters ......................................................... 143
Table 4.5 Simulation Model Parameters ......................................................... 146
Table 4.6 Simulation Model Parameters ......................................................... 149

Table 5.1 Transmission Line Simulation Material Parameters ......................... 177
Table 5.2 Transmission Line Dimensions ....................................................... 179

Table 6.1 Antenna Dimensions ................................................................. 210
Table 6.2 Antenna Performance ................................................................. 210
Table 6.3 Ag on Sapphire Antenna Simulation Parameters ......................... 224

Table C.1 Linear System Electrical Specifications ..................................... 294
1

Introduction
1.1 Motivations

Typical communication links require continually growing power in the transmit band and sensitivity in the receive band. Interference from spurious signals in the receive band of a system increases the level of the minimum detectable signal. In narrowband communications systems, such spurious frequency content manifests itself as nonlinear generation of additional frequency content related to the input signal. This new frequency content can be transmitted or coupled to the receive channel, potentially blocking it from use or substantially reducing the channel signal to interference ratio [1–5].

Increasing coverage areas, bandwidth, and capacity usually leads to a need for larger link budgets. Such high budget links are normally found in shipboard antennas, ground stations, antennas for satellite communications, and wireless telecommunication systems. In these high power systems, signal paths must be kept highly linear. Nonlinear frequency content arises from the active components in these systems, but in general these products can be filtered or reduced through feed-forward methods to produce the required linear signal. Intermodulation distortion generated by passive, linear elements such as filters, transmission lines, connectors, attenuators, and circulators cannot be removed in this manner. Distortion of this type is commonly called Passive Intermodulation Distortion (PIM). Passive distortion is a weakly nonlinear effect, but due to the combination of high power transmitters and collocated low noise receivers in communications systems, it becomes the system limiting source of spurious frequency content and thus one of the major factors limiting the ability to develop highly linear systems.

1.2 Overview

Passive intermodulation distortion is the distortion generated by passive components in the system. The processes responsible for PIM generation are weakly nonlinear and abundant, with many processes thought to produce similar distortion amplitudes [6,7]. Processes thought to be responsible for PIM generation include contact nonlinearities such as metal-metal and metal-oxide-metal contacts and material nonlinearities such as ferromagnetics, piezoelectrics, ferroelectrics, carbon fibers, kovar, non-linear material conductivity, and any other material property leading to nonlinear I-V characteristics in the ma-
Attempts have been made to physically explain the sources of PIM, but due to the sheer number of effects and the similarity of their physical mechanisms coupled with the inherent difficulty in PIM measurement, no definitive PIM mechanism has emerged.

A base physical mechanism of PIM is examined in this work, electro-thermal conductivity modulation. Electro-thermal conductivity modulation is present in every facet of passive communications systems. This dissertation aims to explain and derive electro-thermal distortion across the major design components associated with a communications system including lumped components, distributed structures, and resonant structures. The development of a measurement system with unprecedented dynamic range is presented to facilitate measurement of PIM. The results of this work can be used to design higher linearity and reduced co-site interference communications systems.

1.2.1 Passive Distortion Modeling

Exact physics or mechanisms of passive distortion are not known conclusively, thus very few physically-based models are available. Behavioral approaches are frequently used in lieu of physically-based solutions to model the total observed dynamic of a system [3,11–13]. Behavioral modeling works well for repeatable, time-invariant processes, but PIM distortion can be dynamic and can vary widely under the same measurement conditions [4,10,13]. This behavior could be from measurement, as PIM is extremely difficult to measure due to the weakness of the nonlinearity, or could be a property of the PIM process. Behavioral models give a general idea of the trends of device performance but are not generally accurate for high performance design. Physically-based models, if obtainable, allow accurate prediction of system performance. The foremost objective of this work is the production of physical models for dominant passive distortion mechanisms.

1.2.2 Passive Intermodulation Measurement

Measurement of PIM requires system spurious free dynamic ranges on the order of 100 to 160 dBc at high power, depending on the component under test [10,13]. Most commercial spectrum analyzers provide only 75 dB of dynamic range even at optimal input.
conditions. Adding to the difficulty is the requirement that every active and passive component in the system produce less distortion than the required dynamic range. Attenuation could be introduced to reduce or eliminate receiver nonlinearity, but this will both reduce the magnitude of nonlinearities and increase noise. To reduce both noise and carrier level simultaneously, filtering or active feed-forward techniques are needed. Filtering is an extremely effective technique and is used in most PIM measurements to date. Unfortunately, this method is limited by filter bandwidths and the sharpness of duplexer skirts, and is not tunable without physically changing filters to cover other frequency bands [14]. Measurements are limited in this method to several megahertz separation, which is not adequate for measurement of all amplitude modulated signals. PIM is often frequency and amplitude modulation dependent, thus a method allowing accurate PIM measurement regardless of frequency or signal type is needed. Such a system is developed in this work, employing feed-forward cancellation in order to provide tunable, bandwidth independent measurement capabilities for PIM measurement.

1.2.3 Objective: Explain and Model the physical source of PIM

There are four main aspects of this work:

- The analytic derivation of the physics behind electro-thermal distortion in general lossy microwave components and corresponding simulation models.
- The physical derivation and model development for distributed passive intermodulation including forward and reverse wave growth on a transmission line.
- The determination of PIM mechanisms in antennas and a simulation tool to predict antenna PIM performance.
- The development of a ultra-high dynamic range, broadband PIM measurement system based on feed-forward cancellation.
1.3 Original Contributions

A number of original research initiatives were undertaken and are described in this dissertation. The author’s contributions to the field of RF and microwave engineering are summarized below.

1.3.1 Broadband High Dynamic Range Measurement System

Passive intermodulation distortion is a weakly nonlinear process. It is generated in the presence of high power signals. These signals will either saturate the receiver used for measurement or mask the much smaller signals. A system capable of measuring small signals in the presence of large signals is needed to enable PIM measurement.

A PIM measurement system based on feed-forward cancellation is developed. The system reduces the linear applied signal and corresponding noise, increasing the relative magnitude of the distortion products. A formula for automatic cancellation is developed, resulting in a wideband, high dynamic range system well suited for PIM measurement.

1.3.2 Nonlinear Electro-Thermal Theory

Electro-thermal PIM in microwave systems is a baseline for system performance as it exists in every conductor. Electro-thermal theory as it applies to distortion at RF and microwave frequencies is derived from base physics for the first time to the author’s knowledge for lumped lossy components, distributed structures, and resonant structures.

1.3.3 Fractional Electro-Thermal Circuit Model

Circuit models for coupled physical phenomena are desirable to minimize simulation time in complex circuits. Often finite difference and element methods are too computationally intensive to allow coupling electrical circuit simulation with other physics in nonlinear systems.

A highly accurate physically based circuit model was developed based on heat conduction theory and the fractional derivative. The model allows accurate simulation of electro-thermal coupling in large scale circuit simulators over all frequency.
1.3.4 Foster Expansion Electro-Thermal Circuit Model

Commercial simulators do not possess routines for computation of fractional derivatives. An approximate method is desirable to impart the ability to simulate electro-thermal processes with a slightly diminished level of accuracy. A circuit model based on a foster fit of the electro-thermal response is developed to facilitate approximate simulation in commercial simulators over limited bandwidths.

1.3.5 Distributed PIM Model

Transmission lines and circuit elements constructed from them generate PIM that varies with transmission line structure, dimensions, and materials. Rules of thumb exist to minimize PIM on these structures, but a physically-based theory is needed to accurately model performance. A physically-derived model is presented for distributed PIM generation.

1.3.6 Resonant PIM Model

Antennas generate PIM that varies with structure, dimensions, and materials used to construct the device. Little is known of PIM generation mechanisms in antennas. A base physical mechanism of PIM generation is isolated and a physically-derived model is presented for resonant PIM generation.

1.3.7 Electromagnetic and Acoustic Anechoic Chamber

An anechoic chamber is designed and constructed to allow measurement of both electromagnetic and acoustic signals in a shielded environment with reduced reflection. The chamber allows single and mixed domain measurement.

1.4 Dissertation Outline

Chapter two of this dissertation gives a literature review of PIM processes, PIM research, distortion analysis, distortion measurement, and mathematical techniques necessary for PIM modeling.
Chapter three discusses broadband high dynamic nonlinear measurement. Feed-forward cancellation theory is presented along side techniques for implementing a linear feed-forward system. Measurement of nonlinearities require that the components be more linear than the device under test. Linearity against both forward and reverse waves are presented for each type of component in the developed measurement system. Reflections, finite isolation, and coupling are discussed in the context of linearity effects on the system. System performance for two-tone PIM measurement and in-band distortion in wideband signals is presented.

Chapter four delves into the electro-thermal process responsible for PIM in lossy lumped components. The heat conduction environment for general lumped components and the corresponding electrical and thermal domain coupling is analyzed resulting in a fractional electro-thermal model which correctly accounts for the long tail transients seen in these devices. A circuit model based on the fractional derivative and an approximate model based on foster expansion is presented. Several cases of interest including microwave terminations, platinum attenuators, and integrated circuit polysilicon electro-thermal distortion are explored.

Chapter five takes a look at the PIM in a distributed structure such as a transmission line. The PIM is found to be electro-thermal in nature in each infinitesimal lossy section of the transmission line. The growth and decay of PIM with line length in both forward and reverse propagation directions is presented. Transmission line samples, silver on quartz and silver on sapphire, are manufactured isolating the electro-thermal process. A simulation model for prediction of PIM from material parameters and transmission line dimensions is presented.

Chapter six explores PIM production in resonant structures. Antennas are manufactured on both FR4 and sapphire to isolate the distortion, which is shown to be electro-thermal in nature. A physical model based on current distributions is discussed leading to a simulation model for prediction of PIM on an antenna.

Chapter seven contains a summary of the work performed and lists the significant outcomes of this work.
1.5 Published Works

1.5.1 Journals


1.5.2 Conferences


1.6 Unpublished Works


2

Review of Nonlinear Analysis and Measurement Techniques
2.1 Introduction

Linear system behavior is strived for but in reality can never be achieved due to the intrinsic nonlinearity of all systems. The non-ideal behavior of systems results in the generation of new frequency content. This content is of both harmonic and cross modulation nature. Harmonic distortion is restricted to all new signals at integer multiples of the input signal frequency. This type of distortion is usually not problematic as it can be bandpass filtered from the signal of interest. Unfortunately, cross or intermodulation distortion, is much more difficult to design out of a system due to the proximity in frequency of many cross modulation components to the original signal.

Intermodulation distortion occurs when signals with more than one frequency are input to a system. As implied by the name, each frequency component crosses or mixes with every other component in the system, even other nonlinear frequency content as prescribed by the nonlinear characteristic of the system. The spectra associated with this process yields baseband components, as well as the in-band co-channel and adjacent band frequency components at the fundamental and harmonic responses. While baseband effects can usually be countered through filtering, in-band components are quite problematic. In-band co-channel components are at or near the frequency of the fundamental. Components falling on the fundamental sum with the fundamental as either correlated or uncorrelated noise, reducing the signal to noise ratio of a channel. In-band adjacent distortion is possibly the most dangerous of all, as its mixing products can fall within the receive band of either the generating system or other systems operating in nearby frequency bands. As communications systems continue to add more capacity, commercial spectrum congestion increases proliferating mixing products. Generally adjacent distortion is too close to the carriers to be filtered out, leading to the need for nonlinear system design to eliminate or reduce these components. Nonlinear modeling breaks distortion processes into that of active devices and passive devices due to the difference in the relative strength of the the nonlinearities.

Nonlinear system design has conventionally been limited to active devices, which are known to exhibit strongly nonlinear behavior if operated outside of their linear regime. Nonlinear responses can be avoided in an active system by operating the circuit under small signal conditions, or using feedback or more exotic techniques such as pre-distortion to output a linear response within system requirements. In general active devices are designed
to provide a prescribed linearity and are not the limiting factor in system performance.

2.2 Passive Distortion

Passive nonlinearities, contrary to active devices, are weakly nonlinear. Due to the weakness of the nonlinearity, coupled with modeling difficulty, passive components are conventionally modeled as linear. This assumption holds until system power levels and dynamic range requirements become large, at which point passive nonlinearities can become the ultimate system limit. Unfortunately passive processes are difficult to measure due to the massive dynamic range needed to measure the effects [13]. Unlike active distortion, where the device physics leading to distortion are well modeled, passive devices have many potential mechanisms that can result in distortion. Most of these processes can happen in the same physical situation, causing great difficulty in confirmation of the theory associated with passive processes [1–8, 15]. Enabling systems to be designed for low or no passive intermodulation distortion (PIM) requires definitive knowledge of the dominant processes responsible for PIM. Arriving at any conclusion about the nature of PIM requires a broad knowledge all the possible contributors, which include many different subsets of physics. A review of the prominent factors potentially affecting passive components is given, differentiating between contact and material based nonlinearity and the corresponding possible nonlinear processes underlying each.

2.2.1 Metal-Insulator-Metal and Metal-Metal Contact Nonlinearities

At any contact between components, there exists the potential for distortion generation through a variety of mechanisms. These mechanisms spring from two general physical situations, the metal-insulator-metal contact (MIM) and the metal-metal (MM) contact. Thin insulating layers, either oxides or sulphides, are natural to all metals save gold resulting in the initial formation of MIM structures at low pressure [11]. As the surfaces are more intimately contacted, the insulator can be penetrated leaving a metal to metal connection. The two types of contact can occur in many different manners and concurrently, dependent upon both the surface topography and the pressure of the contact [10–12, 16].

Nominally, the contact area of two conductors in a connector or waveguide would
just be the size of the conductor pin or the surface area of the waveguide. Unfortunately, the nature of such contacts are not as simple as would be preferred, as no material can be completely smooth. Material surfaces do not enjoy the bond stability of the material interior [17]. Minimization of the free energy associated with dangling bonds at the material surface leads to a different lattice structure near the surface, which often grows in islands due to lattice imperfections. These islands grow at various rates and to various heights, producing extremely irregular contact areas. Contacting two surfaces of this nature is closer to contacting many needles of various lengths than a flat surface [11,16]. It becomes apparent that surface topography limits the actual contact area to a tiny percentage of the macroscopic contact area. Decreasing loss and increasing reliability would require the most contact area possible, which can be accomplished to some degree through increasing connection pressure [11].

The variation of connection pressure will cause the surface properties to change over three domains of mechanical deformation, elastic, elasto-plastic, and plastic. In the elastic regime, the pressure is low enough that the surfaces recover their original shape when taken apart. This region could be thought of as a slight bending of the needles, where further contact becomes possible to the angling of the needles. The plastic regime of deformation is the other extreme, forever altering the surface properties of the contact. This region will provide the most contact area possible and can rupture the insulator layer, resulting in a much higher percentage of metal-metal contacts. The elasto-plastic region is a partial permanent deformation of the surface, which could be thought of as a permanent bending or breakage of longer needles while shorter needles are not permanently affected.

Each of the physical structures, MIM and MM contacts, have several of their own, distinct nonlinear mechanisms. Metal-insulator-metal structures are most susceptible to tunneling and thermionic emission. Metal to metal structures can create diode like junctions due to differences in metal work functions as well as nonlinear contact resistances due to thermal processes including thermal expansion of the material and thermal resistance variation.
Tunneling

In structures where two conductors are separated by an insulator, classically it was thought that no current could flow between the conductors as the insulator provides a potential barrier that the electrons did not have enough energy to overcome, shown in Fig. 2.1(a). With the advent of quantum mechanics, the finite probability of an electron being in the second conductor was found to be non-zero due to electron wave function behavior, pictured in Fig. 2.1(b). As long as the electron has enough energy, the wavefunction amplitude can exponentially decay to the other side of the barrier, where it can continue to propagate with a reduced amplitude wavefunction. The transmission and reflection coefficients of a given barrier structure can be calculated by solution of the Schrödinger equation:

\[ i\hbar \frac{\partial \Psi(x,t)}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + V\Psi(x,t). \]  

The solution to this equation is just the appropriate combination of forward and backward waves, given by

\[ \psi(x) = Ae^{ikx} + Be^{-ikx}, \]  

where there exist a wavefunction for each potential region [18]. Continuity of the wavefunction forces the matching of the boundary conditions, resulting in transmitted and reflected waves at each interface. This situation is analogous to scattering in microwave signals, and is exactly what is used to compute the transmission and reflection coefficients. Of course, once the transmission coefficient has been obtained, the total tunneling current is just the total number of electrons approaching the barrier times the transmission coefficient. The transmission coefficient is the square of the transmitted amplitude divided by the square of the incident amplitude, which due to the exponential decay of the wavefunction through the barrier results in an exponential dependance of the tunneling current on the electron energy coming from the applied field.

Many analysis exist of tunneling in contacts and other similar structures. Perhaps one of the most useful models for tunneling is the Simmons model [19], which provides current voltage relationships for MIM structures including image force and irregularities of the insulator. The model allows equations to be derived for low, intermediate, and high fields. The high field case, or field emission, is given by Fowler-Nordheim theory, but is
Figure 2.1: A metal-oxide-metal barrier under: (a) no applied potential where no tunneling occurs, (b) and an applied potential allowing an electron with sufficient energy to tunnel to the metal on the lower energy side of the barrier.
generally not applicable in practical applications. The Simmons’ model is sufficient for low fields and medium fields, which is our area of interest. The Simmons’ model under medium field for a generalized barrier provides the tunneling current, given by [19]

\[ J = \left[ 6.2 \times 10^{10} / (\beta \Delta s)^2 \right] \left\{ \varphi \exp \left( -1.025 \beta \Delta s \varphi^{-5} \right) - (\varphi + V) \exp \left[ -1.025 \beta \Delta s (\varphi + V)^{-5} \right] \right\}, \quad (2.3) \]

where \( \varphi \) is the mean barrier height, \( V \) is the applied voltage, \( \Delta s \) is the barrier thickness at the fermi level, and \( \beta \) is a correction factor that is approximately valued unity.

Appreciable currents, as related to PIM generation at contacts, require the insulator to be quite thin, usually 10 nm or less [11]. This requirement is easily met as native oxide layers usually range from 2—3 nanometers all the way up to 10 nm, assuming no rusting effects have occurred. In [16], the authors suggested that PIM tunneling was too low to appreciably effect systems when insulators were more than 20 angstroms, or 2 nm, in thickness. Analysis of the Simmons’ equation for a single contact point provides insight into this situation. As seen in Fig. 2.2(a) - Fig. 2.2(c), for a 1 \( \mu \text{m}^2 \) area contact, the power of third order distortion due to tunneling reduces approximately 100 dB per 5 angstrom increase oxide thickness as the thickness of the oxide is increased from 10 angstroms to 20 angstroms.

The effective contact area of a connector is limited to a small percentage of the actual contact area due to conductor surface roughness [16]. Much of the contact area will be metal-metal contacts, which will compete with the tunneling process for current transport. The small contact area of tunneling contacts and the requirement of a 1—2 nm thin oxide layer for appreciable tunneling currents make tunneling a minor contributor to passive intermodulation distortion in most metal contacts that do not use aluminum.

**Thermionic Emission**

Another process capable of appreciable currents in a MIM structure is thermionic emission. This process considers the possibility of thermal energy causing electrons to jump the potential wall formed by the insulator into the other conductor, shown in Fig. 2.3. Originally this phenomena was evaluated by Richardson and Dushman at a metal-vacuum interface, which exhibited an exponential dependence on temperature. In the case of metal-metal contacts, the dependence is also exponential in the applied voltage. This process was
Figure 2.2: Simulated fundamental, third order, and fifth order power resulting from the tunneling current of a 1 MHz tone separation two-tone signal applied to a single Al-Al$_2$O$_3$-Al contact of area 1 $\mu$m$^2$ in a 50 $\Omega$ connector for: (a) a 10 angstrom Al$_2$O$_3$ layer; (b) a 15 angstrom Al$_2$O$_3$ layer; (c) and a 20 angstrom Al$_2$O$_3$ layer. Results are calculated with (2.3) and a Fourier analysis of the current density.
thought to be independent of barrier size, but Simmons showed not only that thermionic current is dependent upon barrier thickness, but also that it decreases when the potential applied is more than the barrier height [20].

Thermionic emission is important to the accuracy of tunneling currents and clearly demonstrates exponential dependencies on both temperature and applied voltage. This process is secondary to tunneling and represents a correction necessary to obtain correct tunneling currents. The percentage increase in tunneling current due to thermionic emission is small [20] and is not considered further in this dissertation.

**Contact Potential**

When two metals are in contact, if their work function is not exactly the same, a contact potential is formed which acts as a small diode. At the interface between the metals the fermi levels must be equal. Alignment of the fermi levels requires charge to transfer from the high work function face to the low work function face. Since charge transfer has occurred between the metals, a field must exist at the interface which defines a contact potential. This contact potential formation requires only a small difference in work functions. While this is an obvious feature of dissimilar metals, what may be less obvious is the work function
dependence on surface topography even of the same metal. Even in surfaces of the same metal, work functions can vary up to 5% based on surface configurations [17, 21], making weak diode formation possible, if not probable, for any metal-metal interface. In a real interface, work function differences can arise from island formation and random surface roughness. It should be noted that work function differences are very small for low PIM materials such as copper, silver, and gold, which respectively demonstrate mean values of work functions at 2.44, 2.43, and 2.43 electron-volts [17].

**Contact Resistance**

Metal to metal contacts naturally have a resistance associated with them. In Fig. 2.4, a generic contact situation is shown depicting the concentration of current into the contact zones. Further confinement of the current density is forced by the skin effect, resulting in elevated losses in a small contact region on the exterior of the contact. Due to this increase in current density within these zones, nonlinear resistance from thermal expansion and dynamic resistance modulation becomes possible. Temperature variation at a contact can result in the constriction or expansion of the contact, or even thermally modulated resistance of the contact, although this has not been studied in detail [22, 23].

![Figure 2.4: Constriction of current at metal-metal contacts due to the decreased area of a contact region compared to the bulk metal.](image)
2.2.2 Ferromagnetic Material Nonlinearity

Ferromagnetic materials such as iron, steel, cobalt, and nickel are often used for mechanical structures supporting antennas and construction or plating of connectors and cables as well as in the production of transformers and circulators. The metals are known to produce much higher levels of distortion than structures composed of diamagnetic metals [24–27] such as copper, silver, and gold. The basis for the nonlinear nature of these materials lies in the magnetic field dependance of their permeability. A brief review of magnetization is necessary in order to explain the nonlinear behavior of ferromagnetic metals.

The origin of magnetic moments arises on an atomic level from both the spin and orbit of an electron. An atom's total magnetic moment is the summation of electron spin-spin, spin-orbit, and orbit-orbit interactions. Since total spin must be maximized in order for magnetization to occur, it is only witnessed in materials with incomplete electron shells, which is strongest when near half full. A further restriction of ferromagnetism is the necessity for the atomic structure to allow the magnetic dipole moments to align in parallel. Of course the energy from this process, coulomb repulsion, must be minimized along with all addition magnetic energies in the material, which include magnetostatic, magnetostrictive, and magnetocrystalline energies [21,28].

Magnetorestrictive energy is the external magnetic field generated around the material. Since this field can and will perform work, it must be minimized. If all magnetic dipole moments in the material align in a single direction, the largest possible external field would be generated while at the same time minimizing coulomb repulsion. This statement suggest minimization of the potential can be accomplished by increasing the number of magnetic domains in such an order and directionality as to limit the external field. Magnetostrictive energy also prefers smaller domains in order to minimize the elastic strain induced on the lattice. Counterbalancing these requirements is the magnetocrystalline energy, which desires large domains in order to minimize the total number of domain walls. Domain wall minimization arises from the crystal property of unequal magnetization impedance in various crystallographic directions. As the domain magnetic alignment will be in a low impedance direction, the domain walls must necessarily be in high impedance directions, increasing system energy. The balancing of these energies leads to a structure with many domains oriented with 90 or 180 degree walls with respect to each other [21,28].
Figure 2.5: Magnetic hysteresis loop defined by saturation magnetization, $B_s$, remnant magnetization, $B_r$, and the coercive magnetic field, $H_c$.

The nonlinearity of ferromagnetic materials results from these magnetic domains, specifically from the irreversibility of the magnetization and demagnetization process, known as hysteresis. As a magnetic field is applied to the material, the domains in that direction will grow at the cost of differently oriented domains, overtaking crystal imperfections until there is only one domain. Upon removal of the external forcing field, a demagnetizing field from the material will result in the reformation of multiple domains. Hysteresis occurs because the demagnetization field is not strong enough to overcome the crystal defects, effectively increasing the number of domain walls and magnetizing the material. This hysteresis loop is shown in Fig. 2.5, and is characterized by the remnant magnetization, $B_r$, saturation magnetization, $B_s$, and the coercive magnetic field, $H_c$. Domain walls expand and contract in a manner consistent with this hysteresis curve when electromagnetic signals are applied to the material, producing passive distortion [24, 27].

2.2.3 Piezoelectric Material Nonlinearity

Piezoelectric materials respond to exterior forces on the crystal structure by changing shape. Shape change results in dipole moment formations within the material, possibly with a net polarization. They are useful for ultrasound production as well as acoustic
mode SAW and BAW filters. Ferroelectric materials are a special subset of piezoelectric materials, and are of great interest as they allow for tunable capacitors, which can provide dynamic matching networks. Ferroelectric materials are also increasingly being integrated as filler materials in circuit board manufacture. Unlike standard dielectrics that do not keep a residual polarization, ferroelectric materials exhibit polarization memory with reversible polarity. Much like ferromagnetic devices, the electric field dependance on the polarization vector leads to nonlinear behavior in these devices [29,30].

In a piezoelectric material, each unit cell of the material is non-centrosymmetric. If the unit cell had a structure that possessed a center of symmetry, any dipole moment generated by one cell would cancel the corresponding cell by symmetry. The center atom of the unit cell of this material must thus be able to occupy a non-equilibrium position due to local dipole moments throughout the material. This dipole moment can either be spontaneous, as in the case of ferroelectric materials where a net polarization accumulates, or can occur from stress in non-spontaneous polarization, as occurs in quartz [29].

Much like the case of ferromagnetic materials, ferroelectric nonlinearity is based on domain growth and reversal. In this case stray electric field energy, instead of magnetic field, must be minimized. The material, before field application, will have domains that are aligned such that dipole moments of individual domains meet at 90 and 180 degree angles in a manner that minimizes all excess field energy outside the material. When a field is applied to one of these directions, the direction of parallel alignment with the field is reduced in energy while all other components are increased in energy. Minimization of internal energy will cause the parallel domains to grow, overtaking other domains and crystal defects, eventually resulting in only domains directed with the field. When the field is released, the material will once again seek to minimize the energy by forming new domains aligned in other directions. Since the material cannot return to a neutral position in the cells, a net polarization will remain which is irreversible, although application of a reverse field can change the polarity of the polarization [29–32]. This process results in a hysteresis loop, shown in Fig. 2.6, and is characterized by the remnant polarization, $P_r$, saturation polarization, $P_s$, and the coercive electric field, $E_c$.

Another property of the lack of centrosymmetry in piezoelectric material unit cells is the direct, linear conversion of material stress to electric field and electric field to material stress. This property is quite often used in acoustic mode filters, sonar, and ultrasonics.
This property is often overlooked, but still important to passive distortion as it provides a direction mechanism for acoustic and mechanical distortion to couple into electrical systems, with only some applied stress on the material present. Since coming into contact with any other material provides stress, any material, even electrically neutral materials, that can exhibit non-spontaneous dipole moments under stress can be potential medium for vibrational and acoustic coupling. This dependence can result in either modulation of electrical signals, or the more troubling possibility of the introduction of acoustic nonlinearities.

2.2.4 Acoustic Nonlinearities

Acoustics are increasingly integrated with electrical systems through piezoelectrics as transducers, filters, and even memory components. These systems, while valuable, also present new challenges in nonlinear system design. Nonlinearities and spurious signals from acoustic systems are often converted with high efficiencies into the electrical domain, demanding their analysis as distortion generators. Acoustic nonlinear processes due to inelasticity is briefly reviewed here.

Inelasticity is a natural consequence of the energy of gases, liquids, and solids. Due to interatomic forces, as atoms or molecules are moved closer and closer together, the
repelling force between them increases exponentially [17, 21]. The forces required to reach this level of repulsion are usually quite high, leading to the linearization that is elastic behavior under the condition the system is operated at a small power level. Elasticity predicts a linear relation between the applied force and the resultant motion. When the system operates outside of this small force requirement, motion becomes a nonlinear function of the applied force, resulting in generation of harmonic and intermodulation distortion.

These effects are inherent in the atomic structure of materials, but barring a unique coupling as in piezoelectrics, is quite small. Inelastic effects are amplified, as compared to the perfect lattice or volume, through defects in materials such as grain boundaries, cracks, contacting surfaces, and bonded surfaces. The nonlinearity in these cases arises from the inelasticity of a surface or coupled elasticities that differ [33–37].

2.2.5 Electrical Conductivity Nonlinearity

Electrical conductivity has long been known to have a dependence on the temperature of the metal. Every metal has a finite conductivity due to electron interactions with lattice imperfections, dopants, the lattice itself, and other electrons. Temperature affects the situation by increasing the kinetic energy of each component. For each of the lattice sites this translates to an increase of the amplitude of random thermal variation. The cross section of scattering interference conduction electrons feel is dependent on this random thermal variation, resulting in a thermal dependence of the electrical conductivity. Most metals exhibit this phenomena as an increase in resistance with increasing temperature, usually linearly, except for a few metals and alloys such as Nichrome which exhibit nonlinear dependencies on temperature [38]. Semiconductors also exhibit similar dependencies on temperature, but instead of an increase in resistivity a decrease in resistivity occurs. In semiconductors the increase in conductivity occurs from the thermal excitation of electrons to the conduction band, offsetting the increase in the scattering cross section. Active nonlinearities in semiconductors are more strongly nonlinear than thermal processes, thus thermal nonlinearities in metals will receive the most consideration.

The transfer of electrical energy through a metal will always result in some energy loss due to the finite conductivity of the metal. This energy is transferred through scattering to the lattice, which then transfers the energy as phonons through the lattice until it can
be expelled as heat out of the material. This process is simply represented by the thermo-
resistance effect [39]:

\[ \rho_e(T) = \rho_{e0}(1 + \alpha T + \beta T^2 + \ldots). \]  

(2.4)

Here \( \rho_{e0} \) is the static resistivity constant and \( \alpha \) and \( \beta \) are constants representing the temper-
ature coefficients of resistance (TCR). The heating through the material is only dependent
on the resistivity and current density through that resistivity in the relationship

\[ Q = J^2 \rho_e, \]  

(2.5)

where \( J \) is the current density vector.

While there is no dispute that this process is nonlinear, thermal processes are
quite slow, especially compared to any signal above a few kilohertz. This fact has steered
most research away from this phenomena, as it suggested that thermal processes could
only have average effects on any appreciably high frequency signal. The exception to this
train of thought concerning passive intermodulation distortion was suggested by Wilcox
and Molmud [23]. They ignored the slow nature of thermal signals and assumed they would
follow the applied microwave signals instantaneously due to the confinement of heat to the
conducting layer of the metallization. No basis was offered for this assumption. The result
of their study was the prediction in coax cable of intermodulation distortion on the order of
\( -150 \) to \( -140 \) dBm with a two-tone input of 45 dBm per tone, approximately 185 dBC below
the carrier. Such low levels of distortion would suggest other processes to be dominant,
but further analysis of current confinement and surface roughness suggests substantially
higher distortion. As material nonlinearities represent the ultimate performance a system
can obtain, further research of such nonlinearities is warranted. While ferromagnetic and
ferroelectric sources can be eliminated by material choice, electro-thermal nonlinearity can
never be eliminated. Assuming proper design and connection of components, thermal based
distortion is the limiting factor in any component.

2.3 Distortion Analysis Methods

One of the primary problems associated with passive intermodulation distortion
is the lack of accurate models of passive distortion phenomena. Measurement is virtually
impossible to match exactly with theory as distortion can invariably be attributed to many different sources. In general, modeling of PIM processes is either based upon the derivation of the constitutive equations and their subsequent nonlinear expansion, or the use of a general nonlinear element fit to measurement to represent all possible processes in a given situation. Expansion in either case is usually accomplished through a series expansion of the nonlinear equation, unless analytic solution of the nonlinear equation set is possible. Expansion of constitutive equations is done either through a power series or volterra series method, while a general nonlinear element implies a behavioral model, which is a measurement based fit of the nonlinear behavior of the element.

2.3.1 Power Series

The power series of a nonlinear equation is an effective tool for modeling of a nonlinear system due to the inherent prediction of harmonic and intermodulation distortion. Particular frequency components of the output signal are given by the corresponding order of the series expansion, greatly contributing to intuition of the system mechanics. The power series is represented by,

\[ y(t) = \sum_{n=1}^{N} a_n x^n(t), \]

where \(a_n\) are the series coefficients, which can be either real or complex, and are determined by a Taylor series expansion of the nonlinear constitutive equations [40, 41]. Since a Taylor series expansion is used to obtain the coefficients, the method is inherently localized around the point it is expanded about. Due to the localization, the solution of the nonlinear system is only applicable for a given input range, which may or may not be problematic, dependent on the possible input range [42]. The inherent lack of memory limits or eliminates completely their applicability in multi-physics problems where memory can occur between processes.

2.3.2 Volterra Series

In linear theory, systems can be described by their transfer functions, which tell exactly what the response to a given input will be. A natural extension of this premise to nonlinear systems is the Volterra series, which provides an impulse response for each order
of the expansion, referred to as a Volterra kernel $h_n$ for arbitrary order $n$. The output of a system is then given by the $n^{th}$ order convolution integral of the order dependently delayed input signal and the Volterra kernel. This operation can be expressed as

$$y_n(t) = \int_0^t \int_0^t \ldots \int_0^t h_n(t - v_1, t - v_2, \ldots, t - v_n)x(v_1)x(v_2)\ldots x(v_n) \partial v_1 \partial v_2 \ldots \partial v_n.$$

The Volterra series has advantages and disadvantages, as with any technique. The greatest advantage of this technique is that it provides approximate analytical solutions to mildly nonlinear systems, which allows conclusions about the mechanisms effecting a system to be drawn. The approximate nature of the analytic solution can be increased in accuracy like every other series, by simply increasing the number of terms. Kernels up to $5^{th}$ order have been extracted for electro-thermal simulation [43], but in general the Volterra series suffers from convergence problems prohibiting its use past a fifth order expansion [42]. The complexity of finding the next kernel is of factorial complexity, dependent on the combinations of all the previous solutions. The need for intuition about the physical processes producing PIM greatly limit the use of this technique in PIM modeling, although it is still implemented for some well defined processes.

### 2.3.3 Behavioral Models

Behavioral models attempt to circumvent the need for a physically linked process in favor of a model that simply approximates system performance and characteristics over a given range. These models are usually a mathematical fit of measurement data, where no thought is given to the physics of the situation. The promise of this type of model is the quick cycle time and speedup in simulation potentially gained. Behavioral models have been implemented for PIM processes to provide some level of prediction capability for desperate system designers. Waveguide contacts and electro-thermal simulation are two examples of PIM behavioral models based on power series and volterra series expansions, respectively [11,43]. The goal of this work is to provide knowledge of the physical processes producing PIM, eliminating behavioral modeling as a feasible method for describing system behavior.
2.3.4 Physical Models

It would be desirable, in any situation, to know exactly what causes every effect in a system. Although it is improbable, if not impossible to have complete knowledge of a process, physics based models get closer to providing this than any other technique. Only by understanding the physics of each possible process can the dominant contributors to an effect be determined. Physical models are derived from the differential equations for each process in the system. In many cases these processes are actually coupled, as is the case for both electro-thermal distortion and tunneling currents. When processes even with a linear dependence are coupled, nonlinear systems result which can be very difficult to solve. Two options present themselves for solution to such a problem, analytic simplification and numerical methods.

Attempting the solution of nonlinear differential equations can be quite a tedious task. Unlike most circuits which are ordinary differential (ODE) equations of arbitrary order, passive distortion processes often involve coupling of an ordinary differential equation to that of a partial differential equation (PDE). Each equation itself could be nonlinear, or the system can become nonlinear from the coupling itself. Actual use of any derived formula depends on its ability to interact with other external equations, or circuits. A complete solution over a whole domain for the PDE is usually untractable without numerical methods, so the system is either simplified into lower order systems that can be modeled or numerical simulation of the complete system is implemented. Simplification into lower order systems can be accomplished through several mathematical methods, of which fractional calculus system reductions are focused on in this work due to their natural application to diffusive and long memory systems. Upon simplification and solution of the linear versions of each process, they can be coupled again, and expanded in terms of the coupling equation. This approach can lead to analytic solutions containing the bulk of the nonlinear behavior of a system. The region of validity of such a solution is that the nonlinear behavior of each process by itself must be weak compared to that of the coupled system, which is the case in many, but not all, passive phenomena coupling.

Numerical solution is a resource intensive process offering tradeoffs between accuracy and speed. Three dimensional modeling is the most resource intensive method but provides a complete solution over a whole domain. Finite difference, finite element,
and boundary element methods are the main workhorses used to accomplish such solutions. These methods break a domain into small pieces, called meshing, then relate the derivatives at the boundaries of the individual cells. Unfortunately, the time it takes such simulators to solve nonlinear equations is cubically dependent on the size and meshing of the device equations. Two dimensional simulation works in the same manner, but reduces the simulation dependence to square law at the cost of accuracy. Both of these methods are too resource intensive to use if more than a few elements need nonlinear solution. In an effort to allow mass simulation, one dimensional models are implemented that represent the lowest accuracy with speedy solution. Such models seek to emulate a reduced version of a PDE by a few circuit elements, usually resistors, capacitors, or inductors. System frequency responses are fit by partial fraction expansions of the system behavior leading to bandlimited models that can be fairly accurate over a given bandwidth. As this is simply a pole-zero fit, the number of components needed to model wider and wider bandwidths grows almost linearly. If wide band signals are used in a system with many components, these models can also become unwieldy.

2.4 Distortion Measurement

Any model of a system is only as good as its correlation with the process it models. Theory alone is not enough to guarantee the model’s correlation with the physical system. Measurement is needed to test theory and provide relevant parameters needed by the model in order to ensure correlation with real devices. The measurement of nonlinear devices must be reviewed to provide insight into the optimum method in a given situation for confirming theory and retrieving model parameters.

2.4.1 AM-AM and AM-PM Characterization

AM-AM measurements characterize the relationship between input signal amplitude and output signal amplitude at the fundamental frequency, shown in Fig. 2.7(a). AM-AM, as the name implies, only accounts for the nonlinear, memoryless amplitude characteristics of a system, usually through a polynomial model. This technique is useful for characterizing gain in active systems but lacks phase information. Phase data can be ob-
tained from AM-PM measurements, which describe the phase change of the output signal with input signal amplitude, shown in Fig. 2.7(b), allowing inclusion of some types of memory.

A measurement setup capable of both AM-AM and AM-PM tests is shown in Fig. 2.8. The stimulus is produced by the signal generator and then split for the separate measurements. During AM-AM measurement, the stimulus is applied to the test device and measured with the spectrum analyzer. AM-PM measurement recombines a phase shifted version of the original stimulus with a variably attenuated version of the test device output. The output is then canceled using the phase shifter, where phase shift will be 180 degrees plus the cancelation phase.

Although useful in active systems, these measurements are not of primary concern in passive systems where there is no gain. Harmonic and intermodulation products cannot be defined by these techniques alone, leading to the need for other measurements to define passive systems.

2.4.2 THD Characterization

Another single tone measurement that considers the system harmonic distortion generation is total harmonic distortion (THD). This technique is a comparison of the measured power at the harmonics of the signal and the power of the fundamental, shown in Fig. 2.9. The AM-AM path of the measurement system of Fig. 2.8 can be used to complete this measurement. While quite useful in audio systems where harmonics can fall within the operational bandwidth prohibiting filtering, in systems where harmonics fall outside of the bandwidth of interest this measurement provides little useful information to designers. Of primary importance are terms generated within the bandwidth of interest, usually arising as the result of intermodulation. As total harmonic distortion cannot account for intermodulation distortion (IMD) products, further measurements are required in order to characterize the system.
Figure 2.7: Measurement output from AM-AM and AM-PM tests.

Figure 2.8: Test configuration needed to implement both AM-AM and AM-PM measurements composed of signal source, splitters, variable attenuator, phase shifter, combiner, switch, and spectrum analyzer.
2.4.3 Two-Tone Characterization

Quantification of in-band distortion, or intermodulation products, requires at least two distinct frequency components to be present at the system input. The minimum stimulus for such an excitation is a two-tone signal, which is simply two sinusoids of different frequency, shown in Fig. 2.10(a). This type of input signal provides all information from a single tone stimulus, plus cross product combinations of the fundamental, harmonics, and other cross products, as pictured in Fig. 2.10(b).

While not as accurate to final system performance as the actual communications signal, two-tone measurements are the foundation for providing understanding of intermodulation processes, as they allow deterministic, analytic analysis of the distortion process. Two-tone tests constitute the standard in distortion measurement for insight into in-band performance and thus the final system performance and requirements. They are relied upon in this work to examine the fundamental mechanisms of distortion and to provide an efficient basis set for the construction of nonlinear analysis. They can be conducted by generating two stimulus signals, usually from separate, isolated signal sources, before being combined and applied to the test device as shown in Fig. 2.11. Great care must be taken in these measurements as two-tone signals interrogate every component in the test chain. Its validity is then limited by the linearity and isolation of system components.
Figure 2.10: Input stimulus and output spectrum of two-tone measurement.

Figure 2.11: General two-tone test configuration.
2.4.4 Multi-Tone and Band-Limited Continuous Characterization

Two-tone measurement techniques fail to provide information on an important parameter in communications systems, co-channel distortion coincident with fundamentals. Distortion components that fall on top of fundamentals affect the signal to noise ratio of the channel, but are impossible to measure in two tone tests due to the strength of fundamentals in comparison to the mixing products at those frequencies. Accounting for these distortion components can be accomplished by using bandlimited continuous signals such as pseudo-randomly digitally modulated carriers, multi-tone signals, and bandlimited noise. Since these signals have many contributions at a given frequency, the non-correlated intermodulation distortion at fundamentals can be determined by notch filtering of the input signal at the intended measurement frequency. This measurement, the noise power ratio (NPR), is unable to determine correlated intermodulation. These components can be found through the co-channel power ratio test (CCPR) which measures both correlated and uncorrelated intermodulation products by removing, or cancelling, the original input stimulus at the output, bandlimited gaussian white noise. Of course the drawback of this approach for PIM process exploration is that these signals are statistically defined, adding considerable complexity to the mathematics of already complicated and unknown systems.

2.4.5 Passive Distortion Measurement

Application of multi-tone and even higher complexity signals seemingly give us the most complete subset of information about the distortion processes of a system. In active systems, the knowledge of nonlinear co-channel contributions, both correlated and uncorrelated, is needed for correct operation and design of the system. Passive systems are not strongly nonlinear as in the active case, lessening the need for knowledge of in-band distortion coincident with fundamentals. More important are the intermodulation products, which can fall in receive bands causing interference leading to degradation of signal to noise ratios and possibly even blocking communication channels. Due to the general lack of knowledge as to what the processes are that produce distortion in passive devices, a deterministic, mathematically tractable signal that provides mixing products is an ideal test candidate. Two-tone testing provides exactly this tractable analysis of test results. It is used throughout this work in order to analyze the physical processes associated
with passive distortion. Unfortunately, measurement of passive distortion is virtually never as simple as application of two tones to a passive device, as passive processes produce such weak distortion that the dynamic range requirements become intensive.

Passive components are often so weakly nonlinear that the distortion they produce is inmeasurable by conventional systems even at several watts of input power to the passive device, due to the extremely large magnitude of the probe stimuli [44]. The output of these devices must be pre-conditioned in some manner in order for spectrum analyzers or vector signal analyzers to be able to measure the products of interest without generating their own distortion, masking the intermodulation signals. This pre-conditioning must necessarily accomplish the reduction of the probe stimuli magnitude while not affecting the desired distortion products. Attenuation could be introduced to reduce or eliminate receiver nonlinearity, but this will both reduce the magnitude of nonlinearities and increase noise. Reducing both noise and carrier level simultaneously to provide the dynamic range needed for passive distortion measurement requires filtering or active feed-forward techniques [45].

Filter Based Methods

Bandpass filtering is normally applied to RF systems in order to remove harmonics from the system. Application of filters to intermodulation measurement is usually limited by the small frequency separation of the IMD products from the fundamental signals. High order duplexers can help bypass this issue, as high isolation and tight separation of frequency bands can be obtained. Proper choice of probing stimuli frequencies can allow the intermodulation products to fall within the receive band of the duplexer. Commercial systems have reported dynamic ranges of 168 dBc in two-tone tests at 20 Watts per carrier [14]. Although these systems present amazing dynamic range, they are limited by their lack of tunability. High isolation from high order filters requires signals to be fixed within the filter bandwidths, leading to very narrow band systems limited to several megahertz tone separations. Many processes have thermal dependencies, which are always dependent on amplitude modulation of a signal. As input stimuli are brought increasingly close in frequency to each other, amplitude modulation increases, thereby characterizing those processes. Filtering based systems cannot measure such a variation and are thus fundamentally limited in exploring any PIM process with a wide band frequency or thermal dependence.
Feed-Forward Techniques

Passive process characterization requires the measurement of at least two-tone signals over wide, tunable bandwidths with massive dynamic range and many variations of frequency separation. These requirements are competing in any filtering based method, leading to the need for a technique not reliant on filtering. Harold Black [46], the inventor of feedback in circuits, supplied this method with the advent of feed-forward techniques. Feed-forward is an unconditionally stable technique, thus is inherently broadband. Feed-forward relies on the cancelation of signals at a reference plane by combining the undesired signal with an amplitude matched version of the same signal with opposite phase and equal delay. Cancelation can be thought of as the vector addition of two signals resulting in destructive interference of the signal [46].

Knowledge of many things are needed to perform cancelation including exact amplitude and phase of the signal to be canceled as well as the loss and delay of the channel.
over frequency. The stringency of this requirement lies in the statement of exact amplitude and phase. In any real system there will be some degree of error in both the amplitude and phase measurements. The deviation from ideal cancelation due to amplitude and phase error follows the well known equation for rejection,

\[ P_O - P_C = -10 \log (\alpha^2 + 2\alpha \cos (\phi) + 1) \]  

(2.8)

Where \( P_O - P_C \) is the difference in amplitude (dB) of the original signal and the canceled signal. The \( \alpha \) and \( \phi \) terms represent the amplitude and phase error, respectively. A contour plot is shown in Fig. 2.12 of cancelation versus phase and amplitude error clearly showing that very small errors result in large deviations from ideal cancelation. Schemes such as power minimization and gradient techniques are effective at canceling tones even under such high accuracy requirements because they iterate towards the true value through an error function instead of requiring exact values for amplitude and phase. These methods often take many iterations to reach reasonable cancelation levels [45].

As long as the signals are continuous wave and not containing transient components, the delay associated with cancelation can be tolerated in order to gain virtually unlimited bandwidths and tone separations. If transient signals are to be measured, cancelation for PIM measurement cannot be used unless periodically pulsed signals are used to determine cancelation parameters a priori to the cancelation or the signal operates much slower than the cancelation control loop. If all these requirements are met, this technique is very attractive for high dynamic range measurements.
A general system accomplishing cancelation in a two-tone measurement is shown in Fig. 2.13. Each signal is split initially and sent through either a variable gain amplifier or a variable attenuator. It is then shifted in phase by a variable phase shifter and recombined with the output of the test device. The two-tone signal at the output of the test device is then canceled, leaving only the distortion products of interest, shown in Fig. 2.14. The dynamic range of the receiver is only limited by the distortion products at this point, meaning that the gain in dynamic range is the smaller of cancelation of the probe signal or the dBc range from the probe stimulus to the distortion products. Passive distortion products are commonly over 100 dBc from the probe signals. If cancelation of 50 dB is obtained in a system with a receiver dynamic range of 75 dB, a system with 125 dB dynamic range can be obtained. Up to 80 dB of cancelation has been obtained [44], making systems with comparable dynamic range to that of filtering systems possible.

2.5 Fractional Calculus

A fundamental problem when working with multi-physics problems is time scale separation of processes. If time scales between two processes are separated by several orders of magnitude, inconsistency in solutions that are localized about a point occur. The reason for this is that the slow function contributes to the value of the fast function, creating a long memory effect [47, 48]. This statement says the derivative of the fast function is no longer dependent just on the previous value of itself, but is instead dependent also on the previous value of the slow function, which may have many periods of the fast function contained.
within it. In most physical cases, the fast process is not coupled to the slow process as it will average to zero before a change in the slow function can occur [48]. However, in multi-physics situations where diffusion processes or diffusion equations are involved, the faster wave process can be tangibly affected by the slower diffuse process, provided that the fast process has appreciable variation within the bandwidth of the slower process. The result of time scale inseparability and long memory is the need to recast the system as a fractional differential equation.

The fractional derivative is a non-local operator that embeds the complete knowledge of the past history of a function [47–51]. Due to this property, it is possible to create physically based reduced order models of phenomena that could not be conveniently described by integer order models. It is exactly this property that makes them so useful. By recasting a partial differential equation as a fractional equation, not only can a reduced order model be obtained, but since the fractional solution contains memory of the entire function it also alleviates the need to solve the entire spatial domain. This property effectively reduces the partial differential equation for any chosen point to a function only dependent on time while still accounting for spatial effects throughout the entire medium. For the equation to be recast as fractional, either an equation can be suggested or the equation can be derived from integer order constitutive equations. The main drawback for the use of fractional calculus is its complexity in implementation, specifically as it applies to commercial simulators. Although there are several ways to circumvent these problems, only extremely limited implementations currently exist.

2.5.1 Fractional Calculus Natural Functions

In any branch of mathematics, a few functions with special properties exist that allow them to be used for the solution of many problems with fractional calculus descriptions. An example of this statement are the transcendental functions and the exponential. Their special properties allow them to be used as a basis set to approximate an incredibly large group of functions and solve an incredibly large group of problems. Several of these functions exist in the fractional calculus as well, the most important of which are the gamma, beta, and Mittag-Leffler functions.
**Gamma Function**

The gamma function was developed by Euler in order to generalize the factorial to all real numbers. It can take the form of an integral and is given by

$$\Gamma (t) = \int_{0}^{\infty} x^{t-1} e^{-x} dx. \quad (2.9)$$

where $t$ can be a complex number. Several important properties are exhibited by this function. When $t$ is positive real, the gamma function can be expressed as

$$\Gamma (t) = (t - 1)!,$$  \hspace{1cm} (2.10)

and for all real numbers

$$\Gamma (t + 1) = t \Gamma (t). \quad (2.11)$$

These relationships allow the extension of the gamma function to negative real numbers. Interestingly, the function goes to infinity at negative integer values but is defined for negative non-integer values [49–51].

**Beta Function**

The Euler beta function is important in fractional calculus due to its similar form to the fractional integrals and derivatives of many polynomials and the Mittag-Leffler function. The function is conveniently given in terms of the gamma function by

$$B (x, y) = \int_{0}^{1} (1 - t)^{y-1} t^{x-1} dt = \frac{\Gamma (x) \Gamma (y)}{\Gamma (x + y)}, \quad (2.12)$$

where $x$ and $y$ are positive real.

**Mittag-Leffler Function**

The exponential function is often used to solve integer order differential equations and may be considered a natural solution to those equations. The fractional analog of a natural solution is the Mittag-Leffler function [49–51] which is represented by
\[ E_{\alpha,\beta}(x) = \sum_{k=0}^{\infty} \frac{x^k}{\Gamma(\alpha k + \beta)}, \quad (2.13) \]

where \( \alpha \) and \( \beta \) are complex parameters, but the real part of \( \alpha \) is restricted to be positive. It is interesting to note that the Mittag-Leffler function contains several other natural functions within it, depending on the values chosen for \( \alpha \) and \( \beta \), including the exponential function, error function, the sum of a geometric progression, and hyperbolic trigonometric functions. The Mittag-Leffler functions interpolate between a purely exponential law and power-like behavior of phenomena, and are natural solutions to random walks, Levy flights, and many more fractional processes [47, 48].

### 2.5.2 Fractional Integral Definition

The fractional integral form most often used is that of Riemann-Liouville. They sought to extend the integer order integral through the use of the Cauchy formula for repeated integration. This formula gives the \( n \)th integration of a function,

\[
f^{[n]}(t) = \int_0^t \cdots \int_0^{y_{n-1}} f(y_n)dy_n \cdots dy_2dy_1 = \frac{1}{(n-1)!} \int_0^t (t-\tau)^{n-1} f(\tau)d\tau. \tag{2.14} \]

The factorial function in (2.14) restricts \( n \) to integer integrations. To reach a fractional integration, the factorial can be expanded for all real numbers, resulting in the gamma function. The fractional integral becomes

\[
f_a(t) = \frac{1}{\Gamma(a)} \int_0^t (t-\tau)^{a-1} f(\tau)d\tau, \tag{2.15} \]

where \( \alpha \) can now be any positive real number. It should be noted that the effect of the integration limit at zero forces the causality of the function [49–51].

The fractional integral can also be derived in a different manner, used by Grunwald and Letnikov. In this definition, instead of repeated integration, repeated differentiation was first defined and then extended to negative order fractional exponents [49]. The result is an equivalent formation to the Riemann-Liouville definition, given by

\[
d^{-\alpha}f(t) = \lim_{h \to 0} h^\alpha \sum_{m=0}^{\frac{t-h}{h}} \frac{\Gamma(\alpha + m)}{m!\Gamma(\alpha)} f(t - mh). \tag{2.16} \]
2.5.3 Fractional Derivative Definition

Fractional Calculus has had many contributors, each with their own definition of the fractional derivative and integral. The three most relevant definitions, Riemann-Liouville, Caputo, and Grunwald-Letnikov are briefly reviewed here along with the Laplace representation of a fractional derivative or integral.

Riemann-Liouville Definition

In Newton’s classical calculus, Cauchy reduced the calculation of an n-fold integral of the function \( f(t) \) into a single convolution integral possessing an Abel power law kernel. Riemann and Liouville continued his work to develop perhaps the most well-known definition of the fractional derivative which is given by:

\[
a D_t^p f(t) = \left( \frac{d}{dt} \right)^{m+1} \int_a^t (t - \tau)^{m-p} f(\tau) d\tau, \quad (m \leq p < m + 1). \tag{2.17}
\]

The concept behind this definition of the fractional derivative is the use of fractional integral first, followed by the subsequent integer order differentiation. This definition is preferable in the mathematical community for the explicit reason of the order of these operations. Since integration is the first operation, as long as the function meets the condition of causality, so must all the subsequent derivatives. While this is a great feature mathematically, it results in the need for fractional derivative initial conditions in order to solve real-world problems [49–51].

Caputo Definition

In engineering, initial conditions must be specified in terms that are physically definable. To solve this practical need, M. Caputo redefined the Riemann-Liouville approach to the fractional derivative by reversing the order of operations. In his definition, differentiation of integer order is performed first, followed by subsequent fractional order integration using the same approach as Riemann and Liouville. His definition of the fractional
derivative is given by:

\[
C_a D^a_t f(t) = \frac{1}{\Gamma(a - n)} \int_t^T \frac{f^{(n)}(\tau)}{(t-\tau)^{a+1-n}} d\tau, \quad (n-1 \leq a < n).
\] (2.18)

Caputo’s definition, unlike the Riemann-Liouville definition, can use standard integer order derivatives as initial conditions, which are physically defined. The derivative of a constant using Caputo’s definition is zero, unlike the Riemann-Liouville definition, which provides a non-zero result. In essence, the Caputo definition allows application of fractional calculus to real problems [49–51].

Grunwald-Letnikov Definition

Unlike the Riemann-Liouville and Caputo definitions of the fractional derivative, the Grunwald-Letnikov definition does not focus on a repeated integration, but instead focuses on repeated differentiation [51]. Starting with the definition of the derivative,

\[
f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h},
\] (2.19)

and applying the formula a second time to obtain the second derivative:

\[
f''(x) = \lim_{h \to 0} \frac{f'(x+h) - f'(x)}{h} = \lim_{h_1 \to 0} \frac{\lim_{h_2 \to 0} \frac{f(x+h_1+h_2) - f(x+h_1)}{h_2} - \lim_{h_2 \to 0} \frac{f(x+h_2) - f(x)}{h_2}}{h_1} \tag{2.20}
\]

\[
f''(x) = \lim_{h \to 0} \frac{f(x+2h) - 2f(x+h) + f(x)}{h^2}. \tag{2.21}
\]

For an \(n\)th-order derivative, the formula is reduced to Grunwald’s definition when the binomial sum is expanded to fractional values using the gamma function. The Grunwald-Letnikov definition of the fractional derivative is thus given by

\[
D^n_{t_0} f(t) = \lim_{h \to 0} \frac{1}{h^n} \sum_{j=0}^{t-t_0} (-1)^j \binom{n}{j} f(t-jh)
\]

\[
= \lim_{h \to 0} \frac{1}{h^n} \sum_{j=0}^{t-t_0} (-1)^j \frac{\Gamma(a+1)}{j!\Gamma(a-j+1)} f(t-jh).
\] (2.22)

The importance of the Grunwald-Letnikov definition, as can be seen, is the fact that it lends itself to direct numerical approximation of the derivative of any order.
Fractional Laplace and Fourier Transform

The Laplace transform is a powerful solution method for differential equations as it allows transformation of the equation to an algebraic equation of the transform variable. This technique can be used in certain situations to eliminate either space or time dependencies in a partial differential equation, turning it into a solvable ordinary differential equation in the other variable. The Laplace transform is quite well known and widely used for engineers working in the frequency domain, alongside the Fourier transform. Representation of the fractional derivative and integral in the frequency domain can be accomplished through the generalization of the Laplace transform to fractional orders.

The premise of the proof of the fractional Laplace transform is the equivalence of a convolution kernel in the time domain to its constituent pieces multiplied in the Laplace domain [51]. The Laplace convolution kernel is given by

\[ f(t) * g(t) = \int_0^t f(t - \tau) g(\tau) d\tau. \quad (2.23) \]

Now define a function

\[ \Phi_\alpha(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)}, \quad (2.24) \]

and convolve that function with an arbitrary function \( f(x) \) to yield

\[ \Phi_\alpha(x) * f(x) = \int_0^t \frac{(x-t)^{\alpha-1}}{\Gamma(\alpha)} f(t) dt, \quad (2.25) \]

which is in fact the fractional integral. The Laplace transform of \( \Phi_\alpha(x) \) is quite simple by choice, and is given by

\[ \mathcal{L}\left\{ \frac{x^{\alpha-1}}{\Gamma(\alpha)} \right\} = s^{-\alpha}. \quad (2.26) \]

Multiplying the respective transforms, the equivalent of the fractional integral in the Laplace domain is obtained, which is

\[ \mathcal{L}\left\{ \frac{dt^{-\alpha}}{d\tau^{-\alpha}} f(x) \right\} = s^{-\alpha} f(s). \quad (2.27) \]

Since fractional differentiation can always be performed by a fractional integral and separate integer order derivative, this formula is easily extended to the fractional derivative, resulting
in
\[
\mathcal{L}\left\{ \frac{d^{\alpha}f(x)}{dt^{\alpha}} \right\} = s^\alpha f(s).
\] (2.28)

Although a proof could also be extended to Fourier analysis, the relationship between the Fourier variable \( \omega \) and the Laplace variable \( s \) allows direct translation to the Fourier domain.

### 2.5.4 Fractional Differential Equations

Wave and diffusive behavior are not as separate as would be preferable from a mathematical standpoint. An example of this fact is low frequency electromagnetic induction. At low frequencies the system acts diffusively, but as the frequency is increased, behavior becomes blended between wave and diffusive behavior before it finally transitions to primarily wave behavior.

In terms of the constituent partial differential equations, this means that the order of the equation lies somewhere between a first and second order process \((1 < \alpha < 2)\) implying a blending of oscillatory and relaxation processes. Diffusion processes are often better represented by fractional systems of less than order one, \((0 < \alpha < 1)\). Using multiple fractional derivatives of various order can include any combination of system behavior desired, making it one of the most versatile mathematical methods available to the engineer. The solutions of these equations follow similar methods to that of integer order equations, but using fractional versions of the methods instead such as the fractional Laplace transform or adapted Green’s functions.

### 2.5.5 Numerical Methods

Even if analytic solution to a problem is possible, it is usually simplified to an extent. Although it provides great intuition to designers, extremely accurate simulations are inevitably needed to create reliable products. In order to simulate these problems quickly and accurately, application of fractional differential equations is paramount. The long memory requirement of fractional solutions has hampered its adoption by making time domain implementation difficult, which is needed for nonlinear simulation. Several methods have been developed, and continue to be developed, to alleviate this concern.
Several methods of interest are reviewed here, including the short memory principle, the 
time scaled method, and memory-less methods.

**Short Memory Principle**

The short memory principle was introduced by Igor Podlubny to shorten the large 
amounts of memory needed to compute fractional derivatives [51]. In looking at the original 
Grunwald-Letnikov definition that for large times, the role of the history of the behavior of 
the function becomes negligible near the lower terminal \( t = a \) under certain conditions. He 
prescribes taking into account only the most recent history of the function, in the interval 
\([t - L, t]\), where \( L \) is the memory length:

\[
a D_t^\alpha f(t) \approx_{t-L} D_t^\alpha f(t), \quad (t > a + L).
\] (2.29)

The derivative is approximated at the lower terminal by a moving lower terminal. The error 
estimate for this simplification is given by:

\[
\Delta(t) = |a D_t^\alpha f(t) -_{t-L} D_t^\alpha f(t)| \leq \frac{ML^{-\alpha}}{\Gamma(1 - \alpha)} \quad (a + L \leq t \leq b).
\] (2.30)

If the desired accuracy \( (\varepsilon) \) is known, it can be used to determine the required memory 
length, given by

\[
\Delta(t) \leq \varepsilon \rightarrow L \geq \left( \frac{M}{\varepsilon \Gamma(1 - \alpha)} \right)^{1/\alpha}.
\] (2.31)

**Time Scaled Method**

All of the aforementioned methods for implementing the fractional derivative have, 
to a degree, lacked a clear time domain representation. In [52] the authors propose a 
time scaled method of differentiation. In their interpretation, a discrete fractional order 
derivative is the derivative of the sampling time scaled discrete integral. The Riemann-
Liouville definition of the derivative is rewritten to yield

\[
0 D_t^\alpha f(t) = \frac{1}{\Gamma(1 - \alpha)} \frac{d}{dt} \int_0^t \frac{f(\tau)}{[t - \tau]^\alpha} d\tau = \frac{d}{dt} \left[ \int_0^t f(\tau) dq_t(\tau) \right],
\] (2.32)
with the integration factor rewritten as a derivative, which is given by
\[
\frac{1}{\Gamma(2-a)} \left[ t^{1-a} - (t - \tau)^{1-a} \right]. 
\]
(2.33)

The scaled sampling time steps are given by:

\[
T'_n(n) = \frac{1^{1-a} - 0^{1-a}}{\Gamma(2-a)} T^{1-a}
\]
(2.34)

\[
T'_n(n-1) = \frac{2^{1-a} - 1^{1-a}}{\Gamma(2-a)} T^{1-a}
\]
(2.35)

\[
T'_n(1) = \frac{n^{1-a} - (n-1)^{1-a}}{\Gamma(2-a)} T^{1-a},
\]
(2.36)

where \( T_n \) is the sampling time at the \( n \)th time step and \( a \) is the order of the derivative.

Finally, application of the trapezoidal integration rule yields

\[
\int_0^{nT} f(\tau) d\tau \approx \sum_{k=1}^{n} \frac{f(kT) + f[(k-1)T]}{2} T'_n(k),
\]
(2.37)

where \( T \) is the constant time step. According to the authors [38,46], the interpretation of discrete fractional order derivatives are the derivatives of fractional \((1-\alpha)\) order integrals. It can be understood geometrically as the changing ratio of the “scaled integral area” due to the scaled sampling time.

**Memoryless Fractional Derivatives**

Very recently, Yuan and Agrawal [53, 54] proposed a method for evaluating fractional derivatives that requires no storage and processing of the function history. This method is based on a transformation of the fractional differential equation to a set of ordinary integer order differential equations. In internal variable theory equivalence between an exponential kernel and an ordinary linear differential equation can be drawn, just as with fractional derivatives. They consider a general fractional differential equation of the form,

\[
mD^2 x(t) + cD^\alpha x(t) + kx(t) = f(t)
\]
(2.38)
with the derivative defined as the Riemann-Liouville fractional derivative. Noting the gamma function properties

\[D^a x(t) = \int_0^t \frac{(t-\tau)^{-a}}{\Gamma(1-a)} D\tau \]

\[\Gamma(a) = \int_0^\infty e^{-z} z^{a-1} dz\]

\[\Gamma(a)\Gamma(1-a) = \frac{\pi}{\sin(\pi a)}\]

The fractional derivative can be rewritten as

\[D^a x(t) = \frac{\sin(\pi a)}{\pi} \int_0^t \left( \int_0^\infty e^{-z} \left[ \frac{z}{t-\tau} \right]^a \frac{dz}{z} \right) D\tau d\tau\]

Substituting \(z = (t-\tau)y^2\) and rewriting the derivative,

\[D^a x(t) = \frac{2 \sin(\pi a)}{\pi} \int_0^t y^{2a-1} \left( \int_0^\infty e^{-(t-\tau)y^2} D\tau d\tau \right) dy\]

as well as introducing the new variable

\[\phi(y, t) = y^{2a-1} \int_0^t e^{-(t-\tau)y^2} D\tau d\tau\]

allows the fractional differential equation to be written, upon taking the time derivative, as

\[D\phi(y, t) + y^2 \phi(y, t) = y^{2a-1} Dx(t)\]

This formulation allows the fractional differential equation to be solved without function history. Only integration of the remaining integer order equation is required.

**Summary of Fractional Calculus**

Several numerical methods to compute the fractional derivative, used to reduce the requirement of full function memory at every time step, were reviewed in this Section. The
short memory principle simply truncates the memory of the fractional derivative computation to a length that causes a pre-determined, user defined amount of error. The time scaled method uses standard trapezoidal integration, but varies the time step of the integration to represent a fractional derivative. This property makes it attractive for implementation into circuit simulators with variable time step capabilities. The memoryless method is ideal for implementation in standard circuit simulators, as it allows a fractional differential equation to be recast into an integer order equation that can be solved with standard numerical integration methods. This review was conducted to allow further implementation of the work in this dissertation. No reduction in complexity using numerical methods was used in this work. Fractional derivatives were numerically computed using the definition of the Grunwald-Letnikov and Caputo fractional derivative definitions.

2.6 Conclusion

Passive components such as resistors, transmission lines, and antennas can produce intermodulation distortion under high power input conditions, which is referred to as passive intermodulation distortion. PIM is generally a much weaker nonlinearity than the nonlinearity of active components such as amplifiers. The weakness of the linearity coupled with the large number of potential nonlinear mechanisms has made physical description of PIM difficult. Many physical mechanisms have been suggested as the cause of PIM, including metal-oxide-metal and metal-metal contacts, ferromagnetic materials, piezoelectric materials, acoustic coupling, and electrical conductivity modulation. Each mechanism was reviewed in this Chapter to allow separation of the mechanisms in this work. The differences in material properties favored electrical conductivity modulation as a dominant PIM mechanism as it must exist in every microwave component.

Several of the PIM generating mechanisms often exist concurrently in microwave devices, limiting the modeling of passive devices to empirical or behavioral models. A review of common distortion analysis and modeling techniques was given including the power series and volterra series. The differences between behavioral modeling and physical models were discussed, with an emphasis on the design benefits of a physical model compared to empirical or behavioral techniques.
Passive intermodulation distortion was described as a weak nonlinearity requiring high dynamic range measurement methods for analysis. Common distortion characterization methods such as AM-AM, AM-PM, total harmonic distortion, two-tone, and multi-tone were reviewed. Two-tone measurements were shown to be the most useful method for this work, as they allow the most simple analytic formulations of PIM. Measurement systems and methods capable of the high dynamic range necessary for PIM testing were reviewed. Although the highest dynamic range is obtained by filter based methods, the system does not allow the test frequencies and separations to be varied substantially. Feed-forward cancellation was shown to be the most relevant measurement method for this work.

A short review of fractional calculus was conducted to allow the modeling of electro-thermal distortion in this dissertation. The definitions used to compute long-tail transients and frequency responses, as well as derive analytic electro-thermal physics were provided. Numerical methods that can be used to reduce the computational complexity of fractional derivatives were presented to allow future expansion of this work into standard circuit simulators.
3

Broadband High Dynamic Range Measurement
3.1 Introduction

Intermodulation distortion can be the limiting factor in high dynamic range communication systems such as satellites and cellular basestations. Although the active system components are designed to meet an industry distortion specification, often passive components such as transmission lines, filters, combiners, and antennas in the transmit path produce distortion after the active components which can then couple to the receive system. Two-tone testing is conventionally used to characterize and identify the sources of nonlinear distortion. This requires the ability to measure distortion components that could be 100 decibels or more below the level of applied tones and separated by a few hertz. Measurements of tones separated by a few hertz is required to distinguish between sources of passive intermodulation distortion (PIM), specifically electro-thermal PIM, nonlinear junction effects, and tunneling [55,56].

In a nonlinear distortion measurement systems, large stimulus tones are applied to the device under test (DUT), and nonlinear distortion components are generated at much lower power levels. The interfering stimulus must be removed before reaching the receiver, as distortion signals would otherwise be masked by distortion generated in the receiver [57]. To reduce both the stimulus signal and large noise levels simultaneously, filtering or active feed-forward techniques [58] are needed.

High dynamic range measurement systems based on filtering typically use diplexers or notch filters to remove system-generated distortion components before application to the DUT. A second filtering stage removes the stimulus signal before final measurement of the distortion components by a spectrum or vector analyzer [14]. This scheme has the highest dynamic range but cannot be used with closely spaced tones because of the limited roll-off of the filter skirts. Overcoming tunability limitations is of utmost importance when testing for passive intermodulation distortion (PIM), as some types of PIM such as electro-thermal distortion must be tested for at very small tone separations where no filter skirt can be sharp enough. The only choice for broadband, spacing independent distortion testing is feed-forward cancellation.

Feed-forward cancellation directly increases the dynamic range of the system by the cancellation achievable until the components of the test system itself begin to generate distortion. Every component will produce PIM at some level, generally dependent on both
power and frequency. Active devices, even if confined to a single frequency, will produce spurious frequency content at PIM frequencies if there is a radiation path in the system. Reflections and finite isolation in the system compound the nonlinearities due to forward and backward wave mixing in components that should be single frequency. All of these effects must be accounted for in order to provide high dynamic range measurement capabilities.

This chapter details the design and test of a high dynamic range measurement system based on feed-forward cancellation. Feed-forward theory and system design concepts are discussed in Section 3.2 and Section 3.3. Component linearity for each type of component used in the system is discussed in Section 3.4. Design concepts for high dynamic range included reflection effects, load matching, limiting amplifier mixing, and radiative coupling are presented in Section 3.5. Section 3.6 discusses system applications and performance for two-tone PIM measurement and digitally modulated signals. Section 3.7 summarizes the results presented in this chapter.

### 3.2 Feed-Forward Cancellation Theory

The process of cancelling a signal is mathematically a simple one, requiring only the summation of that signal with an equal amplitude anti-phase signal at the same frequency. Non-idealities such as small imbalances in amplitude, phase, and group delay must be considered in any system implementation [59–62]. Amplitude and phase imbalance, $\alpha$ and $\Delta\phi$, respectively, limit the depth of cancellation achievable, while group delay imbalance limits both the depth and bandwidth of cancellation [59–63]. Cancellation performance, $CP$, is described by

$$CP = 10 \log \left(1 + \alpha^2 - 2\alpha \cos \left(2\pi \cdot \frac{\Delta\lambda}{\lambda_c} \left(1 - \frac{f}{f_c}\right) + \Delta\phi\right)\right),$$

(3.1)

where $\Delta\lambda$ is the difference in wavelengths between the two paths normalized to the wavelength of the center frequency, $f_c$, and $f$ is the frequency of operation. Group delay mismatch affects bandwidth due to the deviation of phase shift of the cancellation and signal paths as the frequency is shifted away from the center frequency of cancellation. The limiting effect of group delay on the cancellation depth results from the degree of delay mismatch. As the difference in delay grows, so does the rate of phase mismatch with increasing separation from the center frequency. Group delay can be matched to a level acceptable for
bandwidth requirements according to (3.1). Under these conditions, the phase shift and amplitude necessary for cancellation must be obtained. The method developed here automatically tunes the center frequency of the cancellation signal enabling the measurement system to operate over wide bandwidths.

Cancellation methods using power minimization or gradient techniques are effective at eliminating phase error iteratively, but the number of iterations to reach cancellation in excess of 30 dB can be large [58,62,64]. In the work described here a formula is used to determine the phase shift necessary for cancellation given the amplitude of the combined feed-forward and DUT signal, resulting in greater speed and accuracy than can be obtained with iterative search methods. Amplitude matching is described after discussion of phase correction. The formula is based on amplitude measurements facilitating implementation using standard laboratory equipment such as a spectrum analyzer, an oscilloscope, or a vector signal analyzer. Development of a formula based on amplitude measurement can be
obtained by first considering the cancellation of a single sinusoidal signal.

Consider a single sinusoidal signal which is the combination of the original signal and the cancellation signal [65],

\[
\alpha \cos (\omega_1 t + \phi_1) + \alpha \cos (\omega_1 t + \phi_2) = \beta \cos (\omega_1 t + \phi_3)
\]  (3.2)

where \(\alpha\) is the amplitude of both the original and cancellation signal, \(\omega_1\) is the radian frequency of the signals, \(\phi_1\) is the phase of the original signal, \(\phi_2\) is the phase of the cancellation signal, \(\phi_3\) is the combined signal phase, and \(\beta\) is the combined signal amplitude. Since the tones are at the same frequency, this can be rewritten as a function of only the magnitude and phase difference,

\[
\cos (\phi_1) + \cos (\phi_2) = \frac{\beta}{\alpha},
\]  (3.3)

which can be seen by setting \(\phi_3 = 0\) and considering \(t = 0\) in (3.2). From trigonometric identities it can be shown that

\[
\cos (\phi_1) + \cos (\phi_2) 2 \cos \left( \frac{\phi_1 + \phi_2}{2} \right) \cos \left( \frac{\phi_1 - \phi_2}{2} \right).
\]  (3.4)

The phase of the combined signal is related to the phase of the cancellation and original signal by

\[
\frac{(\phi_1 + \phi_2)}{2} = \phi_3 = 0.
\]  (3.5)

Defining \(\phi_1 - \phi_2\) as the phase difference \(\phi\), (3.3) and (3.4) become

\[
\cos \left( \frac{\phi}{2} \right) = \frac{\beta}{2\alpha}.
\]  (3.6)

Solving for the phase difference gives

\[
\phi = n\pi \pm 2\arccos \left( \frac{\beta}{2\alpha} \right); \quad n = 0, 1, \ldots
\]  (3.7)

The phase difference required for cancellation is \(\pi\). Choosing \(n = 0\) and rewriting in terms of the phase shift required for complete cancellation, \(\phi_s\) [65],

\[
\phi_s = \pi \pm 2\arccos \left( \frac{\beta}{2\alpha} \right).
\]  (3.8)

There will be a phase prediction error, in general, which is the difference between the predicted phase shift and the ideal phase shift. This error tends to reduce as the signals
approach 180° phase difference. This can be seen in Fig. 3.1, where the signal cancellation achieved for equal amplitude signals is plotted versus signal phase difference. The absolute phase error of the formula is also shown against phase difference in Fig. 3.1. The accuracy of the formula, assuming no amplitude error, is lowest when the two tones have a phase difference of 3°. At this point the formula has 1.242° of phase error, which limits cancellation to 33 dB. Error reduces as the phase difference of the tones increases, lending itself to greatly increased accuracy upon a second application, in practice often exceeding 50 dB. Although the formula is derived to cancel only a single tone, it is extensible to other signals such as digital modulation and multi-sine, where the cancellation bandwidth can be derived from (3.1).

Equation (3.8) provides only the magnitude of the required phase shift. Thus taking one of the solutions, say the positive phase shift, may not increase cancellation and so the negative phase shift must then be used. With a vector signal analyzer used to measure both the phase and amplitude of the signal component being cancelled, it is possible to determine which phase shift solution to use. Implementation of (3.8) can be accomplished through the establishment of a reference cancellation plane. The reference plane is established by measuring the probe tone power and then measuring the power from the combination of the feed-forward path and the DUT path signals at the bridge point. Their respective phases must also be measured if the sign of the phase shift is to be determined. Only two amplitude measurements are required to determine where the relative phase of the tones is on the unit circle before cancelling those tones.

The DUT signal amplitude is matched automatically at the bridge point using feed-forward path amplitude calibration. The system of Fig. 3.2 automatically calibrates the feed-forward channel transmission characteristic to extract measurement error, channel losses, and vector modulator characteristics. The calibration is used to generate an output power characteristic as a function of power and frequency to enable amplitude correction between the system DUT and feed-forward paths.
Figure 3.2: High dynamic range passive intermodulation distortion measurement system composed of highly linear amplifiers, RF sources, isolators, combiners, vector modulators, and a vector signal analyzer. Four separate digital to analog converters (DACs), represented by one DAC, are used to individually control the positive and negative in-phase and quadrature inputs of the vector modulators, shown with their internal schematics. The LO input of the vector modulators is identified.
3.3 Linear Feed-Forward System Design

Feed-forward cancellation is an application of the bridge method for measuring small variations in signals [66]. Typically the feed-forward and DUT paths of the bridge combine at a reference plane. Only the original signal will be fed forward and cancelled, enabling high dynamic range measurement of the unaffected distortion components beyond the reference plane. A few practical issues arise that govern the implementation of the bridge technique including phase tracking of the cancellation signal and noise coherence. Both noise coherence and phase tracking are governed by the method used to generate the feed-forward signal. The most prominent methods include separate sources, sampling and regenerating the signal with analog to digital and digital to analog converters, and coupling off of part of the original signal, are discussed in this section in the context of measurement system architecture.

Phase instability is an unavoidable issue when employing separate cancellation sources. Modern frequency synthesizer architectures are frequency locked rather than being phase locked [67]. With frequency-locked sources, random phase variations of independent fractional synthesizers are only required to have on average a particular time-varying phase relationship. While this significantly reduces the level of spurious tones, this time-varying phase relationship causes the phase difference between two frequency locked signals to wander. Using separate sources to provide the cancellation signals thus requires constant phase control to retain cancellation [65]. If continuous phase control is not employed, the result is typically a slow variation of a few decibels or complete loss of cancellation. Designs that sample or couple the original signal and use this to form the cancellation signal do not have this phase instability problem as the cancellation signal is inherently phase and frequency locked to the original signal.

Noise coherence only exists when coupling of the original signal is used. When separate sources are used, the noise in the cancellation signal is random with respect to the noise in the applied DUT signal. The noise from a sampled and regenerated signal is random with respect to the DUT channel unless the circuitry can sample both the noise and signal. The signals may be frequency and phase locked, but the noise will be independent and will sum. Noise summation detracts from the dynamic range of the system, and ultimately limits the capability of the measurement system. Feeding forward part of the original signal
Figure 3.3: Measured vector modulator response: (a) I/Q gains versus input LO drive power below specification at $-30$ dBm and $-25$ dBm, and within specification at $-5$ dBm; and (b) output power variation with output phase over input LO drive power from $-45$ dBm to $-20$ dBm in 5 dBm increments.
avoids this problem, as source noise is also fed forward and cancelled, leaving only the noise added by components such as the DUT, active devices, and passive components.

A high dynamic range feed-forward measurement system coupling off part of the original signal is shown in Fig. 3.2. Here the system is implemented for a two-tone signal, with vector modulators used to control the phase and amplitude of the signals in the cancellation channels replacing variable attenuators and delay lines. In the implementation described, the Hittite HMC497LP4 vector modulators have a bandwidth of 3.9 GHz, enabling nonlinear vector signal analysis from 100 MHz to 4 GHz. They are digitally-controlled by digital to analog converters (DACs) and provide up to 360 degrees phase rotation and amplitude control between 20 and 40 dB. Isolators are used to limit the reverse traveling waves in the system. The two branches of the bridge, the cancellation channel and DUT channel, are finally summed at a reference plane before measurement by the vector signal analyzer.

Reflected signals must be minimized at both the input and output ports as small reflected signals at the local oscillator (LO) port of the vector modulator can be devastating to automated control. They can result in offsets in the amplitude of the output signal on the order of several decibels, high enough to cause loss of signal cancellation. The variation in output amplitude occurs due to the summation of the original and reflected waves within the vector modulator. If the reflected signal is much smaller than the applied signal, and the gain is not balanced over input power between the in-phase (I) and quadrature (Q) channels, the smaller signal will experience different gains in the I and Q channels (in the linear mode). Fig. 3.3(a) compares the gains of the I and Q channels. It is seen that the gains are balanced only when the LO signal is large enough to drive both of the I and Q channel mixers into their limiting mode of operation. As the LO level reduces below the limiting gain level, the I and Q channel gains become increasingly unbalanced. This leads to the much smaller reflected LO signal being amplified in linear mode and combined with the limited LO signal at a phase offset dependent on the reflected path characteristics and the linear mode I and Q gain differences. Eliminating this problem requires either suppression of the reflected signal or maintaining constant phase of the reflected wave over frequency to prevent cancellation of the intended output signal of the vector modulator. In Fig. 3.3(b), a small interfering LO signal produces a linearly amplified output that oscillates with output phase due to gain imbalances. The cancellation occurring at the vector modulator output
can be found from the combination of the linearly amplified reflected LO signal and the applied LO according to (3.1). For an applied local oscillator power of $-5 \text{ dBm}$, the reflected wave must be at least 40 dB below the applied signal to impact the output by less than 0.15 dB.

### 3.4 Component Linearity

Components for the transmit channel must be as linear as the receive channel, including feed-forward cancellation, to achieve the desired increase in dynamic range. Even in one tone per channel configurations, reflections and finite isolation will cause multiple frequencies to be applied to the components, often in nonconventional ways. Due to this practical limitation, each device in the system must be analyzed for both forward wave distortion, reverse wave distortion, and combinations of both. Unfortunately most manufacturers specify only forward wave distortion with equal amplitude two-tone signals. Conversely, the signals of interest in a system of this type are one large forward propagating signal and a smaller forward or reverse propagating signal. The forward case will be discussed in radiative coupling, Section 3.5.2. In this section component behavior under a large forward propagating signal and a small reverse propagating signal is discussed.

#### 3.4.1 Amplifiers

Amplifiers effect the system distortion through their linearity and reverse isolation. The linearity of the amplifier can be measured through it’s third order intercept point (IIP3), which gives the maximum drive level for a given amplitude modulated signal at a specified dynamic range. The IIP3 should be as high as possible for both amplitude modulated signals and single channel systems that have signal combination in the system due to the finite reverse isolation through the amplifier. The reverse isolation through the amplifier can lead to the production of distortion products in an amplifier driven by a single frequency if signals of different frequency are present at the output.

Intermodulation distortion is defined as the cross products generated between signal frequencies in a nonlinear device. Any combination of multiple signals will result in some distortion. Only the forward two-tone, multi-tone, or spread spectrum signal is com-
Figure 3.4: Amplifier with a small signal applied at the input and a small reverse wave present at the output generating distortion due to the finite isolation of the amplifier.

Characterization of the nonlinear behavior of the amplifier can be performed by sweeping either the forward or reverse wave power while the opposite wave is held constant. In Fig. 3.5, a test circuit capable of measuring amplifier distortion, composed of two signal generators, two amplifiers, isolators for each channel, a directional coupler, a 10 dB attenuator, and a bandpass filter is shown. The small backward wave channel must be well isolated from the high amplitude forward wave channel to ensure only the channel of interest distortion is measured. Isolators provide reverse signal protection, but due to their own inherent nonlinearity, it is preferable in amplifier testing to reduce the signal amplitude...
applied to them if possible. The directional coupler allows an isolation to be defined even with the reflective nature of the bandpass filter. For this reason, the directional coupler is used instead of a splitter or hybrid combiner. The bandpass filter is used to reject the large stimulus from the measurement channel, increasing the measurement dynamic range by the rejection of the filter. The attenuator is used to reduce large reflected signals sent back into the circulators. The distortion product of interest passes through the passband of the filter and is measured with the vector signal analyzer.

Three frequency sets were used to provide test points at two points of similar isolation and another point with higher isolation. These frequency combinations were determined by the K&L 5DR30-1000/T25-O/O bandpass filter passband frequency edge, 990 MHz, and are given by: 700 MHz and 845.0015 MHz, 798 MHz and 894.0015 MHz, and 900 MHz and 945.0015 MHz. Each channel had three isolators in series to prevent reverse wave interaction from reflections, as opposed to the applied signal, at both amplifier outputs. The effective isolation due to each isolator was frequency dependent and is given by: 31 dB at 700 MHz, 17.3 dB at 798 MHz, 16.3 dB at 845 MHz, 17 dB at 894 MHz, 17.3 dB at 900 MHz, and 20.5 dB at 945 MHz. The isolation provided by the directional coupler was 20 dB across the measurement band. The forward output power of the measurement channel is swept from 16 to 41 dBm in Fig. 3.6, with the reverse wave at the output of the reverse channel amplifier held constant at 41 dBm. The power of the generated distortion increases at approximately twice the rate of the input power as predicted by a general power series expansion.

An analysis of reverse wave effects is shown in Fig. 3.7, where the forward output
power is held constant at 41 dBm and the power at the output of the backward wave amplifier is swept from 16 to 41 dBm. The distortion from the amplifier increases with a one to one relationship to the reverse power, once again as predicted by a general power series expansion. The directivity of the amplifier is 37 dB at 700 MHz, 34 dB at 798 MHz, and 30 dB at 900 MHz. Taking into account reverse isolation from the circulators and the directivity of the amplifier, it can be inferred that the distortion product experiences a small conversion loss of $3 - 5$ dB but is generally close in amplitude to the magnitude of the reverse wave minus the directivity of the amplifier. In high dynamic range measurement, it is imperative to reduce reflections to a minimum as distortion products will be generated at the aforementioned amplitude when the amplifier is close to compression.
Figure 3.7: Amplifier output reverse wave power sweep from 16 to 41 dBm with forward wave from test amplifier held constant at 41 dBm. Reverse wave isolation through circulators and directional coupler is 113 dB at 700 MHz, 72 dB at 798 MHz, and 72 dB at 900 MHz.

3.4.2 Circulators

Circulators are needed to supplement the reverse isolation of active devices in the system for two-tone PIM testing. The ferromagnetic nature of these devices makes them nonlinear, requiring characterization dependent on the port the signal is applied to and the type of signal applied. The circulator allows a signal to flow in one direction around the circulator and must dissipate the signal flowing in the reverse direction. If a single frequency is applied to port one of the circulator, as shown in Fig. 3.8, and a second frequency is applied to port two of the circulator, distortion will be generated related to the isolation and loss of the circulator propagating in the forward direction. Distortion will also be generated in the reverse direction but is of little concern in this measurement system due to the use of series circulators which severely attenuate reverse traveling distortion. The forward traveling distortion is not attenuated in any way and can establish the ultimate limit in dynamic range of a measurement system if all active devices are adequately isolated.
Figure 3.8: A circulator excited by a single tone, $x_1(f)$ at port one, and a second tone, $x_2(f)$, at port two reflects part of the original signal and allows a finite amount of $x_2(f)$ through to port one. Distortion is generated at all ports, but the isolation of subsequent circulators makes only distortion at port two relevant.

A reverse small signal will also propagate through the circulator, which is the combination of the reflection from the terminated port three and the isolation from port two to port one. The reverse signal will continue traveling toward other nonlinear devices such as amplifiers, resulting in effects such as those seen in Section 3.4.1. In the measurement system of Fig. 3.2, the primary concern is reverse wave interaction with the high power forward wave. The nonlinear interaction of circulators is dependent on both the reverse isolation and loss of the isolator, which are both frequency dependent parameters. The reverse isolation, $S_{12}$, and the loss, $S_{21}$, of a 650 – 1000 MHz Raditek circulator are shown in Fig. 3.9, where the third port has been terminated with a 50Ω termination characterized by a return loss, $S_{11}$, of 35 dB.

A distortion analysis of this circulator was conducted for forward power, reverse power, and frequency by sweeping the parameter in question while holding all other parameters constant. The test system of Fig. 3.5 was used with four isolators per channel. Three frequency sets were used to provide test points at two points of similar isolation and another point with higher isolation for the forward power sweeps. These frequency com-
Figure 3.9: Forward loss, $S_{21}$, and reverse isolation, $S_{12}$, for the Raditek 650 − 1000 MHz circulator.

Combinations were determined by the K&L 5DR30-1000/T25-O/O bandpass filter passband frequency edge, 990 MHz, and are given by: 700 MHz and 845.0015 MHz, 798 MHz and 894.0015 MHz, and 900 MHz and 945.0015 MHz. These frequency combinations have respective reverse isolation to the amplifiers of 144 dB, 89 dB, and 89 dB, and respective insertion loss, $S_{21}$, in each isolator of 0.33 dB, 0.36 dB, and 0.41 dB. The nonlinear relationship to forward power with a constant reverse wave of 41 dBm at the output of the reverse channel amplifier is shown in Fig. 3.10. The forward power is swept from 16 dBm to 41 dBm. The distortion generated observes a 2 dB slope with input power as predicted by a general power series expansion but increases nonlinearly with loss. The reverse wave isolation seems to have little effect on the circulator distortion.

Circulator distortion dependence on reverse wave power does not follow the same simple trend as the dependence on forward power. In Fig. 3.11(a), six reverse wave power sweeps are conducted at a reverse wave frequency of 900 MHz and a forward wave frequency of 945.0015 MHz with forward power held constant at 31 dBm, 33 dBm, 36 dBm, 37 dBm, 38 dBm, and 39 dBm. The reverse wave power is swept symmetrically around $41 - P_f$,
Figure 3.10: Forward wave power sweep from 16 to 41 dBm with with reverse wave at reverse channel amplifier output held constant at 41 dBm. Tone frequencies of the reverse and forward wave, respectively, are 700 MHz and 845.0015 MHz, 798 MHz and 894.0015 MHz, and 900 MHz and 945.0015 MHz. Amplitude at isolator of reverse wave is down 20 dB from displayed magnitude.

where $P_f$ is the forward power. The distortion increases almost equally with the reverse wave if the power of the reverse wave power is far below the forward wave power, but reaches an inflection point where further increases in reverse power result in no further increase in distortion magnitude. Further increases in power beyond the inflection point lead to a decrease in distortion magnitude. This behavior is possibly due to the directions of the field within the ferromagnetic material affecting the domain orientations of the material, effectively modulating the material loss. The same behavior is shown in Fig. 3.11(b) but is loss dependent. Here four sweeps at three different forward wave frequencies, 845 MHz, 865 MHz, and 894 MHz, and three different reverse wave frequencies, 700 MHz, 760 MHz, and 798 MHz, respectively, are performed at forward power levels of 36 dBm to 39 dBm. The forward return loss at each reverse wave frequency is 0.12 dB, 0.2 dB, and 0.42 dB, respectively. The reverse wave power is once again swept symmetrically around $41 - P_f$. 
As the reverse wave loss grows, the saturation of distortion with increasing power becomes increasingly prominent. The distortion response of the circulator over frequency, shown in Fig. 3.12, generally increases with loss. Large standing wave patterns in the distortion exist, suggesting the distortion response to waves inserted at different ports in the resonant magnetic material is directionally dependent. The isolation of the circulator is seen to have little effect on the distortion generated within the magnetic material.

Circulators are commonly used for isolation in high dynamic range testing. These devices can greatly reduce the distortion from unintended interaction at active device outputs. As with all passive components, they will generate weak distortion under high power operating conditions. Nonlinear output from inputs at separate ports is of primary concern for high dynamic range measurement. The behavior of a circulator is loss dependent and gains in nonlinear performance by reducing the reverse power are only feasible if the loss of the circulator is small enough that the circulator is not operating in a nonlinearly saturated region.

3.4.3 Terminators and Attenuators

Terminations are used throughout the feed-forward system of Fig. 3.2 at the hybrid combiners, circulators, and attenuators. No termination ever completely eliminates the signal, but rather reduces its amplitude 30 – 50 dB before reflection back into the system. The terminator of Fig. 3.13 dissipates the incoming signal, $x_1(f) + x_2(f)$, as heat, but is not perfectly matched to the transmission line. The original signal is reflected into the system reduced by the terminator return loss, $S_{11}$. The nonlinear signal, $x_{nl}(f)$, is injected into the system in the reflected direction as if the terminator were a source.

Measurement receivers are designed to be linear at a specified input signal level. The high power signals used to test a passive device must be reduced in amplitude to the optimal receiving level. Resistive attenuators provide this functionality, but will reflect a finite amount of signal back into the system as well as producing their own distortion. The attenuator of Fig. 3.14 is excited by a two tone signal, $x_1(f) + x_2(f)$. The attenuator reflects the incoming signal according to the return loss, $S_{11}$, and attenuates the through signal according to the transmission loss $S_{21}$. Distortion is generated in the attenuator that is coupled to both the input and output ports. The amplitude of this distortion at each of
Figure 3.11: Isolator reverse wave sweeps under constant forward power: (a) at 900 MHz tone frequencies with forward wave at 945 MHz and forward power of 31 dBm, 33 dBm, 36 dBm, 37 dBm, 38 dBm, and 39 dBm. Reverse wave power swept symmetrically around the forward output power up to 41 dBm. Amplitude at isolator of reverse wave is down 20 dB from displayed amplitude. (b) and at forward power of 36 dBm to 39 dBm. Reverse wave frequencies given by 700 MHz, 760 MHz, and 798 MHz and forward wave frequencies given by 845 MHz, 865 MHz, and 894 MHz, respectively. Reverse wave power swept symmetrically around the forward output power up to 41 dBm. Amplitude at isolator of reverse wave is down 20 dB from displayed amplitude.
Figure 3.12: Two tone equal amplitude sweep at \(-5 \text{ dBm}\). Frequency of lower tone swept from 700 MHz to 900 MHz. Frequency separation of upper tone determined by \(((990.0015 - f_1)/2 + f_1) \text{ MHz}\). Amplitude at isolator of reverse wave is down 20 dB from forward wave amplitude.

the ports is dependent on where the majority of the power is dissipated.

Lab terminations and attenuators are resistive elements and thus inherently produce electro-thermal distortion \[65\] in relation to the power handling and heat sinking capabilities of the device. Unlike the applied stimulus which is attenuated, distortion generated from within a terminator or attenuator will be injected into the system without attenuation. It is of paramount importance to use low PIM attenuators where high power levels are present. In ultra high linearity applications, an extremely long, low loss cable is the lowest PIM option for termination or attenuation. Return loss of attenuators and terminations greatly impact linearity at signal combination points and should be selected carefully to provide reflections of acceptable levels for the proposed test.
Figure 3.13: A termination excited by a two tone signal, $x_1(f) + x_2(f)$, reflects part of the original signal back into the system and generates distortion traveling in the reverse direction.

### 3.4.4 Cables and Connectors

Cables produce two problems of foremost importance in the measurement system, finite radiation isolation and distortion generation. In Fig. 3.15, a cable is excited by a two tone signal, $x_1(f) + x_2(f)$, that generates distortion along the length of the line. The input signal is attenuated by the transmission loss, $S_{21}$, of the transmission line in the forward direction and is reflected at the return loss, $S_{11}$ of the transmission line in the reverse direction. The nonlinear signal $x_{nl}(f)$, grows preferentially in the forward direction with length and is not necessarily attenuated by transmission loss of the line. In the reverse direction, $x_{nl}(f)$ combines out of phase after $\lambda/4$ and is much smaller than the forward direction distortion. The cable is a weak antenna which radiates and receives signals that propagate in both directions on the line.

Cables produce distortion based on their current density, loss, and length, as well as their end connectors [68]. Thick center conductors for better heat conduction properties combined with single metal rather than braided outer conductors provide the best performance for a given length. As PIM grows with cable length [69, 70], keeping cables very
Figure 3.14: An attenuator excited by a two tone signal, $x_1(f) + x_2(f)$, reflects part of the original signal back into the system and generates distortion coupled to both the input and output port.

Electromagnetic interference (EMI) detrimentally effects system performance through inter-channel coupling before gain elements. Cables are weak radiators which are not commonly a problem in RF systems for linearity, but at dynamic range levels in excess of 100 dB can become a system limiting concern. Cables radiate at levels 50 – 60 dBC from the applied signal and receive at a similar loss, with very small gains achieved by adding shielding layers. Shielding layers from manufacturers always connect at the input to the connector, which leaves a current return path from the outside of the connector back across the cable. This path is very weak but high enough to limit system performance. Reduction of the EMI coupling requires the cable to be put inside of a conductor which is soldered at the ends to the equipment connectors, not the cable itself. This configuration results in a current loop which captures the radiation from the outside return path, severely attenuating the radiation.
The connector can be thought of as an extremely low loss attenuator. Signal reflections and nonlinearities are one for one between a connector and the attenuator of Fig. 3.14. The choice of connector type employed throughout the high power system paths generally specifies the passive distortion generated. According to [68], DIN and N type connectors plated with silver or tri-metal exhibit superior PIM performance over smaller counterparts such as SMA type connectors. The method of connector attachment must be chosen carefully to gain the benefits of larger contact area connectors. Solder connections are always preferable over clamped or crimped connections. Selection of hybrid combiners, directional couplers, attenuators, and cabling all follow the same connector guidelines, and generate much less distortion than is measurable in the system.

### 3.4.5 Summary of Component Linearity

Active and passive components alike generate distortion in high power paths. The distortion they generate is highly dependent on the forward power level and reverse wave at the output of the component. Forward power frequently can not be reduced for a given com-
ponent, stipulating an isolation that must be achieved from the reverse wave for each type of component in order to obtain the dynamic range desired for a particular measurement. System component choice and component matching dictates the lowest level of distortion obtainable under ideal system implementation.

### 3.5 High Dynamic Range Design

This section discusses system design considerations that must be addressed to achieve high dynamic range measurement of distortion signals in the presence of large signals. The major limiting factors are nonlinear distortion generated by test system components and spurious frequency content. Distortion content generated by test system components is largely due to reflections and finite isolation in the system. Spurious frequency content (spurs) can be at the same frequencies as the intended distortion components to be measured, and comes in three particularly troublesome forms: bias spurs, source spurs, and radiatively coupled spurs. The effects of spurs are compounded by active component nonlinearity, especially in amplifiers and vector modulators where compression and limiting occur.

#### 3.5.1 Mixing Effects in Amplifiers

An amplifier driven in the linear region by a signal of interest combined with much smaller spurious content will linearly amplify both the signal of interest and the spurious content. As the amplifier begins to operate in compression, the signal of interest will begin to be limited by the available output voltage. This operation will result in the formation of harmonics of the applied signal in a sinc pattern in the frequency domain. Smaller spurious signals are no longer linearly amplified as the gain of the amplifier approaches zero as the amplifier is driven further into compression by the signal of interest. This process is equivalent to the small signals being multiplied by a rectangular pulse train in the time domain, resulting in a convolution of that frequency content with the signal of interest.

The combination of a signal of amplitude large enough to drive an amplifier into compression (or enhancement) and a signal of amplitude too small to cause compression is applied to the input of an amplifier in Fig. 3.16. The output spectrum of this amplifier
contains a gained version of the large signal, the original small signal, and a mirror image of
the small signal around the large signal and its harmonics. Consider the signal \( x_1 (f_1, t) + x_2 (f_2, t) \), defined by
\[
x_1 (f_1, t) = A_1 \sin (2\pi f_1 t + \phi_1)
\]
\[
x_2 (f_2, t) = A_2 \sin (2\pi f_2 t + \phi_2),
\]
applied to the input of an amplifier. The amplitude, frequency, and phase of the respective
signals are given by \( A_1, A_2, f_1, f_2, \phi_1, \) and \( \phi_2 \), respectively. These signals are graphically
shown in Fig. 3.17. If the amplitude \( A_1 \) is much larger than the amplitude of \( A_2 \) and
larger than the input saturation amplitude of the amplifier, \( A_s \), limiting will occur on the
amplified combined signal \( x_1 (f_1, t) + x_2 (f_2, t) \). The smaller signal, \( x_2 (f_2, t) \), experiences
gain only when the larger signal is less than \( A_s \); its contribution to the output is effectively
reduced to zero when \( x_1 (f_1, t) \) is above \( A_s \), shown in Fig. 3.17. The sudden reduction of
signal amplitude to zero is equivalent to the multiplication of a pulse train, \( P (f_p, t) \), with
\( x_2 (f_2, t) \), where \( P (f_p, t) \) is defined in the period of \( x_1 (f_1, t) \) as
\[
P (f_p, t) = u \left( t - \frac{1}{2\pi f_1} \arcsin (A_s / A_1) \right) - ... \\
u \left( t - \frac{1}{2f_1} + \frac{1}{2\pi f_1} \arcsin (A_s / A_1) \right) + ... \\
u \left( t - \frac{1}{2f_1} - \frac{1}{2\pi f_1} \arcsin (A_s / A_1) \right) - ... \\
u \left( t - \frac{1}{f_1} + \frac{1}{2\pi f_1} \arcsin (A_s / A_1) \right) + A_0,
\]
where \( A_o \) is an arbitrary offset dependent on the amplifier.

Both positive and negative limiting result in the small signal, \( x_2 (f_2, t) \), being
reduced to zero amplitude. The frequency of the pulse train, \( P (f_p, t) \), is then twice the rate
of the frequency of the large limiting signal \( x_1 (f_1, t) \). The mixing spectrum associated with
the multiplication of signal \( P (f_p, t) \) and signal \( x_2 (f_2, t) \) can be obtained by convolving the
Fourier Transform of the respective signals,
\[
P (f_p, t) [G x_2 (f_2, t)] = \Im \{ P \} \ast \Im \{ x_2 \}
\]
\[
= \int_{-\infty}^{\infty} P (\tau) x_2 (\tau - f) d\tau,
\]
(3.12)
Figure 3.16: Amplifier driven by a signal large enough to drive it into compression, \( x_1(f) \), and a much smaller signal \( x_2(f) \). The spectral output of the amplifier contains an amplified version of \( x_1(f) \), a slightly amplified version of \( x_2(f) \), and a version of \( x_2(f) \) symmetrically mirrored around the large signal \( x_1(f) \).

where \( G \) is the gain the small signal experiences, defined by

\[
G = \frac{g}{\frac{1}{\Delta f} \int_0^{1/\Delta f} |P_i(\tau) x_2(\tau)| \, d\tau}.
\]

(3.13)

In (3.13), \( g \) is the linear small signal gain of the amplifier and \( P_i \) is defined as

\[
P_i(f_p,t) = 1 - P(f_p,t).
\]

The spectrum will consist of a mirrored image of the signal \( x_2(f_2,t) \) around the large signal \( x_1(f_1,t) \) and its harmonics at a frequency separation, \( \Delta f \), of \( |f_1 - f_2| \). The extent of mirroring around the harmonics will depend upon the symmetry and offset of the limiting operation. In Fig. 3.18(a), a \(-10.5 \text{ dBm} \) tone at 100 MHz is applied to a BGA2717 amplifier driving it into limiting. A second \(-58 \text{ dBm} \) tone at 99 MHz is summed with the 100 MHz tone, resulting in mirroring of the 99 MHz tone with only 0.4 dB of conversion loss. The output spectrum of the limiting amplifier up to the twenty-sixth harmonic is shown in Fig. 3.18(b). The mirrored tone is converted at each of the harmonics and experiences a varying conversion loss dependent on the order of the harmonic.
Figure 3.17: A large signal, \( x_1 (f) \), drives the amplifier into limiting after the amplitude \( A_s \) is reached. The small signal effectively sees a pulse train where the gain is reduced to zero by the large signal.

### 3.5.2 Radiative Coupling Effects and Sources

Radiative coupling is always an issue in high dynamic range measurement where it is common for nonlinear content from active components to be close in amplitude to the coupled frequency. The coupled signal can easily be on the order of 100 to 120 dBC from the original signal through triple shielded cables. Coupling to the input of the amplifier from cable interaction will result in an output nonlinearity that can easily be within the level of dynamic range the system is to achieve, through the limiting process described in Section 3.5.1. The measurement system can be described by the \( N \)-port of Fig. 3.19, where an extra port represented by port \( N + 1 \) is receiving electromagnetic radiation and transferring to all the other ports in the system.

Particularly troublesome radiative and receiving areas are denoted in Fig. 3.20, and include the inputs to amplifiers, vector modulators, and power supply terminals. Electric field distribution must be considered when configuring a high dynamic range measurement system at all times. Isolation from radiative coupling is best accomplished by not using cables between the DUT and amplifiers. Microwave absorber and reflective surfaces should be employed at crucial points in the system such as the source to cancellation chassis, amplifier connections, and source connections. Battery cables and terminals should never be left unprotected.
Figure 3.18: Amplifier response to a limiting signal of $-10.5 \text{ dBm}$ at 100 MHz combined with a small signal of $-58 \text{ dBm}$ at 99 MHz: (a) the 99 MHz tone is mirrored around the 100 MHz tone with only 0.4 dB conversion loss, (b) and the entire mixing spectrum to the 26th harmonic with the small signal harmonics and mirrored tones amplitude traced.
3.5.3 Spurious Frequency Content

Spurious frequency content is often the limiting factor in system dynamic range. Any non-linearity in a system path will alter the relationship between the spurs and the signal of interest, reducing or eliminating cancellation of the spurious content. The local oscillator in vector modulators undergoes mixing with any spurs coupled to the LO port through the process of hard limiting. Any signal coupling to I and Q inputs will also be upconverted by the mixers within the vector modulator. Spurious content altered by hard limiting or upconverted interference in the cancellation path will not cancel with the linear DUT path spurious content due to the alteration of spur amplitude level and frequency spectra. In this case the spurs will sum and interact, destroying the dynamic range where they reside. When using vector modulators to provide phase shift and attenuation, it is necessary to pre-limit the signal before application to the DUT and cancellation channels in order to achieve suppression of spurious signals. Pre-limiting the signal ensures that the spurious frequency content is mirrored and maintains a defined relationship to the signal.

Several sources of spurious frequency content exist. Spurs from signal generators can be suppressed by pre-limiting the signal, or guaranteeing both the DUT and vector...
modulator paths are linear. Ensuring that the paths are linear by driving the vector modulator far below its intended drive level would seem to be the logical choice. However this would limit the output power effectively diminishing the achievable system dynamic range or would require power amplification. Pre-limiting the signal before both the vector modulator and the DUT path pre-distorts the spurs and allow suppression of spurious content at the cancellation reference plane. In effect this removes amplitude modulated distortion. Spurs from bias circuitry can be removed by the same process, but it is often not feasible to do so. If bias spurs are present in subsequent amplification stages or in the vector modulator power supply, a spectrum of harmonics resulting from AC power rectification will be mixed onto the original signal. Such spurs can only be removed by extracting the signal from farther down the signal path at often undesirable locations such as after amplification. The only other alternatives are heavy EMI filtering or using DC power with clean bias networks.

Due to the effect of limiting within any compressive nonlinearity, coupling from a secondary carrier will act as a spur which is translated directly to a third-order inter-
Figure 3.21: Vector Modulator response: (a) spectrum showing input signal composed of LO, spurious signals, and an injected interferer, (b) spectrum showing mirrored interferer and spurious signals in cancelled output, (c) spectrum with all spurious signals and the interferer removed when input signal is pre-limited.
modulation frequency through mirroring. The signal of Fig. 3.21(a), composed of the LO, spurious content, and an injected interferer is cancelled both without and with pre-limiting in Fig. 3.21(b) and Fig. 3.21(c), respectively. In Fig. 3.21(b), an injected interference tone is symmetrically mirrored around the LO frequency when the signal is not pre-limited before cancellation. All spurious content is also left unaltered due to the mirrored components generated within the vector modulator. When the signal is pre-limited before cancellation, the injected interferer is removed along with all spurious frequency content dBc relative to LO cancellation. The result is the clean cancelled spectrum shown in Fig. 3.21(c). This effect occurs due to the virtual elimination of the small spurious tones during the time period the LO is limited, equivalent to the multiplication of the spurious frequency content by a pulse train. Fig. 3.22 shows the response of an amplifier driven by a signal large enough to cause limiting and a small interferer. The original interferer is mirrored around the large

Figure 3.22: Response of an amplifier driven into limiting by a signal composed of large signal and a small interferer over interferer input power.
signal according to

\[ P(f_p, t)[Gx_2(f_2, t)] = \Im\{P\} \star \Im\{x_2\} \]

\[ = \int_{-\infty}^{\infty} P(\tau)x_2(\tau - f)\,d\tau, \tag{3.15} \]

with the power of the mirrored signal increasing linearly with the interferer. Actual third-order distortion is not seen until the interferer signal becomes appreciable compared to the large signal, but the effective dynamic range is reduced by about 40 dB. Often inter-channel coupling occurs which will be mirrored to intermodulation frequencies by this effect. Inter-channel coupling can result from radiatively coupled components at the input to an amplifier through the finite isolation of cabling. It can also result from a reflected signal traveling through the finite isolation of the amplifier. The coupling levels of these spurs are very different in the DUT and vector modulator paths, and are thus extremely difficult to remove. Isolation must be provided against the radiation coupling of other channels into the gain path of another channel as well as from reflected signals at the output of any active devices. The achievable isolation becomes the effective limit on the dynamic range of the system. Accomplishing radiation isolation requires the complete shielding of every circuit in the cancellation path from any cable containing another channel. In the system of Fig. 3.2, the two channels are broken into two dual shielded rack mountable chassis to ensure adequate isolation.

### 3.5.4 Reflection Effects on Nonlinearities

The nonlinearities in the system are to a large extent dependent on the isolation and reflection of the system. Although the channels are separate in the measurement system, the hybrid combiner has finite isolation between ports one and two in Fig. 3.23. A reverse signal containing the frequency content of the opposite channel, reduced in amplitude from the return loss at port three and port four, will travel toward the output of the opposite amplifier. Nonlinear behavior in the circulators and the amplifier will be caused by the small reverse wave interacting with the high power forward wave. Thus the dynamic range performance of the system is dependent across frequency on the isolation of the combiner and the return loss at all the combiner ports. If adequate reverse wave isolation is assured,
Figure 3.23: A hybrid combiner is shown with a signal $x_1(f)$ incident at port one and a signal $x_2(f)$ incident at port two. The signals are isolated from direct transfer between port one and two, but reflections from port three and port four create a reverse wave into ports one and two. When ports three and four are well matched the isolation between port one and two is maintained, but if either port three or four is not well matched the isolation is reduced significantly.

The isolation of the hybrid combiner is only useful as long as the return loss of the DUT exceeds the combiner isolation. The reflection from the DUT will be split through the combiner and will travel in reverse to the circulators and amplifiers in the system. Antennas and other non-matched devices result in a much lower channel isolation and thus a much lower dynamic range than matched devices. The reverse isolation can be increased by further addition of circulators in order to offset DUT matching, however, any gains in linearity associated with applying a lower magnitude reverse signal to the circulators will
be lost.

3.5.5 Summary of Nonlinear System Design

The DUT path must be distortion free within the dynamic range of the system, with the exception of the DUT itself. Radiative coupling must be kept to an absolute minimum to prevent limiting amplifier mixing to distortion frequencies. Spurious content from sources and supplies must be pre-limited to equalize the feed-forward and DUT paths for cancellation of that content to occur. Key to the architecture is the use of individual linear amplifiers for each test signal and the effective backward wave isolation provided by the isolators. Isolators are necessary to limit the output of the amplifier from nonlinear interaction with other stimuli through the finite isolation of the hybrid combiner. DUT return loss must be considered to assure proper isolation. The isolation is important in the cancellation branch as well, where reflected wave components can interact with the nonlinear junctions contained within the vector modulators. While isolators provide needed isolation, they also limit system bandwidth and produce low level distortion even under ideal use. System bandwidth limitations can be overcome by using high power switches designed for low PIM performance to switch between isolators in different bands.

3.6 System Applications

Two common situations often occur in distortion measurement, low level intermodulation distortion (IMD) detection and high level distortion products within the applied signal bandwidth. Cancellation can facilitate measurement in both of these situations, but alteration of the test setup is required. In the case of low level IMD detection, such as is necessary in PIM testing, extreme dynamic range is accomplished by using multi-channel cancellation and with pre-limiting to enable spur suppression. Digitally modulated signals have both co-channel and adjacent channel interference (ACI). While ACI has received much attention due to its ease of measurement, co-channel distortion is not directly observable from the spectral measurement of the output signal and is instead quantified by waveform signal quality metrics such as signal-to-noise and distortion ratio (SNDR), error vector magnitude (EVM), and the correlation coefficient ($\rho$) [71]. A single channel linearly
driven cancellation system can directly display the co-channel distortion by removing the original output signal. Dynamic range enhancement for PIM applications and digitally modulated signals are discussed in the following subsections.

### 3.6.1 Dynamic Range Enhancement for PIM Measurement

In application of the system to PIM measurements, cancellation levels achieved for a two-tone signal were 35-40 dB over power and frequency for a single iteration of (3.8). This increased to approximately 50 dB upon a second iteration. This level of performance was achieved from 380 MHz to 1 GHz in the system of Fig. 3.2, limited by the power amplifiers and available isolators. Dynamic range of the system was increased to 113 dB at 5 W of output power for 100 Hz signal separations and above by suppressing radiative coupling and reflected signals while providing cancellation in excess of 40 dB. Even at tight signal separations of less than 10 Hz at least 106 dB of dynamic range was obtained, as shown in Fig. 3.24. This limit was established by the uncorrelated phase noise of the reference
oscillators. The finite resolution of amplitude measurements also impacted the dynamic range as it was the prominent source of cancellation error. Amplitude measurement error increased as the receiver became more saturated. Experimentally the cancellation algorithm is more effective at lower power and can be reapplied to compensate for receiver saturation in high power testing, where subsequent iterations return the receiver to linear operation.

3.6.2 Digitally-Modulated Signals

The measurement system can also be used in distortion measurements with digitally-modulated signals with results equal to or exceeding those in [64, 72, 73]. Pre-limiting in both the DUT and feed-forward paths by the limiting amplifiers in Fig. 3.2 allows correlated spurious frequency content to be suppressed along with the probe stimulus. With wideband signals such as wideband code division multiple access (WCDMA), limiting operations such as those occurring in vector modulators result in a large degree of spectral regrowth. If the signal is driven within the LO input specification of the vector modulator, then the
signal is distorted as if it were run through a limiting amplifier. Driving the LO port at a reduced power level results in linear mode operation in which the phase of the wideband signal can be altered without causing increases in the error vector magnitude of the applied signal. Typically the noise associated with wideband signals significantly exceeds spurious frequency content from signal generators eliminating the need for spur suppression. The system of Fig. 3.2 was altered such that only one channel was used and limiting amplifiers were removed for WCDMA testing.

The linear mode of operation, where the limiting amplifiers in Fig. 3.2 are removed, requires calibration as the gain paths are usually unbalanced. This calibration is carried out directly with WCDMA signals instead of discrete tones as in the high dynamic range PIM application case. Once this is accomplished, signal cancellation can be carried out normally, as long as the DUT path is linear except for the DUT itself. A WCDMA signal, shown in Fig. 3.25 and Fig. 3.26, was applied to the vector modulators in order to demonstrate the ability to operate with wideband signals when they are supplied to the vector modulator LO
Figure 3.27: WCDMA signal applied to a Mini-Circuits VLF-400 low pass filter. Cancellation is shown for both equalized and unequalized group delay.

port instead of being upconverted through the I and Q inputs. In Fig. 3.25 cancellation in excess of 40 dB is demonstrated with no measurable spectral regrowth. The measured error vector magnitudes of the original WCDMA signal and the feedforward signal were 1.21\% and 1.27\%, respectively. The applied WCDMA signal, the output signal of a BGA2716 amplifier, and the remaining distortion components after cancellation in excess of 40 dB are shown in Fig. 3.26. The cancelled signal is offset by −5 dB to increase display intelligibility in Fig. 3.26.

Group delay differences of the feed-forward and DUT paths limit the achievable cancellation bandwidth [59–62, 64]. Testing of devices with filters pronounces this effect greatly due to both a larger difference in constant group delay and frequency-dependent group delay response. This effect is illustrated in Fig. 3.27. The shape of the cancelled WCDMA signal in Fig. 3.27 is the result of group delay mismatch when testing a Mini-Circuits VLF-400 low pass filter as the DUT. The limited cancellation is due to the filter group delay response. Enhanced cancellation can be achieved by equalizing the group delay
in the DUT and feedforward path. To examine this a second VLF-400 filter was inserted in
the feed-forward path after the vector modulator. Fig. 3.27 shows the impact of doing this
by comparing the equalized group delay cancelled WCDMA spectrum with the spectrum of
the ”cancelled WCDMA” signal without group delay equalization. Thus approximate group
delay equalization effectively eliminates the bandwidth limitation that otherwise occurs.

3.6.3 Summary of System Applications

The measurement system presented in this chapter is unique in that it is a fully
automated, highly linear feed-forward cancellation system with performance equivalent or
exceeding that of existing manual systems. It provides the capability to measure signals
requiring extremely high linearity, such as PIM signals, as well as inband uncorrelated
distortion in wideband signals. The system presented is analyzed for not only linear effects
on performance including group delay, signal reflections, and mismatch, but also nonlinear
effects from components and spurious frequency content.

3.7 Conclusion

The measurement of small signals in the present of large signals has been per-
formed primarily with filters. Filter based measurement methods allow a very specific and
limited signal bandwidth to be measured due to the obtainable steepness of filter skirts
and the lack of tunability in filters. As passive intermodulation distortion continues to in-
crease in importance in RF communications systems, so too does the need for broadband,
selective high dynamic range measurement systems. Intermodulation products are gener-
ally dependent on amplitude modulation, requiring measurement capabilities down to small
resolutions as low as one hertz. It is not feasible to create filters with such tight passband
to stopband tolerance. Techniques capable of these measurement capabilities such as feed-
forward techniques must be pursued to shed light on PIM processes in RF communications
systems.

In this chapter, feed-forward cancellation theory was presented resulting in formula
analytically predicting the phase shift necessary for feed-forward cancellation from a single
amplitude measurement. Previously feed-forward cancellation has been performed using
iterative methods such as gradient or power minimization techniques, requiring upwards of ten iterations to provide cancellation levels in excess of 20 dB. The presented formula results in cancellation of a minimum of 33 dB under the worst case phase separation conditions and can be reapplied to yield further cancellation in excess of 50 dB. The formula allows saturated receivers to be compensated as each iteration yields greater cancellation.

The construction of an automated vector modulator based feed-forward cancellation system was presented. Vector modulator control and the effects of matching on vector modulator amplitude and phase accuracy were discussed, allowing design of the feed-forward cancellation path. Noise coherence in separate source, digital to analog, and same source coupled implementations are reviewed, suggesting superiority of same source coupled implementations for noise performance. Phase locking of sources is shown to be of utmost importance in cancellation for maintained cancellation.

Although automated linear cancellation is a difficult problem, it is only half of the measurement system design. Measurement of nonlinearities requires that the system be more linear than the distortion of interest. A system is only as linear as its components. Unfortunately passive component linearity is seldom specified; if it is specified, like active components, it is for a forward applied signal. In the developed measurement system, most of the components are exposed to nonstandard signals including large signals combined with extremely small signals at the input port, large signals with a small applied signal at the output port, and a large signals with large applied signals at the output port. All of the general lab components necessary to build the system, including amplifiers, circulators, combiners, attenuators, and terminations were analyzed for their linearity to these nonstandard types of signals. Small forward and reverse signals are mirrored in amplifiers when a large signal is present, resulting in signal content falling on distortion frequencies. Reducing these reverse signals is shown to be the ultimate limit of system linearity.

Each test component reacts in a given manner to applied forward and reverse signals. The forward and reverse wave amplitudes are dictated by the isolation and matching between every component in the system. Radiation into sensitive terminals feeding into active devices and power supplies provide an independent route for these small spurious forward and reverse waves to enter the system. The compressive nonlinearity in amplifiers was shown to mirror this spurious content around the large intended signal, resulting in spurious frequency content falling on distortion frequencies and other non-ideal locations.
across the spectrum. The mirroring effect in amplifiers results in distortion content that increases in a one to one ratio with the small interference signal, thus gains in linearity from reducing power are not possible when this type of distortion is present. Reduction of cables to an absolute minimum along with the use of shielding and appropriate wireline isolation is suggested to be the most effective means for reduction of these effects.

Spurious content from signal sources and power supplies are mirrored by amplifier nonlinearities as well as small radiatively coupled and reverse wireline signals. Vector modulators, used for phase and amplitude control, must drive the local oscillator port of the device which is applied to several amplifiers and mixers within the device. A signal going through a vector modulator will mix all the small spurious frequency content in the local oscillator signal with the large intended signal, mirroring its content symmetrically around the large applied signal. The compressive nonlinearity in the DUT test path is generally not the same as the feed-forward path, thus the spurious content will not cancel at the signal combination plane. Limiting the signal from the sources and using direct current power is shown to pre-distort both signal paths, allowing cancellation of all spurious content without affecting DUT nonlinear output.

The developed measurement system is applied to both broadband, high dynamic range PIM measurement and uncorrelated inband distortion measurement from wideband signals. The wideband signal measurement capability allows automated measurement of uncorrelated inband distortion, which cannot be measured by filtering or other means. Group delay effects on wideband signals are also analyzed, showing that mismatches in group delay severely reduce cancellation bandwidth. Pre-distortion of group delay in the feed-forward path can greatly improve the mismatch and extend the cancellation bandwidth. High dynamic range two-tone PIM measurement can be conducted up to 115 dBc from the probe signal at tone separations approaching 1 Hz. This capability allows the measurement of previously unmeasurable distortion characteristics unique to electro-thermal distortion, which is the focus of the remainder of this dissertation.
Electro-Thermal Passive Intermodulation Distortion
4.1 Introduction

Lossy passive components and the circuits that are constructed from them have traditionally been treated as though they are linear. Under high power excitation, these devices can generate distortion similar to an active device. Distortion from these passive devices in the high power transmit channel of RF communications systems can fall within the receive band of the same system or other nearby systems. These distortion products act as an interfering signal in the receive band, resulting in decreases in range and data transmission capabilities in the communication system.

Many physical mechanisms generating passive intermodulation distortion have been suggested in the literature including metal-oxide-metal contacts, metal-metal contacts, material defects, and dirty contacts [1–8, 15]. Each of these mechanisms theoretically provides a nonlinear response and are difficult to isolate from each other. Nonlinear conductivity has been suggested as a physical process responsible for PIM [74]. Thermal contributions to this process have been largely overlooked due to the difference in time constants between microwave and thermal processes, which often are separated by many orders of magnitude. Thermal and electrical signal interaction can occur when the modulated RF signal has baseband components up to several megahertz and the periods of these baseband signals are comparable to the thermal time constants of the device. These thermal transients cause time-varying resistance resulting in intermodulation components at RF frequencies.

Electrical components operate on voltages and currents, while the heat transfer system they are linked to operates on dissipated electrical power. These systems operate on different orders of the same signal, and it is shown here that this results in non-integer order Laplacian system behavior. The non-integer order Laplacian system behavior manifests in long-tail transients that can not be described by exponentials and frequency dispersion defined by fractional order functions. Long-tail transients and their corresponding fractional order frequency responses can not be described by an integer-order differential equation and instead requires a fractional calculus based differential description.

In this chapter an understanding of thermally-induced nonlinearities is developed by analyzing a general resistive component with a signal whose envelope contains instantaneous power components at baseband frequencies. Heat conduction theory is reviewed
in Section 4.2 for a rectangular resistive element revealing the linkage between electrical and thermal domains as well as the basis for the electro-thermal mixing process. Thermal dispersion, as it affects electrical signals, is then analyzed in Section 4.2.2 to account for time scaling and long-tail memory effects, justifying the use of the fractional thermal model developed in Section 4.2.3. A fractional circuit model for describing the electro-thermal process, based on the developed fractional thermal model, is presented in Section 4.3.1. An approximation method for circuit simulators using solvers for systems of first order differential equations is presented in Section 4.3.2, allowing the developed theory to be used with little modification to current simulation methods. Three separate microwave elements, terminations, resistors in attenuators, and resistive material including gate polysilicon in integrated circuits are measured for thermal properties and electro-thermal distortion in Section 4.4, Section 4.5, and Section 4.6, respectively. The developed theory and simulation methods are favorably compared against measurement of each device. Section 4.6 further shows the impact of heat sinking and methods to control the direction of heat flow through element construction.

4.2 Heat Conduction and Electro-Thermal Distortion

Electro-thermal conductivity modulation generally has not been considered as a dominant PIM process due to the difference between thermal time constants and the period of high frequency electrical signals. Heat conduction typically occurs over a time period of a few milliseconds to a few seconds while RF electrical signals occur in less than a microsecond. Conventional thought would dictate an averaging of the electrical signal power over the thermal time constant, which does indeed happen if there are no baseband components to the signal. However, if baseband components or amplitude modulation exists, the power of the electrical signal can modulate at time periods within the thermal bandwidth of a material. This modulation leads to strong coupling between electrical and thermal domains that generates electro-thermal distortion. The coupling of electrical and thermal domains, resulting in electro-thermal PIM, is discussed in Section 4.2.1. Circuits operate on electric and magnetic fields and their scalar counterparts, voltages and currents, while heat transfer operates on electrical power. The coupling of different order processes results in
memory, manifesting as long tail transients. The behavior of electro-thermal time evolution is discussed in Section 4.2.2. Full thermal finite domain simulation is needed to simulate electro-thermal distortion with current methods as the thermal solution of a system is generally not analytically tractable. A reduced order, fully analytic thermal solution based on a fractional derivative is presented Section 4.2.3. The reduced order thermal solution is motivated by its capability for accurate thermal simulation using compact models and its analytic nature, which results in a closed form solution for electro-thermal PIM.

4.2.1 Electrical and Thermal Coupling

The understanding of electro-thermal distortion begins with knowledge of the coupling between electrical and thermal domains, heat conduction, and the electro-thermal mixing process. Metals exhibit a thermally-based resistance that derives from the thermal dependence of electron scattering by lattice vibrations in that material [38]. This process is termed the thermo-resistance effect, and models the specific resistivity, \( \rho_e \) (units of \( \Omega \cdot m \)) of a material as a function of temperature, \( T \) [39]:

\[
\rho_e(T) = \rho_{e0}(1 + \alpha T + \beta T^2 + \ldots).
\]  

(4.1)

Here \( \rho_{e0} \) is the static resistivity constant and \( \alpha \) and \( \beta \) are constants representing the temperature coefficients of resistance (TCR).

The thermo-resistance equation above couples the thermal domain to the electrical domain. Coupling from electrical to thermal domains results from dissipated electrical power, termed self-heating or joule heating. The heat generated per unit volume, \( Q \) (units of \( W \cdot m^{-3} \)) from self heating is

\[
Q = J^2 \rho_e,
\]  

(4.2)

where \( J \) is the current density vector in units of \( A \cdot m^{-2} \). The heat produced drives the heat conduction equation

\[
\nabla \cdot \left( \frac{\nabla T}{R_{th}} \right) - \rho_d c_v \frac{\partial T}{\partial t} = Q,
\]  

(4.3)

where \( c_v \) is the thermal capacity (units of \( J \cdot K^{-1} \cdot kg^{-1} \)), \( \rho_d \) is the density (units of \( kg \cdot m^{-3} \)), and the thermal resistance, (units of \( K \cdot W^{-1} \)) is

\[
R_{th} = \frac{\Delta T}{P} = \frac{\Delta T}{P^2 R},
\]  

(4.4)
where $\Delta T$ is the change in temperature with an injected thermal power $P$.

Thermal capacity combines the ability of a material to store heat by raising its temperature and the rate that heat is conducted to the surrounding environment. The thermal capacity at constant volume can be expressed as (units of J·K$^{-1}$)

$$C_v = \left( \frac{\partial Q}{\partial T} \right)_v = T \left( \frac{\partial S}{\partial T} \right)_v = c_v \rho_d V. \quad (4.5)$$

Here $S$ is the entropy of the system. The density, $\rho_d$, and volume of the material have been absorbed into the definition of the thermal capacity to represent a system of given dimensions.

The forcing function, in this case joule heating (5.18) with (5.17), can be substituted into the heat conduction equation, (4.3), yielding

$$\nabla \cdot \left( \frac{\nabla T}{R_{th}} \right) - C_v \frac{\partial T}{\partial t} = J^2 \rho_e (1 + \alpha T + \beta T^2 + \ldots), \quad (4.6)$$

which describes a nonlinear system. In practice, the first-order coefficient of the thermo-resistance equation, (5.17), is several orders of magnitude larger than any higher order coefficient ($\alpha >> \beta$) in most metals, leading to its dominance in the distortion spectrum of a resistive device.

The electro-thermal process can be separated into static and dynamic components, with static and dynamic power signals $P_s$ and $P_d$, respectively. The static and dynamic power signals are dissipated in the respective static and dynamic series resistance components $R_s$ and $R_d$. The power dissipated over these resistance components is converted to the heat signal $Q(P_s + P_d)$ and filtered by the material thermal response. When a single RF tone is applied to a resistive element, the electro-thermal process responsible for modulating the device resistance provides a resistance with negligible dynamic variation as the thermal capacity cannot react quickly enough to the high frequency signal to significantly heat or cool the resistive material, resulting in only a step change in the static resistance due to average power dissipation. The situation changes when two or more signals are applied to an electro-thermal system, as a dynamic, periodically varying resistance component becomes possible in addition to the step change in static resistance incurred from average power dissipation.

A two-tone signal $V_i(t)$, the spectrum of which is shown in Fig. 4.3(a), has a time-varying signal envelope, shown in Fig. 4.1. The instantaneous power of this signal
Figure 4.1: A two-tone signal, $V_i(t)$, in the time domain composed of a 400 MHz one volt cosine and a 440 MHz one volt cosine. The signal is amplitude modulated at the difference of the two tone frequencies, 40 MHz.

Figure 4.2: The baseband power component of a two-tone signal, $V_i(t)$, in the time domain composed of a 400 MHz one volt cosine and a 440 MHz one volt cosine.
varies periodically at the beat frequency of the two-tone input to the device, shown in the time domain in Fig. 4.2, and contains both sum and difference frequency components as shown in Fig. 4.3(b). If the beat frequency is within the bandwidth of the lowpass filter shown in Fig. 4.3(c), periodic heating and cooling of the element occurs at baseband frequencies. Consequently, the resistance of the element varies periodically. In effect this periodic oscillation creates a passive mixer producing intermodulation distortion through upconversion of the envelope frequencies at baseband to RF frequencies, resulting in the voltage output spectrum shown in Fig. 4.3(d).

The electrical and thermal domain couple at RF electrical frequencies if baseband components in the power signal exist, governed by the thermal bandwidth of the system. The filtering property shown in Fig. 4.3 of the thermal domain is due to the diffusive nature of heat conduction. The coupling of the wave equation nature of the electrical signal and the diffusive nature of the heat conduction system result in memory which manifests itself in long tail transients in the time domain. Electro-thermal memory determines the time required to generate a solution for a thermal coupled electrical system. It also determines the ability of conventional circuit simulators to find a solution without coupling to a thermal simulator, as most circuit simulators use a set, small number of solution points for a prediction of the next solution point. Such a solution scheme may not be valid for processes with long memory. Electro-thermal memory is discussed in the following section.

4.2.2 Fractional Time Evolution

Thermal coupling not only leads to intermodulation distortion, but also signal dispersion due to the diffusive nature of thermal transport. Thermal dispersion can be analyzed by studying the natural response of the heat conduction equation when applied to a semi-infinite rectangular structure. In a semi-infinite rectangular structure, the heat equation becomes the one-dimensional differential equation:

\[ C_v \frac{\partial T}{\partial t} = \frac{1}{R_{th}} \left( \frac{\partial^2 T}{\partial x^2} \right) (t > 0, -\infty < x < 0). \]

(4.7)

Insight into the nature of thermal coupled distortion products can be obtained from the solution of (4.7) assuming a temperature of zero at \( x = -\infty \) and an initial temperature
Figure 4.3: Passive mixing process inherent in coupled electrical and thermal systems with (a) input spectrum of voltages $V_i(f)$ resulting from two-tone excitation, (b) input power spectrum $P_i(f)$ resulting from a two-tone excitation, (c) component of the input power spectrum $P_i(f, T)$ at baseband able to interact with the thermal response, and (d) output spectrum of voltages $V_o(f, T)$ after electro-thermal mixing has occurred.
The temperature response has an error function solution [75]:

\[ T(t) = T_0 \cdot \text{erf} \left( \frac{x}{2\sqrt{\kappa t}} \right), \quad (4.8) \]

where \( \kappa \) is the thermal diffusivity (units of \( \text{m}^2 \cdot \text{s}^{-1} \)) and

\[ \kappa = \left( R_{\text{th}} C_v \right)^{-1}. \quad (4.9) \]

It is apparent from this simple solution that in diffusive situations time is effectively scaled, sometimes referred to as being dilated. The temperature at a given point in the material progresses with respect to the square root of time. Strong coupling between electrical and thermal signals due to heat conduction by electrons leads to time dilated electrical signals, where the time dilation is dependent on the strength of electrical and thermal domain coupling. The electrical signal of interest becomes dependent on a much slower thermal process, leading to inseparable time scales between electrical and thermal signals. In turn, this results in a non-exponential response, sometimes called a long tail or long memory response [48]. This phenomenon is seen in Fig. 4.4, which shows the measured voltage step response to a current step input of a 100 Ω platinum resistive element experiencing self-heating. The response initially approximates an exponential, but then continues to increase with fractional power law memory.

In the literature [39, 76, 77], this behavior has been referred to as a stretched exponential response or the response of a system possessing many time constants [39]

\[ T(t) = \sum_i R_i \left( 1 - e^{-t/\tau_i} \right). \quad (4.10) \]

This response can also be viewed as that of a filter with an infinite number of poles and zeroes [78, 79]. A new view describing the phenomena comes from the fractional calculus description of a heat conduction system, which exactly describes fractional power law memory. Fractional calculus leads to a reduced-order model for an electro-thermal system accounting for time dilation and fractional memory in the electrical signal, allowing accurate simulation of electro-thermal systems in circuit simulators without full thermal domain simulation. A fractional calculus based thermal model is derived in the following section encompassing these attributes.
Figure 4.4: Normalized power step response of resistive element exhibiting fractional power law behavior.

### 4.2.3 Fractional Heat Conduction System for Lossy Lumped Components

In this section, a reduced-order thermal model of a general lumped lossy element is developed that is suitable for circuit-level models. Dynamic circuit-level models of thermal effects have used full-domain simulations and compact models [39,74], however these models are either too computationally intensive for integration into time-stepping simulators or are not able to fully model long tail memory and dispersive effects [74]. In order to arrive at a compromise between these two simulation extremes, while still maintaining the accuracy of a full-domain solution at each time point in a time-stepping circuit simulator, it is necessary to describe heat conduction from an element as though it were localized at a single spatial point. Such a description of the heat conduction equation is described here, which is the cornerstone of an accurate circuit model suitable for time-stepping simulation and an
analytic derivation of electro-thermal distortion.

In any heat conduction problem, geometry and boundary conditions determine the solution of the heat transfer problem in a given medium. In order to reduce the heat conduction equation to its fractional equivalent, the heat transfer problem must be of a unidirectional nature in a semi-infinite thermal domain [49, 80, 81].

A resistive element can be readily modeled as a semi-infinite heat conduction system, pictured in Fig. 4.5. The resistive element is bounded by an insulator, a thermally conductive substrate, and a thermally conductive metal. The metal and substrate are mounted on a heat sink. The resistive element is distributed with units of ohms per square, thus the loss is distributed across the whole structure and it can be partitioned into infinitesimal elements. The thermally conductive substrate becomes large compared to the infinitesimal size of a resistive element, enabling the system to be modeled as a semi-infinite plane with parallel point heaters. Due to insulation on the top of the heater element and point heater symmetry, heat transfer can be considered to be one dimensional near the actual heater element, except at the end points, which leads to negligible error for electro-thermal distortion calculations. In the electrical domain, infinitesimal resistance elements are in series, thus the combined signal with heating effects is equivalent to the lumped element resistance.

In the thermal system in Fig. 4.5, temperature must be finite and boundary conditions are given by,

\[ T(0, x) = 0 \]  \hspace{1cm} (4.11)

\[ T(t, 0) = T_H(t) \]  \hspace{1cm} (4.12)

where \( T_H(t) \) is the surface temperature of the resistive element, which in steady state is

\[ T_H(t) = Q(t) \cdot R_{\text{th}}. \]  \hspace{1cm} (4.13)

The ambient temperature can be added to the solution through superposition, so no loss of generality is incurred by assuming an initial temperature of zero to simplify the dynamic solution. The one dimensional heat equation is dependant on both time and space, giving rise to the need for separation of the domains. One method to accomplish this is the Laplace transform, which has been shown to be invertible under these boundary conditions [82].
Figure 4.5: Standard microwave chip termination highlighting thermodynamic environment: (a) for the entire structure; and (b) for a single infinitesimal lossy element.
Upon transforming the temperature,
\[
\frac{1}{R_{th}} \left( \frac{\partial^2 T(s, x)}{\partial x^2} \right) - C_v s T(s, x) = 0,
\] (4.14)
an equation dependent only on space, \(x\), is obtained, where \(s\) is the Laplace variable. Imposing boundary conditions at \(x = 0\) and \(x = -\infty\) (where the temperature must be zero), the solution to the ordinary differential equation (4.14), is
\[
T(s, x) = T(s, 0) e^{x \sqrt{s C_v R_{th}}}. \tag{4.15}
\]
Taking the spatial derivative yields
\[
\frac{\partial T(s, x)}{\partial x} = T(s, 0) \sqrt{s C_v R_{th}} e^{x \sqrt{s C_v R_{th}}}. \tag{4.16}
\]
Combining (4.15) and (4.16) and taking the temperature at the surface \([48–51,83]\):
\[
s^{-1/2} \frac{\partial T(s, 0)}{\partial x} = T(s, 0) \sqrt{C_v R_{th}}. \tag{4.17}
\]
The variable \(s\) in (4.17) represents a derivative of first order, thus \(s^{1/2}\) indicates a derivative of half order. A derivative of non integer order requires the use of fractional calculus. Here it is defined in the Caputo sense as \([48–51,84]\),
\[
C_a D^q_t f(t) = \frac{1}{\Gamma(n-q)} \int_a^t \frac{f^{(n)}(\tau)}{(t-\tau)^{q+1-n}} d\tau \tag{4.18}
\]
\((n-1 \leq q < n)\),
where \(C\) denotes Caputo, \(a\) denotes the lower limit of the integral, \(q\) is the order of the derivative, and \(t\) is the variable the derivative is with respect to.

The path back to the time-domain solution, the inverse transform, is found through the inverse Laplace transform of the reduced order system. The fractional Laplace transform, and its inverse, are defined as \([48–51]\):
\[
L \{ C_0 D^q_t f(t); s \} = s^q F(s). \tag{4.19}
\]
Applying the inverse transform to (4.17) yields,
\[
0 D_t^{-1/2} \frac{\partial T(t, 0)}{\partial x} = \sqrt{C_v R_{th}} T_H(t). \tag{4.20}
\]
Rearranging and following [48–51,81],
\[
\frac{1}{R_{\text{th}}} \frac{\partial T(t, 0)}{\partial x} = \sqrt{C_v R_{\text{th}}^{-1}} D_t^{1/2} T_H(t),
\]
which is the fractional form of the one-dimensional heat equation. Other geometries can be shown to follow similar solutions, as long as the problem is unidirectional and semi-infinite.

A standard coaxial microwave termination, shown in Fig. 4.6, has a cylindrical resistive disk element and has a one dimensional thermal description similar to the rectangular chip construction. The fractional description is not exact in the cylindrical case [49,81], but still maintains reasonable thermal accuracy at the point of interest [84] while accounting for the time scaling [85,86] so important to the description of electro-thermal systems. In this case, the resistive element can still be decomposed, and the heat flow is one dimensional and unidirectional radially. The outer metallic conductor acts as a heat sink, once again allowing a model of a semi-infinite plane with parallel point heaters with respect to the inner conductor.

The key to understanding the usefulness of (4.21) comes from realizing the non-locality of the fractional derivative operator. This non-locality implies that the fractional derivative of a point contains the complete knowledge of the past history of a function [48, 51]. Because of this property it is possible to use the fractional derivative as a reduced-order
model where it is not necessary to solve the full thermal domain to obtain the temperature or heat flux at a point. This fractional-order differential equation is only dependant on the heat flux and surface temperature at a single point and accounts for the time scaling [85,86] that occurs in the thermal domain. These aspects of the reduced-order heat conduction equation allow both the derivation of an analytic electro-thermal model and accurate prediction of electro-thermal distortion.

4.2.4 Summary of Electro-Thermal Nonlinearity

Electrical and thermal systems couple together due to low frequency variations in the dissipated electrical power. The loss of a lumped component varies with its temperature at the frequency of the low frequency variations of the electrical power. The two processes are coupled in a square root relationship between the differential equations, resulting in non-integer order Laplacian behavior that must be described by fractional calculus. The non-integer order Laplacian behavior manifests through long-tail transients and frequency dispersion of fractional order in the electrical response. The thermal domain behavior can be represented in the electrical domain by a fractional order differential equation, which provides information that could only be obtained otherwise by large scale 3D simulation. Accurate multi-physics modeling in a compact format as well as an analytic formulation for electro-thermal PIM is made possible through the reduced-order thermal formulation. Electro-thermal compact models and analytic derivations based on the reduced-order thermal model are presented in the next section.

4.3 Electro-Thermal Circuit Models

Thermal models are used to determine the operational temperature of electrical components. This information is used to correct the simulated I-V characteristics of a device and determine corrected maximum device operating conditions. The two most prominent methods for electro-thermal simulation are detailed numerical simulation and compact thermal models [39]. The detailed numerical methods include finite difference, finite element, and boundary element methods. These methods provide the entire temperature distribution of a component but are computationally intensive. To enable simulation on a large
Figure 4.7: A standard one pole electro-thermal model commonly used to approximate the thermal response of a material.

scale, a compact thermal model, usually composed of a small electrical network, is preferred because it is computationally efficient. The compact model of the self heating effect in a resistive element is simply a RC filter with the power dissipated in the element providing a current into the thermal resistance and thermal capacity of the device, as shown in Fig. 4.7.

The basic compact model can not capture the time constant of the thermal process accurately because it is based on the assumption that the thermal process is exponential. Significant deviations from exponential behavior are commonly seen [87], but electrothermal models still commonly use an exponential fit to give the time constant of the compact thermal model. Although this model accurately gives final values of temperature and approximates real circuit time constants, it fails to accurately model the frequency response of a real thermal model in electrical simulators.

In this section, compact simulation models are discussed that accurately model both transient and frequency response behavior of electro-thermal systems. Section 4.3.1 presents a fractional compact circuit model based on the reduced-order thermal solution that accurately models frequency response and time constants of electro-thermal processes over all frequency. An analytic closed form solution of electro-thermal PIM for a general lossy element is derived based on the fractional compact circuit model and reduced-order thermal model. The closed form solution allows electro-thermal distortion products to be predicted from material properties alone. The simulation model of Section 4.3.1 is dependent on complete memory of the solution function, which is not available in most
predictor-corrector methods employed in standard circuit simulators. A model that can be used in standard circuit simulators is needed to allow immediate use of the results presented in this chapter. A method to create an circuit to approximate the fractional derivative response of the model of Section 4.3.1 is presented in Section 4.3.2 for use with standard predictor-corrector methods. The method allows approximation of the response over a limited bandwidth and is useful for transient or bandlimited simulation.

4.3.1 Fractional Compact Circuit Model for Electro-Thermal PIM

A fractional differential equation representation of the heat conduction equation provides for the creation of a semi-compact circuit model embodying the solution to the heat conduction equation at the point of interest. The analytic model developed in this section is suitable for time-stepping simulators, and directly leads to analytic expressions for electrothermal distortion in lossy elements. Unlike frequency corrected compact models which are approximate over a limited bandwidth [88], fractional models link thermal and electrical domains over all frequency. Time scaling, inverse power law memory, and reduction of the solution of the heat equation from an entire structure to a single point are all accomplished in the fractional thermal circuit model.

The structure of the compact model can be determined from the solution of the application of a harmonic signal to the integer order heat equation for a semi-infinite solid. The solution to this problem is given by [75]

\[ T = Ae^{-kx} \cos(\omega t - kx) \]  \hspace{1cm} (4.22)

\[ k = (\omega/2\kappa)^{1/2}. \]  \hspace{1cm} (4.23)

It is clear that higher harmonics are attenuated as they travel into the medium. The obvious electrical analogy to represent this behavior is a lowpass filter. The lowpass filter for a thermal node is shown in Fig. 4.8, where the thermal capacity is now modeled by a fractional derivative based capacitor and the thermal resistance remains unchanged according to the fractional order heat conduction equation. The thermal capacitance of this model is equivalent to

\[ C_{th} = \frac{C_v}{R_{th}} \]  \hspace{1cm} (4.24)
and is the thermal capacitance measured by a curve fit to the long memory transient resultant from a power step applied to the component being modeled. Heat applied to the thermal model is just the dissipated electrical power in the modeled element. The ambient temperature is incorporated through superposition as a voltage source.

Generation of an analytic PIM expression requires reversion to a discussion of the coupling of thermal and electrical systems. The TCR equation of the device in question is the coupling equation between the two domains, and could be any order polynomial, but is generally described by the linear thermo-resistance

\[
R(T) = R_0 \{1 + \alpha [T(t) + T_a]\},
\]

(4.25)

where \(R_0\) is the reference resistance of the thermo-resistance equation (4.25) measured at 273K on the TCR curve, \(T_a\) is the ambient temperature, and \(\alpha\) is the first order thermo-resistance coefficient. Instantaneous temperature is now redefined as

\[
T(t) = Q(t) \cdot R_{th, eq},
\]

(4.26)

where the equivalent thermal resistance is the compact fractional model equivalent resistance over frequency,

\[
R_{th, eq}(j\omega) = \frac{R_{th}}{1 + R_{th} \sqrt{j\omega C_{th}}}.
\]

(4.27)
The generated heat, (5.18), is thus redefined as

\[ Q(t) = I(t)^2 R_0 \{1 + \alpha [T(t) + T_a]\}. \]  

(4.28)

The voltage is related to the current via Ohm’s law as,

\[ V(t) = I(t) \cdot [R_0 + \alpha R_0 T_a + \alpha R_0 T(t)]. \]  

(4.29)

Remembering that the temperature is defined in terms of the generated heat, which is itself in terms of current and temperature dependant resistance,

\[ T(\omega) = Q(\omega) \cdot R_{th,eq}(\omega) \]

\[ = I(\omega)^2 (R_0 + \alpha R_0 T_a + \alpha R_0 T(\omega)) R_{th,eq}(\omega). \]  

(4.30)

Current, voltage, heat, temperature, and equivalent thermal resistance dependence on frequency will be considered inherent within variables going forward, as defined in (4.26)-(4.30). Substituting this relation back into Ohm’s Law a recursion relation is obtained,

\[ V = I(R_0 + \alpha R_0 T_a) + ... \]

\[ I^3 \alpha R_0 R_{th,eq}(R_0 + \alpha R_0 T_a) + ... \]  

\[ I^5 (\alpha R_0 R_{th,eq})^2 (R_0 + \alpha R_0 T_a + \alpha R_0 T_{R_{th,eq}}) + ... , \]  

which can be written in closed form as

\[ V = R_0 (1 + \alpha T_a) I + \]

\[ R_0 (1 + \alpha T_a) \sum_{n=1}^{\infty} I^{2n+1} (R_0 \alpha R_{th,eq})^n. \]  

(4.32)

Removing the DC term, the analytic representation of electro-thermal PIM is obtained (PIM\(_{ET}\), in Volts):

\[ PIM_{ET} = R_0 (1 + \alpha T_a) \sum_{n=1}^{\infty} I^{2n+1} (R_0 \alpha R_{th,eq})^n. \]  

(4.33)

The convergence of this formula is assured through application of the ratio test

\[ \lim_{n \to \infty} \left| \frac{I(\omega)^{2n+2} (R_0 \alpha R_{th,eq}(\omega))^{n+1}}{I(\omega)^{2n+1} (R_0 \alpha R_{th,eq}(\omega))^n} \right| = \]

\[ |I(\omega) (R_0 \alpha R_{th,eq}(\omega))|. \]  

(4.34)
Absolute convergence requires that the limit of the series ratio be less than one, ensuring decay rather than growth of the series with higher order terms. Convergence is guaranteed for the condition

$$|I(j\omega)(R_0\alpha R_{th,eq}(j\omega))| < 1.$$ (4.35)

The analytic formulation of electro-thermal PIM includes only material and environmental parameters with the fractional derivative embedded within the equivalent thermal resistance. The analytic formulation includes the inverse power law memory contained within the diffusive heat conduction equation, accurately modeling electro-thermal distortion over frequency as well as the transient long tail effects associated with the electro-thermal process. This formulation is useful for design and analysis of single devices, while the circuit model presented in this section is useful for simulation of large designs containing many devices. Unfortunately many circuit simulators cannot implement the fractional derivative based circuit model for transient simulation without additional solver methods. The required solver methods must predict the next solution point based on all previous solution points. Circuits can be described by a system of first order differential equations, which predisposes the solution methods in standard circuit simulators to predictor-corrector schemes based on a small subset of solution points. Overcoming this hurdle necessitates a model that approximates the fractional compact model response through a system of first order differential equations. A model approximating the response of the fractional compact model and a method to synthesize it is presented in the next section.

### 4.3.2 Foster Approximation

In circuit simulators solution methods are generally tailored to methods for the solution of a set of first order differential equations. These methods work on a small subset of past or future solution points as opposed to predicting the next solution point from a method using all previous solution points. A fractional derivative used as a predictor is a method that utilizes all previous solution points. The fractional differential equation can not be broken into a system of first order differential equations, but can be approximated over a limited bandwidth by them. The system of first order differential equations used to approximate the fractional response must be both physically consistent and implemented
in a circuit simulator. In this section, a method to synthesize a circuit approximating the transient and frequency response of the fractional compact circuit model over a limited bandwidth is presented.

The transfer function of the fractional compact circuit model is given by (4.33). It has the form of a lowpass filter with a slope of 10 dB per decade, where the thermal bandwidth of the component, determined by material thermal parameters, defines the passband of the thermal filter. The bandwidth of the passband determines the first pole of this approximation, with subsequent zeroes and poles alternated to give the required filter roll off. This process is demonstrated in Fig. 4.9, where the approximation function oscillates around the desired response.

The bandwidth of the thermal passband is determined by the thermal resistance, $R_{th}$, and capacity, $C_{th}$, of a device, including any response from materials in direct thermal contact with the device. The 3 dB bandwidth of the thermal passband is given in terms of these parameters by

$$f_{3dB} = \frac{1}{2\pi R_{th} C_{th}}.$$  \hfill (4.36)

The desired approximation transfer function can be obtained by alternating poles and zeroes

![Figure 4.9: A series of alternating poles and zeroes approximating a filter slope over a finite bandwidth. The accuracy of the approximation is dependent on the spacing of poles and zeroes. Increases in accuracy or bandwidth greatly increase the order of the approximation function.](image)
at frequencies given by

\[ f_{ci} = f_{co} \left( \frac{f_{c(N-1)}}{f_{co}} \right)^{\frac{i}{2N}}, \]  

(4.37)

where \( f_{co} \) is the lowest frequency of interest, \( f_{c(N-1)} \) is the highest frequency of interest, \( N \) is the order of the approximation, and \( f_{ci} \) is the crossover frequency where the estimated response crosses the intended response. The necessary poles and zeroes can be found according to

\[ f_{pk} = -f_{c2k} \left( \frac{f_{c(N-1)}}{f_{co}} \right)^{\frac{2k-\alpha}{2N}}, \]  

(4.38)

\[ f_{zk} = -f_{c2k} \left( \frac{f_{c(N-1)}}{f_{co}} \right)^{\frac{2k+\alpha}{2N}}, \]  

(4.39)

where \( \alpha \) is the desired slope of the filter. A detailed description can be found in [78]. The admittance transfer function of the desired filter can be expressed as

\[ \frac{Y(s)}{s} = \frac{(s + p_1)(s + p_2) \ldots (s + p_N)(s + p_{N+1})}{s(s + z_1)(s + z_2) \ldots (s + z_N)}. \]  

(4.40)

The approximation function \( Y(s)/s \) can now be partial fraction expanded to yield

\[ Y(s) = g \left( k'_0 + \sum_{i=1}^{n} \frac{s k'_i}{s + \sigma_i} \right), \]  

(4.41)

where \( k'_i \) are the zero locations, \( \sigma_i \) are the pole locations, \( g \) is the gain, and \( k'_0 \) is an amplitude offset. The original response of the fractional model is that of a resistor in parallel with a capacitor. The form of the circuit synthesized must provide the same zero frequency response. In this approximation method, the pole-zero networks are synthesized into resistors in series with capacitors to prevent zero frequency current flow in the circuit, save the final thermal resistance. This format provides an intuitive feel as the pole and zero networks simply are added on to the standard compact thermal network of Fig. 4.7.

The desired circuit can be synthesized according to [5] by

\[ R_i = \frac{1}{k'_i}, \quad C_i = \frac{k'_i}{\sigma_i}. \]  

(4.42)

where \( R_i \) and \( C_i \) are the respective resistor and capacitor values of each series RC network. According to the thermal resistance equation

\[ \Delta T = PR_{th}, \]  

(4.43)
the value of resistance in the model at zero frequency must be the physical thermal resistance to accurately model the temperature. This implies that the expansion must be multiplied by the gain factor

\[ g = \frac{1}{R_{\text{th}k_0}}. \] (4.44)

To guarantee that the filter magnitude continues to roll off, a capacitor generating a pole at the end of the approximated filter response must be added in parallel with the rest of the model. Due to constraints on the time constants of the circuit, this capacitor is no longer the thermal capacity of the original compact model of Fig. 4.7. It must be selected to not significantly add to the time constant of the filter. In this model, the thermal capacity was computed according to

\[ C_\infty = \frac{k_1 R_{\text{th}}}{z_N}, \] (4.45)

where \( k_1 \) is a constant that shifts the pole to the end of the approximation response and \( z_N \) is the highest frequency zero of the approximation function. The final synthesized form of the approximate electro-thermal model is shown in Fig. 4.10. The resistance of the circuit at zero frequency is that of the physical thermal resistance of the device. The physical thermal capacity of the device has been broken into \( N \) RC branches and a stability capacitor, approximating a fractional response over a limited bandwidth. Outside the approximation bandwidth the model returns to a 20 dB per decade slope representing the response of a first order derivative. The accuracy of this model is determined by both the bandwidth and allowed oscillation around the desired function. The order of the approximation circuit grows significantly with increases in bandwidth or accuracy.

4.3.3 Summary of Electro-Thermal Modeling

The temperature of a circuit element determines the bias point of its I-V characteristics in both passive and active devices. The bias point of device I-V characteristics changes due to both an average temperature rise from electrical power dissipation, and more subtly, with temperature oscillation due to low frequency electrical signal amplitude modulation. The average and transient nature of device temperature change due to electrical power dissipation is generally modeled by a simple RC circuit, which can not describe
the non-integer order Laplacian behavior of the coupled electro-thermal system. Accurate representation of this behavior requires coupling a circuit simulator to a 3D thermal solver, increasing required computational resources.

An electro-thermal compact model based on the fractional derivative was developed in this section that gives the benefits of accurate 3D methods of thermal simulation for electrical systems, including non-integer order Laplacian behavior such as long tail transients and fractional filter responses. Fractional derivative based solver methods, which require memory of the entire solution for prediction of further solution points, are not readily available in circuit simulators. Circuit simulators are suited for solution of a system of first order differential equations which generally require only a small subset of solution points for prediction of future solution points. A limited bandwidth circuit model for approximating the fractional electro-thermal compact model was presented for use in simulation tools possessing first order differential equation solver methods. Design optimization of components is well suited to simulation, but design insight is more easily obtained from analytic representations. An analytic closed-form representation of electro-thermal PIM for lossy elements, developed from the presented compact model, was presented allowing accurate prediction of distortion products and temperature transients using only material parameters. Several case studies on lossy microwave components including terminations, attenuators, and integrated circuit materials are presented in the following sections showing good agreement between the presented theory, simulation models, and measurement.
4.4 Case Study: Microwave Terminations

Microwave terminations are a commonly used lossy laboratory component that serves as a matched one port resistive load in test systems. High power signals can be dissipated in terminations through signal summation devices in which one port is terminated, as in a combiner, or in measurements employing a terminated coupler. The heat generated in the termination will cause the resistance to vary, resulting in PIM. In this section microwave terminations made of platinum, nichrome, ruthenium dioxide, and tantalum nitride are examined. They are characterized for thermal properties and electro-thermal interaction.

Accurate modeling of thermal induced distortion requires knowledge of the material thermo-resistance equation parameters, thermal capacity, and thermal resistance, while model and theory validation requires two-tone distortion characterization of the device. Three separate measurements are required to determine and validate all of the necessary model parameters. Device TCR can be found by characterizing resistance change over temperature. Thermal capacity and resistance are obtained through the application of a power step and subsequent measurement of the ensuing thermal transient. Electro-thermal distortion can be characterized through a tone spacing sweep, which provides model validation and inverse extraction of modeling parameters. Each of the measurements necessary for model identification and validation are detailed in the subsequent TCR, thermal transient, and two-tone IM characterization subsections.

4.4.1 Thermal Coefficient of Resistance Characterization

The thermo-resistance of a device relates the temperature change of a device to the resistance change in the device. This parameter dictates the coupling strength of the thermal and electrical domains in a device. It is frequency independent and does not include transient thermal effects, which are accounted for by the thermal resistance and capacity of the material. The magnitude of distortion generated from a resistive element is highly dependent on the magnitude of the TCR.

The thermo-resistance effect was characterized using a thermal chamber, a current source accurate to 1 nA, and a voltage meter accurate to 5 µV to measure the change in resistance over the operational temperature of the device under test. The measurement
configuration for a lossy element is shown in Fig. 4.11(a), where a HP4142B direct current source supplies a current to two long wires that are soldered to the test component. The test component and most of the length of the wire leads are inside the thermal chamber, while a voltage meter is attached in parallel with the current source. The chamber temperature is raised in 5°C increments from 0°C to 125°C with 30 minutes allowed at each temperature for transient stabilization. The voltage is measured down to 5 µV at each temperature and used to obtain resistance and change in resistance data accurate to 5 ppm. This procedure is repeated, shown in Fig. 4.11(b), for the test leads soldered together. This test is used to compensate the measured test component resistance for the test lead thermo-resistance. Measured thermo-resistance curves are shown in Fig. 4.12 for several terminations composed of nichrome (NiCr), tantalum nitride (TaN), and ruthenium dioxide (RuO₂) as well as a platinum element (Pt). The platinum curve is divided by 10 due to its magnitude in comparison to the other materials for display purposes. The slope of each curve represents the TCR for a given material, each order of which is the coefficient of the corresponding term in a polynomial fit to the curve. The α term used in the fractional compact model is always the first order coefficient.

The platinum element is thermally the most linear and has the highest linear TCR. Distortion levels for this device are also the highest per unit power of all devices examined. Several of the terminations exhibit thermo-resistance curves which are strongly non-linear over the operational temperature range, yet still possess temperature coefficients of resistance that are relatively small in magnitude. These elements produce distortion levels immeasurable in the system. The distortion is orders of magnitude lower than that of the linear thermo-resistance devices even though they exhibit much more non-linear thermo-resistances. Clearly the slope of the thermo-resistance over the operation range is the dominant effect on device distortion, corresponding directly to the strength of electrical and thermal coupling. For devices with strongly non-linear thermo-resistance, the temperature dependence on distortion coefficients must be considered for accurate modeling. With the strength of domain coupling obtained, transient thermal parameters must now be characterized to model the device. Thermal transient characterization to extract thermal resistance and capacity is discussed in the following section.
Figure 4.11: Measurement configuration composed of a voltage meter, HP4142B direct current source, and a thermal chamber for determining the temperature coefficient of resistance of a lossy component for: a) the device under test and b) de-embedding the test leads.

4.4.2 Thermal Transient Characterization

The electro-thermal process produces long-tail responses to transients that are directly related to the thermal resistance and capacity of a material. The distortion frequency response of a device can likewise be described by the same thermal parameters governing transient response. Frequency response characterization of these parameters requires dynamic range higher than commercially available test equipment can provide and is generally difficult to perform. Transient response characterization of thermal parameters is presented here for model parameter extraction with minimal required equipment.

Several methods can be used to determine thermal resistance and capacity, including the three omega method [89], thermal cameras, and current step tests [90]. As distortion will depend on the heat sinking condition of the entire system, a current step applied to the device in its environment of interest provides an accurate portrayal of device operational characteristics when exact construction is not known. By measuring voltage response and converting to power, the power step response can be fit as if it were a simple lowpass filter.
Figure 4.12: TCR curves of microwave chip terminations including nichrome, platinum, tantalum nitride, and ruthenium oxide showing linear behavior of most high TCR components. The coefficients of a polynomial curve fit represent each order TCR coefficient. The platinum curve is divided by ten for display purposes.

The test configuration to generate an accurate self-heating response requires an accurate, fast current pulse source and an oscilloscope to capture the resulting long-tail response. The test system used in this work is shown in Fig. 4.13, and is composed of an HP4142B pulsed current source operated in one shot mode and a triggered agilent oscilloscope. The probe impedance of the oscilloscope is 1 MΩ. The length of the trace captured is on the order of several seconds, increasing with current step amplitude, as the entirety of the self-heating event must be captured.

Thermal resistance is the difference in the final value of resistance for a given power step minus the ambient resistance, normalized to the power, TCR, and ambient resistance

\[
P(t) = P_0 \left(1 - e^{-t/R_{th}C_{th}}\right). \tag{4.46}
\]

response, given by [90]
of the device in the relationship

$$R_{th} = \frac{(R_1 - R_2)}{PR_1 \alpha}$$

(4.47)

Thermal capacity is given by fitting the transient response of the power step and employing the measured value of thermal resistance. Care must be taken in performing such a fit, as devices exhibiting significant electro-thermal distortion will have long thermal memory. Fitting the initial exponential response of the long tail effect and the end value of the transient, combined with the corresponding thermal resistance provides reasonably accurate results with minimal fitting effort. Thermal cameras can be used in the same manner, with a temperature step and curve fit of the response. The thermal parameters determined from this method combined with device TCR provide the necessary parameters for nonlinear device modeling. Two-tone characterization of the devices is performed in the following section and compared to simulation results.

### 4.4.3 Two-Tone PIM Characterization

Sweeping the spacing of a two-tone signal is effective in characterizing thermally induced distortion in RF systems, as it produces a sinusoidal thermal signal within the thermal response bandwidth. Low PIM components in the measurement system must be used to guarantee measurement integrity. Commercially available spectrum analyzers are limited to less than 80 dB dynamic range, and the vector signal analyzer used in this test was limited to approximately 75 dB. In order to circumvent this limitation at low tone spacing where filtering is not an option, the active cancellation system described in Chapter 3 was used to make accurate, high dynamic range measurements of approximately 106 dB. In the
cancellation based test system of Fig. 3.2, the DUT component is replaced by the directional coupler and test termination of Fig. 4.14 to allow measurement of a termination.

Components measured include 50 Ω chip terminations composed of tantalum nitride, platinum, and nichrome. Further measurements of typical 50 Ω lab terminations in N-type and SMA-type configurations were also conducted. Device TCR ranged from 150 ppm to 3850 ppm. Tantulum nitride, platinum, and coaxial lab terminations all produced measurable distortion in direct agreement with the fractional electro-thermal model. Power levels during testing were at one quarter or less of rated device power for N-type and SMA type terminations. Nichrome terminations possessed extremely low TCR, putting their distortion at levels below measurement capability. Using measured values for thermal resistance, capacity, and device TCR, measurement results for IM3 are compared to the electro-thermal analytic model in Fig. 4.15. Fig. 4.15 compares the analytic electro-thermal model with measured results of distortion generated from applying two sinusoidal signals of equal amplitude centered at 400 MHz to the various devices where the signal frequency spacing is swept from 1 Hz to 100 Hz separation. Measurement and model results coincide for all measurable materials for both cylindrical and planar semi-infinite geometries. Agreement between model and measurement demonstrates the necessity of a semi-derivative formulation for heat transfer to account for thermal time scaling in the electrical domain.

The intermodulation distortion magnitude in Fig. 4.15 is dependant on tone spacing and consequently signal bandwidth for a modulated signal. Fortunately, several other
<table>
<thead>
<tr>
<th>Frequency Offset (Hz)</th>
<th>Power (dbm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.1623</td>
</tr>
<tr>
<td>10</td>
<td>31.6228</td>
</tr>
<tr>
<td>100</td>
<td>316.2278</td>
</tr>
<tr>
<td>1000</td>
<td>1000</td>
</tr>
</tbody>
</table>

Figure 4.15: Analytic prediction and measured distortion for platinum, tantalum nitride, SMA type terminator, and N-type terminator for two tone spacing swept excitation at 20, 30, 27, and 25 dBm input power per tone, respectively.

Factors can reduce distortion levels to a bare minimum regardless of the signal applied. From the analytic model presented and matching measurement data, two factors become obvious for minimum electro-thermal PIM design when signal bandwidth cannot be controlled. First, and most important, is the device TCR, which must be reduced to a minimum over the operational range. Operational temperature must be kept away from any sharp sloped areas of the thermo-resistance curve. This can be accomplished through material manufacture methods, careful design of power dropped over resistive elements, or heat sinking to guarantee a maximum dynamic temperature rise over operational power. Thermal resistance and thermal capacity are material and size dependant parameters that control the bandwidth of the thermal filter. As the knee of the thermal filter increases in frequency, distortion levels due to electro-thermal distortion will increase toward their maximum level. Generally, increasing the size of the device will decrease distortion by increasing overall thermal capacity and reducing system thermal resistance. Choice of materials should focus...
on highest thermal conductivity for minimum distortion.

4.4.4 Summary of Electro-Thermal Distortion in Terminations

Microwave terminations are a commonly used laboratory component that generate electro-thermal distortion under high power conditions. The prediction of distortion from thermal coupling has been largely neglected due to microwave and thermal process time constant differences. In this section, microwave terminations were characterized for thermal parameters and electro-thermal distortion. Distortion data for microwave terminations has been unavailable in the literature due to measurement limitations imposed by the required high dynamic range and extremely narrowband nature of electro-thermal processes. The electro-thermal process was indeed shown to occur in the presented measurements of microwave terminations, matching the derived electro-thermal theory closely. The measurement techniques necessary to obtain the thermal parameters necessary for accurate modeling of electro-thermal distortion for a general lossy element were presented, enabling modeling of devices of known and unknown construction. These techniques are employed again in the following section to analyze another standard microwave component, an attenuator.

4.5 Case Study: Platinum Attenuator

Attenuators are common signal conditioning devices used in RF and microwave systems to attenuate signals a specified amount. They come in many configurations but are always composed of some type of resistive element which dissipates excess power through heat. Under amplitude modulation conditions, these resistive elements will experience periodic heating and cooling resulting in nonlinear resistivity and thus PIM generation. In this case study, platinum and nichrome resistors are configured into π attenuators, characterized for thermal parameters, measured for electro-thermal distortion, and modeling using the Foster expansion model of Section 4.3.2. The behavior of these devices is virtually the same as microwave terminations, further supporting electro-thermal distortion as a dominant PIM contributor in lossy passive devices.
4.5.1 Thermal Parameter Characterization and Foster Model

An attenuator is composed of a number of resistive elements with each one of these elements generating its own electro-thermal distortion. The thermal parameters of each element experiencing significant power must be found for accurate nonlinear modeling. Thermal parameters are obtained using the thermal transient characterization method presented in Section 4.4.2. These parameters are used to generate an approximate fractional compact model suited for standard circuit simulation.

Formulating an electro-thermal circuit model for an attenuator requires the device thermal parameters in the assembled configuration. The foster model approximation could be fit to the measured electro-thermal frequency response, but without the thermal parameters will not be physically consistent. The physical thermal parameters can be obtained through the method described in Section 4.4.2. A current step is applied to the attenuator in Fig. 4.16, resulting in a self heating transient that contains the thermal capacity and thermal resistance as prescribed by (4.46). The thermal resistance is determined through (4.47) and is found to be 74 Ω. The thermal capacity is determined by the current step fit to be .022 F.

The thermal dispersion characteristic can be predicted using thermal parameters and (4.33). The foster fit of this response following the method outlined in Section 4.3.2 results in the circuit of Fig. 4.17, implemented in Advanced Design System 2006A (ADS). The transient response of the foster model is self consistent with final values of heating as predicted in theory and is approximately accurate for transients, as shown in Fig. 4.16. The accuracy of the transient is dependent on the order of the approximation, which was eighth order in this study. The frequency response of the distortion and the approximate model is discussed in the following section.

4.5.2 Platinum Electro-Thermal Dispersion

A resistive attenuator π network, shown in Fig. 4.18, was built and used to evaluate attenuator electro-thermal PIM in a RF test system. The attenuator is composed of two 100 Ω platinum resistors in parallel on the input at port one, a 1 kΩ platinum bridge resistor, and two 100 Ω platinum resistors on the output at port two. Due to equipment limitations, only spacing frequencies of 1 Hz to 100 Hz could be measured. For each power
level, the two tone spacing was swept logarithmically, ten points per decade. The through signal was measured from the 32.5 dB attenuator, with PIM up to 58 dBc down from the carrier signal level at 23 dBm input power. PIM magnitude was observed to decrease with an approximate 10 dB per decade slope as the carrier frequency separation was swept in the measurement range, the defining characteristic of the electro-thermal process. Simulation results produced a very close replication of the observed PIM magnitude over the measurement range with respect to tone spacing, as shown in Fig. 4.19, with no significant deviations. The electro-thermally induced spectrum is shown in Fig. 4.20, clearly demonstrating third order intermodulation products.

Two different types of attenuators were built for comparison. A platinum resistance temperature detector based attenuator was built because of its high, almost linear temperature coefficient of resistance (TCR), 3850 ppm/°C, and complete lack of any ferromagnetic materials. A nickel chromium based attenuator was also built because of its extremely low TCR of 5 ppm/°C, which averages over the operational temperature range.
Figure 4.17: ADS implementation of an 8\textsuperscript{th} order foster fit electro-thermal model for the platinum resistors used in the $\pi$ attenuator.
Figure 4.18: A 32.5 dB attenuator composed of two 100 Ω platinum resistors in parallel on the input at port one, a 1 kΩ platinum bridge resistor, and two 100 Ω platinum resistors on the output at port two.

to almost zero parts per million. The components of this attenuator were exactly the same as the components from the platinum attenuator, except composed of nichrome. PIM from the platinum attenuator was measurable, but due to its low TCR, no PIM products in the measurement range were observable for the nickel chromium based attenuator.

The magnitude of electro-thermally generated PIM from an element is strongly linked to the TCR of that element due to thermal baseband resistive mixing with the input signal. Lowering TCR effectively minimizes the thermal based resistance variation and thus the PIM products. It follows that a possible method for reducing electrothermal PIM is to use low TCR components in circuit design.

The intermodulation distortion magnitude is also clearly dependant on tone spacing, as seen in Fig. 4.19. The closer the tones are spaced the longer the time period will be for heating and cooling of the material, which leads to wider resistivity swings. If the input signal was devised such that the power variation in the signal was much faster than the thermal filter bandwidth, the power would average out and no effective periodic heating and cooling of the material would occur. This type of signal would have the same effect as a widely spaced two-tone signal and would minimize thermally induced distortion.

4.5.3 Summary of Electro-Thermal Distortion in Attenuators

Attenuators are a commonly used laboratory component that generate electro-thermal distortion under high power conditions. The prediction of distortion from thermal coupling has been largely neglected due to differences of microwave and thermal time constants. In this section, attenuators were characterized for thermal parameters and electro-
Figure 4.19: Foster fit approximate model prediction of IM3 versus tone separation overlaid on measured IM3 for platinum at 12, 20, and 23 dBm input power.

Figure 4.20: Platinum attenuator frequency spectrum at 400 MHz center frequency and tone spacing of 5 Hz.
thermal distortion. Distortion data for microwave attenuators has been unavailable in the literature due to measurement limitations imposed by the required high dynamic range and extremely narrowband nature of electro-thermal processes. The electro-thermal process was indeed shown to occur in the presented measurements of microwave attenuators, matching the derived electro-thermal theory closely. The use of devices that were enhanced for temperature variation and the large increase in distortion magnitude at similar power levels over microwave terminations strongly supports electro-thermal distortion as the physical mechanism responsible for distortion in both microwave terminations and attenuators. Terminations and attenuators, while providing support for electro-thermal theory, did not allow exact control over device construction and thus thermal configuration. In the following section, devices with tightly controlled thermal configuration are built and measured on an integrated circuit.

4.6 Case Study: Integrated Circuit Distortion

In the studies of microwave terminations and attenuators, the electro-thermal models predicted a passband at very low frequencies in the thermal dispersion. This passband is not predicted by conventional analysis of the homogeneous heat conduction equation. The predicted response was not observed due to the high thermal capacity and thermal resistance of the materials measured. In this section, resistors are developed with tightly controlled thermal configurations by specifying device dimensions and materials.

The thermal configuration of resistive material can be controlled by material choice and element dimensions. Resistor material choice determines the heat density generated for a given resistance, while the surrounding materials determine the thermal resistance and capacity of the structure. In this section, five resistors are designed with varying thermal properties to control the thermal passband. The resistive materials are used to control the dimensions of the device. The surrounding materials are used to control the thermal resistance and capacity of the structure, channeling the heat produced in the resistive material in prescribed directions.

The resistors are manufactured in the IBM CMRF7SF process, which is a 7 metal layer RF integrated circuit (IC) process. Dielectric layer thicknesses and process materials
are well defined for parasitic modeling. Six different types of resistive materials are available in the process. Three of these options for resistive material include either p+/n+ doping or compensation. Doped material is not ideal for isolating the electro-thermal effect as the process of doping could introduce separate unrelated nonlinearities. Resistive polysilicon, silicided gate polysilicon, and tantalum nitride were chosen as the resistive materials due to their lack of doping.

The measurement configuration used to characterize the IC’s for electro-thermal distortion is presented in Section 4.6.1. Resistors were developed to have a thermal passband that is either wideband or narrowband. The respective materials and design of each resistor is discussed categorized by its thermal passband. Wideband devices are discussed in Section 4.6.2 while narrowband devices are discussed in Section 4.6.3.

4.6.1 Electro-Thermal Dispersion Measurement

Electro-thermal dispersion is defined by its fractional derivative slope. Thermal theory predicts that as the frequency of thermal modulation is decreased, the temperature oscillation will approach infinity. Physically, the temperature oscillation must approach a finite amplitude as the oscillation frequency goes to zero. This condition manifests as the passband of the lowpass thermal filter. An amplitude modulation sweep, implemented as a two-tone separation sweep, will produce the thermal passband characteristic in the third order distortion.

Testing of on chip resistors requires either a probe station or wirebonding the chip. The linearity of probes is unknown and not easy to obtain. Probes present added difficulty in the material damage they cause the test device and the metal-metal junction or possibly metal-oxide-metal junction they create. As several new physical mechanisms for distortion generation may stem from the use of probes, wirebonding was selected for measurement of on chip devices. A printed circuit board (PCB) was designed to interface the chip resistors as an attenuator. The IC resistors were wirebonded with aluminum using the Westbond 7476E directly to the silver immersion coating of the PCB. The wirebonded devices were connected to act as a 50 Ω resistor at port one of the attenuator of Fig. 4.21. This measurement setup was then inserted into the DUT position in the measurement system of Fig. 3.2, allowing direct comparison with previously measured attenuators. A two-tone separation sweep,
Figure 4.21: A 32.5 dB attenuator composed of a wirebonded 50 Ω integrated circuit resistor at port one, a 1 kΩ 5 ppm TCR bridge resistor, and one 50 Ω high power microwave termination on the output at port two.

centered at 400 MHz, was conducted from 4 Hz to 10 MHz at 24 dBm input power. The thermal dispersion characteristics are presented with each resistor design in the following sections.

4.6.2 Wide Thermal Bandwidth Devices

Theoretically the thermal passband of a resistive structure can be designed based on material properties and dimensions. Three resistors were designed using high resistivity polysilicon, low resistivity polysilicon, and tantalum nitride with attached heat sinks to produce a wide thermal bandwidth. Each of these devices have different thermo-resistance coefficients (TCR), providing further clarification of the effects of TCR on the electro-thermal process. The design and construction of each device is discussed in the context of the resulting thermal filter response. Thermal bandwidths in excess of 1 MHz, highly dependent on heat sinking, with appreciable distortion extending beyond 10 MHz, is shown in each case.

High Resistivity Polysilicon

The resistor of Fig. 4.22(a) was designed and manufactured in high resistance 1600 Ω per square polysilicon material using IBM standard cells to generate a 50 Ω resistor. It is composed of a single wide polysilicon resistive element to maximize heat transfer to the heat sinking metallization. The resistive element is interfaced with aluminum metallization over four times the size of the element on the input and output, creating a high thermal conductivity path for heat transfer. The aluminum heat sinking metallization is connected
Table 4.1: Process Metallization Thicknesses

<table>
<thead>
<tr>
<th>Metal</th>
<th>Thickness</th>
<th>Dielectric Thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>—</td>
<td>5.36 µm</td>
</tr>
<tr>
<td>P</td>
<td>Cu</td>
<td>0.20 µm</td>
</tr>
<tr>
<td>M1</td>
<td>Al</td>
<td>0.41 µm</td>
</tr>
<tr>
<td>M2</td>
<td>Al</td>
<td>0.48 µm</td>
</tr>
<tr>
<td>M3</td>
<td>Al</td>
<td>0.48 µm</td>
</tr>
<tr>
<td>M4</td>
<td>Al</td>
<td>0.48 µm</td>
</tr>
<tr>
<td>M5</td>
<td>Al</td>
<td>0.48 µm</td>
</tr>
<tr>
<td>M6</td>
<td>Al</td>
<td>0.48 µm</td>
</tr>
<tr>
<td>M7</td>
<td>Al</td>
<td>2.0 µm</td>
</tr>
</tbody>
</table>

directly to bondpads for external interface. The material configuration of this resistor is shown in Fig. 4.22(b) and the dimensional thicknesses are defined in Table 5.2. The dielectric thicknesses in Table 5.2 represent the thickness to the top metal layer from the current material layer. The polysilicon resistive element is isolated from the substrate by a thermal insulator, silicon dioxide. Aluminum metallization, surrounded by silicon dioxide, contacts the polysilicon on both sides providing a heat conduction path that has over 200 times lower thermal impedance than the heat conduction path into the silicon substrate. The heat flows up through the via metallization into the large metal heat sink on the top layer, as does the electrical signal.

The simulation of this device was conducted using the model of Section 4.3.1. Thermal capacity and resistance were that of the aluminum metallization in the process, calculated to be unitless for the actual size of the structure. The material properties and simulation parameters for this resistor are defined in Table 4.2. The measured distortion characteristics of the resistor from a two-tone separation sweep from 4 Hz to 10 MHz, shown in Fig. 4.23, matched simulated results well using only device dimensions and material parameters. It was originally thought that the resistivity of the material would significantly effect the distortion generated due to the relative size differences for the same resistance. However, the thermal bandwidth in this experiment, which shows a passband of approximately 1 MHz, is only dependent on the length of metallization directed outward from the
Figure 4.22: High resistivity (1600 Ω per square) 50 Ω polysilicon resistor: (a) A 800 µm by 28.04 µm resistive element is bounded by metallization of dimension 800 µm by 115 µm on each side. The TCR is $-1360 \text{ ppm/°C}$ for the bulk and $-790 \text{ ppm/°C}$ for the end, (b) and a cross section of the resistor denoting heat flow from the polysilicon resistive material through the via metallization stack and into the aluminum metallization heat reservoir. The via and metallization stack is 6 layers high, expanding in width with each layer.
resistive material. The TCR determines the maximum of the distortion generated and was highly dependent on not only the bulk TCR, TCR\(_b\), but also the TCR of the end of the material, TCR\(_e\).

**Low Resistivity Polysilicon**

The resistor of Fig. 4.28(a) was designed and manufactured in low resistance 165 \(\Omega\) per square polysilicon material using IBM standard cells to generate a 50\(\Omega\) resistor. It is composed of a single small square of polysilicon attached to a aluminum heat sink metallization of approximately 2.5 times the resistive element size, creating a high thermal conductivity path for heat transfer. Input and output connections were directly to bondpads for external interface. The material configuration of this resistor is shown in Fig. 4.24(b) and the dimensional thicknesses are defined in Table 5.2. The polysilicon resistive element is isolated from the substrate by a thermal insulator, silicon dioxide. Similar to the previous resistor, aluminum metallization surrounded by silicon dioxide contacts the polysilicon
on both sides, providing a heat conduction path that has over 200 times lower thermal impedance than the heat conduction path into the silicon substrate. The heat flows up through the via metallization into the large metal heat sink on the top layer, as does the electrical signal.

The simulation of this device was conducted using the model of Section 4.3.1. Thermal capacity and resistance were that of the aluminum metallization in the process, calculated to be unitless for the actual size of the structure. The material properties and simulation parameters for this resistor are defined in Table 4.3. The measured distortion characteristics of the resistor from a two-tone separation sweep from 4 Hz to 10 MHz, shown in Fig. 4.25, matched simulated results well using only device dimensions and material parameters. It was originally thought that the resistivity of the material would significantly affect the distortion generated due to the relative size differences for the same resistance. However, the thermal bandwidth in this experiment, which shows a passband of approximately 1 MHz, was once again only dependent on the length of metallization directed outward from the resistive material. The TCR determines the maximum of the distortion generated and was highly dependent on not only the bulk TCR, TCR\textsubscript{b}, but also the TCR of the end of the material, TCR\textsubscript{e}. Notably, the distortion magnitude was virtually the same for this resistor and the previous one even though the bulk TCR was an order of magnitude larger for the previous resistor. The end TCR of a material, which is seen at contacts, must dominantly contribute to the generated distortion for these magnitudes to match in the thermal passband of the two resistors.

<table>
<thead>
<tr>
<th></th>
<th>Al</th>
<th>Polysilicon</th>
<th>SiO\textsubscript{2}</th>
<th>Si</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$ (W·m\textsuperscript{-1}K\textsuperscript{-1})</td>
<td>250</td>
<td>125</td>
<td>1.3</td>
<td>150</td>
<td>—</td>
</tr>
<tr>
<td>$c_v$ (J·Kg\textsuperscript{-1}K\textsuperscript{-1})</td>
<td>890</td>
<td>753</td>
<td>670</td>
<td>700</td>
<td>—</td>
</tr>
<tr>
<td>$\rho$ (Kg·m\textsuperscript{-3})</td>
<td>2700</td>
<td>2230</td>
<td>2330</td>
<td>2330</td>
<td>—</td>
</tr>
<tr>
<td>$R_{th}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>17.39 Ω</td>
</tr>
<tr>
<td>$C_{th}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>178.9 nF</td>
</tr>
<tr>
<td>TCR\textsubscript{b}</td>
<td>0.0039 K\textsuperscript{-1}</td>
<td>-0.001360 K\textsuperscript{-1}</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TCR\textsubscript{e}</td>
<td>0.0039 K\textsuperscript{-1}</td>
<td>-0.000790 K\textsuperscript{-1}</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
Figure 4.24: Low resistivity (165 Ω per square) 50 Ω polysilicon resistor: (a) A 47 µm by 150 µm resistive element is bounded by a 115 µm by 150 µm bondpad on each side. The resistivity of the element is 165 Ω per square and the TCR is 210 ppm/°C for the bulk and −1960 ppm/°C for the end, (b) and a cross section of the resistor denoting heat flow from the polysilicon resistive material through the via metallization stack and into the aluminum metallization heat reservoir. The via and metallization stack is 6 layers high, expanding in width with each layer.
Figure 4.25: Fractional compact model prediction of IM3 versus tone separation overlaid on measured IM3 for the small, medium resistivity single element polysilicon resistor of Fig. 4.24(a) at 24 dBm input power. The measurement and simulation data is not corrected for attenuation, which was 30 dB after the attenuator.

**Tantalum Nitride**

The resistor of Fig. 4.26(a) was designed and manufactured in low resistance 61 Ω per square tantalum nitride material using IBM standard cells to generate a 50Ω resistor. It is composed of a five resistive elements in series connected by aluminum, in parallel with five duplicate branches, to create a device of similar size to the small polysilicon resistor of Fig. 4.24(a). The resistive structure is connected to aluminum heat sink metallization of approximately half the outward length of the previous structures, creating a heat conduction path that is approximately twice the thermal impedance of the two previous resistors. Input and output connections were directly to bondpads for external interface. The material configuration of this resistor is shown in Fig. 4.26(b) and the dimensional thicknesses are defined in Table 5.2. The tantalum nitride resistive element is isolated from the substrate by a thermal insulator, silicon dioxide, which is also over twice as thick as in the previous cases. Similar to the previous resistors, aluminum metallization surrounded by silicon dioxide contacts the resistive element on both sides, providing a heat conduction path that has over
Table 4.3: Simulation Model Parameters

<table>
<thead>
<tr>
<th></th>
<th>Al</th>
<th>Polysilicon</th>
<th>SiO₂</th>
<th>Si</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k\ (W \cdot m^{-1} K^{-1}))</td>
<td>250</td>
<td>125</td>
<td>1.3</td>
<td>150</td>
<td>—</td>
</tr>
<tr>
<td>(c_v\ (J \cdot Kg^{-1} K^{-1}))</td>
<td>890</td>
<td>753</td>
<td>670</td>
<td>700</td>
<td>—</td>
</tr>
<tr>
<td>(\rho\ (Kg \cdot m^{-3}))</td>
<td>2700</td>
<td>2230</td>
<td>2330</td>
<td>2330</td>
<td>—</td>
</tr>
<tr>
<td>(R_{th})</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>13.33 Ω</td>
</tr>
<tr>
<td>(C_{th})</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>102.1 nF</td>
</tr>
<tr>
<td>(TCR_b)</td>
<td>0.0039 K^{-1}</td>
<td>-0.001960 K^{-1}</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(TCR_e)</td>
<td>0.0039 K^{-1}</td>
<td>0.000210 K^{-1}</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

200 times lower thermal impedance than the heat conduction path into the silicon substrate. The heat flows up through the via metallization into the large metal heat sink on the top layer, as does the electrical signal.

The simulation of this device was conducted using the model of Section 4.3.1. Thermal capacity and resistance were that of the aluminum metallization in the process, calculated to be unitless for the actual size of the structure. The material properties and simulation parameters for this resistor are defined in Table 4.4. Tantalum nitride thermal parameters were not available, and are also not necessary. The measured distortion characteristics of the resistor from a two-tone separation sweep from 4 Hz to 10 MHz, shown in Fig. 4.25, matched simulated results well using only device dimensions and material parameters. The thermal bandwidth in this experiment, which shows a passband of approximately 1 MHz, was once again only dependent on the length of metallization directed outward from the resistive material. The combined end and bulk TCR of this device is nearly half of the TCR of the previous devices, resulting in a 3 dB drop in distortion magnitude. The passband of this device is more narrow than the two previous resistors due to the decrease in heat sink metallization outward length, which effectively increased the device thermal resistance.

### 4.6.3 Narrow Thermal Bandwidth Devices

The thermal passband of a resistive structure was shown to be based on material properties and dimensions in the previous section. The design of a thermal filter with no
Figure 4.26: Low resistivity (61 Ω per square) 50 Ω tantalum nitride resistor: (a) A 150 µm by 172 µm resistive element composed of 5 series elements in parallel with 5 duplicate branches is bounded by metallization of dimension 68 µm by 172 µm on each side. The TCR is $-387$ ppm/°C for the bulk and 600 ppm/°C for the end, (b) and a cross section of the resistor denoting heat flow through the tantalum nitride resistive material and the via metallization stack into the aluminum metallization heat reservoir. The via and metallization stack is 5 layers high, expanding in width with each layer.
passband is theoretically possible, just as a wide passband thermal filter. Two resistors were designed using high resistivity polysilicon and silicided gate polysilicon with minimal heat sinking to produce a narrow thermal bandwidth. Each of the devices have different thermo-resistance coefficients (TCR), providing further clarification of the effects of TCR on the electro-thermal process. The design and construction of each device is discussed in the context of the resulting thermal filter response. Thermal passbands are completely eliminated in each resistor. Thermal heat flow is shown to be forced by layout to travel through low thermal conductivity routes rather than heat sinking.

High Resistivity Polysilicon

The resistor of Fig. 4.28(a) was designed and manufactured in high resistance 1600 Ω per square polysilicon material using IBM standard cells to generate a 50 Ω resistor. It is composed of six resistive elements of high resistance polysilicon in parallel, allowing a much more compact layout than the resistor of Fig. 4.22(a) with the same resistance. Heat
Table 4.4: Simulation Model Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Al</th>
<th>TaN</th>
<th>SiO₂</th>
<th>Si</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$ (W·m⁻¹K⁻¹)</td>
<td>250</td>
<td>—</td>
<td>1.3</td>
<td>150</td>
<td>—</td>
</tr>
<tr>
<td>$c_v$ (J·Kg⁻¹K⁻¹)</td>
<td>890</td>
<td>—</td>
<td>670</td>
<td>700</td>
<td>—</td>
</tr>
<tr>
<td>$\rho$ (Kg·m⁻³)</td>
<td>2700</td>
<td>2230</td>
<td>—</td>
<td>2330</td>
<td>—</td>
</tr>
<tr>
<td>$R_{th}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$C_{th}$</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TCR₀</td>
<td>0.0039 K⁻¹</td>
<td>-0.000387 K⁻¹</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>TCRₑ</td>
<td>0.0039 K⁻¹</td>
<td>-0.000600 K⁻¹</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

transfer to the heat sinking metallization is minimized through the interdigitated layout. The thin interconnect between resistive elements does not have the thermal capacity to sink the heat generated in the resistive elements. The temperature of the interconnect quickly rises to the temperature of the resistive elements. The heat sinks at the ends of the device have a very small contact area with the interconnect metallization, which raises the thermal resistance of this path above that of the silicon dioxide and silicon substrate. Although the resistive element is interfaced with aluminum metallization approximately the size of the element on the input and output, the aluminum metallization has no impact on the thermal transfer.

The material configuration of this resistor is shown in Fig. 4.28(b) and the dimensional thicknesses are defined in Table 5.2. The polysilicon resistive element is isolated from the substrate by a thermal insulator, silicon dioxide. Aluminum metallization, surrounded by silicon dioxide, contacts the polysilicon on both sides. The contact to the heat sinking metallization is only a few microns wide rather than the entire length of the resistive element in this case. The thermal resistance through this path is greatly increased and prompts the majority of the heat to flow through the silicon dioxide into the substrate.

The simulation of this device was conducted using the model of Section 4.3.1. Thermal capacity and resistance were that of the silicon dioxide and silicon substrate in the process, calculated to be unitless for the actual size of the structure. The material properties and simulation parameters for this resistor are defined in Table 4.5. The measured distortion characteristics of the resistor from a two-tone separation sweep from 4 Hz
Figure 4.28: High resistivity (1600 Ω per square) 50 Ω polysilicon resistor: (a) A 125 µm by 27 µm resistive element is in parallel with six duplicate elements. The metallization from signal to ground is shared between two devices and is 1 µm wide at the resistive material and 6 µm wide on the top layer. The layout assures no high thermal conductivity route for heat to the bondpads. The TCR is $-1360$ ppm/$^\circ$C for the bulk and $-790$ ppm/$^\circ$C for the end, (b) and a cross section of the resistor denoting heat flow from the polysilicon resistive material into the substrate due to the small contact area with the heat sinking reservoir.
Figure 4.29: Fractional compact model prediction of IM3 versus tone separation overlaid on measured IM3 for the small, high resistivity multi-element polysilicon resistor of Fig. 4.28(a) at 24 dBm input power. The measurement and simulation data is not corrected for attenuation, which was 30 dB after the attenuator.

to 1 KHz, shown in Fig. 4.25, matched simulated results reasonably well using only device dimensions and material parameters. The device slope was thought to be impacted by system generated distortion near 1 khz, as this is the limit of the dynamic range of the measurement system. The thermal bandwidth in this experiment is unknown as the device shows no passband, as predicted in theory from material parameters and heat flow directions. The heat flow is significantly impacted by decreasing the contact area with the heat sink, greatly decreasing the electro-thermal bandwidth and generated distortion. Device reliability was not considered in this experiment.

Low Resistivity Silicided Gate Polysilicon

The resistor of Fig. 4.30(a) was custom designed and manufactured in low resistance 8 Ω per square silicided gate polysilicon material to generate a 50 Ω resistor. It is composed of two resistive elements in series connected by metallization on the end one
The simulation of this device was conducted using the model of Section 4.3.1. Thermal capacity and resistance were that of the silicon dioxide and silicon substrate for one filter and the aluminum metallization for the other filter, calculated to be unitless for the actual size of the structure. The material properties and simulation parameters for this resistor are defined in Table 4.6. The measured distortion characteristics of the resistor from a two-tone separation sweep from 4 Hz to 5 MHz, shown in Fig. 4.31, matched simulated results reasonably well using only device dimensions and material parameters. The thermal bandwidth is approximately 100 Hz in this experiment as predicted in theory from material parameters and heat flow directions. Notably the slope of the filter response
Figure 4.30: Low resistivity (8 Ω per square) 50 Ω silicided polysilicon resistor: (a) A 471 μm by 150 μm resistive element is connected to a second duplicate element in series by metallization of dimension 310 μm by 32 μm. Each end is connected to bondpad by metallization of dimension 150 μm by 16 μm. The TCR is 210 ppm/°C for the bulk and −1960 ppm/°C for the end, (b) and a cross section of the resistor denoting heat flow from the silicided polysilicon resistive material through the via metallization stack and into the aluminum metallization heat reservoir. The heat reservoir is not large enough to dissipate the heat so the heat must travel through the silicon dioxide insulator and into the substrate. The via and metallization stack is 6 layers high, expanding in width with each layer.
Figure 4.31: Fractional compact model prediction of IM3 versus tone separation overlaid on measured IM3 for the large, low resistivity single element gate polysilicon resistor of Fig. 4.30(a) at 24 dBm input power. The measurement and simulation data is not corrected for attenuation, which was 30 dB after the attenuator.

is only 5 dB per decade at low frequencies before it transitions to a 10 dB per decade response. This effect coincides directly with two separate electro-thermal filters in series as it concerns the heat flow, demonstrating that heat flow and direction can be directly controlled through metallization dimensions and contact area. The wide contact to the polysilicon provides a high thermal conductivity route to the heat sink, which dissipates the heat into the low thermal conductivity silicon dioxide as it cannot transmit heat back into the high temperature resistive material. The heat flow is significantly impacted by reducing the metallization size, greatly decreasing the electro-thermal bandwidth but not significantly effecting the maximum generated distortion.

4.6.4 Summary of Electro-Thermal Distortion in IC’s

Electro-thermal distortion was shown to exist in microwave terminations and attenuators, but did not exhibit the thermal passband predicted in the presented electro-thermal theory due to large thermal capacity in each measured device. Integrated circuit
resistors were manufactured in the IBM CMRF7SF process, allowing direct control of device thermal bandwidth. Dimensions were controlled through the use of high, medium, and low resistivity materials, specifying the volume of heat generation and area of heat transfer. Heat sinking metallization interfaced the area defined by layout and device dimension, effectively defining the thermal resistance and capacity of each device.

Each resistor was characterized for electro-thermal distortion through two-tone separation sweeps. The dispersion characteristics of these resistors demonstrated that device size and metallization can significantly alter distortion generation for the same resistivity material both in magnitude and in thermal bandwidth. Thermal dispersion characteristics confirmed the direction of heat transfer can be controlled through the contact area and length of metallization to a heat generation zone. Resistors possessing small metallization contact areas transferred heat preferentially to the substrate through low thermal conductivity silicon dioxide. This effect occurred even when large metallization was nearby, implying that the thermal resistance a device sees is controlled by material contact area and dimensions. In one instance, heat was confirmed to travel up through metallization due to a high thermal conductivity path, but had to travel through low conductivity silicon dioxide to get to the substrate due to the inadequate thermal capacity of the metallization.

Thermal passbands were generated in heat sunk devices, resulting in large, almost constant distortion for signal bandwidths up to approximately 10 MHz. This result
shows that electro-thermal distortion is not limited to very large peak to average ratio signals. Thermal heat sinking, while increasing power handling, can greatly increase distortion generation in magnitude and thermal bandwidth. The results presented here have implications on active device nonlinearity as well, as distortion generated on the gate polysilicon will be amplified by the active device.

### 4.7 Conclusion

In this chapter, electro-thermal conductivity modulation was shown to be a dominant PIM mechanism. The impact of electro-thermal conductivity modulation was demonstrated by the close agreement between experimentally observed distortion and the presented theory of distortion based on electro-thermal effects. Generally the large time scale differences between thermal and electrical processes was thought to reduce thermal contributions to microwave distortion. Baseband components of the power from electrical signals were shown to interact with the thermal system with appreciable coupling, which coupled back to the electrical system through conductivity modulation at the baseband rate. Analysis of heat conduction theory showed that due to the multi-physics nature of the problem, specifically the coupling between electrical and thermal domains, non-integer order Laplacian behavior results that generates long-tail transient behavior in electrical signals that cannot be described with exponential functions. The non-integer order Laplacian behavior is a memory process which can be represented through the use a fractional differential equation. Lumped element distortion, in general, is shown to follow the fractional differential equation formulation for heat conduction in both frequency and transient behavior.

Knowledge of the thermal process combined with the fractional differential equation formulation allowed the construction of a reduced-order circuit model based on a fractional derivative, giving near the performance of 3D thermal simulation. This model accurately reproduces both electro-thermal distortion and linear electro-thermal transients in circuit simulators. Circuit analysis of the model led to a closed form representation of electro-thermal distortion employing only material parameters of the device in question. Unfortunately, most circuit simulators, which employ solvers limited to a system of first order differential equations, do not possess the capability at this point for fractional derivative
solver methods, which requires use of all previous solution points for future solution point prediction. A circuit model that approximates the fractional model response over a limited bandwidth was presented based on foster expansion, allowing currently available simulation tools to use the developed theory with no modification.

The generated theory was compared against microwave terminations, attenuators, and integrated circuit resistors. PIM was characterized using two-tone sweeps in a high dynamic range measurement system for each of these elements, providing experimental results previously unavailable due to dynamic range and tunability limitations. These measurements contained a unique non-integer order Laplacian dispersion trend characteristic of electro-thermal distortion. The dispersion trend matched extremely well with the presented theory for all devices tested. Thermal measurement techniques were presented to provide all necessary parameters for electro-thermal modeling, including devices of unknown thermal configuration. Methods for minimizing distortion, controlling heat flow direction, and controlling thermal parameters and bandwidths in integrated circuits were presented using the analytic formulation of electro-thermal PIM as a design guide.
Distributed Passive

Intermodulation Distortion
5.1 Introduction

Passive intermodulation distortion (PIM) is of concern in any communications system, as the spurious content generated can fall in the receive or transmit bands and detrimentally effect the dynamic range of that system. PIM has been observed to be produced by many passive components including ferrite circulators [25], waveguides [91], cable connectors [92,93], duplexers [94], attenuators [74], terminations [65], ferromagnetic metals [95], and antennas [3,5,8,96]. Many physical mechanisms have been suggested including ferromagnetism [93,95], tunneling [16,91,97], constriction resistances [93], and nonlinear conductivity [98], but due to the difficulty in isolation of PIM mechanisms on transmission lines, no physical mechanism has been identified as the dominant effect.

The distributed nature of PIM on transmission lines has been suggested in [99], where PIM fields were postulated to grow with the length of the line to a maximum before decaying with increasing distance due to losses. Near field probing has recently shown growth of forward wave PIM to exist on transmission lines [69,70]. Point sources of PIM such as solder droplets, scratches, and debris still exist but decay rather than grow with line length [69]. The physical phenomena responsible for the intrinsic performance of transmission lines must be distributed, requiring its existence down the entirety of the line.

Ferromagnetic materials, conductor surface roughness, tunneling, and nonlinear conductivity could exist at all points on a transmission line. Recent literature has tended to model PIM generation as a current related nonlinearity [68,93,98]. In [93], SMA connectors were evaluated for PIM performance by creating a standing current wave over them. The increased current density resulted in an 18 dB increase in distortion output as compared to the matched case. Following the same principle, connectors were evaluated in [68] in a resonator cavity. Maximums in the generated distortion products were found at the current magnitude peaks corresponding to voltage magnitude minimums. Minimums in the generated distortion products were found at the voltage magnitude peaks corresponding to current magnitude minimums. Tunneling is a voltage induced nonlinearity, greatly decreasing its likelihood as a dominant contributor [100]. The dependency on current density seems to exist with and without ferromagnetic materials [93]. Nonlinear conductivity can exist in tandem with ferromagnetic material, supporting it as an intrinsic mechanism for PIM generation. In [65], the authors showed that the temperature dependence of conductivity
can produce appreciable distortion in lossy lumped microwave elements. Electro-thermal conductivity modulation will exist in every conductor down the entire length of a distributed structure and cannot be removed or isolated. PIM from other mechanisms such as ferromagnetic materials and surface roughness can be removed through manufacturing techniques but in general will exist on a structure in combination with electro-thermal distortion. Electro-thermal distortion must be analyzed to de-embed any other PIM generating mechanism, and thus represents the baseline of physical performance a transmission line can achieve under optimum manufacturing conditions.

In this chapter, electro-thermal conductivity modulation as a physical mechanism for PIM generation on distributed structures is presented. The heat conduction system of a transmission line is explored in Section 5.2.1, yielding the temperature in the conductive layer of the metal. The PIM generated by an infinitesimal element of the conductor due to heating is discussed in Section 5.2.2. PIM coupling, growth, and loss along the transmission line is presented for forward and reverse propagation modes in Section 5.2.3. Isolation of the electro-thermal mechanism from other physical mechanisms through manufacturing is discussed in Section 5.3.1. Section 5.3.2 presents the design of electro-thermally isolated transmission lines while Section 5.3.3 shows their interface to an RF test system. The electro-thermal dispersion measurement of two fully assembled transmission line samples, silver on sapphire and silver on quartz, is discussed in Section 5.4.1. The effect of thermal material parameters and current bunching on thermal dispersion and distortion generation is analyzed in Section 5.4.2. Line dimensional dependencies are discussed in Section 5.4.3, resulting in design guidelines for low PIM transmission lines.

5.2 Distributed Electro-Thermal Theory

Distributed structures have many different possible physical mechanisms responsible for PIM, including ferromagnetic materials, surface roughness, tunneling, dielectric loss, dissimilar metals, and nonlinear conductivity. Electro-thermal distortion generally has not been considered as a dominant PIM mechanism in distributed structures due to the low power dissipation over the transmission line as well as the difference between thermal time constants and the period of high frequency electrical signals. A distributed element is gener-
ally designed to be low loss in order to transport maximum signal to other circuit elements. Temperature variation of a distributed structure is usually not considered further than the requirements for power handling dictate. Under moderate to high power conditions, significant thermal variation can occur over distributed structures just as in the lumped element case. The distortion generated by this temperature variation is not localized and combines along the length of the line. The thermal domain of a general distributed structure composed of a metal on top of a substrate is analyzed in Section 5.2.1, yielding the temperature distribution in the conductive layer of the metal due to the metal and substrate thermal properties. Using the developed thermal solution, the electro-thermal distortion from an infinitesimal element of the metal is derived in Section 5.2.2. The propagation, combination, and decay of both fundamental applied signals and infinitesimal element generated distortion is discussed in Section 5.2.3.

5.2.1 Heat Conduction on Transmission Lines

Heat dissipation occurs in conductors due to the finite conductivity of any real metal. As the frequency of operation increases, the skin effect reduces the effective area the current in the conductor flows through. The conductive loss, or heat generated, not only increases but is also confined to a smaller volume. At radio frequency this effect becomes significant enough to alter the resistivity of the transmission line sinusoidally over time. Periodic variations of the resistivity due to thermal effects are known to produce distortion, as shown previously in microwave terminations and attenuators [65]. The derivation of the distortion generated by electro-thermal processes requires analysis of heat conduction over the length of the transmission line. This section derives the temperature of the interior of the transmission line based on current distributions to enable the derivation of the PIM generated due to electro-thermal effects.

Two cases exist for heat conduction on a transmission line dependent on the conductor thickness relative to the current skin depth at a given operational frequency. When the conductor thickness is less than a few skin depths, the substrate thermal properties will dominate that of the metallization. If the conductor thickness is significantly greater than a few skin depths, the thermal properties of the metallization will increasingly contribute to the thermal conduction in combination with the substrate. The conductor of
the transmission line can be modeled as a semi-infinite plane due to the relatively uniform
temperature across the resistive plane in comparison to the interior of the conductor or the
substrate. This heat flow can be considered one dimensional from each conductor wall, as
each resistance element of the conductor functions as an independent parallel source heater
on the surface.

The transmission line of Fig. 5.1 can readily be modeled as a semi-infinite heat
conduction system. Here a conductor is bounded by air on all sides save the interface
with the dielectric. The resistive loss of the conductor is distributed across the structure
intrinsically and thus is naturally partitioned into infinitesimal elements where the depth of
the heating cell is given by the skin depth. Current flows through each of these distributed
elements generating heat. The air surrounding the structure can be thought of as thermal
insulation as the thermal conductivity is at least an order of magnitude below the thermal
conductivity of most electrical substrates. A conductor that is thinner than a few skin
depths will be approximately evenly heated through the conductor, requiring heat to flow
into the substrate. In a conductor that is significantly larger than a few skin depths,
the unheated interior of the metal will contribute to the heat sinking capabilities of the
structure in tandem with the substrate. In either case, the heated layer of the conductor
interfaces the heat sinking layer on a plane where every point in the heated conductor layer
functions as a point heater for each corresponding point in the heat sinking layer. Air
insulation in combination with point heater symmetry allow heat transfer to be considered
one dimensional at the interface with the heater element.

The temperature distribution within a few skin depths is responsible for the conductivity modulation that results in electro-thermal distortion. Derivation of the temperature in this region begins with the one dimensional heat conduction equation,

\[ k \frac{\partial^2 T(x,t)}{\partial x^2} - \rho_d c_v \frac{\partial T(x,t)}{\partial t} = g(x,t), \quad (5.1) \]

where \( c_v \) is the thermal capacity (units of \( J \cdot K^{-1} \cdot kg^{-1} \)), \( \rho_d \) is the density (units of \( kg \cdot m^{-3} \)), and \( k \) is the thermal conductivity (units of \( K^{-1} \cdot W \)). The forcing term, \( g(x,t) \) (units of \( W \cdot m^{-3} \)), is the heat generated within the metal by current dissipation.

The electrical signal power dissipation generates heat within the conductor, resulting in a temperature profile that is dependent both on the generated heat and the material thermal parameters. Due to this source condition, it becomes useful to express the heat conduction equation in terms of heat rather than temperature. Expression of the heat conduction equation in heat can be readily accomplished by taking the spatial derivative of the equation and using the definition of heat flux,

\[ q(x,t) = -k \frac{\partial T(x,t)}{\partial x}. \quad (5.2) \]

The heat conduction equation in terms of heat flux becomes

\[ \frac{\partial^2 q(x,t)}{\partial x^2} - \frac{\rho_d c_v}{k} \frac{\partial q(x,t)}{\partial t} = \frac{\partial g(x,t)}{\partial x}. \quad (5.3) \]

The heat generated in the material, \( g(x,t) \), must necessarily flow through the same unit area as the heat flux, \( q(x,t) \). If the heat flowing is confined to be only the heat generated within the material, then the two terms are proportional by the corresponding dimensions, effectively transforming the nonhomogeneous equation (5.1) to the homogenous equation

\[ \frac{\partial^2 q(x,t)}{\partial x^2} - \frac{\rho_d c_v}{k} \frac{\partial q(x,t)}{\partial t} = A \frac{\partial q(x,t)}{V \partial x}. \quad (5.4) \]

Here \( V \) is the volume of the thermal system where the heat is generated and \( A \) is the area the heat flux must travel through. The utility of (5.4) lies in the frequency response of the solution, which is bounded for all frequencies. The solution to (5.3) is not bounded at low frequencies, where the temperature is predicted to approach infinity.

The solution of (5.4) is needed for the case of constant flux and periodic flux. The constant flux case yields the effective bias point of the electrical and thermal conductivity,
while the periodic flux case determines the nonlinear behavior of the electro-thermal system.

The boundary conditions of the constant flux system over the semi-infinite domain \(-\infty < x < 0\), are given by

\[
\begin{align*}
q(0, t) &= q_a \quad t > 0 \\
q(x, t) &= 0 \quad t = 0 \\
T(x, t) &= T_a \quad t = 0
\end{align*}
\] (5.5)

where \(q_a\) is the average heat produced in the material and \(T_a\) is the ambient temperature in the material. The temperature is, by the definition of heat flux, given by

\[
T(x, t) = \frac{1}{k} \int q(x, t) \, dx = \frac{q_a}{k} + T_a.
\] (5.6)

This temperature represents the temperature rise due to the average electrical power dissipation superimposed with the ambient temperature of the material. With the bias point of the thermal conductivity obtained, the solution for a periodic flux must now be obtained in order to describe the time dependent thermal modulation of electrical conductivity. The boundary conditions for this one dimensional system over the domain \(-\infty < x < 0\) are given by

\[
\begin{align*}
q(0, t) &= q_p \cos(\omega t - \phi) \quad t > 0 \\
q(x, t) &= 0 \quad t = 0
\end{align*}
\] (5.7)

where \(q_p\) is the magnitude of the generated periodic heat, \(\omega\) is the radian frequency, and \(\phi\) is generated heat phase. The heat signal may be dependent on space as well, but here is assumed to decrease exponentially at a rate exceeding the decrease of the heat flux if applied at the boundary \(x = 0\) alone. In [99], the thermal penetration depth was predicted to be smaller than the skin depth of the conductor assuming the same frequency for both electrical and thermal systems. The frequency of the thermal system cannot be taken to be the same or higher than the electrical system when amplitude modulation is involved, as it produces low frequency terms resulting in thermal depths much larger than the skin depth of the applied electrical signal.

A periodic heat flux applied to the conductor will have a temperature solution that is periodic in time and attenuated as it propagates in space. Such a profile implies the use of a solution of the form

\[
q(x, t) = X(x) e^{i(\omega t - \phi)},
\] (5.8)
where the function $X(x)$ describes the heat profile in the material.

Substituting this solution into (5.4) yields

$$\frac{\partial^2 X(x)}{\partial x^2} - \frac{A}{V} \frac{\partial X(x)}{\partial x} - \frac{j\omega \rho c_v}{k} X(x) = 0,$$

an equation dependent only on space and periodic in time which has the solution

$$X(x, t) = Be^{r_1 x} + Ce^{r_2 x},$$

with roots, $r_1$ and $r_2$ given by

$$r_1 = 1 + \frac{\sqrt{1 + 4j\omega \rho c_v k^{-1}V^2A^{-2}}}{2VA^{-1}}$$

$$r_2 = 1 - \frac{\sqrt{1 + 4j\omega \rho c_v k^{-1}V^2A^{-2}}}{2VA^{-1}}.$$  

The term which is finite as $x \to -\infty$ and $\omega > 0$ is

$$X(x) = Be^{r_1 x}.$$  

The solution which has the value of the source at $x = 0$ is

$$q(x, t) = q_p e^{\Re(r_1)x} \cos(\omega t - \Im(r_1)x - \phi).$$

The temperature distribution is obtained by the integration of the heat flux over space, which is given by

$$T(x, t) = \frac{1}{k} \int_{0}^{\infty} q(x, t) dy$$

$$= \frac{1}{k} \int_{0}^{\infty} q_p e^{\Re(r_1)x} \cos(\omega t - \Im(r_1)x - \phi) dy.$$  

Integration by parts and application of the original boundary conditions yields the solution for the temperature in the conductor,

$$T(x, t) = \frac{2VA^{-1}k^{-1}q_p e^{\Re(r_1)x} \cos(\omega t - \Im(r_1)x - \phi)}{1 + \sqrt{1 + 4j\omega \rho c_v k^{-1}V^2A^{-2}}}.$$  

This solution is similar to that in Section 4.3.1, but it is derived from the heat conduction equation rather than from a compact model. The solution to the heat conduction
equation with only boundary conditions rather than a forcing term results in a function that is unbounded as the frequency approaches zero. The homogenous equation is thus limited to only the high frequency case and can not predict the low pass filter response of the thermal system. This discrepancy does not exist in the solution to the non-homogenous equation, which is appropriately bounded over all frequencies.

The solution given by (5.16) represents a low pass thermal filter with a slope of 10 dB per decade in its frequency response applied to the electrically generated heat signal. The thermal bandwidth is determined by both the thermal parameters of the material and its dimensions. The volume of heat generation and the area that heat must flow through, assuming one dimensional heat transfer, largely determines the bandwidth of the thermal filter. The fractional derivative of Section 4.3.1 is contained within the solution, but has not been converted back to the time domain.

5.2.2 Electro-Thermal PIM of a Finite Element

The nonlinear nature of electro-thermal distortion stems from the dependence of the metal conductivity on temperature. This effect occurs due to the thermal dependence of electron scattering by lattice vibrations in the metal [38]. It is by nature a distributed effect, as it exists throughout every part of the material. The process is called the thermo-resistance effect, and models the electrical resistivity, $\rho_e$, (units of $\Omega \cdot m$), of a material as a function of temperature, $T$ [39]:

$$\rho_e(T) = \rho_{e0}(1 + \alpha T + \beta T^2 + \ldots).$$  \hspace{1cm} (5.17)

Here $\rho_{e0}$ is the static resistivity constant and $\alpha$ and $\beta$ are constants representing the temperature coefficients of resistance (TCR). The temperature in (5.17) is determined by the heat in the conductor, which is a function of the electrical power. The heat generated over any lossy element, $Q$, (units of $W \cdot m^{-3}$) is equivalent to the power dissipation over that element, given by

$$Q = J^2 \rho_e(T),$$  \hspace{1cm} (5.18)

where $J$ is the current density vector in units of $A \cdot m^{-2}$. The electric field in the conductor due to the current density $J$ is given by

$$E = J \rho_{e0} [1 + \alpha (T_a + T_p)].$$  \hspace{1cm} (5.19)
Here $T_a$ is the ambient temperature and $T_p$ is the periodic temperature due to sinusoidal heating. Further terms in the series have been dropped due to the magnitude difference between $\alpha$ and $\beta$ ($\alpha \gg \beta$) in most metals.

The nonlinear electric field in (5.19) is exclusively determined by the current density in the conductor of interest. The current density in that conductor is defined by the solution to Maxwell’s equations for a given transmission line configuration. The current density over a finite cell of the conductor, which could be due to propagating fields in any direction, can be specified by a two-tone spatially independent signal to derive nonlinear effects. The total current density at a point is defined as

$$J = J_1 \cos(\omega_1 t + \phi_1) + J_2 \cos(\omega_2 t + \phi_2).$$

The coefficients $J_1$ and $J_2$ are the magnitude of the current density, defined by the transmission line configuration. The radian frequencies are represented by $\omega_1$ and $\omega_2$ while $t$ is the time and $\phi_1$ and $\phi_2$ are the respective phases.

The periodic temperature, $T_p$, is given by equation (5.16) with $q_p$ defined by the expansion of the dissipated power,

$$Q = \frac{1}{2} \rho_e (J_1^2 + J_2^2) + \ldots$$

$$\frac{1}{2} \rho_e J_1^2 \cos(2\omega_1 t + 2\phi_1) + \ldots$$

$$\frac{1}{2} \rho_e J_2^2 \cos(2\omega_2 t + 2\phi_2) + \ldots$$

$$\rho_e J_1 J_2 \cos[(\omega_2 - \omega_1)t + (\phi_2 - \phi_1)] + \ldots$$

$$\rho_e J_1 J_2 \cos[(\omega_2 + \omega_1)t + (\phi_2 + \phi_1)].$$

The form of equation (5.16) is that of a low pass filter. The interaction from heat signal components with frequencies significantly above the 3 dB point of the thermal filter will have a negligible impact upon the distortion. The condition on frequencies that will contribute to the distortion negligibly is

$$f \gg \left| \frac{kA}{8\pi\rho_c V} \right|,$$

where $f$ is the electrical frequency. The term $\omega_2 - \omega_1$ is very low frequency in comparison to the high frequency terms and is the only term in the expansion that can appreciably
contribute to the distortion. The heat dissipated can be reduced to

\[ Q = \frac{1}{2} \rho_e (J_1^2 + J_2^2) + ... \]

\[ \rho_e J_1 J_2 \cos [(\omega_2 - \omega_1) t + (\phi_2 - \phi_1)] , \]

which represents the average and periodic heat dissipated. In (5.16), the heat flux was integrated over the semi-infinite domain. The temperature inside the heat generation layer is the variable of interest in (5.17). The heat generation layer is defined by the skin depth in the conductor, in units of meters,

\[ \delta = \sqrt{\frac{\rho_e(T)}{\pi f \mu_0 \mu_r}} . \] (5.24)

The periodic temperature, \( T_p \), over the skin depth \( \delta \) where \( x \) in (5.16) is approximately zero must necessarily be

\[ T_p(0, t) = \]

\[ \frac{2 \rho_e (1 + \alpha T_a) J_1 J_2 k^{-1} V A^{-1} \delta \cos (\omega_d t - \phi_d)}{1 + \sqrt{1 + 4 \omega \rho_c k^{-1} V^2 A^{-2}}} , \]

where \( \omega_d = \omega_2 - \omega_1 \), \( \phi_d = \phi_2 - \phi_1 \), and \( r_1 \) is defined in (5.11). The third-order PIM generated can be determined by expansion of (5.19). The upper and lower products generated by each finite element are given by

\[ E_{2\omega_1 - \omega_2} = \beta_1 \cos ((2\omega_1 - \omega_2) t - (2\phi_1 - \phi_2)) \]

\[ E_{2\omega_2 - \omega_1} = \beta_2 \cos ((2\omega_2 - \omega_1) t - (2\phi_2 - \phi_1)) \]

\[ \beta_1 = \frac{2 V A^{-1} \rho_e^2 (1 + \alpha T_a) J_1 J_2^2 \alpha \delta k^{-1}}{1 + \sqrt{1 + 4 \omega \rho_c k^{-1} V^2 A^{-2}}} \]

\[ \beta_2 = \frac{2 V A^{-1} \rho_e^2 (1 + \alpha T_a) J_2 J_2^2 \alpha \delta k^{-1}}{1 + \sqrt{1 + 4 \omega \rho_c k^{-1} V^2 A^{-2}}} \]

The expansion of (5.19) can be continued to reach a closed form for higher order distortion products, given by

\[ E = \rho_e (1 + \alpha T_a) J + \rho_e (1 + \alpha T_a) \cdot \]

\[ \sum_{n=1}^{\infty} j^{2n+1} \left( \frac{2 V A^{-1} \rho_e^2 \alpha \delta k^{-1}}{1 + \sqrt{1 + 4 \omega \rho_c k^{-1} V^2 A^{-2}}} \right)^n . \] (5.30)
Convergence of the series is guaranteed by the ratio test for the condition

\[
\left| J(j\omega) \left( \frac{2VA^{-1}\rho_0\alpha_0\delta k^{-1}}{1 + \sqrt{1 + 4j\omega\rho_c k^{-1}V^2A^{-2}}} \right) \right| < 1.
\] (5.31)

The distortion generated over a finite cell can be used to derive the nonlinear electric field based on a current density defined by the electromagnetic modes of a given transmission line structure. The fields described here will propagate down the transmission line summing with each subsequent finite cell. It is necessary to describe summing effects for a given structure in order to fully describe the generated fields at any point on the line. The interference patterns of electro-thermal point sources are examined in the following section.

5.2.3 Distributed PIM Interference

A transmission line can be viewed as a line of non-linear point generators extending the length of the transmission line. Each of these point sources is fixed in space, generating isotropic PIM. The metal conductivity is modulated by heat producing distortion at each point generator. The resulting distortion, as predicted by hyugen’s principle, travels in both the forward and reverse directions on the line with respect to the point generator origin. Two cases of interest emerge from this viewpoint, the distortion at the output from the forward traveling PIM wave and the distortion at the input from the reverse traveling PIM wave. The cumulative PIM at each port can be found from the spatial summation of the individual nonlinear waves from each point generator. This section discusses the summation of a series of electro-thermal point sources down the length of a transmission line in both the forward and reverse directions. The phase relationship of the point source generated distortion to a sinusoidal wavefront propagating down a transmission line is analyzed to determine the interference pattern of the point source generators along the length of the transmission line.

When a forward propagating electric field is initially applied to the beginning of the line, each nonlinear point generator is energized sequentially in time and space. The generation of distortion along the transmission line is shown by the line of point generators separated by \( \Delta z \) in Fig. 5.2(a), which sequentially produce distortion due to the impinging RF energy. As each incremental non-linear cell is heated, a wave at the third order frequency
is generated and propagates down the line at the phase velocity of that frequency. The wave front phase stays constant as it initially progresses down the line, which can be seen by analyzing the phase shift due to the propagation of the wave over an incremental time $\Delta t$ and the corresponding distance $\Delta z$. The two-tone electric field, if a TEM mode of propagation is assumed, can be described by

$$
E = E_1 \cos (\omega_1 \Delta t - \kappa_1 \Delta z + \phi_1) + E_2 \cos (\omega_2 \Delta t - \kappa_2 \Delta z + \phi_2),
$$

(5.32)

where $\omega_1$ and $\omega_2$ are the radian frequencies, $\kappa_1$ and $\kappa_2$ are the wavenumbers, and $\phi_1$ and $\phi_2$ are the respective signal phases. The phase change per forward increment in space and time of the wave front, $\Delta \phi_{\omega_1}$ and $\Delta \phi_{\omega_2}$ is given by

$$
\Delta \phi_{\omega_1} = 2\pi f_1 \Delta t - 2\pi f_1 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}}(f_1)} = 0
$$

(5.33)

$$
\Delta \phi_{\omega_2} = 2\pi f_2 \Delta t - 2\pi f_2 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}}(f_2)} = 0.
$$

(5.34)

where $f_1$ and $f_2$ are the respective signal frequencies.

Each non-linear generator will see the same wavefront at sequentially incremental times in the case of a dispersionless medium. The total distortion at point $b$ at time $\Delta t$ in Fig. 5.2(c) will be the combination of the distortion generated at point $a$ at time 0 and the distortion generated at point $b$ at time $\Delta t$. The distortion at point $b$ at time $\Delta t$ is generated at a phase directly related to the wavefront phase, assumed to be zero here. The distortion generated at point $a$ in Fig. 5.2(c) has traveled to point $b$ at time $\Delta t$. In the case of a dispersionless medium, the phase of the distortion propagating from point $a$ at time $\Delta t$ is given by

$$
\Delta \phi_{\omega_3} = 2\pi f_3 \Delta t - 2\pi f_3 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}}(f_3)} = 0,
$$

(5.35)

where $f_3$ is the third order distortion frequency. The electric field at point $b$ in Fig. 5.2(c), excluding line losses, is then in phase with the distortion that has propagated from point $a$. The magnitude of the forward-propagating third order electric field is simply given by the summation of the magnitude of the fields from each point generator,

$$
E_{\omega_3} = E_{\omega_3} (a) + E_{\omega_3} (b).
$$

(5.36)
Figure 5.2: Distributed PIM generation from an encroaching wavefront of a sinusoidal signal applied at \( t = 0 \) at three incremental points in time \( \Delta t \) and space \( \Delta z \): (a) distortion sources at each point coupling forward and reverse propagating waves, (b) the encroaching wavefront at point a at \( t = 0 \), (c) the encroaching wavefront at point b at \( t = \Delta t \), and (d) the encroaching wavefront at point c at \( t = 2\Delta t \).

The reverse propagating distortion signal does not experience constructive interference, as the phase increases with the opposite sign of the distortion generating forward-traveling wave due to the difference in propagation direction. This can be shown by summing the individual signals at point a, Fig. 5.2(b), at an incremental time where a third order distortion product is generated at point b, Fig. 5.2(c), and travels to point a, Fig. 5.2(b). The reverse propagating third order distortion wave is given by,

\[
E_{\omega_3} (a) = E_{\omega_3} (a) + E_{\omega_3} (b) \cos (\omega_3 \Delta t + \kappa_3 \Delta z),
\]

(5.37)

where \( \kappa_3 \) is the wavenumber of the third order distortion. The generated third order distortion from point c, Fig. 5.2(d), has not had sufficient time to reach point a, Fig. 5.2(b). Now the phase of the third order distortion signal from point b, Fig. 5.2(c), at time \( 2\Delta t \) is the incremental phase shift given by

\[
\Delta \phi_{\omega_3} (b) = 2\pi f_3 \Delta t - 2\pi f_3 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}} (f_3)} = 0.
\]

(5.38)
The third order distortion signal at point \( a \) generated at time \( 2\Delta t \), Fig. 5.2(b), has the phase of the original applied signal two spatial increments behind the wave front, given by

\[
\Delta \phi_{\omega 3}(a) = 2 \left[ 2\pi f_3 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}}(f_3)} \right] .
\]  
(5.39)

The total phase shift of each incremental progression of the reverse distortion wave is given through the difference of the two phases

\[
\Delta \phi_{\omega 3} = 2 \left[ 2\pi f_3 \Delta z c^{-1} \sqrt{\varepsilon_{\text{eff}}(f_3)} \right] .
\]  
(5.40)

Substitution of \( \Delta z = \lambda/4 \) into (5.40) shows that each distortion point source combines progressively farther out of phase with the previous point sources until complete destructive interference is reached when the phase shift between point sources, \( \Delta \phi \), reaches \( \pi \) at the length \( \Delta z = \lambda/4 \).

On a real transmission line, dispersion occurs, and will result in a finite phase shift between the generating wave and the distortion components due to the difference in their propagation velocity. The phase contribution to the distortion wave can be found by taking the phase difference of the closest carrier frequency wave with the distortion wave over an incremental interval yielding

\[
\Delta \phi_f = \pm \kappa_3 \left[ \frac{\sqrt{\varepsilon_{\text{eff}}(f_1)}}{\sqrt{\varepsilon_{\text{eff}}(f_3)}} - 1 \right] \Delta z
\]  
(5.41)

for a forward wave and for a backward wave

\[
\Delta \phi_r = \pm \kappa_3 \left[ \frac{\sqrt{\varepsilon_{\text{eff}}(f_1)}}{\sqrt{\varepsilon_{\text{eff}}(f_3)}} + 1 \right] \Delta z.
\]  
(5.42)

The sign of the propagation is dependent upon which wave, the linear signal or distortion, is traveling with higher phase velocity.

Considering the phase progression of forward and backward propagating distortion waves, dispersion phase shift, and line loss on both the fundamentals and distortion products, the distributed PIM contributions can be summed over the line to give an expression for the PIM at any point on the line. Considering the power loss of the linear electric field, the third order electric field amplitude generated at each point along the line can be expressed

\[
E_3 = \frac{2VA^{-1} \rho_e^2 \left( 1 + \alpha T_a \right) \alpha \delta k^{-1} \left[ J_0 e^{\gamma p \rho L (2MWZ_0)^{-1}} \right]^3}{1 + \sqrt{1 + 4\rho \rho c k^{-1} V^2 A^{-2}}}
\]  
(5.43)
with the forward wave given by

\[
P_{\text{IM}}^f = \sum_{n=0}^{M} E_3 \cdot e^{(M-n)\rho_e L(2MW\delta Z_0)^{-1}} \cdot \cos \left( \omega_3 n \Delta t \Delta z + \kappa_3 n \Delta z \pm \kappa_3 n \Delta z \Delta \phi_f \right).
\]  
(5.44)

The backward traveling PIM is similarly given by

\[
P_{\text{IM}}^r = \sum_{n=0}^{\lambda/4} E_3 \cdot e^{(M-n)\rho_e L(2MW\delta Z_0)^{-1}} \cdot \cos \left( \omega_3 n \Delta t \Delta z + \kappa_3 n \Delta z \pm \kappa_3 n \Delta z \Delta \phi_f \right).
\]  
(5.45)

The expressions for forward and backward PIM, (5.44) and (5.45), are not dependent on matching conditions. The matching condition is implicitly contained in the specification of the applied current amplitude and phase. Reflections from any load or source conditions can be analyzed by specifying the amplitude of reflection and reapplying the summations of (5.44) and (5.45). It is apparent that output port matching will appreciably effect the input port distortion, both through new generation of forward distortion in the reverse direction and from the partial reflection of previously generated forward wave distortion.

The dispersion predicted in (5.41) and (5.42) has little effect on the summation if the signal is narrowband. As the signal grows in bandwidth and the transmission line length increases, the summation of distortion products is reduced in amplitude and with high enough dispersion can completely destructively interfere. Line loss reduces the generated distortion magnitude with increasing line length competing with the dispersion effect.

5.2.4 Summary of Distributed Electro-Thermal PIM Theory

Small thermal variations in the conducting layer of a metal result in PIM generated at each point on the line. The generated distortion is dependent on line dimensions, especially width and length of the line, as well as the thermal properties of the substrate. The thermal properties of the substrate determine both the peak amplitude of the distortion generated and the thermal bandwidth of the transmission line. The width of the transmission line is seen to directly effect the heat generated per unit power, increasing distortion with thinner line width. The distortion generated at each point on the line couples back into the modes of the line, combining in a constructive manner in the forward propagation direction and destructively after \(\lambda/4\) in the reverse direction. This behavior agrees
extremely well with previous PIM near field measurements along the length of transmission lines in the literature.

5.3 Nonlinear Conductivity Isolation

Several physical mechanisms are thought to produce distributed PIM including ferromagnetic metals, conductor surface roughness, tunneling, weak junction effects at metal-oxide-metal junctions, dielectric nonlinearities, and nonlinear conductor resistivity. In order to confirm or discount a particular physical mechanism as the responsible effect, the mechanisms must be isolated from each other. However, very few physical mechanisms can be completely eliminated from the transmission line. The mechanism of interest must be enhanced while every other effect is suppressed. The manufacturing process described in this section was designed to minimize or eliminate every physical mechanism save nonlinear conductor resistivity. Nonlinear resistivity was chosen as the physical mechanism to isolate due to its existence in every transmission line. The manufacturing process developed to isolate the nonlinear conductivity specifically from electro-thermal processes is presented in Section 5.3.1. The design of several transmission lines using this process is discussed in Section 5.3.2. The interface of the transmission line to a RF measurement system to allow high dynamic range measurement of the samples is presented in Section 5.3.3.

5.3.1 Materials Design for Process Isolation

The manufacturing process described here minimized each of the physical mechanisms possibly responsible for distributed PIM generation, save electro-thermal conductor resistivity modulation. Ferromagnetic materials were avoided in both sample design, connectorization, and test equipment. Minimization of dielectric loss is accomplished by using low loss substrates, single crystal sapphire and quartz, having respective loss tangents of 0.00002 and 0.0001. The 100 mm diameter substrates were epi-polished to < 15 angstroms surface roughness for quartz and < 5 angstroms surface roughness for sapphire. Metallization annealing, combined with the epi-polished substrates, effectively eliminate surface roughness contributions to conductivity. The lack of surface structures eliminates the possibility of tunneling and by only using single metal, metal-metal and metal-oxide-metal
structures are avoided. The electro-thermal mechanism is enhanced through line thickness control and guaranteed to be the strongest nonlinear process in these samples.

The wafers were first washed with acetone and methanol to remove any film, then dried for five minutes at 500 °C. A seed metal layer with excellent adhesion such as chromium is normally used when sputtering other metals onto a substrate. In this experiment, a seed layer would form both a dissimilar metal-metal junction and result in a distributed ferromagnetic structure. To prevent this, the seed layer was avoided and silver was directly sputtered onto the substrate, 1.8 µm thick with a maximum 0.2 µm variation from wafer center to wafer edge. Adhesion was achieved by annealing the sample in air for 30 minutes at 500 °C. Photo-resist was applied, UV patterned, and used to etch the transmission lines with CR-7 etchant. The photo-resist was then removed and the wafer cleaned with acetone, methanol, and a deionized water rinse. The wafer was sputtered on the backside and annealed again at 500 °C for 30 minutes to provide a ground plane. The device was washed again and stored in a nitrogen atmosphere to prevent oxide or sulphide formation.

The procedure to produce these test samples results in a single metal transmission lines with bottom side roughness equivalent to the smoothness of the epi-polished substrate. The bottom side smoothness is shown in the cross section of the transmission line pictured in the SEM image of Fig. 5.3. Top side surface roughness is controlled by annealing the device to form extremely flat grains without grain merging, pictured in the SEM image of Fig. 5.4. Conductivity losses are directly controlled through the thickness of the metallization, enabling the enhancement of this effect to a degree. At a minimum thickness of 1.1 µm, grains begin to grow together during annealing resulting in large constriction resistances, pictured in the SEM image of Fig. 5.5. The test samples used here were free of constriction resistances, shown in Fig. 5.4.

The process described results in silver structures that are single metal and are directly wirebondable. All commonly suggested physical nonlinear mechanisms are minimized or eliminated in this process. Silver is the highest conductivity metal and is commonly used as plating in low PIM components. This manufacturing process generates the lowest possible PIM components obtainable when conductor thickness is not minimized, and represents the best achievable performance of transmission lines.
5.3.2 Transmission Line Design

A transmission line must be designed for a given impedance, usually 50 Ω, for it to be matched to most RF measurement systems without a matching network. Matching networks, although they allow test of arbitrary impedance transmission lines, also provide another source of distortion at the input of the transmission line. The transmission lines manufactured in this research were designed to be 50 Ω to allow connection to the measurement system of Fig. 3.2 without matching networks.

Three separate 50 Ω transmission lines were designed and manufactured, two on a sapphire substrate and one on a quartz substrate. The sapphire substrate is an anisotropic, high permittivity material, $11.58\epsilon_r$ perpendicular to the surface and $9.3\epsilon_r$ parallel to the surface. A transmission line on this substrate can be designed to be very thin, greatly enhancing conductive loss. ADS2006a linecalc was used to obtain the width required, 411 µm, for a 50 Ω transmission line. In the literature, [69, 70], it was shown that PIM magnitude grows with the length of the transmission line. The length of line, 1.26 m, was designed to be approximately three wavelengths at the test frequency of 480 MHz, as in the literature, to guarantee adequate length for PIM growth. The substrate is a circular 100 mm wafer, requiring the line to be serpentine to fit on the substrate, shown in Fig. 5.6. Line segments were spaced at five times the width of the line to ensure minimum reverse wave coupling of the signal along line segments. Reflections were minimized by using semi-
circle bends to connect segments, ensuring only smooth transitions with no discontinuities on the line. The ends of the lines were slightly flared to allow multiple wire bonds directly to the transmission line. Two designs were done on sapphire to allow comparison of both physical and electrical length between different substrates.

The third 50 Ω transmission line was manufactured on quartz substrate, a $3.8\varepsilon_r$ isotropic permittivity medium. A transmission line on this substrate can be designed to be approximately three times wider than a transmission line on a sapphire substrate, reducing the loss by approximately a factor of three. The length of line, 1.23 m, was again designed to be approximately three wavelengths at the test frequency of 480 MHz, as in the literature, to guarantee adequate length for PIM growth. The substrate is also a circular 100 mm wafer, requiring the line to be serpentine to fit on the substrate, shown in Fig. 5.7. Line segments were spaced at five times the width of the line to ensure minimum reverse wave coupling of the signal along line segments. Reflections were minimized by using semi-circle bends to connect segments, ensuring only smooth transitions with no discontinuities on the line. Line ends were not required to be flared as the line is adequately wide to easily apply several wirebonds.

The different permittivities of the quartz and sapphire substrates allow transmission lines of different line loss at the same characteristic impedance. Both substrates are of extremely low dielectric loss, 0.00002 for sapphire and 0.0001 for quartz, isolating the
loss of the conductor alone. Control of the metal thickness further enhances the metal loss, allowing isolation of conductivity modulation as opposed to other physical mechanisms such as dielectric loss.

5.3.3 Transmission Line Interface

The test sample was interfaced to the test system through a printed circuit board (PCB) mounting assembly, shown in Fig. 5.8. The mounting assembly is comprised of a transmission line on 100 mm wafer mounted and wirebonded to a connectorized PCB board. The PCB designed for sample mounting and interface, shown in Fig. 5.9, used Rogers 6002 substrate material with $\geq 5 \, \mu m$ of electroplated gold on all copper metallization. Interfacial metal layers such as chromium and nickel were avoided between the gold and copper to prevent ferromagnetic effects. The traces on the PCB board were coplanar waveguide at wirebonding interfaces to further reduce any current in the copper metallization. Silver conductive paint, 50% weight, was used to provide a connection from the ground plane of the sample to the PCB board. Additional wirebonding connections were designed to minimize cost associated with testing small wafer samples for process refinement. They are grounded for large sample measurement and result in no significant alteration of field patterns around the transmission line.
Figure 5.6: Photolithography mask used to pattern silver on sapphire transmission line.

The silver transmission line for both substrates was thermosonically ball wire-bonded to the PCB board coplanar line with at least five 25.4 µm diameter gold wirebonds at each port using a Kulicke and Soffa 4524AD gold wirebonder. Gold was chosen for wire-bonding due to its lack of oxide formation and favorable alloy properties with silver. Gold and silver form an isomorphous alloy system due to their FCC lattice structure and comparable size resulting in no metal-oxide-metal regions at the wirebonds and a continuous metal-metal contact system [101]. Tunneling and constriction resistance from wirebonding is virtually eliminated in this wirebonding arrangement.

The PCB board was connectorized with a custom made assembly from a Spinner ATL low PIM N-male to N-male cable, guaranteed to exhibit −160 dBc or less distortion at an input signal of two 44 dBm carriers. The helical copper encased cable was cut to dimension with a bandsaw. The connector was trimmed and the dielectric shaved from the center conductor before being cleaned with isopropyl alcohol. The helical copper ground conductor was split to connect to both the surface and bottom of the PCB both electrically and mechanically. The end launch connector was soldered to the PCB using 60/40 tin-lead
solder. The resulting input and output return loss for both fully mounted samples is in excess of 30 dB. The complete test configuration has no ferromagnetic materials, metal-oxide-metal structures, minimum dielectric loss, minimum surface roughness, and enhanced conductive loss.

5.3.4 Summary of Transmission Line Sample Design

Transmission lines manufactured with standard processes possess many physical mechanisms that are thought to produce distortion including nonlinear surface roughness, tunneling, nonlinear conductivity, and ferromagnetic materials. Isolation of any one of these mechanisms requires careful manufacturing and process design to avoid surface roughness and material nonlinearities. A manufacturing process for single metal silver structures on sapphire and quartz substrates was presented that minimizes or isolates all nonlinear physical mechanisms save electro-thermal conductivity modulation. Several transmission lines with varying loss were designed and manufactured using this process. Interface methods presented here allow these samples to be tested using standard RF equipment without gen-
erating additional distortion even under high dynamic range requirements. The developed process represents the highest linearity transmission lines obtainable, and directly relates to cable and connector design where silver is commonly used to plate components for enhanced linearity.

5.4 Case Study: Electro-Thermal Distortion on Transmission Lines

In the theoretical analysis of Sections 5.2.1, 5.2.2, and 5.2.3, the theory of electro-thermal distortion in lumped components was extended to distributed elements. The heat conduction system and even the strength of the coupling between electrical and thermal domains is wildly different in lumped and distributed devices. The commonality in theory between lumped and distributed devices is the characteristic thermal dispersion associated with the generated PIM. The dispersion inherent to the electro-thermal process uniquely defines the physical mechanism responsible for the distortion whether it is distributed or localized. In this section, a two-tone sweep is applied to the transmission lines designed in Section 5.3.2 to characterize the thermal dispersion characteristic of the generated distortion. The measured distortion relation is shown to coincide well with both the predicted
Figure 5.9: Printed circuit board designed on Roger’s 6002 dielectric for interfacing a 100 mm wafer. Coplanar waveguide transmission line connects the wirebond to a N-type connector. Six SMA connections, each possessing five jumper ready wirebonding pads, surround the wafer. Their presence was to lower the cost of testing many small samples for manufacturing process development and did not effect testing of the presented transmission lines.

growth of distortion down the transmission line and the characteristic non-integer Laplacian dispersion predicted in the developed theory. The measurement of electro-thermal distortion from silver, the highest conductivity metal, is discussed followed by an analysis of both thermal parameter and line dimension impact on PIM.

5.4.1 Electro-Thermal Dispersion Measurement

Electro-thermal distortion has been shown in [65] to be characterized in lossy lumped components by sweeping the beat of an amplitude modulated signal. Under this condition, the material exhibits a periodic temperature based upon the material properties and dimensions of the medium. The period of the amplitude modulation of the test signal can be readily swept in a two-tone characterization by altering the tone separation. The resultant third order distortion, if electro-thermal, will exhibit a PIM dispersion characteristic
Table 5.1: Transmission Line Simulation Material Parameters

<table>
<thead>
<tr>
<th>Material</th>
<th>Ag</th>
<th>Al$_2$O$_3$</th>
<th>SiO$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$ (W·m$^{-1}$K$^{-1}$)</td>
<td>429</td>
<td>31</td>
<td>1.3</td>
</tr>
<tr>
<td>$c_v$ (J·Kg$^{-1}$K$^{-1}$)</td>
<td>232</td>
<td>740</td>
<td>670</td>
</tr>
<tr>
<td>$\rho$ (Kg·m$^{-3}$)</td>
<td>10490</td>
<td>3980</td>
<td>2230</td>
</tr>
<tr>
<td>$\alpha$ (K$^{-1}$)</td>
<td>0.0038</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>tan(δ)</td>
<td>—</td>
<td>0.00002</td>
<td>0.0001</td>
</tr>
<tr>
<td>$\sigma$ (MΩ$^{-1}$)</td>
<td>—</td>
<td>18.903</td>
<td>23.3137</td>
</tr>
</tbody>
</table>

in which the response is non-integer Laplacian and is dictated by the thermal parameters of the material.

The sapphire and quartz transmission line samples, interfaced as described in Section 5.3.3, were placed in the location of the DUT in the system of Fig. 3.2 with a 32.3 dB low PIM cable attenuator at the output of the sample. The two transmission line samples, silver on sapphire and silver on quartz, were characterized for electro-thermal dispersion by sweeping the tone separation, $\Delta f$, of a two-tone signal centered at 480 MHz. The silver on sapphire sample was tested at 33 dBm input power from 4 Hz to 10 kHz $\Delta f$ while the silver on quartz sample was tested at 30 dBm input power from 4 Hz to 200 Hz $\Delta f$.

The measured response to the 4 Hz to 10 kHz $\Delta f$ sweep for the silver on sapphire transmission line, shown in Fig. 5.10, exhibits a low pass, approximately 10 dB per decade response, the defining characteristic of electro-thermal distortion. The silver on sapphire transmission line contains the thermal knee predicted in the theory and previously seen only in the integrated circuit resistors of Section 4.6. The 3 dB bandwidth of the thermal dispersion characteristic is approximately 400 Hz. The theoretical model, based solely on material properties and electric field distributions for the transmission line, agrees well with the measured data containing both the same thermal bandwidth and magnitude. The transmission line thermal properties used for simulation of (5.44) are given in Table 5.1 and the dimensional properties are given in Table 5.2.

The measured response to the 4 Hz to 100 Hz $\Delta f$ sweep of the silver on quartz transmission line, shown in Fig. 5.10, exhibits a slow transition low pass characteristic
approaching 10 dB per decade. The thermal parameters of the quartz substrate create a much lower thermal dispersion bandwidth with no visible knee or transition region. The presented distributed electro-thermal theory predicts a sub-hertz thermal bandwidth for this system and accurately predicts the distortion generated in both amplitude and dispersion characteristic. Once again the model for this transmission line is based solely on material properties and electric field distributions. The transmission line thermal properties used for simulation of (5.44) are given in Table 5.1 and the dimensional properties are given in Table 5.2.

5.4.2 Material and Current Bunching Thermal Effects

Electro-thermal distortion is heavily dependent upon the heat conduction properties of the materials in the system. The silver on quartz transmission line, although the loss
Table 5.2: Transmission Line Dimensions

<table>
<thead>
<tr>
<th></th>
<th>Al₂O₃</th>
<th>SiO₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>433 μm</td>
<td>1112 μm</td>
</tr>
<tr>
<td>L</td>
<td>1.2644 m</td>
<td>1.2344 m</td>
</tr>
<tr>
<td>T</td>
<td>1.7 μm</td>
<td>1.7 μm</td>
</tr>
<tr>
<td>S₁</td>
<td>500 μm</td>
<td>500 μm</td>
</tr>
<tr>
<td>VA⁻¹</td>
<td>90 μm</td>
<td>200 μm</td>
</tr>
</tbody>
</table>

is three times lower than the silver on sapphire transmission line, exhibits distortion that would exceed that of the sapphire substrate line with equal applied power. The thermal dispersion bandwidth is also much lower for the silver on quartz transmission line, sub-hertz, compared to the 400 Hz bandwidth of the silver on sapphire transmission line. Decreased loss and thermal bandwidth would suggest lower distortion, but is not seen in the silver on sapphire and silver on quartz two-tone sweeps. This section discusses the effect of thermal material properties and current confinement on distortion generation.

The material properties of sapphire, quartz, and silver are given in Table 5.1. The material properties of silver, substituted into (5.30), can not reproduce the distortion measured in Fig. 5.10 for the transmission line on either substrate. The electro-thermal distortion that would be produced from the silver thermal properties alone is shown in Fig. 5.11. The material properties of the silver yield a thermal bandwidth of over 10 MHz and a peak distortion of −118 dBm at 33 dBm input power for a 50 Ω 433 μm wide transmission line. The silver line does not possess a large enough material volume, and thus thermal capacity, to absorb the heat generated. The heat that is not absorbed by the conductor must flow into the substrate. The thermal material properties of the sapphire and quartz substrate provide a thermal bandwidth of 400 Hz and < 1 Hz, respectively. The magnitude of the thermal response is much greater in both of these materials than the metal due to the large difference in thermal conductivity, which is separated by at least one order of magnitude for each material. The thermal properties of the silver metallization can be ignored as the thermal bandwidth of the substrate dominate, as shown in the measured electro-thermal distortion characteristics of Fig. 5.10.

The thermal bandwidth is defined further by current bunching on the transmission
line through the $V A^{-1}$ scaling factor in (5.30). Current bunching confines the majority of the heat generated to a volume significantly smaller the line width on each edge of the line. The current density for a 50 Ω, 433 μm wide silver on sapphire transmission line, computed in Ansoft HFSS at 480 MHz for a 30 dBm signal, is shown for line width and length in Fig. 5.12(a) and against width only in Fig. 5.12(b). The volume the heat is confined to directly follows the current density function, decreasing with increasing electrical frequency. In Fig. 5.12(b), the peak current density is confined to approximately a 70 μm width on each side of the conductor. The heat generated in each of these volumes spreads and flows through the entire area of the transmission line interface with the substrate. The total distortion generated in each finite electro-thermal cell is then two times that of a single side of the conductor, while the thermal filter bandwidth $V A^{-1}$ scaling parameter is due to the volume from only one side of the conductor. The thermal dispersion characteristic, considering current bunching, is dependent on material thermal properties, electrical frequency, and
5.4.3 PIM Dependency on Line Dimensions

Transmission line distortion is dependent on thermal material properties, current bunching, substrate dimensions, and conductor dimensions. Conductor dimensions such as width and thickness determine the characteristic impedance and loss of the line, in turn defining the heat generated at a given input power. The length of a transmission line effects the generated distortion through spatial summation and loss. This section discusses the effects of each dimensional parameter resulting in design guidelines for low PIM transmission lines.

Growth of PIM over the length of a transmission line, as described in [69, 70], is described by the presented theory with growth over length highly dependent on line loss, shown in Fig. 5.13 for a two-tone signal with a 10 Hz tone separation. PIM grows along the line for both quartz and sapphire samples with each finite cell summing until the loss on the fundamental tones and previously generated PIM is overcome by line loss. PIM decay occurs down the remainder of the length of the line as required by line loss. The measured PIM for both samples matches well with the distributed model at the end of the line.

The effect of line width on PIM over line length is shown in Fig. 5.14, where the width of the transmission line is doubled for each simulation from 102.75 µm to 1644 µm while the characteristic impedance, $Z_0 = 50 \, \Omega$, remains the same. The thermal properties of the substrate are held constant save the width variation, representing a decrease in the permittivity of the substrate. The corresponding distortion is predicted to fall with width in a cubic relationship in (5.30) for a finite electro-thermal cell. However, the growth of the distortion down the transmission line also experiences a cubic width effect from loss along the length of the transmission line as predicted in (5.44). The total width contribution is then sixth order, and for the transmission line of Fig. 5.14 approximately an 18 dB drop in maximum distortion amplitude is obtained for each doubling of the line width.

Similarly the width can be altered while the substrate permittivity is kept constant, representing a sweep of the characteristic impedance of a line. The characteristic impedance increases the current on the line for increasing width, while the width decreases the line loss. Electro-thermal distortion is a current based nonlinearity, leading to a slightly competing
Figure 5.12: The current density distribution of a 433 µm wide silver on sapphire transmission line at 480 MHz: (a) along the width and length of the line, and (b) focused on the width of the line.
trend where the improvement in generated distortion slowly decreases with increasing width. This effect is shown in Fig. 5.15, where the width of the transmission line on a sapphire substrate is doubled each simulation from 102.75 μm to 1644 μm while the characteristic impedance, $Z_0$, is 82.4 Ω, 66.1 Ω, 50 Ω, 34.9 Ω, and 22.3 Ω at each respective width. The initial improvement in distortion is approximately 16 dB when doubling the width of an 82.4 Ω characteristic impedance line. Subsequent improvements in distortion with line width doubling drop to 14 dB as the characteristic impedance is further reduced to 22.3 Ω at a line width of 1644 μm.

The heat generation volume to area ratio of (5.30), $VA^{-1}$, and the line width can be altered through choice of substrate height. Both the magnitude and thermal bandwidth of the thermal dispersion characteristic can be controlled with this height. In Fig. 5.16, the thickness of the substrate, $S_t$, is doubled in each simulation from 125 μm to 2000 μm, while the characteristic impedance is maintained at 50 Ω by doubling the line width each simulation from 102.75 μm to 1644 μm. The dispersion bandwidth is altered significantly by the
change in substrate thickness, decreasing continuously as substrate thickness is increased. The lower distortion on thin substrates in Fig. 5.16 is line length dependent and can be attributed to heavier loss at the given simulation length of 1.26 m. Generally increasing substrate thickness leads to lower distortion over bandwidth.

Thickness of the line primarily effects the conductive loss of the metal until it becomes several times thicker than the skin depth. At this point the additional metal begins to sink heat, effectively increasing the thermal conductivity of the system and reducing the distortion. Metal possesses thermal conductivity one to three orders of magnitude greater than that of electrical substrates. PIM can be significantly reduced if there is enough metal to sink all generated heat. Predicted dimensional dependencies suggest the dynamic range achievable on a transmission line will always be greatest on thick, short, wide lines with thick, high thermal conductivity substrates.
Figure 5.15: PIM generation along length of transmission line for a sapphire substrate where characteristic impedance is varied from 82.3608 Ω to 22.3 Ω from line width variation in multiples of two from 102.75 µm to 1644 µm.

5.4.4 Summary of Distributed PIM Measurements

Distortion from transmission lines has been attributed to many physical mechanisms due to the difficulty in isolating single mechanisms and the lack of a defining characteristic for the distortion generated from each mechanism. Electro-thermal conductivity modulation was isolated and enhanced to be the dominant physical mechanism through careful design and manufacturing enabled by the presented electro-thermal theory. Two-tones sweeps of the tone separation to modify the frequency of amplitude modulation were applied to the samples, generating a non-integer Laplacian dispersion characteristic unique to electro-thermal distortion. This dispersion characteristic matched the presented distributed electro-thermal PIM theory very well for both silver on sapphire and silver on quartz transmission lines, supporting the electro-thermal mechanism as a dominant PIM contributor in distributed structures. The effects of thermal material parameters and device dimensions were analyzed, resulting in design insight and guidelines for low PIM structures.
Figure 5.16: PIM magnitude and thermal dispersion bandwidth of a 50 Ω transmission line on sapphire at sapphire thickness, $S_t$, of 125 µm, 250 µm, 500 µm, 1000 µm, and 2000 µm. The width of the silver metallization is 102.75 µm, 205.5 µm, 411 µm, 822 µm, and 1644 µm, respectively, at an applied tone power of 33 dBm.

5.5 Conclusion

Many physical mechanisms of PIM production on transmission lines have been suggested including ferromagnetism, tunneling, constriction resistances, and nonlinear conductivity. These mechanisms could all be present on a transmission line at the same time, making a single mechanism difficult to isolate. Theoretical treatment of these mechanisms often predicts similar amplitude of nonlinearities, lending no further clarification to the problem. Transmission lines with only one of these physical mechanisms are needed to determine which mechanisms are dominant PIM contributors.

In this chapter, electro-thermal conductivity modulation was suggested as a dominant mechanism producing PIM in distributed structures. Previously conductivity modulation was suggested as being due to surface roughness on the transmission line. Thermal effects were overlooked as the relaxation times between thermal and electrical systems are
offset by several orders of magnitude. Amplitude modulation produces power components of the electrical signal within the thermal relaxation time and was pursued as a PIM contributor due to this coupling. A theoretical analysis of heat conduction and electro-thermal interaction was presented resulting in a closed form solution for electro-thermal PIM based on current density in infinitesimal sections of the distributed element.

PIM has been shown to grow along the length of distributed structures, suggesting its nature as a distributed nonlinearity rather than a point nonlinearity such as a solder droplet or connector. Electro-thermal conductivity modulation is present throughout the conductor of the transmission line. Each infinitesimal element of the transmission line is a nonlinear point generator, which couples distortion into the propagating modes of the line. This distortion is theoretically predicted here to sum constructively in the forward propagation direction and destructively in the reverse propagating direction after one-quarter wavelength.

The theoretical analysis presented enabled transmission line samples isolating electro-thermal conductivity modulation to be designed and manufactured. Two transmission lines made of single metal silver were manufactured on epi-polished sapphire and quartz substrates, resulting in surface roughness on the order of two atoms. Contacts to the samples were made with a silver and gold isomorphous alloy system that creates a continuous transition between the metals, alleviating constriction resistances and contact nonlinearities. Conductor thickness was controlled through sputtering, allowing conductive losses to be enhanced, while separate substrates provided different heat conduction and current density cases for analysis and confirmation of the electro-thermal nonlinearity.

Electro-thermal theory and measurement, both in very good agreement, were used to analyze the effect of material properties and dimensions on transmission line linearity. Transmission line distortion was found to have a sixth order dependency on the width of the transmission line from both current density and the loss rate along the length of the line. Distortion grows along the length of the line, the rate of which is determined by line loss, making line length important for low PIM applications. Thermal properties of the substrate were shown to be dominant over metal thermal properties for thin metallization, and can be controlled by substrate depth. Low PIM transmission lines can be manufactured by designing thick, short, wide lines with thick, high thermal conductivity substrates. Analysis of any design can be conducted using the presented theory, giving simulation capabilities
for predicting PIM in distributed structures.
Passive Intermodulation Distortion in Resonant Structures
6.1 Introduction

Passive intermodulation distortion (PIM) is of concern in any communications system, as the spurious content generated can fall in the receive or transmit bands and detrimentally effect the dynamic range of that system. Many physical mechanisms including ferromagnetic effects, thermionic emission, field emission, tunneling, and micro-discharge have been suggested as physical mechanisms generating PIM on antenna structures [3, 5, 8, 102–104]. Tunneling in reflector antennas has been the focus of the majority of antenna PIM research, where the use of aluminum panels creates Al-Al$_2$O$_3$-Al junctions of the correct dimension to permit tunneling [96]. In [105], high current densities and ferromagnetic materials were shown to increase PIM in antenna structures, although no physical mechanism for the PIM dependency on current density was suggested. In antennas that do not have aluminum junctions, PIM mechanisms are generally lumped together and explained through heuristic models rather than physically based models.

Ferromagnetic effects, thermionic emission, field emission, tunneling, and micro-discharge could all exist on an antenna. Tunneling is a dominant PIM source on aluminum structures [96, 100] due to the thin 2-3 nm oxide natively formed on aluminum, which is exactly the size needed to produce appreciable distortion. Although tunneling can produce appreciable distortion in antenna structures, antenna structures such as dipoles which do not possess junctions still produce PIM [106]. Ferromagnetic effects can be eliminated through material choice. Thermionic and field emission as well as micro-discharge can not be eliminated but can be suppressed by operating under medium power conditions. Electro-thermal conductivity modulation, although not previously suggested as an antenna PIM source, will exist in every metal antenna and possesses the current density dependence suggested in [105]. PIM from other mechanisms such as ferromagnetic materials and surface roughness can be removed through manufacturing techniques but in general will exist on a structure in combination with electro-thermal distortion. Electro-thermal distortion must be analyzed to de-embed any other PIM generating mechanism, and thus represents the baseline of physical performance an antenna can achieve under optimum manufacturing conditions.

In this chapter, electro-thermal conductivity modulation as a physical mechanism for PIM generation on resonant structures is presented. The heat conduction system of a
rectangular patch antenna is explored in Section 6.2.1, yielding the temperature in the conductive layer of the metal. The PIM generated by an infinitesimal element of the conductor due to heating from a standing wave current and the summation of the resulting electric field is discussed in Section 6.2.2. Isolation of the electro-thermal mechanism from other physical mechanisms through manufacturing is discussed in Section 6.3.1. Section 6.3.2 discusses the manufacture of a silver on sapphire rectangular patch antenna and describes its interface to an RF test system. The design of the manufactured rectangular patch antenna is described in Section 6.4.1. In Section 6.4.2 the developed design is simulated in Ansoft HFSS 11, showing the validity of the assumed current distribution and the coupled nonlinear electric field modes. Effects on the measurement system of chapter 3 in both transmission and reflection measurements of the antenna, which constitutes a load with frequency dependent return loss, are discussed in Section 6.4.3. Section 6.4.3 also compares the electro-thermal dispersion of the fully assembled antenna for both transmission and reflection measurements to the predicted distortion of the electro-thermal model with good agreement.

### 6.2 Resonant Structure Electro-Thermal Theory

Electro-thermal distortion generally has not been considered as a dominant PIM mechanism in resonant structures due to the predominately metal structure of antennas as well as the difference between thermal time constants and the period of high frequency electrical signals. In the previous chapter, single metal transmission line structures were shown to experience thermal conductivity modulation within the electrically conducting skin depth. Similarly, currents on antennas are confined to a few skin depths and experience the same thermal conductivity modulation. Thermal conductivity modulation is enhanced in resonant structures due to the formation of a standing wave, which increases the heat produced several times higher than a transmission line. The additional metallization and standing wave properties of a resonant structure significantly alter the heat conduction environment from that of a transmission line. The heat conduction environment for a microstrip rectangular patch antenna is analyzed in Section 6.2.1, yielding the temperature distribution in the conductive layer of the metal due to thermal conduction to the low
current regions of the antenna metallization. Using the developed thermal solution, the electro-thermal distortion from an infinitesimal element of the metal is derived and applied to a sinusoidal current distribution over a rectangular patch antenna in Section 6.2.2.

6.2.1 Heat Conduction on a Rectangular Patch Antenna

Heat dissipation occurs in conductors due to the finite conductivity of any real metal. As the frequency of operation increases, the skin effect reduces the effective area the current in the conductor flows through, increasing the conductive loss and thus the heat generated. At radio frequencies this effect becomes significant enough to alter the resistivity of the transmission line sinusoidally over time. Periodic variations of the resistivity due to thermal modulation are known to produce distortion in microwave terminations, attenuators, integrated circuits, and transmission lines [65]. The derivation of the distortion generated by electro-thermal processes in a given structure requires analysis of heat conduction in that structure. This section presents an analysis of the heat conduction environment of a rectangular patch antenna. A solution for the temperature of the interior of the antenna metallization is presented based on current distributions, allowing description of electro-thermal PIM on the antenna.

The rectangular patch antenna of Fig. 6.1 can readily be modeled as a semi-infinite heat conduction system. Here the antenna metallization is bounded by air on all sides save the interface with the dielectric. The air surrounding the structure can be thought of as thermal insulation as the thermal conductivity is at least an order of magnitude below the thermal conductivity of most electrical substrates and several orders of magnitude below metals. The resistive loss of the conductor is distributed across the structure intrinsically and thus is naturally partitioned into infinitesimal elements where the depth of the heating cell is given by the skin depth. The current flow through each of these distributed elements generates heat according to the spatial distribution of current. The boundary conditions of the antenna require that the current and thus the heat be zero at two of the antenna edges. This boundary condition results in a high thermal conductivity path through the metallization to an effective heat sink composed of the low current regions of the antenna. The heat flows through the metallization due to the difference in thermal conductivity between the heat conduction path through the metallization and the electrical substrate. The heat
conduction direction will alter with increasing metallization thickness in proportion to the magnitude of the thermal conductivity of the x-direction and z-direction heat conduction paths in Fig. 6.1. If the metallization thickness is an order of magnitude less than the antenna dimensions, the dynamic heat flow will be confined to the x-direction.

The temperature distribution within a few skin depths is responsible for the conductivity modulation that results in electro-thermal distortion. Derivation of the temperature in this region begins with the one dimensional heat conduction equation,

\[ k \frac{\partial^2 T(x, t)}{\partial x^2} - \rho_d c_v \frac{\partial T(x, t)}{\partial t} = g(x, t), \]

where \( c_v \) is the thermal capacity (units of J·K\(^{-1}\)·kg\(^{-1}\)), \( \rho_d \) is the density (units of kg·m\(^{-3}\)), and \( k \) is the thermal conductivity (units of K\(^{-1}\)·W). The forcing term, \( g(x, t) \) (units of W·m\(^{-3}\)), is the heat generated within the metal by current dissipation.

The electrical signal power dissipation generates heat within the conductor, result-
ing in a temperature profile that is dependent both on the generated heat and the material thermal parameters. Due to this source condition, it becomes useful to express the heat conduction equation in terms of heat rather than temperature. Expression of the heat conduction equation in heat can be readily accomplished by taking the spatial derivative of the equation and using the definition of heat flux,

\[ q(x, t) = -k \frac{\partial T(x, t)}{\partial x}. \] (6.2)

The heat conduction equation in terms of heat flux becomes

\[ \frac{\partial^2 q(x, t)}{\partial x^2} - \frac{\rho_d c_v}{k} \frac{\partial q(x, t)}{\partial t} = \frac{\partial g(x, t)}{\partial x}. \] (6.3)

The heat generated in the material, \( g(x, t) \), must necessarily flow through the same unit area as the heat flux, \( q(x, t) \). If the heat flowing is confined to be only the heat generated within the material, then the two terms are proportional by the corresponding dimensions, effectively transforming the nonhomogeneous equation (5.1) to the homogenous equation

\[ \frac{\partial^2 q(x, t)}{\partial x^2} - \frac{\rho_d c_v}{k} \frac{\partial q(x, t)}{\partial t} = \frac{A}{V} \frac{\partial q(x, t)}{\partial x}. \] (6.4)

Here \( V \) is the volume of the thermal system where the heat is generated and \( A \) is the area the heat flux must travel through. The utility of (6.4) lies in the frequency response of the solution, which is bounded for all frequencies. The solution to (6.3) is not bounded at low frequencies, where the temperature is predicted to approach infinity.

The solution of (6.4) is needed for the case of constant flux and periodic flux. The constant flux case yields the effective bias point of the electrical and thermal conductivity, while the periodic flux case determines the nonlinear behavior of the electro-thermal system. The boundary conditions of the constant flux system over the semi-infinite domain \(-\infty < x < 0\), are given by

\[ q(0, t) = q_a, \quad t > 0 \]
\[ q(x, t) = 0, \quad t = 0 \]
\[ T(x, t) = T_a, \quad t = 0 \] (6.5)

where \( q_a \) is the average heat produced in the material and \( T_a \) is the ambient temperature in the material. The temperature is, by the definition of heat flux, given by

\[ T(x, t) = \frac{1}{k} \int q(x, t) \, dx = \frac{q_a}{k} + T_a. \] (6.6)
This temperature represents the temperature rise due to the average electrical power dissipation superimposed with the ambient temperature of the material. With the bias point of the thermal conductivity obtained, the solution for a periodic flux must now be obtained in order to describe the time dependent thermal modulation of electrical conductivity. The boundary conditions for this one dimensional system over the domain \(-\infty < x < 0\) are given by

\[
q(0, t) = q_p \cos(\omega t - \phi) \quad t > 0
\]
\[
q(x, t) = 0 \quad t = 0
\]  
(6.7)

where \(q_p\) is the magnitude of the generated periodic heat, \(\omega\) is the radian frequency, and \(\phi\) is generated heat phase. The heat signal may be dependent on space as well, but here is assumed to decrease exponentially at a rate exceeding the decrease of the heat flux if applied at the boundary \(x = 0\) alone. In [99], the thermal penetration depth was predicted to be smaller than the skin depth of the conductor assuming the same frequency for both electrical and thermal systems. The frequency of the thermal system cannot be taken to be the same or higher than the electrical system when amplitude modulation is involved, as it produces low frequency terms resulting in thermal depths much larger than the skin depth of the applied electrical signal.

A periodic heat flux applied to the conductor will have a temperature solution that is periodic in time and attenuated as it propagates in space. Such a profile implies the use of a solution of the form

\[
q(x, t) = X(x) e^{i(\omega t - \phi)},
\]  
(6.8)

where the function \(X(x)\) describes the heat profile in the material.

Substituting this solution into (6.4) yields

\[
\frac{\partial^2 X(x)}{\partial x^2} - \frac{A}{V} \frac{\partial X(x)}{\partial x} - \frac{j \omega \rho c_v}{k} X(x) = 0,
\]  
(6.9)

an equation dependent only on space and periodic in time which has the solution

\[
X(x, t) = Be^{r_1 x} + Ce^{r_2 x},
\]  
(6.10)

with roots, \(r_1\) and \(r_2\) given by

\[
r_1 = \frac{1 + \sqrt{1 + 4j \omega \rho c_v k^{-1} V^2 A^{-2}}}{2VA^{-1}}
\]  
(6.11)
\[ r_2 = \frac{1 - \sqrt{1 + 4\omega \rho c_e k^{-1} V^2 A^{-2}}}{2VA^{-1}}. \]  \hspace{1cm} (6.12)

The term which is finite as \( x \to -\infty \) and \( \omega > 0 \) is

\[ X(x) = Be^{\rho_1 x}. \]  \hspace{1cm} (6.13)

The solution which has the value of the source at \( x = 0 \) is

\[ q(x, t) = q_pe^{\rho(r_1)x} \cos(\omega t - \Im(r_1)x - \phi). \]  \hspace{1cm} (6.14)

The temperature distribution is obtained by the integration of the heat flux over space, which is given by

\[ T(x, t) = \frac{1}{k} \int_{0}^{\infty} q(x, t) \, dy \]
\[ = \frac{1}{k} \int_{0}^{\infty} q_pe^{\rho(r_1)x} \cos(\omega t - \Im(r_1)x - \phi) \, dy. \]  \hspace{1cm} (6.15)

Integration by parts and application of the original boundary conditions yields the solution for the temperature in the conductor,

\[ T(x, t) = \frac{2VA^{-1}k^{-1}q_pe^{\rho(r_1)x} \cos(\omega t - \Im(r_1)x - \phi)}{1 + \sqrt{1 + 4\omega \rho c_e k^{-1} V^2 A^{-2}}}. \]  \hspace{1cm} (6.16)

This solution is similar to that in Section 4.3.1, but it is derived from the heat conduction equation rather than from a compact model. The solution to the heat conduction equation with only boundary conditions rather than a forcing term results in a function that is unbounded as the frequency approaches zero. The homogenous equation is thus limited to only the high frequency case and can not predict the low pass filter response of the thermal system. This discrepancy does not exist in the solution to the non-homogenous equation, which is appropriately bounded over all frequencies.

6.2.2 Electro-Thermal PIM of a Finite Element

The nonlinear nature of electro-thermal distortion stems from the dependence of the metal conductivity on temperature. This effect occurs due to the thermal dependence of electron scattering by lattice vibrations in the metal [38]. It is by nature a distributed
effect, as it exists throughout every part of the material. The process is called the thermo-
resistance effect, and models the electrical resistivity, \( \rho_e \), (units of \( \Omega \cdot \text{m} \)), of a material as a function of temperature, \( T \) [39]:

\[
\rho_e(T) = \rho_{e0}(1 + \alpha T + \beta T^2 + \ldots). \tag{6.17}
\]

Here \( \rho_{e0} \) is the static resistivity constant and \( \alpha \) and \( \beta \) are constants representing the temperature coefficients of resistance (TCR). The temperature in (6.17) is determined by the heat in the conductor, which is a function of the electrical power. The heat generated over any lossy element, \( Q \), (units of \( \text{W} \cdot \text{m}^{-3} \)) is equivalent to the power dissipation over that element, given by

\[
Q = J^2 \rho_e(T), \tag{6.18}
\]

where \( J \) is the current density vector in units of \( \text{A} \cdot \text{m}^{-2} \). The electric field in the conductor due to the current density \( J \) is given by

\[
E = J \rho_{e0}(f) (1 + \alpha(T_a + T_p)). \tag{6.19}
\]

Here \( T_a \) is the ambient temperature and \( T_p \) is the periodic temperature due to sinusoidal heating. Further terms in the series have been dropped due to the magnitude difference between \( \alpha \) and \( \beta \) (\( \alpha \gg \beta \)) in most metals.

The nonlinear electric field in (6.19) is exclusively determined by the current density in the conductor of interest. The current density in that conductor is defined by the solution to Maxwell’s equations for a given transmission line configuration. The current density over a finite cell of the conductor can be specified by a two-tone standing wave to derive nonlinear effects. This signal is defined here by a sinusoidal standing wave, given by

\[
J = J_1 \sin(\beta_1 x) \cos(\omega_1 t + \phi_1) + J_2 \sin(\beta_2 x) \cos(\omega_2 t + \phi_2). \tag{6.20}
\]

The coefficients \( J_1 \) and \( J_2 \) are the magnitude of the current density, defined by the antenna structure. The radian frequencies are represented by \( \omega_1 \) and \( \omega_2 \) while \( t \) is the time and \( \phi_1 \) and \( \phi_2 \) are the respective phases. The propagation constants \( \beta_1 \) and \( \beta_2 \) are given by

\[
\beta_1 = \omega_1 \sqrt{\mu \varepsilon} \tag{6.21}
\]

\[
\beta_2 = \omega_2 \sqrt{\mu \varepsilon}. \tag{6.22}
\]
The periodic temperature, \( T_p \), is given by equation (6.16) with \( q_p \) defined by the expansion of the dissipated power,

\[
Q = \frac{1}{2} \rho_e (f) \left( J_1^2 \sin (\beta_1 x)^2 + J_2^2 \sin (\beta_2 x)^2 \right) + ... \\
\frac{1}{2} \rho_e (f) J_1^2 \sin (\beta_1 x)^2 \cos (2\omega_1 t + 2\phi_1) + ... \\
\frac{1}{2} \rho_e (f) J_2^2 \sin (\beta_2 x)^2 \cos (2\omega_2 t + 2\phi_2) + ... \\
\rho_e (f) J_1 J_2 \sin (\beta_1 x) \sin (\beta_2 x) \cos [(\omega_2 - \omega_1) t + (\phi_2 - \phi_1)] + ... \\
\rho_e (f) J_1 J_2 \sin (\beta_1 x) \sin (\beta_2 x) \cos [(\omega_2 + \omega_1) t + (\phi_2 + \phi_1)].
\] (6.23)

The form of equation (6.16) is that of a low pass filter. The interaction from heat signal components with frequencies significantly above the 3 dB point of the thermal filter will have a negligible impact upon the distortion. The condition on frequencies that will contribute to the distortion negligibly is

\[
f >> \left| \frac{kA}{8\pi\rho_c V} \right|,
\] (6.24)

where \( f \) is the electrical frequency. The term \( \omega_2 - \omega_1 \) is very low frequency in comparison to the high frequency terms and is the only term in the expansion that can appreciably contribute to the distortion. The heat dissipated can be reduced to

\[
Q = \frac{1}{2} \rho_e (f) \left( J_1^2 \sin (\beta_1 x)^2 + J_2^2 \sin (\beta_2 x)^2 \right) + ... \\
\rho_e (f) J_1 J_2 \sin (\beta_1 x) \sin (\beta_2 x) \cos [(\omega_2 - \omega_1) t + (\phi_2 - \phi_1)],
\] (6.25)

which represents the average and periodic heat dissipated. In (6.16), the heat flux was integrated over the semi-infinite domain. The temperature inside the heat generation layer is the variable of interest in (6.17). The heat generation layer is defined by the skin depth in the conductor, (units of meters),

\[
\delta = \sqrt{\frac{\rho_e(T)}{\pi f \mu_0}}.
\] (6.26)

The periodic temperature, \( T_p \), over the skin depth \( \delta \) where \( z \) from (6.16) is approximately zero must necessarily be

\[
T_p (0, t) = \frac{2 \rho_e (1 + \alpha T_a) J_1 J_2 \sin (\beta_1 x) \sin (\beta_2 x) k^{-1} V A^{-1} \delta \cos (\omega_d t - \phi_d)}{1 + \sqrt{1 + 4 j \omega_d \rho_c k^{-1} V^2 A^{-2}}},
\] (6.27)
where $\omega_d = \omega_2 - \omega_1$, $\phi_d = \phi_2 - \phi_1$, and $r_1$ is defined in (6.11). The third order PIM generated can be determined by expansion of (6.19). The upper and lower products generated by each finite element are given by

$$E_{2\omega_1 - \omega_2} = \zeta_1 \cos [(2\omega_1 - \omega_2) t - (2\phi_1 - \phi_2)] \quad (6.28)$$

$$E_{2\omega_2 - \omega_1} = \zeta_2 \cos [(2\omega_2 - \omega_1) t - (2\phi_2 - \phi_1)] \quad (6.29)$$

$$\zeta_1 = \frac{2VA^{-1} \rho_e^2 (1 + \alpha T_o) J_1 J_2^2 \sin (\beta_1) \sin (\beta_2)^2 \alpha \delta k^{-1}}{1 + \sqrt{1 + 4j\omega \rho_e c k^{-1} V^2 A^{-2}}} \quad (6.30)$$

$$\zeta_2 = \frac{2VA^{-1} \rho_e^2 (1 + \alpha T_o) J_2 J_1^2 \sin (\beta_2) \sin (\beta_1)^2 \alpha \delta k^{-1}}{1 + \sqrt{1 + 4j\omega \rho_e c k^{-1} V^2 A^{-2}}} \quad (6.31)$$

As the current flows across the antenna during each cycle, the total field at any point on the antenna, referenced to the zero current edge, must be equal to the current density that has flowed through the resistance from the edge to that point. Thus the nonlinear field at any point inside the antenna is the summation of the nonlinear field along the direction of current flow from the zero current edge to the point of interest, defined by

$$PIM = \sum_{n=0}^{L} \zeta \cos (\omega_3 t - \phi_3). \quad (6.32)$$

where $\omega_3$ is the third order radian frequency, $\phi_3$ is the third order phase, and $\zeta$ is the defined in (6.30) and (6.31). The field of (6.32) is the x-component of the radial field due to the charge at a given point on the antenna. The field of any group of point charges is radially equal, thus the x-component and z-component of the field must be equal. The voltage over a microstrip rectangular patch antenna can be found by the line integral of the electric field of (6.32) to the ground plane. Radiated fields can be found using the radiating slots method and the vector potential defined by the third order current density [107].

### 6.2.3 Summary of Electro-Thermal PIM in Antennas

Small thermal variations in the conducting layer of a metal result in PIM generated along the direction of current flow on an antenna. The generated distortion is dependent
on antenna dimensions as well as the thermal properties of the metallization, as the heat preferentially transfers to the low current regions of the antenna. The current density distribution on the antenna combined with its conductive loss completely defines the distortion that is generated by the antenna. The prediction of summation of the nonlinear field across the structure results in a field distribution that couples into the dominate radiative mode of the antenna. In the next section, a manufacturing process for a rectangular patch antenna is described that isolates the electro-thermal conductivity modulation in the antenna metallization from other nonlinear mechanisms.

6.3 Isolation of Physical Electro-Thermal Process

Several physical mechanisms are thought to produce PIM in resonant structures such as antennas including ferromagnetic metals, tunneling, weak junction effects at feed point connections, and nonlinear conductor resistivity. In order to confirm or discount a particular physical mechanism as the responsible effect, the mechanisms must be isolated from each other. The mechanism of interest must be enhanced while every other effect is suppressed. The manufacturing process described in this section was designed to minimize or eliminate every physical mechanism save nonlinear conductor resistivity. Nonlinear resistivity was chosen as the physical mechanism to isolate due to its existence in every metal antenna. The manufacturing process developed to isolate the nonlinear conductivity specifically from electro-thermal processes for a microstrip rectangular patch antenna is presented in Section 6.3.1. The interface of the transmission line to a RF measurement system to allow high dynamic range measurement of the antenna is presented in Section 6.3.2.

6.3.1 Materials Design for Process Isolation

The manufacturing process described here minimized each of the physical mechanisms possibly responsible for PIM generation on a resonant structure, save electro-thermal conductor resistivity modulation. Ferromagnetic materials were avoided in both sample design, connectorization, and test equipment. Minimization of dielectric loss is accomplished by using a low loss substrate, single crystal sapphire, having a loss tangents of 0.00002. The 100 mm diameter substrate was epi-polished to < 5 angstroms surface roughness. Metal-
lization annealing, combined with the epi-polished substrate, effectively eliminates surface roughness contributions to conductivity. The lack of surface structures eliminates the possibility of tunneling within the antenna structure while the use of single metal inset fed antenna designs prevents feed point junction effects. The electro-thermal mechanism is enhanced through metallization thickness control and guaranteed to be the strongest nonlinear process in these samples.

The wafers were first washed with acetone and methanol to remove any film, then dried for five minutes at 500 °C. A seed metal layer with excellent adhesion such as chromium is normally used when sputtering other metals onto a substrate. In this experiment, a seed layer would form both a dissimilar metal-metal junction and result in a distributed ferromagnetic structure. To prevent this, the seed layer was avoided and silver was directly sputtered onto the substrate, 1.8 µm thick with a maximum 0.2 µm variation from wafer center to wafer edge. Adhesion was achieved by annealing the sample in air for 30 minutes at 500 °C. Photo-resist was applied, UV patterned, and used to etch the transmission lines with CR-7 etchant. The photo-resist was then removed and the wafer cleaned with acetone, methanol, and a deionized water rinse. The wafer was sputtered on the backside and annealed again at 500 °C for 30 minutes to provide a ground plane. The device was washed again and stored in a nitrogen atmosphere to prevent oxide or sulphide formation.

This procedure results in a single metal antenna with bottom side roughness equivalent to the smoothness of the epi-polished substrate. The bottom side smoothness is shown in the cross section of the antenna pictured in the SEM image of Fig. 6.2. Top side surface roughness is controlled by annealing the device to form extremely flat grains without grain merging, pictured in the SEM image of Fig. 6.3. Conductivity losses are directly controlled through the thickness of the metallization, enabling the enhancement of this effect to a degree. At a minimum thickness of 1.1 µm, grains begin to grow together during annealing resulting in large constriction resistances, pictured in the SEM image of Fig. 6.4. The antenna developed here was free of constriction resistances, shown in Fig. 6.3.

A silver on sapphire rectangular patch antenna with an inset feed line was designed and manufactured using this process. The sapphire substrate is an anisotropic, high permittivity material, 11.58 εr perpendicular to the surface and 9.3 εr parallel to the surface. An antenna on this substrate can be designed to be very thin, greatly enhancing conductive
loss. The antenna was designed using the procedure of [108] which is described in Section 6.4.1. The resulting designed was optimized in Ansoft HFSS 11 to achieve 42 dB return loss at the center frequency of 989 MHz including wirebonding contributions. The substrate is a circular 100 mm wafer, which set the minimum operation frequency allowable. The layout of the antenna is shown in Fig. 6.5. The feed inset design matches the antenna to a 50 Ω system without a matching network and without introducing a soldered feed point. The process described results in silver structures that are single metal and are directly wirebondable. All commonly suggested physical nonlinear mechanisms are minimized or eliminated in this process. Silver is the highest conductivity metal and is commonly used as plating in low PIM components. This manufacturing process generates the lowest possible PIM components obtainable when conductor thickness is not minimized, and represents the best achievable performance of a microstrip antenna. The interface of the antenna to a test system is discussed in the following section.
Figure 6.3: SEM image of the top side of the silver metallization at 1.8 um showing extremely flat silver with no merged grains.

Figure 6.4: SEM image of the top side view of the silver metallization at 1.1 um thickness showing merged grains leading to large constriction resistances.
6.3.2 Transmission Line Interface

The test sample was interfaced to the test system through a printed circuit board (PCB) mounting assembly, shown in Fig. 6.6. The mounting assembly is comprised of a rectangular patch antenna on 100 mm wafer mounted and wirebonded to a connectorized PCB board. The PCB designed for sample mounting and interface, shown in Fig. 6.7, used Rogers 6002 substrate material with > 5 µm of electroplated gold on all copper metallization. Interfacial metal layers such as chromium and nickel were avoided between the gold and copper to prevent ferromagnetic effects. The traces on the PCB board were coplanar waveguide at wirebonding interfaces to further reduce any current in the copper metallization. Silver conductive paint, 50% weight, was used to provide a connection from the ground plane of the sample to the PCB board. Additional wirebonding connections were designed to minimize cost associated with testing small wafer samples for process refinement. They are grounded for large sample measurement and result in no significant alteration of field patterns around the antenna.

The silver feed line for the substrate was thermosonically ball wirebonded to the PCB board coplanar line with five 25.4 µm diameter gold wirebonds at each port using a Kulicke and Soffa 4524AD gold wirebonder. Gold was chosen for wirebonding due to its lack of oxide formation and favorable alloy properties with silver. Gold and silver form an isomorphous alloy system due to their FCC lattice structure and comparable size resulting
Figure 6.6: Rectangular patch antenna assembly where a silver antenna on test substrate is mounted on a gold electroplated PCB board on Roger’s 6002 dielectric. Gold wirebonding is directly from the feed line, shown as the top silver layer, to the gold surface of the PCB board.

Figure 6.7: Printed circuit board designed on Roger’s 6002 dielectric for interfacing a 100 mm wafer. Coplanar waveguide transmission line connects the wirebond to a N-type connector. Six SMA connections, each possessing five jumper ready wirebonding pads, surround the wafer. Their presence was to lower the cost of testing many small samples for manufacturing process development and did not effect testing of the presented antenna.
in no metal-oxide-metal regions at the wirebonds and a continuous metal-metal contact system [101]. Tunneling and constriction resistance from wirebonding is virtually eliminated in this wirebonding arrangement.

The PCB board was connectorized with a custom made assembly from a Spinner ATL low PIM N-male to N-male cable, guaranteed to exhibit $-160$ dBc or less distortion at an input signal of two 44 dBm carriers. The helical copper encased cable was cut to dimension with a bandsaw. The connector was trimmed and the dielectric shaved from the center conductor before being cleaned with isopropyl alcohol. The helical copper ground conductor was split to connect to both the surface and bottom of the PCB both electrically and mechanically. The end launch connector was soldered to the PCB using 60/40 tin-lead solder. The resulting return loss for the fully mounted antenna is in excess of 42 dB at the center frequency. The complete test configuration has no ferromagnetic materials, metal-oxide-metal structures, feed point contacts, minimum dielectric loss, minimum surface roughness, and enhanced conductive loss.

6.3.3 Summary of Antenna Sample Design

Antennas manufactured with standard processes possess many physical mechanisms that are thought to produce distortion including nonlinear surface roughness, tunneling, feed point contacts, nonlinear conductivity, and ferromagnetic materials. Isolation of any one of these mechanisms requires careful manufacturing and process design to avoid surface roughness and material nonlinearities. A manufacturing process for single metal silver structures on sapphire substrates was presented that minimizes or isolates all nonlinear physical mechanisms save electro-thermal conductivity modulation. A rectangular patch antenna at approximately 989 MHz was manufactured using this process. Interface methods presented here allow the antenna to be tested using standard RF equipment without generating additional distortion even under high dynamic range requirements. The developed process represents the highest linearity microstrip patch antenna obtainable. In the next section, the design and electro-thermal characterization of this antenna is discussed along with several measurement methods specific to antennas.
6.4 Case Study: Microstrip Rectangular Patch Antenna

The electro-thermal distortion generated by an antenna is determined by the current density of the antenna. An electro-thermal analysis requires an analytically well defined current distribution over the antenna, which is generally only available for a few types of antennas. The antenna feed point, feed line, and any matching networks will introduce further undesirable sources of nonlinearity. In Section 6.4.1, an inset fed rectangular patch antenna is designed on a sapphire substrate to eliminate feed point nonlinearities. The feed line nonlinearities are not eliminated but can be accurately described by the results of the Chapter 5. An inset fed rectangular patch antenna on FR4 is also designed in Section 6.4.1 to allow transmission measurement of the manufactured silver on sapphire antenna.

The model of Section 6.2.2 was derived from heat conduction theory and the assumption of a sinusoidal current distribution over the antenna. Section 6.2.2 further asserted that (6.32) acts as an integration of the current distribution on the antenna providing the correct z-direction electric field at each point of the antenna due to the electric field symmetry of slow moving point charges. The validity of the current distribution and resulting electric field distributions of the electro-thermal model are compared against simulations in Ansoft HFSS 11 in Section 6.4.2.

In Chapter 3, the effects of matching on the cancellation measurement system were discussed. The antenna presents an unmatched load at various test frequencies to the system that is further compounded by resonant frequency shifting due to heating. The measurement system modification necessary to test this antenna over an appreciable bandwidth using a sweep of tone separation in a two-tone test is discussed in Section 6.4.3. The simulation results of the electro-thermal antenna model of Section 6.2.2 are compared against the measured electro-thermal dispersion from the manufactured silver on sapphire antenna, showing thermally induced dispersion in the measured distortion products from both the antenna element and the feed line.

6.4.1 Linear Inset Fed Microstrip Rectangular Patch Antenna Design

A rectangular patch antenna can be interfaced through a feed point where the input impedance of the antenna at that point is the characteristic impedance of the transmission line used to feed the antenna. Feed points are drilled in antennas on FR4 or other
similar substrates with the feeding transmission line soldered to the antenna at the feed point. This arrangement produces a solder drop which has been shown in [69] to generate distortion on transmission lines. A continuous feed arrangement can be produced by edge feeding a transmission line to the rectangular patch antenna to a location possessing the characteristic impedance of the transmission line. Due to the high impedance at the edge of the patch, a $\lambda/4$ impedance transformer or similar matching method is commonly used to match the antenna to the feed line. The matching network or impedance transformer can generate distortion before the antenna which cannot be de-embedded from the antenna distortion. An inset feed line provides a matching method that, unlike soldering a feed point, is continuous and provides a good match to the system impedance without a matching network or impedance transformer. The distortion generated by the inset feed line can be analyzed directly by the methods of Chapter 5, allowing de-embedding of the distortion from the antenna element and the feeding element.

An inset feed line rectangular patch antenna can be designed using the transmission line model of [108]. A brief overview of the method is given here. A practical width of the antenna can be found according to

$$W = \frac{1}{2f_r \sqrt{\mu_0 \epsilon_0}} \sqrt{\frac{2}{\epsilon_r + 1}}. \quad (6.33)$$

where $f_r$ is the radiation frequency, $\epsilon_0$ is the free space permittivity, $\mu_0$ is the free space permeability, and $\epsilon_r$ is the relative permittivity of the substrate. The effective dielectric constant of the antenna must be found to determine the length of the antenna. The effective dielectric constant, $\epsilon_{\text{reff}}$, when the width, $W$, of the antenna is much greater than the height, $h$, of the antenna is given by

$$\epsilon_{\text{reff}} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[ 1 + 12 \frac{h}{W} \right]^{-1/2}. \quad (6.34)$$

A practical approximate relation for the normalized extension of the length, $\Delta L$ due to fringing effects is given by

$$\Delta L = 0.412h \frac{(\epsilon_{\text{reff}} + 0.3) \left( 0.264 + \frac{W}{h} \right)}{(\epsilon_{\text{reff}} - 0.258) \left( 0.8 + \frac{W}{h} \right)}. \quad (6.35)$$

The actual length of the antenna is given by

$$L = \frac{1}{2f_r \sqrt{\epsilon_{\text{reff}} \sqrt{\mu_0 \epsilon_0}}} - 2\Delta L. \quad (6.36)$$
The inset distance, \( x_0 \), necessary to match the antenna to the system impedance was estimated here by a sinusoidal current distribution as

\[
x_0 = \frac{L \cdot \arccot (Z_0)}{\beta}.
\]

This design method gives the width, length and an estimate of the inset distance of the feed line necessary for a match to the system impedance but does not provide an estimate of the inset width. The inset width \( W_i \), shown in Fig. 6.1, is an important parameter for tuning the input reactance of the antenna. A capacitance is generated between the feed line and the antenna within the inset that is dependent on the inset width. Here the inset width was chosen to be ten times the size of the feed line width as a starting point, then successively tuned in Ansoft HFSS 11 to optimize the input reactance for maximum return loss.

This design method can be summarized as follows [108]:

- Determine the width of the antenna using (6.33).
- Determine the effective permittivity of the antenna using (6.34).
- Find the change in length due to fringing fields, given by (6.35).
- Obtain the actual length of the antenna through the use of (6.36).
- Estimate the feed inset distance assuming a sinusoidal current density (6.37).
- Estimate the inset width to be ten times the width of the line on each side of the feed line.
- Tune the inset depth to adjust the input impedance of the antenna.
- Tune the inset width to adjust the input reactance of the antenna.

Two antennas on different substrates, sapphire and FR4, were designed using this method. The FR4 antenna was designed in HFSS 11 to have 27 dB of return loss at a center frequency of 985 MHz, shown in Fig. 6.8(a). The actual center frequency of the antenna upon manufacturing was 1.006 GHz with a return loss of 28 dB. The permittivity of FR4 varies from batch to batch accounting for the 2.13 % shift in the center frequency of the antenna. The silver on sapphire antenna was designed in HFSS 11 to have a \(-0.2 j \Omega\)
Table 6.1: Antenna Dimensions

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Width</th>
<th>Length</th>
<th>Inset Width</th>
<th>Inset Length</th>
<th>Loading</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR4</td>
<td>100 mm</td>
<td>72 mm</td>
<td>16 mm</td>
<td>8 mm</td>
<td>—</td>
</tr>
<tr>
<td>FR4 Feed Line</td>
<td>100 mm</td>
<td>64 mm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>FR4, Ag Load</td>
<td>100 mm</td>
<td>72 mm</td>
<td>16 mm</td>
<td>6 mm</td>
<td>9 mm x 3 mm</td>
</tr>
<tr>
<td>Sapphire</td>
<td>60 mm</td>
<td>45 mm</td>
<td>5.5 mm</td>
<td>9 mm</td>
<td>—</td>
</tr>
<tr>
<td>Sapphire Feed Line</td>
<td>411 µm</td>
<td>27.5 mm</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Table 6.2: Antenna Performance

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Design Frequency</th>
<th>Manufactured Frequency</th>
<th>% Error</th>
<th>Return Loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR4</td>
<td>985 MHz</td>
<td>1.006 GHz</td>
<td>2.13 %</td>
<td>28 dB</td>
</tr>
<tr>
<td>FR4, Ag Load</td>
<td>985 MHz</td>
<td>989 MHz</td>
<td>0.4 %</td>
<td>26.5 dB</td>
</tr>
<tr>
<td>Sapphire</td>
<td>985 MHz</td>
<td>989 MHz</td>
<td>0.4 %</td>
<td>42 dB</td>
</tr>
</tbody>
</table>

reactance to counteract the anticipated bond wire inductance from 5 parallel 1 mm long 1 mil diameter gold bond wires. Simulation of this antenna predicted a 34 dB return loss at a center frequency of 985 MHz, shown in Fig. 6.8(b). The center frequency of this antenna was actually 989 MHz, representing a 0.4 % shift in the center frequency that could be due to the bond wire compensation or the anisotropic nature of the sapphire permittivity. The large center frequency shift of the FR4 antenna compared to the silver on sapphire antenna made them incompatible for a transmission test. The FR4 antenna was shifted down in frequency using silver paint, 50% weight, to paint capacitive loading on each feed side edge of the antenna and to change the inset distance to rematch the antenna. The dimensions of both antennas are detailed in Table 6.1 and their performance is detailed in Table 6.2.

6.4.2 Antenna Electric Fields and Surface Currents

The derivation of electro-thermal PIM in Section 6.2.2 was based on an assumed sinusoidal current distribution. Analytically the current distribution over a complex structure is difficult to obtain and often not tractable for arbitrary structures. Computational electromagnetics allows current and electric field distributions to be obtained for such ar-
Figure 6.8: Ansoft HFSS 11 $S_{11}$ simulation and vector network analyzer $S_{11}$ measurement for: (a) The rectangular patch antenna on FR4. The $S_{11}$ measurement of the rectangular patch with silver paint loading shows the shift of the center frequency to the center frequency of the sapphire substrate antenna with only 1.5 dB reduction in return loss, (b) and the rectangular patch antenna on sapphire showing a return loss of 42 dB.
bitrary structures. In this section, Ansoft HFSS 11 is used to simulate the surface current distribution over the rectangular patch antenna structure to validate the assumed current distribution of the electro-thermal model. The electric field distribution at resonance is compared to the nonlinear electric field predicted by the electro-thermal model, demonstrating the nonlinear field coupling into radiating antenna modes.

The surface current on a rectangular patch antenna can be broken into two chief components, the top side antenna current and the bottom side antenna current. The top side antenna current is shown in Fig. 6.9(a) for a silver on sapphire antenna simulated in Ansoft HFSS 11 at 1 Watt input power. The surface current is approximately $0.1 \text{ A} \cdot \text{m}^{-1}$ on the top side of the antenna, increasing exponentially at the radiating slots to approximately $300 \text{ A} \cdot \text{m}^{-1}$ due to some of the charge being forced from the bottom to the top of the antenna. The surface current is approximately $250 \text{ A} \cdot \text{m}^{-1}$ on the bottom side of the antenna, shown in Fig. 6.9(b). The current on the bottom side decreases nearly sinusoidally to approximately $0.5 \text{ A} \cdot \text{m}^{-1}$. The majority of the current, and thus heat, follows a near sinusoidal distribution over the silver on sapphire antenna validating the assumption of a sinusoidal current distribution for the electro-thermal patch antenna model. Simulation suggests that the edge of the antenna generates little heat compared to the middle of the antenna, causing the heat to flow toward the edge of the metallization.

The surface current magnitude of a sinusoidal distribution should be near two times that of an equivalent 50 Ω transmission line on the same substrate. The magnitude of the current density is compared to a 50 Ω transmission line on the same sapphire substrate as the rectangular patch antenna in Fig. 6.10. The surface current due to bunching at the edges of the transmission line is approximately $125 \text{ A} \cdot \text{m}^{-1}$, which is half that of the bottom side of the silver on sapphire rectangular patch antenna. A sinusoidal standing wave relative to the feeding transmission line is a valid approximation for the silver on sapphire antenna.

The electro-thermal model was predicted in Section 6.2.2 to give radiating electric field modes due to the summation of the field across the antenna. This assertion can be verified by comparison of the linear electric field at resonance with the predicted nonlinear field from the electro-thermal model. In Fig. 6.11(a), the linear vector electric field for a single tone on the silver on sapphire antenna structure is shown for a 1 Watt input power. The substrate is square rather than cylindrical for graphics rendering purposes, but still represents the silver on sapphire antenna. The electric field distribution is near that of a
Figure 6.9: Ansoft HFSS 11 simulated current distributions over a silver on sapphire antenna at 989 MHz: (a) The surface current distribution for the top side of the antenna showing no current in the middle of the antenna and exponentially increasing current at the radiating edges, (b) and the surface current distribution for the bottom side of the antenna showing a near sinusoidal current distribution that is approximately twice the magnitude of the current on a 50 Ω transmission line on the same substrate.
Figure 6.10: Surface current over a matched 50 Ω microstrip 411 µm wide transmission line on a sapphire substrate simulated in Ansoft HFSS 11.

cosine function, and has a magnitude equivalent to 10 V at the antenna edges for a 500 µm substrate thickness. The electro-thermal model predicts a field distribution near that of a cosine function, shown in Fig. 6.11(b) for a 33 dBm two-tone input at 10 Hz tone separation. The magnitude of the field is equivalent to 850 µV or $-51.5$ dBm referenced to the ground plane. The field predicted in Fig. 6.11(b) represents the potential difference across the antenna, where the actual zero potential reference would be the center of the antenna at 22.5 mm. The model parameters and implementation are discussed in Section 6.4.3.

The rectangular patch antenna on FR4 follows similar surface current patterns to the silver on sapphire antenna, but as not as well defined. The top side antenna current is shown in Fig. 6.12(a) for the FR4 antenna simulated in Ansoft HFSS 11 at 1 Watt input power. The surface current is approximately 0.1 A·m$^{-1}$ on the top side of the antenna, increasing exponentially at the radiating slots to approximately 10 A·m$^{-1}$ due to some of the charge being forced from the bottom to the top of the antenna. The surface current is approximately 10 A·m$^{-1}$ on the bottom side of the antenna, shown in Fig. 6.12(b). The current on the bottom side decreases almost sinusoidally to approximately 0.5 A·m$^{-1}$, but is not nearly as well defined as the silver on sapphire antenna. Simulation suggests
Figure 6.11: Simulated electric field distribution over a silver on sapphire antenna at 989 MHz: (a) Ansoft HFSS 11 simulated linear vector electric field distributions showing a near cosine function distribution of electric field over the antenna, (b) and electro-thermal model simulated third order electric field distributions showing a near cosine function distribution of electric field over the antenna.
that the edge of the antenna generates little heat compared to the middle of the antenna, causing the heat to flow toward the edge of the metallization. In this case the electro-thermal model would need to be augmented with the correct current distribution in order to provide accurate distortion product prediction.

The electric field generated by the summation of the field across the antenna need to be partitioned into smaller problems to accurately model the fields produced by asymmetric current distributions. In Fig. 6.13, the linear vector electric field for a single tone of the FR4 antenna structure is shown for a 1 Watt input power. The electric field pattern is slightly irregular and asymmetric compared to the silver on sapphire antenna. The electro-thermal model will not predict this behavior from an analytic implementation. Accurate modeling under irregular field and current patterns requires the model to be implemented as a finite element method.

The electro-thermal model assumption of a sinusoidal current distribution has been compared to 3D electromagnetic simulation in this section. The assumption compares favorably for high permittivity substrates and is less accurate for low permittivity substrates. The electro-thermal nonlinear electric field distribution over the antenna was shown to agree with allowed radiated modes for frequency spacings within the antenna bandwidth. In the next section, the electro-thermal simulation model is detailed and compared against measured electro-thermal distortion.

6.4.3 Antenna Measurement

Electro-thermal distortion has been shown in [65] to exhibit a PIM dispersion characteristic which is non-integer Laplacian and is dictated by the thermal parameters of the material. The current distribution and predicted nonlinear field of the electro-thermal model were discussed in the previous section under the assumption that the distortion was electro-thermal. The electro-thermal nature of the distortion for the silver on sapphire rectangular patch antenna can be validated by sweeping the separation frequency of a two-tone signal within the resonant bandwidth of the antenna. A transmission measurement is first detailed to verify the radiation of the distortion from the antenna. The transmission measurement could not be conducted over an appreciable bandwidth due the rapid change in resonant frequency of the antenna with applied two-tone signal and the reduction in
Figure 6.12: Ansoft HFSS 11 simulated current distributions over a FR4 antenna at 989 MHz: (a) The surface current distribution for the top side of the antenna showing no current in the middle of the antenna and exponentially increasing current at the radiating edges, (b) and the surface current distribution for the bottom side of the antenna showing a semi-sinusoidal current distribution.
dynamic range due to the resulting unmatched load. A reflection measurement is discussed that alleviates the dynamic range reduction by using the return loss of the antenna to reduce the probe stimulus without effecting the distortion. The section is concluded with an explicit definition of the electro-thermal simulation model for the silver on sapphire antenna and comparison of the model with a two-tone reflection measurement of the antenna.

**Antenna Transmission Measurement**

The electro-thermal model of Section 6.2.2 predicts a field distribution in the dominant radiating mode of the silver on sapphire rectangular patch antenna. A transmission measurement of a two-tone sweep would verify the electro-thermal distortion is indeed radiating from the antenna as predicted by the electro-thermal model. A separation sweep of an applied two-tone signal will produce a varying amplitude modulation frequency which will follow a thermal dispersion trend if the distortion is electro-thermal. In this section, a two-tone test signal is applied to the silver on sapphire antenna that shifts the resonant frequency of the antenna due to periodic heating. The effects of this phenomenon on the
measurement system of Chapter 3 are discussed.

An antenna transmission measurement requires two antennas, one in transmit and one in receive mode. The transmitting antenna will have a much higher power density than the corresponding receive antenna. The linearity of the receive antenna need not be considered if the transmission loss is enough to drop the signal level to a few milliwatts, which can be easily done by varying the distance between the two antennas. The distortion generated, if the test system is linear, will be from the transmitting antenna only. Two rectangular patch antennas were developed in this work to allow a transmission measurement. Both antennas are resonant at 989 MHz and have similar radiation patterns at this frequency. The silver on sapphire antenna radiation pattern is shown in Fig. 6.14(a) and the FR4 antenna radiation pattern is shown in Fig. 6.14(b). The gain is approximately equal for both antennas at $-10$ dB, which corresponds well with the measured 25 dB transmission loss at a 1 m separation between the antennas. The antennas are placed inside an anechoic chamber and spaced to give 41 dB of transmission loss, as shown in Fig. ???. The DUT in the measurement system of Fig. 3.2 is replaced by Fig. ?? for transmission measurement.

The silver on sapphire antenna was observed to rapidly shift center frequency upon applying a two-tone signal. The center frequency returned to the original center frequency upon cessation of the two-tone signal and did not shift for a single tone input. Due to this phenomenon, the DUT in the cancellation system of Fig. 3.2 becomes unmatched quickly, even with a return loss of 42 dB. The dynamic range of the measurement system decreases in a one to one relationship with the decrease in isolation when the return loss of the antenna drops below the isolation of the hybrid combiner, which was 31 dB at the test frequency of 989 MHz. At a total measurement bandwidth of 1 MHz, which corresponds to a tone spacing of 200 Khz, the system dynamic range for this test would be approximately 90 dBC. The rapid shift of the center frequency further reduces the isolation such that the system dynamic range reduces to 85 dBC at 1 MHz test bandwidth. The test bandwidth for transmission measurement due to the limited bandwidth of the silver on sapphire antenna and the rapid center frequency shift due to heating was confined to 10 Hz to ensure at least 100 dBC dynamic range over the measurement bandwidth. The next section discusses a reflection measurement method that extends the dynamic range for measurement of PIM on antennas. The results of both transmission and reflection measurements are discussed in their own subsection following the reflection measurement method.
Figure 6.14: Ansoft HFSS 11 simulated radiation pattern for: (a) a silver on sapphire rectangular patch antenna at 989 MHz, (b) and a FR4 rectangular patch antenna at 989 MHz.
Antenna Reflection Measurement

The transmission measurement of the previous section was severely hampered by the reverse waves created by the decreasing antenna return loss. Antenna electro-thermal dispersion could not be measured over an appreciable bandwidth due to this measurement concern. The fact that reverse waves have such an impact on measurement system linearity gives a hint as to how to overcome the limitation. Distortion sources from passive devices are localized to that structure and produce both forward and reverse distortion relative to the propagation mode, as shown in Chapter 5 for transmission lines. An antenna does not produce forward and reverse distortion, but rather a standing wave of distortion. If the distortion source couples into not only the radiating mode of the antenna but also the matched input port of the antenna, an increase in the dynamic range equivalent to the return loss of the antenna can be obtained. The applied stimulus, assumed to be within the antenna resonant bandwidth, will be radiated away from the antenna and reflected back into the system at a power reduction equivalent to the antenna return loss while the distortion source will couple into the antenna input port at full strength. Any forward traveling distortion products will experience the same directional filtering, allowing separation of antenna distortion and system distortion over a wider bandwidth than in the transmission measurement case.

Due to the dynamic range enhancement obtainable from the antenna return loss, the measurement system of Chapter 3 was not used for this measurement. The system was linearized to the maximum extent and effective cancellation was obtained from directional
filtering due to the antenna return loss. Extra isolation was added in the form of two additional isolators over the original two per channel to counteract the decreasing isolation between signal channels from reflected waves. After channel linearization, all available isolators at the test frequency of 989 MHz were used, eliminating the feed forward cancellation for further dynamic range enhancement. The isolators begun to generate PIM due to increased reflected waves, as described in Section 3.4.2. Transmission lines on FR4 were used as 3.5 dB attenuators to attenuate the reverse wave from the opposite channel and the desired forward wave, resulting in a 2 dB gain in dynamic range per 1 dB attenuation due to the decreased signal power in the isolators. The dynamic range of the transmit part of the system was increased to 115 dBc at 33 dBm output power per tone using this method.

The modified measurement architecture is shown in Fig. 6.16. Each channel is amplified and then isolated by four isolators per channel. The output of the final isolator is further isolated through a 3.5 dB low PIM transmission line attenuator on FR4 and 0.4 dB of loss through the 3 m cables transporting the signals into the anechoic chamber. The signals are combined in a hybrid combiner where one port is terminated in a low PIM cable termination and the other port is connected to the transmitting antenna. The insertion loss of the combiner above 3 dB is 0.3 dB. The reflected signal is coupled off by a directional coupler at 20 dB down from the reverse signal. An additional 22 dB of attenuation is placed between the directional coupler and the vector signal analyzer. The dynamic range of this measurement setup changes over the measurement bandwidth with antenna return loss, which was increased using microwave absorber loading to approximately 50 dB. At 1 Hz tone separation the dynamic range exceeds 125 dBc for a two-tone 33 dBm stimulus, while at 100 KHz tone separation the dynamic range has been reduced to 103 dBc. This measurement method is used to characterize electro-thermal distortion from the silver on sapphire rectangular patch antenna at up to 100 KHz tone separations in the following section.

**Electro-Thermal Dispersion**

In the previous two sections the measurement methods used to perform a general distortion measurement on an antenna, including transmission and reflection measurements, were discussed but not applied to the manufactured silver on sapphire antenna. These
measurements are detailed in this section as applied to the rectangular patch antenna. The electro-thermal model simulation parameters are defined and the results of the defined model are compared against the dispersion characteristics from the transmission and reflection measurements.

The electro-thermal model for the rectangular patch antenna on sapphire can be determined from material parameters alone. In the analysis of transmission lines in Chapter 5, all the heat traveled into the substrate due to the almost even heating of the metallization. The rectangular patch antenna, due to the standing wave behavior, has large regions of unheated or slightly heated metallization. This metallization provides a thermal conductivity route that is over two orders of magnitude higher than substrate. The heat will flow towards the ends of the antenna requiring the use of metallization thermal parameters rather than the substrate thermal parameters for the antenna element. Table 6.3 details the thermal parameters necessary to simulate the manufactured rectangular patch antenna on a sapphire substrate and its associated feed line.

In both the transmission and reflection measurements, the silver on sapphire rectangular patch antenna was probed with a two-tone 32.3 dBm input power signal. The
transmission measurement was conducted at only two tone separation frequencies, 4 Hz and 10 Hz, at a center frequency of 989 MHz. The measurement could not be conducted outside of 10 Hz tone separation due to previously discussed dynamic range limitations from reverse wave interaction. Further transmission measurements were avoided due to the observation that if the two-tone signal was applied longer than a few seconds the antenna center frequency shifted irreversibly.

The reflection measurement was conducted from 3 Hz to 100 KHz tone separations at a center frequency of 989 MHz. The increased dynamic range of this method allowed a much shorter record time to resolve the distortion products when acquiring frequency spectrum data. The heating induced shift in antenna resonant frequency occurred slower than the decreased data acquisition time for most measurements. At frequency separations below 10 Hz spectrum averaging was not used to reduce record time due to the alteration of frequency content once resonant frequency shifting begins. The sources for this measurement were cut on and off directly with the vector signal analyzer to minimize resonant frequency shift. No further irreversible frequency shifting occurred using this scheme.

The distortion measured using both transmission and reflection measurement, shown in Fig. 6.17, agree with each other over the limited subset of points below 10 Hz. The antenna simulation using the electro-thermal model also predicts distortion in good agreement with the measured results with only the metallization thermal parameters and dimensions. The reflection measurement distortion includes not only a dispersion characteristic from the antenna but also a dispersion characteristic from the feed line. The transmission line model of Chapter 5 was used to simulate the distortion created over the feed line, also shown in Fig. 6.17. The simulated feed line distortion agrees well with the measured distortion from the antenna out to 100 KHz, where dynamic range again limits measurement. The measured distortion for both the antenna element and the feed line bears the characteristic signature of electro-thermal dispersion, which is as high as 79 dBC.

### Table 6.3: Ag on Sapphire Antenna Simulation Parameters

<table>
<thead>
<tr>
<th>Device</th>
<th>$k$</th>
<th>$c_v$</th>
<th>$\rho$</th>
<th>$V/A$</th>
<th>$J$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patch</td>
<td>429 W·m$^{-1}$K$^{-1}$</td>
<td>232 J·Kg$^{-1}$K$^{-1}$</td>
<td>10490 Kg·m$^{-3}$</td>
<td>0.06 m</td>
<td>0.8 GA·m$^{-2}$</td>
</tr>
<tr>
<td>Feed</td>
<td>25 W·m$^{-1}$K$^{-1}$</td>
<td>740 J·Kg$^{-1}$K$^{-1}$</td>
<td>3980 Kg·m$^{-3}$</td>
<td>70 $\mu$m</td>
<td>0.4 GA·m$^{-2}$</td>
</tr>
</tbody>
</table>
Figure 6.17: A silver on sapphire rectangular patch antenna is two-tone tested in both a transmission and reflection configuration at 32.3 dBm input power. The frequency separation is swept from 3 Hz to 100 KHz for the reflection measurement and is conducted at 4 Hz and 10 Hz tone separation for the transmission measurement. Both measurements are compared against electro-thermal model simulation results for both the antenna element and the feed line.

from the carrier at 32.3 dBm input power. The measurements and model of this section clearly show that both the antenna and feed line or feed point are PIM producers that both radiate out of the antenna and couple back into the system through the input port.

6.4.4 Summary of Rectangular Patch Antenna

Distortion from antennas has been attributed to many physical mechanisms due to the difficulty in isolating single mechanisms and the lack of a defining characteristic for the distortion generated from each mechanism. Electro-thermal conductivity modulation was isolated and enhanced to be the dominant physical mechanism through careful design and manufacturing enabled by the presented electro-thermal theory. Two-tones sweeps of the tone separation to modify the frequency of amplitude modulation were applied to the antenna, generating a non-integer Laplacian dispersion characteristic unique to electro-
thermal distortion for both the antenna element and the associated feed line. Transmission measurements and reflection measurements were performed showing electro-thermal distortion both radiates from the antenna and couples back into the system through the input port. The reflection measurement presented used the return loss of the antenna to greatly enhance the dynamic range of the measurement. The increase in dynamic range facilitated resolution of distortion products before heat induced resonant frequency shifts could take place. The dispersion characteristic matched the presented resonant electro-thermal PIM theory well for the presented silver on sapphire rectangular patch antenna in both antenna element distortion and feed line distortion.

6.5 Conclusion

In this chapter, electro-thermal conductivity modulation was suggested as a dominant mechanism producing PIM in resonant structures. Thermal effects were overlooked in previous research as the relaxation times between thermal and electrical systems are offset by several orders of magnitude. Amplitude modulation produces power components of the electrical signal within the thermal relaxation time and was pursued as a PIM contributor due to this coupling. A theoretical analysis of heat conduction and electro-thermal interaction was presented resulting in an expression for electro-thermal PIM based on current density distribution over the resonant structure.

The presented theoretical analysis enabled a silver on sapphire antenna to be built isolating electro-thermal conductivity modulation. The single metal silver rectangular patch antenna was manufactured on an epi-polished sapphire substrate, resulting in surface roughness on the order of two atoms. Contact to the antenna was made with a silver and gold isomorphous alloy system that creates a continuous transition between the metals, alleviating constriction resistances and contact nonlinearities. Conductor thickness was controlled through sputtering, allowing conductive losses to be enhanced. The high permittivity substrate confined the fields tightly to the antenna, further enhancing current density and providing a uniform and almost sinusoidal current distribution for antenna analysis.

Current distributions validating the use of the electro-thermal model were confirmed by 3D electromagnetic simulation using Ansoft HFSS 11. The nonlinear electric
field generated by the electro-thermal model was compared to the dominate linear resonant mode of the antenna with excellent agreement. Transmission measurements using the high dynamic range measurement system of Chapter 3 and an anechoic chamber were performed to confirm the radiation of antenna generated distortion. Reflection measurements enhancing the dynamic range of the test by using the filtering properties of the antenna return loss were performed extending the measurement bandwidth from 10 Hz in the transmission measurement to over 100 KHz. Electro-thermal PIM was shown to both radiate from the antenna and couple back into the system at the antenna input port.

Electro-thermal theory and measurement, both in good agreement, were used to analyze the effect of material properties and dimensions on rectangular patch antenna linearity. Antenna distortion, similar to transmission line distortion, is defined completely by the current distribution and loss of the metallization. Thermal properties of the metallization were shown to be dominant over substrate thermal properties for thin metallization, and are directly linked to antenna dimensions and thus frequency. Low PIM microstrip antennas can be manufactured by increasing metallization thickness and dimensions through low permittivity, low loss substrates and avoiding dissimilar or high contact resistance feed point connections. Analysis of any design can be conducted using the presented theory, giving simulation capabilities for predicting PIM in resonant structures.
7

Conclusion
7.1 Summary of Research and Original Contributions

This dissertation has shown that electro-thermal conductivity modulation is a dominant passive intermodulation source that is inherent in every metallic microwave component. Electro-thermal conductivity modulation in lumped microwave components, transmission lines, and antennas were shown to generate distortion at levels large enough to easily impact sensitive communications equipment such as satellites and cellular base stations. Analytic models accounting for non-integer order Laplacian behavior in the electrical domain due to electro-thermal conductivity modulation were presented for microwave lumped elements, distributed elements, and resonant structures. A high dynamic range measurement system was developed allowing measurement of thermal dispersion in the passive intermodulation, uniquely defining electro-thermal conductivity modulation as a dominant PIM contributor for the first time.

A high dynamic range measurement system based on feed-forward cancellation was developed that is both automated and 40 dB higher dynamic range than the existing state of the art measurement system. A formula predicting the phase shift necessary for exact cancellation was developed based on only two amplitude measurements using a standard spectrum analyzer. This method both enhances the maximum achievable cancellation by 15-20 dB over previously published results and reduces the number of iterations required to achieve cancellation by over 80%. The system is linearized to 113 dBc dynamic range in the feed-forward and DUT test paths through characterization of forward and reverse traveling waves in the system and exhaustive test of test component linearity.

Thermal effects have been previously overlooked as a dominant passive intermodulation contributor as the relaxation times between thermal and electrical systems are offset by several orders of magnitude. Electro-thermal coupling was shown to occur when amplitude modulation exists in the electrical signal due to low frequency components in the signal power envelope. The coupling resulted in long-tail transient behavior and fractional order frequency dispersion that can not be explained with integer order differential equations. A fractional calculus based description of the thermal environment was developed that allowed the non-integer order Laplacian behavior to be described. The fractional calculus solution of the heat conduction equation allowed the distortion from the electro-thermal process to be derived analytically for lumped microwave elements including integrated circuit resistors,
microwave attenuators, and microwave terminations. A simulation model was developed based on the fractional derivative which accurately reproduces both the transient and frequency response of a lossy component, along with an approximate circuit model that can be used in standard circuit simulators. High dynamic range measurements were conducted on integrated circuit resistors, polysilicon, microwave terminators, and microwave attenuators that exhibited electro-thermal dispersion that matched the presented theory well.

Transmission lines are known to generate passive intermodulation distortion that grows in magnitude with the length of the line. Many physical mechanisms have been suggested as a probable cause including tunneling, ferromagnetic materials, surface roughness, and conductivity modulation. Once again, thermal effects were overlooked as a dominant passive intermodulation contributor due to the difference in relaxation times between thermal and electrical processes. It was shown here that the metal conductivity in a transmission line is a temperature dependent parameter which experiences dynamic modulation when amplitude modulation exists in the propagating signal. Thermal conductivity modulation occurs down the length of the line generating distortion summing coherently in the propagation direction and destructively after a quarter wavelength in the reverse direction. Several transmission lines were manufactured to isolate all known possible passive intermodulation mechanisms save electro-thermal conductivity modulation. High dynamic range measurements were conducted on the transmission lines showing thermal dispersion trends in the measured passive intermodulation that confirmed electro-thermal conductivity modulation as the source of the distortion and agreed well with the presented electro-thermal theory.

Resonant structures such as antennas are also known to generate passive intermodulation distortion, which has been largely attributed to electron tunneling. Antennas with aluminum contact have been shown to exhibit tunneling nonlinearities in the literature. Antennas not made from aluminum with no metal-metal junctions have also been shown to produce passive intermodulation, but only heuristic models from near field scans exist for description of passive intermodulation distortion on these antennas. It was theoretically shown here that electro-thermal conductivity modulation will occur on the metallization of the antenna, producing distortion based on the conductivity of the metal, the current distribution over the antenna, and the size of the antenna. A rectangular patch antenna was manufactured to isolate the electro-thermal effect from other possible nonlinear mechanisms. Electromagnetic simulation was performed to validate the current distributions
assumed in the electro-thermal model and validate the nonlinear electric field pattern was in a radiating mode. Transmission and reflection measurements were conducted showing that electro-thermal distortion is indeed produced over the antenna element and the feed line, both of which radiate from the antenna and couple back into the system through the input port of the antenna. The developed electro-thermal theory predicts this behavior and agrees well with both transmission and reflection measurements.

The electro-thermal theory in this dissertation shows for the first time electro-thermal conductivity modulation is a dominant nonlinearity at microwave frequencies. The simulation tools and analytic formulations necessary to describe electro-thermal passive intermodulation distortion in any lossy, non-ferromagnetic microwave element are presented, including lumped, distributed, and resonant structures. A measurement system and measurement methods are developed which are specifically designed to characterize electro-thermal distortion.

7.2 Future Research

Several different research opportunities exist as a direct result of this work in thermal control, measurement, and passive intermodulation distortion. The suggested opportunities are given based on the chapter of their introduction.

Chapter 3 discussed a high dynamic range measurement system based on feed-forward cancellation. A natural extension of this technique is co-site suppression in multi-antenna systems. Co-site interference occurs from the finite isolation between co-located receive and transmit antennas. The bridge technique applied in the measurement system of Chapter 3 can be applied to suppressing the co-site electric field over the co-located receive antenna, greatly increasing isolation between co-located antennas. Chapter 3 also discussed the behavior of a LNA driven into limiting as a direct conversion mixer. RF front end design employing this effect could allow reduction in circuit complexity.

Chapter 4 discussed electro-thermal distortion in lumped microwave elements. The developed electro-thermal model is powerful and accurate for describing both thermal and electro-thermal behavior, but is difficult to implement in standard circuit simulators due to the inherent fractional derivative. Methods to implement or approximate the fractional
derivative would allow its widespread use. The electro-thermal measurement of integrated circuit materials in Chapter 4 showed a method to force the heat generated in a circuit preferentially through a given material or direction. Further research in this area would allow heat to be channeled away from critical circuit elements allowing higher circuit density and operation frequency.

Chapter 5 discussed electro-thermal distortion in transmission lines. The effects of matching and various line configurations such as striplines and waveguides were not explored. Any electromagnetic guiding structure will possess electro-thermal distortion, leaving many guiding structures to be analyzed. Substrate effects and the result of increased surface roughness will also impact transmission line distortion, leaving a wealth of research to perform.

Chapter 6 discussed electro-thermal effects on a microstrip antenna, but did not explore constriction resistances at feed point contacts. Constriction resistances are a dominant PIM mechanism in connectors and would be a likely contributor that needs to be considered in antennas with non-continuous feed points. Many other types of antennas exist each possessing their own thermal conduction configuration and current distribution. Each type of antenna must be separately analyzed to determine the distortion it generates.
Bibliography


Appendices
A

Electromagnetic and Acoustic

Anechoic Chamber
A.1 Introduction

Anechoic chambers can provide isolation from the environment that is vital to high dynamic range testing of strongly radiating microwave elements such as antennas and beneficial to weak radiators such as transmission lines on microstrip. Acoustic signals are generally not considered in the testing of radiative microwave elements. However, microwave substrates and conductors are not immune to acoustic interference. Acoustic interference provides another potential signal port that can interact with high power electrical signals. It is desirable to eliminate such interference when trying to isolate a single effect such as electro-thermal conductivity modulation. In order to measure acoustic, electromagnetic, and acoustically modulated electromagnetic phenomenon, an anechoic chamber attenuating both acoustic and RF energy was built. The chamber is a modular design measuring 8 feet in width, 6 feet in height, and 12 feet in length. Measurements show that the chamber provides significant insertion and return loss as well as dynamic ranges of 114 dB and 130 dB for acoustic and electromagnetic signals as compared to the local environment. This appendix details the construction of the dual electromagnetic and acoustic anechoic chamber that allows high dynamic range characterization of microwave nonlinearities.

A.2 Anechoic Chamber Construction

Overall external chamber dimensions of the chamber, shown in Fig. A.1, are 96 inches in width, 72 inches in height, and 144 inches in length. Due to the thickness of absorbent materials used in construction, the internal usable dimensions are 76 inches in width, 52 inches in height, and 120 inches in length. Chamber construction began by building a raised floor to provide wire runs to various test equipment located within the chamber. The floor measures 109 inches in width by 156 inches in length and rests on top of 9 4 in. x 4 in. posts. Starting from the lower most layer, the floor is comprised of one layer 3/4 in. plywood, one layer cement board, one layer copper mesh, and two layers of 6.0 mm thick Acoustiblok.

The walls and ceiling of the anechoic chamber of Fig. A.1 are constructed using the same process. Support for the walls is via an extruded aluminum space frame manufactured by 80/20 Inc. The outermost layer of each wall and ceiling panel is comprised of copper
Figure A.1: Exploded view of chamber frame and floor. Magnified cross section shows wall construction detail. The Acoustiblok and copper mesh are held in place with a 0.25 x 1.5 in. nylon flat head bolts and nuts attaching these layers to each angle bracket. The Melamine/Quiet Board panels are held in place with 0.25 x 4 in. nylon carriage bolts. Finally, RF tiles are glued to the Quiet Board using contact cement.
mesh, manufactured by TWP Inc., forming a faraday cage around the entire chamber. The square gaps in the copper mesh are 1/4 x 1/4 mm. The faraday cage provides significant isolation from the environment to approximately 5 GHz.

Forming the foundation of each wall and ceiling panel is a layer of 3.0 mm thick Acoustiblok. This high-density rubberized material provides almost 2/3 of the throughwall attenuation above 1 kHz. The Acoustiblok sandwiches the copper mesh against the supporting frame and is held in place with 1.5 in. nylon bolts spaced 4 in. apart. At every seam in the Acoustiblok, acoustical sealant and tape is used to further improve soundproofing.

Attached to the inside surface of the Acoustiblok are 2 ft by 4 ft panels of Quiet-Board glued to Melamine foam, both manufactured by American Micro Industries. Each panel is attached by six nylon bolts and forms the surface to which the RF absorbing foam tiles are glued using contact cement. At every point where these panels are attached, Acoustiblok sealant is again used to ensure soundproofing.

The innermost layer consists of carbon impregnated electromagnetic absorbing tiles from emerson and cuming. Two types of this tile were donated to the research program, each measuring 2 ft square. The pyramidal style tile was used in the most sensitive areas such as along the back wall where the most intense energy will accumulate, as its absorption characteristics are much better at low frequencies. Eggshell tiles were used to fill in where there were not enough of the previous type. All of the tiles were arranged in a manner that reduces the possibility of generating standing waves and attenuates propagating signals the most at the lowest reflection paths to the device under test.

**A.3 Anechoic Chamber Acoustic Characterization**

To evaluate the acoustic performance of the anechoic chamber, experiments were conducted to measure the chambers insertion and return loss. If the wall were modeled as a two-port network, this would correspond to values for $S_{21}$ and $S_{11}$ respectively.

Transmission measurements were performed at both low frequencies from 0.1 to 20 kHz, as well as at ultrasonic frequencies from 50 to 70 kHz. Fig. A.2 shows the acoustic insertion loss is approximately 90 dB above 10 khz linearly derating to 70 dB at 2 khz. The acoustic insertion loss drops off almost exponentially to 20 dB from 2 khz to 100 hz. Low
frequency signals were generated using a PXI-4461 DAQ and transmitted with an Event Electronics TR-8 Studio Monitor. High frequency signals were generated with a Marconi 2024 signal generator and transmitted using an 18 in. Audio Spotlight transducer/amplifier combination customized to accept ultrasonic input. Data was recorded using a PXI-5922 high-speed digitizer connected to PCB Piezotronic condenser microphones. For transmission measurements, one microphone inside the chamber recorded the incident sound pressure level, while another outside the chamber recorded the transmitted sound pressure level. Insertion loss in decibels was calculated using

\[ L_i = 20 \cdot \log_{10} \left( \frac{P_t}{P_i} \right), \]  

where \( P_t \) is the transmitted pressure amplitude and \( P_i \) is the incident pressure amplitude.

The lowest possible sound pressure amplitude that can be measured with our microphones is 6 dB SPL (sound pressure level). This noise floor was calculated using

\[ X_k = \frac{2}{N} \sum_{n=0}^{N-1} X_n \cdot e^{-\frac{2\pi i kn}{N}}. \]  

The discrete Fourier transform was computed at the frequency of interest, over a range of 500,000 time samples and a sampling range of 500 kHz. The maximum sound pressure amplitude that can be generated with our transmitters is 120 dB SPL. This gives an acoustic dynamics range of 114 dB SPL.

Reflection measurements were also performed at low frequencies from 0.1 to 20 kHz and high frequencies from 50 to 70 kHz, showing reflected power in Fig. A.3 of approximately 40 dB above 10 kHz linearly decreasing to 15 dB at 100 Hz. For reflection measurements, all signals were generated using a PXI-4461 DAQ. The low frequency signals were transmitted using the Event Electronics TR-8 Studio Monitors while the high frequency signals were transmitted using the Audio Spotlight. One microphone recorded the incident sound pressure amplitude while another recorded the reflected sound pressure amplitude. Significant amounts of acoustic shielding were required to isolate the reflected microphone from the incident one, as well as using a geometry that separated the incident and reflected microphones by approximately 1.0 meter. The incident and reflected microphones were aligned using a laser so that they each resided within the loudest portion of the sound beam. Return loss in decibels was then calculated using

\[ L_r = 20 \cdot \log_{10} \left( \frac{P_r}{P_i} \right), \]  

(A.3)
Figure A.2: Acoustic insertion loss of back wall of anechoic chamber. Above 50 kHz where higher power signals can be generated, the periodicity of the attenuation indicates that a resonance is being generated.

where $P_r$ is the reflected pressure amplitude and $P_i$ is the incident pressure amplitude. Given the increased path distance for reflection measurements, a correction was made to account for free space path loss as the sound wave traveled from the incident microphone to the reflected microphone.

### A.4 Anechoic Chamber RF Characterization

Both transmission and reflection measurements were taken using S-parameters. An HP8510C network analyzer was used to generate a 0.0 dBm signal at Port 1 to an ETS-Lindgren 3164-03 horn antenna through a frequency range of 400 MHz to 6 GHz. A copper conical-hat receiving antenna connects to Port 2. For insertion loss measurements, the transmitting antenna points directly to the receiving antenna at a distance of 1.55 meters with polarization aligned. A calibration was initially performed to account for beam spreading across this range. Two $S_{21}$ measurements are made, one with the receive antenna
outside the chamber, and one with both antennas inside the chamber at the same distance. The first measurement is divided by the second to obtain a transmission coefficient.

Initial transmission measurements only captured 80 dBm noise with a 0.0 dBm input. The insertion loss was so great that the receiving antenna outside the chamber picked up only noise. To increase the dynamic range, an Ophir 5164 28-watt RF Amplifier was placed at Port 1 to amplify the transmitted signal by 40 dB. The insertion loss in Fig. A.4 is raw data and does not include the 40 dB added by amplification. The insertion loss averages approximately 105 dB.

For return loss measurements, the transmitting antenna points toward the back wall at a distance of 0.86 m, while the receiving antenna is aligned to receive the signal following the most direct ray bounce at a distance of 1.58 m from the wall. A thick stack of RF absorber is then placed between the two antennas to isolate side lobe interference. The $S_{21}$ measurement is then divided by a calibrating $S_{21}$ measurement to account for antenna gain and beam spreading of the two antennas. The calibrating measurement was performed with the transmit and receive antennas placed at a distance of $0.86 + 1.58 = 2.44$ m. The
frequency range of the conical antenna had a bandwidth sufficient for measurement from 0.4 to 4 GHz while the horn antenna has a frequency range of 0.4 to 6 GHz.

The RF return loss shown in Fig. A.5 oscillates between 20 to 40 dB, indicating an adequate lack of reflection for our measurements. The oscillation corresponds to a half wavelength of approximately 30 cm at 50 MHz, which is approximately the thickness of our chamber wall. Furthermore, the comparative drop in return loss around 2.6 GHz corresponds to a drop in the gain of our transmitting antenna in the same range. This drop in antenna gain should theoretically be cancelled out. The calibration of the antenna gain did not cancel due to the fact that the calibration was not performed on an open range. The actual RF reflection is most likely significantly lower than the measurement suggests, due to the lack of proper calibration. Return loss data for the RF absorbent foam from the manufacturer indicates that the foam itself should have higher return loss than the measured data for the constructed anechoic chamber.
A.5 Conclusion

An anechoic chamber attenuating both acoustic and electromagnetic energy has been presented in this appendix. Measurements show that the constructed anechoic chamber significantly attenuates both transmitted and reflected acoustic and RF energy. For acoustic measurements, the chamber offers up to 100 dB insertion loss and 45 dB return loss. The chamber has been shown to achieve an acoustic dynamic range of over 110 dB. RF measurement data show insertion loss of up to 110 dB, and a return loss of up to 50 dB. Return loss data for the RF absorbent foam from the manufacturer indicates that the foam itself should have higher return loss than the measured data for the constructed anechoic chamber. Other measured data collected indicated that interactions between the two antennas used for measurement may have been interacting and affecting the measured return loss performance of the chamber. The RF return loss measurement lack of accuracy was due to the lack of calibration on an open range.
B

Matlab Functions
B.1 PIM in Lossy Lumped Components

% Analytic Electro-Thermal PIM

% Generate 3rd order Analytic PIM
clear all
TonePwr = 27; % Tone Power in dbm
f1 = 400e6; f2 = 0;
%span = [:.0001:.001:.001-.0001],...
[.001:.001:.01-.001],[.01:.01:.1-.01]...
%[.1:.1:.1-[1],[1:10-1],[10:10:100],...
[100:100:1000],[1000:1000:10000]];
span = logspace(0,3,50);
%[1:1:100];
Ro = 50;
%Ro = 42;52/3.5;50.71;
%Ro = 15 % RRPolyResPCS Ro was wasy unbalanced, confirmed in ADS
%Ro = 35 % SilPolyRes

% Ta = 27;
% Rth = 97;
% Cth = .00007;
% %Ta = 100;
% %alpha = .00035;
Ta = 27; Rth = 70; Cth = .00222;
%Ta = 100;
alpha = .002;
%alpha = .00021;%.00385%.0017685;%.0001691;%.00012685;
%alpha = .000647; %RRPolyResPCS metallization subtracts from TCR of Poly
%alpha = .0006; %TanRes
%alpha = .0017; %PrecisionRes
%alpha = .0021; %RRPolyRes
%t = [0:1/(800e9):1/(800e5)];
%Cth = .0155;
%Rth = 74%;35;
%Cth = 700*2320*32e-6*310e-6*1000e-6+740*2320*32e-6*310e-6*6e-6;
%Cth = 1.5*900*2700*32e-6*310e-6*.4e-6;
%Cth = 700*2320*170e-6*125e-6*1000e-6+740*2320*170e-6*125e-6*6e-6;
% RRPolyResPCS
%Cth = 900*2700*350e-6*300e-6*.4e-6; % PrecisionRes
%Cth = 2*900*2700*800e-6*115e-6*.4e-6; % RRPolyRes
%Cth = 2*900*2700*172e-6*150e-6*.4e-6; % TanRes

%Rth = 1/(100*1000e-6)+1/(1.3*6e-6);%8.7;35;
```markdown
% Rth1 = 1/(250*32e-6);
% Rth = 1/(100*1000e-6)+1/(1.3*6e-6); % RRPolyResPCS
% Rth = 1/(250*68e-6*2); % TanRes
% Rth = 1/(250*150e-6*2); % PrecisionRes
% Rth = 1/(250*115e-6*2); % RRPolyRes

Is = 0; Vs = 0;
Rtheq = Rth./(1+Rth*sqrt(Cth)*sqrt(span)*sqrt(i));%*sqrt(span));
%Rtheq1 = Rth1./(1+Rth1*sqrt(Cth1)*sqrt(span)*sqrt(i));
%Rtheq = Rtheq+Rtheq1;
% Calculate Current Magnitude
% Watts = .001*10^(TonePwr/10)
% A = 2*sqrt(Watts)/((Ro+Ro*alpha*Ta)); % Current Amplitude
% A = 2*sqrt(Watts)/((50))
A = dbmtov(TonePwr,50)/50
% frequency sweep
for k=1:length(span)
    f2 = f1+span(k);
    Vs = .75*(A^3)*(Ro*alpha*Rtheq(k))*(Ro+Ro*alpha*Ta);
    Ps(k) = (Vs.*conj(Vs)/(2*50));
end Pwr = 10*log10(Ps/.001);
%figure(2)
%semilogx(span,10*log10(Ps/.001)-32-30,'r--')
plot(10*log10(span),10*log10(Ps/.001)-32-30,'r--')
%save Ptanalytic23.dat Pwr -ascii

B.2 PIM on Transmission Lines

% P1 - Power of f1 in dBm.
% P2 - Power of f2 in dBm.
% L - Total length of the transmission line.
% T_L - User defined length of thermal cells.
% W - Width of the transmission line.
% SH - Line Thickness.
% f1 - Tone one frequency.
% f2 - Tone two frequency.

function [x,wave,a_alt,phase] = 
SSPforward_fieldsex2(P1,P2,f1,f2,L,W,T_L,SH,Zo); 

% DC electrical conductivity of material (1/(ohm-m))
% cond = length/(Resistivity*Area)
cond = .30*63.01e6;

% Skin Depth (m)
SD = 1.7e-6; sqrt(1./(pi.*cond.*f1.*4e-7*pi))

% Line loss in nepers/meter
nep = (1/cond)/(2*Zo*W*SD)

% Volume of Thermal Cell
V = T_L.*W.*(SH)

% Electrical Cross Sectional Area
CSA = SD.*W

% Surface Area of Thermal Cell
SA = W.*T_L

% Scaling parameter found from current density simulation in HFSS
VA = (.18*W*T_L*500e-6)/(W*T_L);

% TCR in ppm ohm/deg C
alpha = .0038;

% Thermal Conductivity of Material (W/(m-k))
% del_Q/(Area*del_t)*Length/(del_Temp)
k = 31;

% Volumetric Thermal Capacity of Material (J/(kg*K))
Cv = 740;

% Density of material kg/m^3
rho = 3980;

% Relative Permittivity
er = (11.58+1)/2;

% Effective Thermal Cell Electrical Conductance
e_c = cond.*(CSA./T_L)

% Effective Thermal Cell Electrical Resistance
e_res = 1./e_c
% Effective Thermal Cell Heat Conductance
e_k = k.*(SA./SH)

% Effective Thermal Cell Heat Resistance
e_rth = 1./e_k

% Effective Thermal Cell Heat Capacity
e_Cv = Cv.*V.*rho

% Kappa of f1
kap_1 = 2.*pi.*sqrt(er).*f1./2.98e8;

% Kappa of f2
kap_2 = 2.*pi.*sqrt(er).*f2./2.98e8;

% Loss of material in Watt/cell
sigma = 8.686.*(1./cond)./(2.*Zo.*CSA)-1.7

% Time interval of travel for f1
delt_1 = 2.*pi.*f1.*sqrt(er)./2.98e8;

% Time interval of travel for f2
delt_2 = 2.*pi.*f2.*sqrt(er)./2.98e8;

% Number of cells line is partitioned into.
n = L./T_L;

% Voltage of f1 tone
V1 = dbmtov(P1,Zo);

% Voltage of f2 tone
V2 = dbmtov(P2,Zo);

% Start Current Density on Line of f1
J1 = V1./(Zo);

% Starting Current Density on Line of f2
J2 = V2./(Zo);

% Current Density on Line of f1 as a function of x
J1_x(1) = J1./CSA;

% Current Density on Line of f2 as a function of x
J2_x(1) = J2./CSA;

x(1) = T_L; loss(1) = .5*e_res.*(J1.^2); V(1) =
2*500e-6.*VA.*SH.*((1./cond).^2).*alpha.*(2./k)./ ...
(1+sqrt(1+4.*i.*2.*pi.*abs(f1-f2).*Cv.*rho.*(VA.^2)/k)).*((J1_x(1)).^3)
a_alt(1) = cos(delt_1*T_L-kap_1*T_L-kap_1*(T_L)*(3.4/3.5-1));
phase(1) = kap_1*(T_L)*(3.4/3.5-1);

% Analyze Line
for j = 2:n

% Current Density on Line of f1 adjusted for line loss
J1_x(j) = (J1/CSA)*exp(-nep*T_L*j);

% Current Density on Line of f2 adjusted for line loss
J2_x(j) = sqrt((J2)^2-2*loss(j-1)/(Zo))/CSA;

loss(j) = .5*e_res.*((J1_x(j).*CSA).^2) + loss(j-1); % in watts

% Amplitude of third order distortion
E3 = 2*500e-6.*VA.*SH.*((1./cond).^2).*alpha.*(2./k)./ ...
(1+sqrt(1+4.*i.*2.*pi.*abs(f1-f2).*Cv.*rho.*(VA.^2)/k)).*((J1_x(j)).^3);

% Calculate generated wave at end of line assuming starting phase = 0
V(j) = E3;
a_alt(j) = cos(delt_1*T_L*j-j*kap_1*T_L-j*kap_1*(T_L)*(3.4/3.5-1));
phase(j) = j*kap_1*(T_L)*(3.4/3.5-1);

% Point on line
x(j) = j.*T_L;
percent_complete_1 = j/n*100;
end

for j = 1:n
throwaway = 0;
sig_add = 0;
for k = 1:j
throwaway = throwaway+V(k)*exp(-nep*T_L*(j-k));
sig_add = sigma.*T_L.*(j-k);
end
loss_add(j) = sig_add;
wave(j) = throwaway;
percent_complete = j/n*100;
end

B.3 PIM on Antennas

% In this script the loss is only correct for 411e-6 wide section of line,
% thus the T_L and W settings are important, unlike with the transmission
% line. A more accurate way to do it would be to find the antenna loss
% in a different manner than making it equivalent to an appropriate sized
% transmission line at the right current density.
%
% P1 - Power of f1 in dBm.
% P2 - Power of f2 in dBm.
% L - Total length of the transmission line.
% T_L - User defined length of thermal cells.
% W - Width of the transmission line.
% SH - Thermal Thickness.
% f1 - Tone one frequency.
% f2 - Tone two frequency.

function [x,wave,I,loss] = 
AntFieldsSim(P1,P2,f1,f2,L,W,T_L,SH,Zo,V_A);

% DC electrical conductivity of material (1/(ohm-m))
% cond = length/(Resistivity*Area)
cond = .30*63.01e6;

% Skin Depth (m)
SD = 1.7e-6;

% Line Loss in nepers
nep = (1/cond)/(2*Zo*W*SD);

% Volume of Thermal Cell
V = T_L.*W.*(SH);

% Electrical Cross Sectional Area
CSA = SD.*W;
% Surface Area of Thermal Cell
SA = W.*T_L;

% TCR in ppm ohm/deg C
alpha = .0038;

% Thermal Conductivity of Material (W/(m-k))
% del_Q/(Area*del_t)*Length/(del_Temp)
k = 429;

% Volumetric Thermal Capacity of Material (J/(kg*K))
Cv = 232;

% Density of material kg/m^3
rho = 10490;

% Relative Permittivity
er = (11.58+1)/2;

% Effective Thermal Cell Electrical Conductance
e_c = cond.*(CSA./T_L);
% e_c = e_c./CSA % Conductance per unit cell for line width/height

% Effective Thermal Cell Electrical Resistance
e_res = 1./e_c;

% Effective Thermal Cell Heat Conductance
e_k = k.*(SA./SH);

% Effective Thermal Cell Heat Resistance
e_rth = 1./e_k;

% Effective Thermal Cell Heat Capacity
e_Cv = Cv.*V.*rho;

% Kappa of f1
kap_1 = 2.*pi.*sqrt(er).*f1./2.98e8;

% Kappa of f2
kap_2 = 2.*pi.*sqrt(er).*f2./2.98e8;

% Loss of material in Watt/cell
sigma = 8.686.*(1./cond)./(2.*Zo.*CSA);
% Time interval of travel for f1
delt_1 = 2.*pi.*f1.*sqrt(er)./2.98e8;

% Time interval of travel for f2
delt_2 = 2.*pi.*f2.*sqrt(er)./2.98e8;

% Number of cells line is partitioned into.
n = L./T_L;

% Voltage of f1 tone
V1 = dbmtov(P1,Zo);

% Voltage of f2 tone
V2 = dbmtov(P2,Zo);

% Start Current Density on Line of f1
J1 = V1./(Zo);

% Starting Current Density on Line of f2
J2 = V2./(Zo);

% Current Density on Line of f1 as a function of x
J1_x(1) = J1./CSA;

% Current Density on Line of f2 as a function of x
J2_x(1) = J2./CSA;

% Init
x(1) = T_L; loss(1) = .5*e_res.*(J1.^2); V(1) =
500e-6.*sin(f1^2/(2.98e8*(sqrt((11.5+1)/2))^2)*0).*1000e-6.*SH.* ...
((1./cond).^2).*alpha.*(2./k)./(1+sqrt(1+4.*i.*2.*pi.*abs(f1-f2) ...
.*Cv.*rho.*(10.^2)./k)).*((J1_x(1)).^3);

% Analyze Finite Cells
for j = 2:n

% Amplitude of third order distortion
E3 = 500e-6.*V_A*1.7e-6.*((1./cond).^2).*alpha.*(2./k)./ ... 
(1+sqrt(1+4.*i.*2.*pi.*abs(f1-f2).*Cv.*rho.*(V_A^2)./k)) ...
.*(J1_x(1).^3*sin((2*pi*f1)^2/((2.98e8*sqrt((11.5+1)/2))^2)) ... 
*T_L*(j-1))*sin((2*pi*f2)^2/((2.98e8*sqrt((11.5+1)/2))^2)) ... 
*T_L*(j-1))*sin((2*pi*f1)^2/((2.98e8*sqrt((11.5+1)/2))^2)*T_L*(j-1));
% Calculate generated wave at end of line assuming starting phase = 0
V(j) = E3;

% Point on line
x(j) = j.*T_L;

percent_complete_1 = j/n*100;
end

% Sum Distortion

for j = 1:n
    throwaway = 0;
    for k = 1:j
        throwaway = throwaway+V(k);
    end
    wave(j) = throwaway;
    percent_complete = j/n*100;
end

B.4 IMD Extractor with Drift

% Used with a shell to feed the spectrums in and
% process externally.
% @ input data - The spectrum array
% @ input startf - The lower carrier tone frequency
% @ input spacing - Tone Spacing between carriers
% @ input splow - Start frequency of spectrum
% @ input spup - End frequency of spectrum
% @ input num_pts - Number of Points in Spectrum
% @ input order - order of IM products to extract
% @ output IMarray - Array of IM products from 3rd l,h
% to upper order l,h
function IMarray =
ExtractRFMDImodDrift(data,startf,endf,spacing,...
splow,spup,num_pts,df,order);

span = spup-splow; postfile_str = '.txt'; pad_str = ' ';
%limit = 1/increment - 1;
%index = (endf - startf)/increment;
thirdorderl = []; fifthorderl = []; seventhorderl = []; ninthorderl = [];
eleventhorderl = []; thirdorderu = []; fifthorderu = []; seventhorderu = []; ninthorderu = [];
eleventhorderu = []; position = 0;

%LwrCarrierIndex = round(num_pts/2-spacing/(2*df));
Index = find(data(1:floor(end/2)) == max(data(1:ceil(end/2))))
%Index = find(data > -78);
LwrCarrierIndex = Index(1)
thirdorderlindex = LwrCarrierIndex-round(spacing/df);
thirdorderuindex = LwrCarrierIndex+2*round(spacing/df);
thirdorderl = max(data(thirdorderlindex-3:thirdorderlindex+3));
thirdorderu = max(data(thirdorderuindex-3:thirdorderuindex+3));
IMarray = [thirdorderl,thirdorderu];
if(order == 5)
    fifthorderl = max(data(LwrCarrierIndex-2*round(spacing/df)-3:LwrCarrierIndex-2*round(spacing/df)+3));
    fifthorderu = max(data(LwrCarrierIndex+3*round(spacing/df)-3:LwrCarrierIndex+3*round(spacing/df)+3));
    IMarray = [thirdorderl,thirdorderu,fifthorderl,fifthorderu];
elseif(order == 7)
    fifthorderl = max(data(LwrCarrierIndex-2*round(spacing/df)-3:LwrCarrierIndex-2*round(spacing/df)+3));
    fifthorderu = max(data(LwrCarrierIndex+3*round(spacing/df)-3:LwrCarrierIndex+3*round(spacing/df)+3));
    seventhorderl = max(data(LwrCarrierIndex-3*round(spacing/df)-3:LwrCarrierIndex-3*round(spacing/df)+3));
    seventhorderu = max(data(LwrCarrierIndex+4*round(spacing/df)-3:LwrCarrierIndex+4*round(spacing/df)+3));
    IMarray = [thirdorderl,thirdorderu,fifthorderl,fifthorderu,...
              seventhorderl,seventhorderu];
end

B.5 IMD Extractor from Spectrum Shell

% RFMD Data Analysis Script
for j=1:num_files
    sprintf('%d',j);
    file = strcat(filepath,file_descr,pwr,num2str(spacing(j)),'.txt');
    data = load(file);
    splow = data(end-1);
    df = data(end);
    data = data(1:end-2);
    num_pts = length(data);
    spup = num_pts*df+splow;
    temp = ExtractRFMDImodDrift(data,startf,endf,spacing(j), ...%
    splow,spup,num_pts,df,order);
    IMarray = [IMarray,temp];
    file ='';
end
thirdorderl = IMarray(1:order-1:end); thirdorderu =
IMarray(2:order-1:end); fifthorderl = IMarray(3:order-1:end);
fifthorderu = IMarray(4:order-1:end); seventhorderl =
IMarray(5:order-1:end); seventhorderu = IMarray(6:order-1:end);
ninthorderl = IMarray(7:order-1:end); ninthorderu =
IMarray(8:order-1:end); eleventhorderl = IMarray(9:order-1:end);
eleventhorderu = IMarray(10:order-1:end);

figure
switch order
    case 11
plot(spacing,[thirdorderl;thirdorderu;fifthorderl;...fifthorderu;seventhorderl;seventhorderu;ninthorderl;...;ninthorderu;eleventhorderl;eleventhorderu])
title('Peak 3rd - 11th Order Intermodulation Amplitude')
xlabel('Frequency (MHz)')
ylabel('Amplitude (dBm)')
legend('3rd order lower', '3rd order upper', ...
'5th order lower', '5th order upper', '7th order lower', ...
'7th order upper', '9th order lower', '9th order upper', ...
'11th order lower', '11th order upper')
case 9
plot(spacing,[thirdorderl;thirdorderu;fifthorderl;...fifthorderu;seventhorderl;seventhorderu;ninthorderl;ninthorderu])
title('Peak 3rd - 9th Order Intermodulation Amplitude')
xlabel('Frequency (MHz)')
ylabel('Amplitude (dBm)')
legend('3rd order lower', '3rd order upper', '5th order lower', ...
'5th order upper', '7th order lower', '7th order upper', ...
'9th order lower', '9th order upper')
case 7
plot(spacing,[thirdorderl;thirdorderu;fifthorderl;...]fifthorderu;seventhorderl;seventhorderu])
title('Peak 3rd - 7th Order Intermodulation Amplitude')
xlabel('Frequency (MHz)')
ylabel('Amplitude (dBm)')
legend('3rd order lower', '3rd order upper', ...
'5th order lower', '5th order upper', ...
'7th order lower', '7th order upper')
case 5
plot(spacing,[thirdorderl;thirdorderu;fifthorderl;fifthorderu])
title('Peak 3rd and 5th Order Intermodulation Amplitude')
xlabel('Frequency (MHz)')
ylabel('Amplitude (dBm)')
legend('3rd order lower', '3rd order upper', ...
'5th order lower', '5th order upper')
otherwise
semilogx(spacing,[thirdorderl;thirdorderu])
title(strcat('LPfilter-','27dbm','-IM3'))
xlabel('Spacing (Hz)')
ylabel('Amplitude (dBm)')
legend('3rd order lower', '3rd order upper')
end
save_name = strcat('E:\', file_descr, pwr, 'IM3.txt');
save(save_name, 'thrdorderl', 'thrdorderu', 'spacing', '-ASCII');

B.6 Voltage to dBm Converter

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function vtodbm() %
% Description: changes voltage values to dbm, referenced to 50 ohms. %
% inputs: x - voltage %
% outputs: output - power level in dbm %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function output = vtodbm(x,Zo)
output = 10.*log10(x.^2/(.001*Zo*2));
return

B.7 dBm to Voltage Converter

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% Function dbmtov() %
% Description: changes dbm values to voltage, referenced to 50 ohms. %
% inputs: x - power level in dbm %
% outputs: output - voltage level %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function output = dbmtov(x,Zo)
output = sqrt((.001*2*Zo).*10.^(x./10));
return

B.8 Cancellation Formula Analysis
f = 400e6; phi = [0:1:180].*pi/180; x =
cos(2*pi*f.*[0:1/(100*f):100/f-1/(100*f)]); y = []; beta = []; shift
= zeros(1,length(phi)); for j=1:length(phi)
    y = cos(2*pi*f.*[0:1/(100*f):100/f-1/(100*f)]+phi(j));
    beta = max(x+y);
    shift(j) = pi-2*acos(beta/2);
end CP = 10*log10(1+1-2*cos(phi(end:-1:1)-shift)); figure(1) hl1 =
line(phi.*180./pi,CP,'Color','r','LineStyle','--'); ax1 = gca;
%axis([0 180 -140 -20]);
set(ax1,'XColor','k','YColor','k','FontName','Times New
Roman','FontSize',14) ax2 = axes('Position',get(ax1,'Position'),...
    'XAxisLocation','top','YAxisLocation','right','Color','none',...
    'XColor','k','YColor','k');
hl2 =
line(phi.*180./pi,(phi(end:-1:1)-shift).*180./pi,'Color','b','Parent',ax2);
set(ax2,'FontName','Times New Roman','FontSize',14)
%axis([0 180 -0.2 1.4]);
xlabel(ax2,'Phase Difference (deg)','FontSize',16)
ylabel(ax2,'Phase Error (deg)','FontSize',16)
ylabel(ax1,'-Cancellation (dB)','FontSize',16)
%clear all;

B.9 Multi-Spectrum File Analysis

% General purpose script for limiting amp analysis and isolator analysis

data = load('C:\Documents and Settings\jrwilker\My
Documents\Siames\Cancellation
Project\Calibration\VM0_a3_-2dBmLO_spectrums_B0.txt');
% data = load('C:\Abominationfiles\CirculatorData ...
\Isolator_900M_945.00015M_-5dBm_RWSweep.txt');
%data = load('C:\Abominationfiles\CirculatorData ...
\Isolator4_FreqSweep_-11dbm_700M_900M_fine.txt');
%data = load('C:\Abominationfiles\CirculatorData ...
\Isolator4_700M_845.00015M_-7dbm_BW.txt');
%data = load('C:\Abominationfiles\CirculatorData ...
\Isolator4_700M_845M_BSSW.txt');
g = 2; ind = find(data == -1000); pwr(1) = max(data(1:ind(1)-3));
for j=2:length(ind)
    pwr(j) = max(data(ind(j-1)+1:ind(j)-3));
end figure(4) h1 = axes('FontName','Times New Roman','FontSize',14);
plot([data(ind(g)-2):data(ind(g)-1):data(ind(g)-1) ...
*length(data(ind(g)-1)+1:ind(g)-4)+data(ind(g)-2)]- ...
data(ind(g)-2),data(ind(g)-1)+1:ind(g)-3))
axis([-4500 4500 -120 -20]);
set(h1,'FontName','Times New Roman') xlabel('Frequency (Hz)'),'FontName','Times New Roman','FontSize',16) ylable('Output Power (dBm)', 'FontName','Times New Roman','FontSize',16) data = load('C:\Documents and Settings\jrwilker\My Documents\Siames\Cancellation Project\Calibration\VM0_a3_-2dBmLO_B0.txt');
data = load('C:\Documents and Settings\jrwilker\My Documents ... \Siames\Cancellation Project\Calibration\VM0_S12_L0-20_a3.txt');
Intdrive = []; Intmeas = []; LO = []; UIM3 = []; LIM3 = [];
for j = 1:ceil(length(data)/5)-1
    Intdrive(j) = data(j*5-4);
    Intmeas(j) = data(j*5-3)+30;
    LO(j) = data(j*5-2)+30;
    UIM3(j) = data(j*5-1)+30;
    LIM3(j) = data(j*5)+30;
end
figure(9)
% plot(46-length(ind)-5+[1:length(ind) ],pwr,'b: ')
%plot(700:1:900,pwr,'r')
figure(5) h2 = axes('FontName','Times New Roman','FontSize',14);
plot(Intdrive,Intmeas,'b',Intdrive,LO,'r',Intdrive,...
UIM3,'g',Intdrive,LIM3,'k')
axis([-120 -5 -90 10]); set(h2,'FontName','Times New Roman')
xlabel('Spur Input Level (dBm)','FontName','Times New Roman','FontSize',16) ylable('Output Power (dBm)', 'FontName','Times New Roman','FontSize',16)
legend('Interferer','LO','2f2-f1','2f1-f2') clear all;

B.10 Sum a Vector

function sum = SumVector(x); sum = 0; for j = 1:length(x)
sum = sum + x(j);
end

B.11 LimiterModel

x1 = dbmtov(3) * cos(2*pi*200e6.*[0:1/800e8:1/100000-1/800e8]); x2 = dbmtov(-50) * (cos(2*pi*510e6.*[0:1/800e8:1/100000-1/800e8])); x_2 = dbmtov(-53) * (cos(2*pi*420e6.*[0:1/800e8:1/100000-1/800e8])); pwm = zeros(1,length(x2)); pwm_inv = zeros(1,length(x2));

% A = dbmtov(0);
for j = 1:2:length(pwm)/50 pwm(50*j+1:50*(j+1))=1;
pwm_inv(50*(j-1)+1:50*(j))=1; end for j = 3:4:length(pwm)/50 pwm_inv(50*(j-1)+1:50*(j))=1; end

x3 = x1 + x2; % + dbmtov(3);
for j=1:length(x3) if (x3(j) > dbmtov(0))
x3(j) = dbmtov(0);
elseif (x3(j) < -dbmtov(2))
    x3(j) = -dbmtov(2);
else
    x3(j) = x3(j); % + dbmtov(-35); % - x2(j);
end end
x3 = x3 + dbmtov(3);

%fNorm = 220e6 /(800e7/2);
% [b, a] = butter(10, fNorm, 'low');
% x4 = filtfilt(b, a, x3);

y = fft(x3); % * pwm_inv;
m = abs(y); clear y; m = m / (length(m)/2); f = (0:length(m)-1).*800e8/length(m);

% y1 = fft((x2).* pwm_inv);
% m1 = abs(y1);
% clear y1;
% m1 = m1 / (length(m1)/2);
% f1 = (0:length(m1)-1).*800e8/length(m1);

figure(1)
plot([0:1/800e8:1/100000-1/800e8], x3, % f1, vtdbm(m1))
% figure(2)
% plot(x4)
figure(2) plot(f, vtdbm(m))
% figure(4)
%freqz(b,a,128,800e7);

B.12 LimiterGain

function [v_out,v_in] = LNA(in)

v_out = -8.604e-18 + 1.*(6.542.*(in) + 2.823e-014.*(in.^2) - 2191.*(in.^3)+ ... -2.099e-11.*(in.^4) + 4.753e5.*(in.^5) + 3.25e-9.*(in.^6) ... -5.143e7.*(in.^7)-9.472e-8.*in.^8 +2.148e9.*(in.^9));

v_in = in;

B.13 Function to Find S-parameters

function [S11a,S11p,S21a,S21p,S12a,S12p,S22a,S22p,freq] = spfind(data); S11a = []; S11p = []; S21a = []; S21p = []; S12a = []; S12p = []; S22a = []; S22p = []; freq = [];

    temp = FindSparams(data(:,1),data(:,2),data(:,3),j*1e6);
    S11a = [S11a,temp(1)];
    S11p = [S11p,temp(2)];
    temp = FindSparams(data(:,1),data(:,4),data(:,5),j*1e6);
    S21a = [S21a,temp(1)];
    S21p = [S21p,temp(2)];
    temp = FindSparams(data(:,1),data(:,6),data(:,7),j*1e6);
    S12a = [S12a,temp(1)];
    S12p = [S12p,temp(2)];
    temp = FindSparams(data(:,1),data(:,8),data(:,9),j*1e6);
    S22a = [S22a,temp(1)];
    S22p = [S22p,temp(2)];
    freq = [freq,j];
end
B.14 S-parameters Interpolation Routine

function output = FindSparams(freq,amp,phase,f);

index_1 = find(freq < f); index_2 = find(freq > f);

m = (amp(index_1(end))-amp(index_2(1))) / (freq(index_1(end))-freq(index_2(1)));
b = amp(index_1(end))-m*freq(index_1(end));

output = [0 0]; output(1) = m*f+b;

m = (phase(index_1(end))-phase(index_2(1))) / (freq(index_1(end))-freq(index_2(1)));
b = phase(index_1(end))-m*freq(index_1(end));

output(2) = m*f+b;

B.15 Optimized Grunwald-Letnikov Fractional Derivative Routine

% @ TS - Time Step
% @ q - IntegroDifferential Operation Order
% @ T - Current Time
% @ f - Function to IntegroDifferentiate

function fd = fracd(q,TS,T,f)

% IntegroDifferential Result
Deriv = [];

% Intermediate Calculation array
Da = [0]; y = 1;

for k=1:(T/TS)
    clear Da;
    % This algorithm needs N+1 points but only uses N points -
% thus the first point is not used.
if (k < 3)
    if (k > 1)
        % Two Points
        Da(1) = -q*f(1)+f(2);
        Deriv(y,1) = Da(1)*2^q/((2*TS)^q);
    end
else
    % Three or More Points
    Da(1) = f(1)*(k-q-2)/(k-1)+f(2);
    for j = 2:k-1
        Da(j) = Da(j-1)*(k-q-j-1)/(k-j)+f(j+1);
    end
    Deriv(y,k-1) = Da(end)*k^q/(((k)*TS)^q);
    % Scale by the (Number of Points)^q/(Time Step)^q
end
end

fd = (Deriv(1:end)-Deriv(1:end-1))/2;

**B.16 Forward Euler Algorithm**

% Standard Forward Euler
% @ x = current point x axis
% @ y = current point y axis
% @ delx = step size
% @ derx = derivative of function

function npf = FEuler(x,y,delt,derx)

npf = y + delt*derx;

**B.17 Fractional Simulator**

% Code to solve basic electrothermal system. Want
% to solve attenuator but with only one element
% being electrothermally active.
% Error Tolerance
errtol = 10e-6;

% Power Level in dbm
P = 30;

% Tone Spacing
span = 1;

% Frequencies
f1 = 50e3; f2 = 50e3 + span;

% Simulation time
df = 20 * f1; delt = 1 / df; rbw = 1.0; length = df / (.1 * rbw);
C = [0:delt:delt*length];

% Source vectors
A = dbmtov(P, 50);
Vsrc_1 = A * cos(2 * pi * f1 * t);
Vsrc_2 = dbmtov(P) * cos(2 * pi * f2 * t);
Vsrc_3 = (dbmtov(1) * cos(2 * pi * 100 * t))^-2;
freque = [49.995e3:0.1:50.006e3]; frequeT = [0.1:0.1:2]; rer_ex = 0;
rei_ex = 0; rer_exT = 0; rei_exT = 0;

% Constants
Ro = 50; Ta = 27; R2 = 1000; R3 = 50; a = .00385; T = 0; Rth = 74;
Cth = .025;

% Attenuator Matrix
Y = [1/(Ro*(1+a*Ta+a*T))+1/R2,-1/R2;-1/R2,1/R2+1/R3]; Yin = inv(Y);

% Thermal Node
Yth = [1/Rth+j*w*Cth];

% Variable vectors
T = 75; V = [0]; J = [0]; Rnl = 0; Output = [0]; derx1 = 0; derx2 = 0;
npf = 0; npc = 0; Pdis = 0; t = 0; OT = 75; T2 = 0; Pdis2 = 0;
Time = 0; errvec=0; tic temp = 0; for k=2:length
  t = (k-1)*delt;
  J = [(A*cos(2* pi * f1 * t) + A*cos(2 * pi * f2 * t)) / (50); 0];
  % Initial network equation solution for attenuator
  Rnl = Ro/(1+a*Ta+a*T);
\[
\begin{align*}
Y(1,1) &= \frac{1}{R_1 + 1/R_2}; \\
Y_{in} &= \text{inv}(Y); \\
V &= Y_{in} \cdot J; \\
P_{dis} &= V(1)^2/(2R_1); \\
T_{ot} &= T;
\end{align*}
\]

\% Forward Euler Solve network equation for thermal circuit
derx1 = \(\frac{P_{dis}}{C_{th}} - T/(R_{th} \cdot C_{th})\);
npf = FEuler((k-1) \cdot \text{delt},T,\text{delt},\text{derx1});

\% Trapezoidal correction for thermal network
\%
derx2 = \(\frac{P_{dis}}{C_{th}} - \frac{npf}{R_{th} \cdot C_{th}}\)
\%
cmp = Trap((k) \cdot \text{delt},OT,T,\text{delt},\text{derx1},\text{derx2});
T = npf;

\% for d=1:3

\%
\%
\%
\% Resolve attenuator network with updated estimate
\%
\% \(R_1 = R_0 \cdot (1 + a \cdot T_a + a \cdot T)\);
\% \(Y(1,1) = \frac{1}{R_1 + 1/R_2}\);
\% \(Y_{in} = \text{inv}(Y)\);
\% \(V = Y_{in} \cdot J\);
\% \(P_{dis} = V(1)^2/(2R_1)\);
\%
\%
\% Trapezoidal correction for thermal network using correction
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\% errvec(k,d) = npc-T;
\%
\% end
\%
\% T = npc;
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\%
\% R1 = R0 \cdot (1 + a \cdot T_a + a \cdot T);
\% Y(1,1) = (1/R1 + 1/R2);
\% Yin = inv(Y);
\% V = Yin*J;
\% Pdis = V(1)^2/(2*R1);
\% rer_ex = rer_ex + V(2) \cdot \text{cos}(2*pi*freqe*\text{t});
\% rei_ex = rei_ex + V(2) \cdot \text{sin}(2*pi*freqe*\text{t});
\% rer_exT = rer_exT + T \cdot \text{cos}(2*pi*freqeT*\text{t});
\% rei_exT = rei_exT + T \cdot \text{sin}(2*pi*freqeT*\text{t});
\% Pdis2 = Pdis2 + Pdis;
\% if(T>T2)
\% \quad T2 = T;
\% end
if(k == 1000000)
    toc
    Time = Time+toc
    tic
elseif(k == 2000000)
    toc
    Time = Time+toc
    tic
elseif(k == 3000000)
    toc
    Time = Time+toc
    tic
elseif(k == 4000000)
    toc
    Time = Time+toc
    tic
elseif(k == 5000000)
    toc
    Time = Time+toc
    tic
elseif(k == 6000000)
    toc
    Time = Time+toc
    tic
elseif(k == 7000000)
    toc
    Time = Time+toc
    tic
elseif(k == 8000000)
    toc
    Time = Time+toc
    tic
elseif(k == 9000000)
    toc
    Time = Time+toc
    tic
elseif(k == 10000000)
    toc
    Time = Time+toc
    tic
elseif(k == 11000000)
    toc
    Time = Time+toc
    tic
tic
elseif(k == 12000000)
toc
Time = Time+toc
tic
elseif(k == 13000000)
toc
Time = Time+toc
tic
end
toc
Time = Time+toc
AvgPower = Pdis2/k
MaxTemp = T2
rer_ex = 
2*rer_ex/length; rei_ex = 2*rei_ex/length;
ampinc=sqrt(rei_ex.\^2+rer_ex.\^2); figure(1)
plot(freque,10*log10(ampinc))
rei_exT = 2*rei_exT/length; rei_exT = 
2*rei_exT/length; ampincT=sqrt(rei_exT.\^2+rer_exT.\^2); figure(2)
plot(frequeT,10*log10(ampincT))

B.18 Riemann-Liouville Fractional Derivative

% RL Approximate fractional derivative from Nonlinear Circuits
% and Systems Jonathan R. Wilkerson

t = 1; TS = .001; q = .5; f = []; deriv = []; Da = []; Dasum = [];
D = 1/TS;

% Make Time Array
T=[0:TS:t+1];

for m=1:((t+1)*1/TS+1)
    \%f(m) = ((m)*TS)^2;
    \%f(m) = sqrTS(TS*(m));
    f(m) = sin(2*pi*10*(m)*TS);
end

for y=1:10
    q = .1*y;
    deriv(1) = f(1);
    for k=1:((t+1)*1/TS)
        for j1=0:k-1
            if q<1
                Da(j1+1)=(f(j1+2)-f(j1+1))*(((k-j1)^\%(1-q)) ...
            end
        end
    end
end
-((k-j1-1)^(1-q));
else
    if j1==0
        Da(j1+1)=0;
    else
        Da(j1+1)=(f(j1+2)-(2*f(j1+1)+f(j1)) ... 
                   *(((k-j1)^(2-q))-((k-j1-1)^(2-q)));
    end
end
end
Dasum=sum(Da);
clear Da
if q<1
    deriv(y,k+1)=((TS^(-q))/gamma(2-q)) ... 
                 *(((1-q)*f(1))/((t/TS)^q)+Dasum);
else
    deriv(y,k+1)=((TS^(-q))/gamma(3-q)) ... 
                 *(((1-q)*(2-q)*f(1))/((t/TS)^q)+((2-q)*(f(2)-f(1))) ... 
                  /((t/TS)^((q-1))))+Dasum);
end
end
%
% x^2 Analytic Solution
deriv25=(8*(T.^1.5))/(3*sqrt(pi));
figure(3)
plot(T,deriv25(1,:),T,deriv25(2,:),T,deriv25(3,:),T,deriv25(4,:), ... 
     T,deriv25(5,:),T,deriv25(6,:),T,deriv25(7,:),T,deriv25(8,:), ... 
     T,deriv25(9,:),T,deriv25(10,:)) title('RL IntegroDifferentiated
Function -> q=.1...1, sqrt(x)')
legend('q=.1','q=.2','q=.3','q=.4','q=.5',... 
       'q=.6','q=.7','q=.8','q=.9','q=1') 
ylabel('Magnitude') xlabel('time') grid

B.19 Riemann-Liouville Short Memory Fractional Derivative

% Approximate fractional derivative from Nonlinear Circuits and Systems
% book

\( t = 1; \ T = .001; \ qd = .5; \ g = []; \ x1 = []; \ f4 = []; \ h4 = []; \ D = 1/T; \ ML = 15; \) for \( m=1:((t+1)*D+1) \)
\( g(m) = ((m-1)*T)^2; \)
\( \% g(m) = \cos(2*\pi*10*(m-1)*TS); \)
end

\( x1(1) = g(1); \) for \( k1=1:((t+1)*D) \)
if \( k1<=ML \)
for \( j1=0:k1-1 \)
if \( qd<1 \)
\( f4(j1+1)=g(j1+2)-g(j1+1)\) \( \times (((k1-j1)^(1-qd)) \ldots -((k1-j1-1)^(1-qd)))); \)
else
if \( j1==0 \)
\( f4(j1+1)=0; \)
else
\( f4(j1+1)=(g(j1+2)-(2*g(j1+1))+g(j1)) \) \( \times (((k1-j1)^(2-qd))-(k1-j1-1)^(2-qd)))); \)
end
end
end
else
for \( j1=k1-ML:k1-1 \)
if \( qd<1 \)
\( f4(j1+1-k1+ML)=(g(j1+2)-g(j1+1)) \) \( \times (((k1-j1)^(1-qd))-(k1-j1-1)^(1-qd)))); \)
else
if \( j1==0 \)
\( f4(j1+1)=0; \)
else
\( f4(j1+1)=(g(j1+2)-(2*g(j1+1))+g(j1)) \) \( \times (((k1-j1)^(2-qd))-(k1-j1-1)^(2-qd)))); \)
end
end
end
end
h4=sum(f4);
clear f4
if \( k1<=ML \)
if \( qd<1 \)
\( x1(k1+1)=((T^(-qd))/\gamma(2-qd))\times (((1-qd) \ldots \ldots)) \)
\*g(1)/((t/T)^{qd})+h_4); 
else 
\begin{align*}
x_{1}(k1+1) &= (((T^{-(q-1)}))/\gamma(3-q)) \times (((1-q) \times \frac{g(1)}{(t/T)^{qd}}) + ((2-qd) \times \frac{g(2)-g(1)}{(t/T)^{(qd-1)}}) + h_4); 
\end{align*} 
end 
else 
if \( q<1 \) 
\begin{align*}
x_{1}(k1+1) &= (((T^{-(q-1)}))/\gamma(2-q)) \times (((1-q) \times \frac{g(1)}{(t/T)^{qd}}) + h_4); 
\end{align*} 
else 
\begin{align*}
x_{1}(k1+1) &= (((T^{-(q-1)}))/\gamma(3-q)) \times (((1-q) \times \frac{g(1)}{(t/T)^{qd}}) + ((2-qd) \times \frac{g(2)-g(1)}{(t/T)^{(qd-1)}}) + h_4); 
\end{align*} 
end 
end 
end 
end

\texttt{inte}_{25} = [0:T:t+1]'; \texttt{deriv}_{25} = (8*(\texttt{inte}_{25}.*1.5))/(3*\sqrt{\pi}); \texttt{deri}_{25} = \texttt{x1}'; \texttt{length(inte}_{25}) \texttt{length(deri}_{25}) \texttt{figure}(3) \texttt{plot(deriv}_{25}(1:end-1)) \texttt{length(deriv}_{25}) \texttt{figure}(4) \texttt{plot(deri}_{25}(1:end-1)) \texttt{length(deri}_{25}) \texttt{deri}_{25} \texttt{figure}(5) \texttt{plot(g(1:1000)) length(g(1:1000))}

\section*{B.20 Grunwald-Letnikov Fractional Derivative}

\begin{verbatim}
\% Algorithm for computation of fractional derivative 
\% with complete memory. No explicit use of the gamma 
\% function is accomplished through a recursive 
\% multiplication addition scheme developed by Spanier 
\% & Oldham in the Fractional Calculus(1974). Algorithm 
\% was coded by Jonathan R. Wilkerson. 
\% Time Step 
TS = .1; 
\% IntegroDifferential Operation Order 
q = .5; 
\% Length of Simulation 
Time = 100; 
\% Time array for initial function
\end{verbatim}
T = TS:TS:Time+1;

% Function to IntegroDifferentiate
f = [];

% Generate function samples
for m=1:((Time+1)*(1/TS)+1)
    %f(m)=((m)*TS)^2;
    f(m) = sin(2*sqrt(TS*(m)));
    %f(m) = 1/(5*(1.01+(cos(2*pi*10*m*TS*.1)+
        cos(2*pi*10*m*(TS+TS*.1)))^2) *(cos(4*pi*10*(m)*TS)+
        cos(4*pi*10*m*(TS+TS*.1)))+cos(4*pi*10*(m)*TS*.1)+
        cos(4*pi*10*(m)*(2*TS+TS*.1))); end

% IntegroDifferential Result
Deriv = [];

% Intermediate Calculation array
Da = [0]; y = 1;

for k=1:((Time+1)/TS)
    clear Da;
    % This algorithm needs N+1 points but only uses N points -
    % thus the first point is not used.
    if(k<3)
        if(k>1)
            % Two Points
            Da(1)=-q*f(1)+f(2);
            Deriv(y,1)=Da(1)*2^q/((2*TS)^q);
        end
    else
        % Three or More Points
        Da(1)=f(1)*(k-q-2)/(k-1)+f(2);
        for j=2:k-1
            Da(j)=Da(j-1)*(k-q-j-1)/(k-j)+f(j+1);
        end
        Deriv(y,k-1)=Da(end)*k^q/(((k)*TS)^q);
    % Scale by the (Number of Points)^q/(Time Step)^q
    end
    end
figure(1) plot(T,besselj(0,T),T,besselj(1,T),T,besselj(2,T))
title('Original Function - sin(2*pi*10*t)') legend('Original Function - sin(2*pi*10*t)') ylabel('Magnitude') xlabel('Time') grid
figure(2)

% plot(T(2:end), Deriv(1,:), T(2:end), Deriv(2,:), T(2:end), ...
% Deriv(3,:), T(2:end), Deriv(4,:), T(2:end), Deriv(5,:), T(2:end), ...
% Deriv(6,:), T(2:end), Deriv(7,:), T(2:end), Deriv(8,:), T(2:end), ...
% Deriv(9,:), T(2:end), Deriv(10,:))
% title('G1 IntegroDifferentiated Function -> ...
% q=.1...1, sin(2*pi*10*t)')
% legend('q=.1','q=.2','q=.3','q=.4', ...
% 'q=.5','q=.6','q=.7','q=.8','q=.9','q=1')
plot(T(2:end), Deriv(1,:)) ylabel('Magnitude') xlabel('Time') grid

B.21 Short Memory Grunwald-Letnikov Fractional Derivative

% Short Memory Algorithm for computation of fractional derivative.
% Algorithm was coded by Jonathan R. Wilkerson.

% Time Step
TS = .001;

% IntegroDifferential Operation Order
q = .5;

% Length of Simulation
Time = 1;

% Time array for initial function
T = TS/2:TS/2:Time;

% Function to IntegroDifferentiate
f = [];

% Sample Memory Length
ML = 100;

% Generate function samples
for m=1:((Time+1)*(1/TS)+1)
    f(m) = sqrt(TS*(m));
end
\%f(m) = \sin(2\pi5*(m-1)\cdot TS); 
end 

\% IntegroDifferential Result
Deriv = []; 

\% Intermediate Calculation array
Da = [0]; 

\% Short Memory Weights
A = []; 

k=((Time+1)/TS); 

\% Calculate the short memory weights
A(1) = 1/gamma(1); for j=1:ML+1 
\hspace{1em} A(j+1)=A(j)*(j-q-1)/(j); 
\hspace{1em} \%A(j+1)=gamma(j-q)/(gamma(-q)*gamma(j+1)); 
end 

for k=1:((Time+1)/TS) 
\hspace{1em} % Use last ML or less points 
\hspace{2em} if(k<ML) 
\hspace{3em} mem=k; 
\hspace{2em} else 
\hspace{3em} mem=ML; 
\hspace{2em} end 
\hspace{1em} for j=0:mem-1 
\hspace{2em} Da(j+1)=f(k-j)*A(j+1); 
\hspace{1em} end 
\hspace{1em} Deriv(k)=sum(Da)\cdot(k^q)/(k\cdot TS)^q; 
end 

figure(1) plot(A) title('Original Function') figure(2) title('IntegroDiffernetiated Function') plot(Deriv) 

B.22 Foster Model Synthesizer

\% This m-file takes an admittance filter, does a foster 2 
\% expansion, and computes the RC network necessary to implement it.
R_{thermal} = 35; SPF = .03; SZF = (SPF*10)/(4.6); R = []; C = [];
order = 8; HF = 8600;
% Impedance based
num = conv([1 SZF],[1 SZF*3]); num = conv(num,[1 SZF*10]); num =
conv(num,[1 SZF*30]); num = conv(num,[1 SZF*100]); num = conv(num,[1
SZF*300]); num = conv(num,[1 SZF*1000]); num = conv(num,[1
SZF*3000]);
den = conv([1 SPF],[1 SPF*10/2]); den = conv(den,[1
SPF*10]); den = conv(den,[1 SPF*100/2]); den = conv(den,[1
SPF*100]); den = conv(den,[1 SPF*1000/2]); den = conv(den,[1
SPF*1000]); den = conv(den,[1 SPF*10000/2]); den = conv(den,[1
SPF*10000]);

% Multiply denominator by 's' to expand Y(s)/s
num = conv(num,[1 0]);

% Get residue form
[r,p,k] = residue(den,num)

% Convert to component values
for j=1:order
    R = [R,1/r(j)];
    C = [C,r(j)/(-p(j))];
end R = [R,1/r(order+1)]; scale = R_{thermal}/R(order+1) R = R*scale; C
= C/scale; C = [C,((1/(HF/30))/R_{thermal})]; R C save
81-3001BModelParams.txt C R -ascii
C

Broadband High Dynamic
Measurement System User Guide
C.1 Introduction

This appendix details the hardware configuration necessary to operate the feed-forward cancellation system of Fig. 3.2 as well as the software configuration necessary to run it. Section C.2 discusses the control software organization and theory of operation, as well as the settings important to run the program. Equipment configuration and electrical specifications are discussed in Section C.3.

C.2 Software

The software for the high dynamic range measurement system is organized into three main sections: initialization, cancellation, and data storage. Initialization includes user defined configuration of the spectrum analyzer, signal source sweep settings, data storage settings, and cancellation unit calibration settings. Cancellation and data storage sections of the software are internal to the system software. The user only interacts with the initialization routines. This section discusses the program organization and theory of operation, followed by standard initialization settings.

C.2.1 Software Organization

The software starts by taking input from the user defining the spectrum collection settings and test type. Two tests are currently supported, a two-tone constant tone separation frequency sweep and a two-tone variable tone separation sweep. This arrangement is shown in the program flow chart of Fig. C.1. The program then proceeds to calibrate the feed-forward channels of the system unless the user has specified otherwise. The data collection branch of the program is then reached, shown in Fig. C.2. The stimulus at the output of the DUT is suppressed before a final spectrum is taken at each measurement frequency. The carrier power levels, third order intermodulation products, and fifth order intermodulation products are stored as an excel, txt, or csv file based on user settings.

The data collection portion of the program starts by collecting a spectrum to find the power level of the stimulus tones at the output of the DUT channel, as shown in Fig. C.3. The amplitude of the individual tones is used to compute an I and Q voltage where Q is
Figure C.1: The program acquires user settings to configure the spectrum analyzer and test type. A two-tone constant tone separation frequency sweep and a two-tone tone separation sweep are available.

set to 100 % and I is set to zero. This voltage is supplied to the vector modulator inputs by a USB controlled digital to analog converter. The I and Q voltages are computed using a gain function for each I and Q channel based on the channel transmission characteristic calibration files. The spectrum resulting from the combination of the feed-forward probe signal and the output of the DUT channel is used to compute the phase shift necessary in the feed-forward channel for cancellation. The phase shift is converted to I and Q channel percentages. Each I and Q percentage is corrected for channel losses, I and Q gain differences, and matching offset from the limiting amplifiers. The USB controlled DAC supplies the vector modulator with the corrected I and Q voltages, cancelling the post DUT signal at the reference plane.

C.2.2 User Settings

User settings include initialization data, calibration input, cancellation spectrum settings, data collection settings, and cancellation initialization data. Initialization data represents the hardware control addresses of the signal sources and vector signal analyzer equipment. Calibration input allows the user to calibrate the system for a particular test configuration before a measurement. Cancellation initialization data allows the user to add or subtract a static offset to the transmission characteristic of the cancellation channel if an attenuator or amplifier is added into the test configuration without the need to recalibrate. Cancellation spectrum settings controls the speed and accuracy of both the cancellation
Figure C.2: The program calibrates the feed-forward channels based on user settings then conducts the user specified test. Data is stored as arrays of the power level of the carriers and intermodulation products.

and the collected data separately. Data collection settings determines what type of test, constant $\Delta f$ or variable $\Delta f$, and over what frequency range the test is conducted. Each of these user setting categories is discussed in this section.

Initialization data, shown in the top left of Fig. C.4, is composed of two GPIB input numbers, a digitizer resource number, a downconverter resource number, and a reference clock selection. The GPIB inputs represent each of the marconi 2025 source addresses, usually 21 and 22. SRC1 GPIB is always associated with channel zero and SRC2 GPIB is always associated with channel one. The physical connection of the sources must match the software associations or the system will not work. The reference clock is completely user configured for a particular measurement. The digitizer and downconverter inputs can be
Figure C.3: The internal algorithm of the cancellation system. The uncancelled spectrum at the output of the DUT is measured and used to set the I and Q gain of the vector modulators. A combined spectrum is then measured and used to compute the I and Q settings necessary for cancellation. Finally, the cancelled spectrum is measured.

determined through the National Instruments measurement and automation explorer if the drivers are ever updated. Currently the downconverter device number is DAQ::6 and the digitizer reference number is 2.

Calibration input, situated just under initialization data in Fig. C.4, controls how much time and how accurate the system calibration will be. In general, the system should be calibrated at the intended resolution bandwidth of the measurement. The program works off relative accuracy rather than absolute, but the absolute accuracy will increase with lower resolution bandwidth and increased averaging. The higher the accuracy the slower the program will run. The calibration routine does not report in real time its progress. A general setting of 100 Hz resolution bandwidth with 5 averages over 1 GHz bandwidth will result in calibrations less than 30 minutes long. Higher resolution bandwidths will decrease the calibration time significantly. The reference level for calibration should always be set to +10 dBm to guarantee accurate measurement. The LO level should never be set above
Figure C.4: The user interface of the cancellation software composed of several user inputs including initialization data, calibration input, cancellation initialization data, cancellation spectrum settings, and data collection settings. Real-time spectrum is displayed along with processed carrier and intermodulation distortion data. Processed carrier and intermodulation distortion data is written to excel, text, or .csv files.
−10 dBm. The optimal LO setting is always −11 dBm. The attenuation should always be set to 30 dB to minimize harmonic generation in the receiver.

Cancellation initialization data, located directly to the right of initialization data in Fig. C.4, simply allows the user to modify the calibration data with external gain or loss without re-calibrating the system. The calibration filepath must be where the calibration files are stored for the program to function. The program mode should always be set to cancellation.

Cancellation spectrum settings, positioned directly below cancellation initialization data in Fig. C.4, determines the spectrum settings for the cancellation and data collection routines. Spectrum averages and resolution bandwidth are applied to the collected data. All other parameters affect the speed and accuracy of the cancellation routine. The reference level should be set at least 7 dB above the maximum power level of the signal to be cancelled. The attenuation should be set to guarantee a signal at the receiver of less than −30 dBm optimally. The program outputs estimated maximum output powers for each feed-forward channel. The post-DUT signal must be at least one decibel below this level at the signal combination point. The optimal span for both routines is determined automatically in the program. The optimal reference level and attenuation are determined and applied before the final data is acquired.

Data collection settings, located directly below cancellation spectrum settings in Fig. C.4, controls what type of test is being performed over what frequency range. Constant tone separation sweeps and variable tone separation sweeps are available. Constant tone separation sweeps are specified by selecting the results display tab labeled constant df sweep. Offset count specifies the number of tone offsets to sweep with one being the minimum. Offset array specifies each tone offset to be swept. Each sweep will start at the frequency specified in start/center frequency and will continue at the frequency interval specified by increment until the frequency specified in end frequency is reached. The SRC power should be the same as that specified in the file calibration, usually −11 dBm. Variable tone separation sweeps are chosen by selecting the results display tab labeled variable df sweep. In this mode, the sweep starts at a center frequency specified in start/center frequency and progresses outward until one of the tones reaches the frequency specified in end frequency. The sweep can be either logarithmic or linear, selected by the log sweep button. In logarithmic sweep mode, position zero in offset array determines the start spac-
ing of the tones. The program sweeps outward at the specified points per decade until the
frequency specified in end frequency is reached. If linear sweep is selected, the program
sweeps outward linearly by the spacing defined in position zero of the offset array until the
frequency specified in end frequency is reached. Increment is not used in variable df sweep
mode.

Although most of the important settings are automatically controlled by the pro-
gram, it is important to know the impact of the parameters in case the need for manual
control arises. The resolution bandwidth setting determines the frequency resolution, noise
floor, and speed of a measurement. In general the resolution bandwidth needs to be no
more than 30 % of the lowest tone separation in the measurement. This setting produces
the fastest, highest noise floor, lowest dynamic range, and the lowest power accuracy mea-
surement. The slowest, lowest noise, highest dynamic range, and highest power accuracy
measurement is given by the minimum resolution bandwidth allowable, 1 Hz. The user is
recommended to use an intermediate setting for cancellation and a high accuracy setting for
final data collection. The receiver in the vector signal analyzer is optimally driven at −33
to −45 dBm. The receiver will be heavily distorted by signals as little at 5 dB above the
optimal input. As signal levels prior to cancellation are far outside this range, the receiver
must be protected with attenuation. The attenuation is internally set to the maximum.
Once signals have been reduced in power, the attenuation protecting the receiver needs to
be set to zero to reduce the noise floor.

C.3 Hardware

The canceller hardware in the high dynamic range measurement system must be
configured properly and driven with appropriate signal levels to operate correctly. The con-
figuration of the hardware and equipment connection details are provided in Section C.3.1.
Section C.3.2 discusses the electrical specifications of the system with a focus on optimal
performance.
C.3.1 Equipment Operational Configuration

The measurement system equipment is composed of two cancellers, two signal sources, an amplifier box, and a vector signal analyzer. The connection of this equipment is straightforward, but the isolation of cabling is not as apparent. This section details equipment connection and the system points susceptible to EMI.

The signal sources are connected directly to the each canceller box at the RF\text{in} port. The reference is set to external for each source. The power level of each source is optimally set to $-11$ dBm and cannot be set above $-10$ dBm less damage occur to the limiters within the cancellers. The frequency of the sources is controlled in software. The power supply can be AC or DC, but DC power further reduces spectrum spurious frequency content.

The Amp\text{in} ports on the rear of the amplifier box are connected to the RF\text{out} port of each canceller through an attenuator large enough to drop the signal level of the limiting amp to the required amplifier input level. Since the RF\text{out} of the canceller produces near the maximum allowable amplifier input level, the attenuation usually ranges between 0 to 13 dB. A box of mini-circuits SMA attenuators is available for this purpose. This connection is an extremely susceptible point for EMI coupling. Any EMI at this port will result in distortion from both the canceller and the amplifiers. The amplifiers must be driven with DC power due to spurious frequency content. The batteries and power cables are not shielded and are also susceptible to EMI. Any EMI coupled into the DC power input will be mixed into the spectrum producing spurious frequency content.

The output of the amplifiers is connected to an EMI shielded test box. These outputs are high power and are the main cause of EMI into other parts of the system. These outputs must be isolated in excess of 160 dB, which is far more than cables can provide. Current return path loops using secondary conductors and well as EMI shielding using either electromagnetic absorber or ferrite tape is necessary for this connection. The outputs must be isolated from each other inside of the EMI shielding box before the isolators are reached as well. Reverse waves at the amplifier outputs will reduce the dynamic range of the system by generating distortion products in each channel.

The outputs of the amplifiers are connected to isolators inside of the EMI shielded box followed by a hybrid combiner. One side of the hybrid combiner is fed to a low PIM
Table C.1: Linear System Electrical Specifications

<table>
<thead>
<tr>
<th></th>
<th>RF$_{in}$</th>
<th>RF$_{out}$</th>
<th>Cancellation$_{out}$</th>
<th>Amp$_{in}$</th>
<th>Amp$_{out}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{max}}$ (dBm)</td>
<td>$-10$</td>
<td>$-5$</td>
<td>$+5$</td>
<td>$-3$</td>
<td>$+43$</td>
</tr>
<tr>
<td>$P_{\text{min}}$ (dBm)</td>
<td>$-15$</td>
<td>$-5$</td>
<td>$-25$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$P_{\text{opt}}$ (dBm)</td>
<td>$-12.5$ to $-10.5$</td>
<td>$-5$</td>
<td>$-10$ to $+2$</td>
<td>$-16$ to $-9$</td>
<td>$+30$ to $+37$</td>
</tr>
</tbody>
</table>

The output of the DUT is connected to another low PIM cable attenuator, which reduces the signal level to the optimum for cancellation. Further attenuation may be required outside the EMI shielded box to further reduce signal levels to that required for optimum cancellation. These levels are defined in the following section.

The output of each canceller, Cancellation$_{out}$, is connected to isolators before being fed to a hybrid combiner. This signal port must be isolated from reverse waves and EMI as any present interfering signal will be converted to spurious frequency content in the same manner as with the high power amplifiers. The output of the hybrid combiner is then combined with the DUT output from the EMI shielded box in another hybrid combiner before connection directly to the vector signal analyzer.

### C.3.2 Electrical Specifications

The system can cancel signals over a wide range of DUT output power levels. However, maximum dynamic range is only obtained for a small subset of this power range. Several competing factors including system noise, vector modulator linearity, and feed-forward amplitude and phase accuracy are impacted by signal power levels. Each of these factors are discussed in this section.

The linear electrical system specifications for each channel are shown in Table C.1. The values denoted by $P_{\text{max}}$, $P_{\text{min}}$, and $P_{\text{opt}}$ represent the maximum, minimum, and optimal powers applied to a given signal port. It should be noted that a computer controlled variable attenuator was never implemented for the system. The input power to the power amplifiers must then be controlled by attenuators inserted in the path from the RF$_{out}$ of the canceller to the Amp$_{in}$ port of the amplifier.
The vector modulators are always driven at the optimum level for noise considerations by the limiting amplifier inside the canceller. If the limiter is ever removed, the signal level applied to the local oscillator port of the vector modulator needs to remain between −6 and 0 dBm. Any lower level of signal drive will result in an increase in the system noise floor that will limit dynamic range. Output levels of less than the optimum level in Table C.1 will also result in an increase in noise relative to the signal level. The output signal level of the canceller must always be above −10 dBm for noise considerations.

Feed-forward amplitude and phase accuracy reduces as the power output of the vector modulator exceeds or falls short of the optimal dynamic range. Optimal cancellation only occurs in this signal range. The signal drive level of the canceller should always be kept within the range specified by $P_{\text{opt}}$ in Table C.1 to guarantee maximum cancellation.

The vector modulators are inherently nonlinear as they are composed of several amplifiers and mixers. Reverse wave signals or electromagnetic interference (EMI) at the output port of the vector modulators or the input port of the limiting amplifiers will result in intermodulation products. The signal level of distortion products generated will be given approximately by the signal level of any reverse waves or EMI at RF$_{\text{out}}$ less the directionality of the vector modulator. Distortion products from EMI coupling to the RF$_{\text{in}}$ will be approximately the power level of the EMI.