ABSTRACT

FONG, REY-YIE. Vision-based Large-area Damage Inspection via Ultrasonic Phase Estimation in Laser Speckle Interferometry. (Under the direction of Dr. Fuh-Gwo Yuan).

Laser speckle interferometry (LSI) techniques such as digital speckle pattern interferometry (DSPI), shearography (SG) and laser speckle photometry (LSP) rely on laser speckles as a sensing cloud in contrast to digital image correlation (DIC) using painted speckles to quantify the deformation. All above techniques are vision-based optical metrologies that can be applied for large-area damage inspection. LSI has the benefits to adjust to the laser emission power, coherence, field of view (FOV), and have the choice to correlate the images optically or numerically. It is more suitable for qualitative damage featuring while DIC is a more quantitative tool for strain measurement and structure characterization.

A direct phase estimation (DPE) via Riesz transform from LSI speckles was proposed for large-area damage inspection. Commercial LSI systems rely on optical setups to extract phase information from intensity in which damage feature is encoded. Depending on the stress loading conditions, either temporal phase shifting (TPS) if quasi-static or steady-state prevails or spatial phase stepping (SPS) when transient state occurs needs to be implemented beforehand in the hardware for phase estimation. In contrast, DPE via Riesz transform numerically evaluates the phase instantly and can be combined with either a filter bank for multi-scale pyramid imaging or monogenic filters for multi-dimensional feature extraction.

This research explores almost entire speckle related techniques for large-area damage inspection. The full-field vision-based LSI system was demonstrated and verified on a honeycomb composite plate and C-17 aileron composite stiffened panel using this proposed DPE algorithm under two classes of loading: (I) low-frequency thermal stress loading (~Hz) and (II) high-frequency steady-state vibration in ultrasonic frequency range for large-area inspection in near real-time. The visualized damage images agreed with images from the baseline of pulse laser/LDV, ultrasonic C-scan and X-ray CT scan. The DPE algorithm has a promising potential to apply on any optical metrology for instantaneous phase estimation and is capable of combining with selected imaging conditions for multi-damages blind detection.

KEYWORDS

Riesz transform, log-Gabor bandpass (bp) filter, direct phase estimation (DPE), temporal phase shifting (TPS), spatial phase shifting (SPS), correlation, laser speckle interferometry (LSI), digital speckle pattern interferometry (DSPI), laser speckle photometry (LSP), shearography (SG), digital image correlation (DIC), barely visible impact damage (BVID).
Vision-based Large-area Damage Inspection via Ultrasonic Phase Estimation in Laser Speckle Interferometry

by
Rey-Yie Fong

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

Mechanical Engineering

Raleigh, North Carolina
2021

APPROVED BY:

Dr. Fuh-Gwo Yuan
Committee Chair

Dr. Richard Keltie

Dr. Chih-Hao Chang

Dr. Yong Zhu

Dr. Hans Hallen
DEDICATION

To my kindly grandma, Tsao-Chih Huang Hung,

my inspirational granduncle, William Hung,

my diligent uncle, Wen-Xiang Huang,

my supportive parents, Chin-Shih Fong and Wen-Ling Huang,

my considerate uncle and aunt, Chin-Ming Lin and Wen-Zhen Huang,

my little aunt, I-Lun Hung,

my lovely sister, Ya-San Fong,

and my cousins, Li-Ying Lin, Hung-Chang Lin, Kuan-Ru Huang and Kuan-Cheng Huang.
BIOGRAPHY

Rey-Yie (Abel) Fong attended NTU and graduated with a Presidential Awards in BS and MS in Mechanical Engineering in 2007 and 2009. After graduation, he worked for United Ship Design & Development Center and started his own business on biomechanics consulting service. During the time, he was preparing the application for the NC State Mechanical Engineering PhD Program and enrolled in 2014 supervised by Dr. Fuh-Gwo Yuan. In 2021, he was awarded NCSU CoE Summer Graduate Merit Award (GMA) as a milestone just before his graduation.

His research interests include physics-informed data-driven modeling (PDM), multi-sensor data fusion (MSDF), vision-based laser speckle interferometry (V-LSI), wave and vibration-based non-destructive damage inspection (WV-NDI), integrated system health management (ISHM), noise, vibration & harshness (NVH), multiphysics CAE and vehicle dynamics (VD).
ACKNOWLEDGEMENTS

Firstly, I would like to express my appreciation to Dr. Fuh-Gwo Yuan for the academic support with patient guidance, profound insight, professional attitude, and dedication during my entire graduate studies from 2014 Spring to 2021 Summer. Thanks for bringing the opportunities of Graduate Research Assistantship (GRA) in National Institute of Aerospace (NIA) under the NASA Langley Professor Program to envision the fields of sensor technologies, non-destructive inspection (NDI) and structural health monitoring (SHM). I would like to thank each of my committee members, Dr. Richard Keltie, Dr. Chih-Hao Chang, Dr. Yong Zhu as internal committees, and Dr. Hans Hallen as an external committee in Physics introduced by Dr. Alexandra Duel-Hallen in ECE, for their technical guidance, academic instruction, and time spent serving on my dissertation.

I would like to extend my appreciation to Dr. Richard Keltie for bridging profound theories with comprehensive physical insights and making them very easy to understand in MAE 361 Dynamics & Control and MAE 789 Advanced Analytical Methods in Structural Vibration. I learned a lot from you how to be a good storyteller and make a good story to enlighten others. The way you describe waves, the incoming and out-going, so concise and I will never forget how they behave; to Dr. Chih-Hao Chang for the engaging and challenging course in MAE 589 Optical Engineering, the derivation, the insights of Fraunhofer and Fourier optics, the flexible choice of project topics, and the open questions for brainstorming, every single element polishes my engineering intuition in a very different angle.

I would like to thank Dr. James E. Hubbard for the special lectures of ENAE 788Z TITAN Aerial Explorer and ENAE 757 Advanced Structural Dynamics & Control in the University of Maryland, from an exploration angle to discuss the cutting-edge technologies and climate change – the vital signs of the planet; to Dr. Yizhen Zhu for the in-depth derivation of EM wave mode theory in ECE Advanced Fiber Optics & Application in Virginia Tech; to Dr. John Strenkowski for the informative lecture in MAE 533 Finite Element Analysis I and the discussion of SIMULIA Isight & SEE portfolio; to Dr. Xiaoning Jiang for insightful lecture of sensor technologies and their working principles in MAE 532 Smart Structures and Micro-Transducers; to Dr. Alexandra Duel-Hallen and Dr. Huaiyu Dai for the consecutive lectures of ECE 515 & ECE 762 Digital Communication as the foundation of one of my most important skillsets in probability, stochastic and signal processing.;

I would like to thank Director Robin Harris for opening the choreography course, DAN 272 Dance Composition, the emerging teaching, peer feedback, the sharing session of enlightening objects – structure, content and contest, the movement invention and development, the contemplation and solo work. You’re
like a magician to release our potential and let our inner spirits to control our body. Beside academia research, dancing becomes an essential activity/process for me to talk to my body, release the pressure and reorganize the next stage planning. I would like to thank you for bringing such an impressive lecture through entire composing and developing process.

I would like to thank David Tsai, as a good friend and skating partner who is also from Taiwan and as a mechanical engineer graduated from NCSU. I always enjoy the skating session and practicing with you in the early morning every week. To Natalia Efimova and Oleg Efimova for coaching me with figure skating moves in the field, gestures, jumps and rotating skills. All because of you, I had a very different campus life in Raleigh when pursuing the PhD program.

I would also like to thank the members, graduates and friends either in our research group or other groups: Tyler Hudson, Patrick Leser, Sri Vikram Palagummi, Mohammad Harb, Donato Girolamo, Karthik Reddy Lyathakula, Sakib Ashraf Zargar, Xiang Yan, Ni Sui, Jiaze (Jason) He, Yuling Liu, Chao Wan, Shaohan Wang, Dongwon Lee, Bryce Abbott, Auston Gray and Jack van Welzen; the three Chang’s: Che-Yuan (Joik) Chang, Huan-Yu (Tony) Chang, Yu-Sheng (Sam) Chang; the research scholars: Dr. Shaorui Yang and Dr. Rui Zhu; good friends in other groups: Terry Huang, Betty Deng, Howuk Kim, Rachel Hyo Son, Yi-De Chen, Yi-Ann (Andy) Chen, Yi-Shin (Daniel) Yeh, Weiyi Chang, Hui-jie Lee, Eric Huang, Chin-Jung Hsu, Angela Chen, Tan Li Hsu, Huy Nguyen, thank you for all the helpful supports I had with you.
TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................................................ xiii
LIST OF FIGURES ....................................................................................................................................... xiv
I. Preface .................................................................................................................................................... 1
  1 Technology Brief ............................................................................................................................ 2
  2 Introduction ..................................................................................................................................... 6
    2.1 NDI/SHM ..................................................................................................................................... 7
    2.2 Point-based Laser NDI ............................................................................................................. 8
    2.3 Vision-based Laser NDI .......................................................................................................... 9
    2.4 The Transition of Sensing Technologies ............................................................................. 10
  3 Motivation and Plan of Works ...................................................................................................... 11
II. Vision-based Large-area Damage Inspection .................................................................................. 12
  4 Full-field Optical NDI Techniques ............................................................................................. 13
    4.1 Digital Speckle Pattern Interferometry (DSPI) ................................................................. 13
    4.2 Laser Speckle Photometry/Photography (LSP) ................................................................. 14
    4.3 Shearography (SG) .............................................................................................................. 15
    4.4 Digital Image Correlation (DIC) ......................................................................................... 16
    4.5 Summary ............................................................................................................................... 19
  5 Working Principles of Laser Speckle Interferometry ................................................................. 20
    5.1 Digital Speckle Pattern Interferometry (DSPI) ................................................................. 21
    5.2 Laser Speckle Photometry/Photography (LSP) ................................................................. 24
    5.3 Shearography (SG) .............................................................................................................. 25
      5.3.1 Strain Components Isolation ....................................................................................... 28
      5.3.2 Interaction of EM and Lamb Waves ........................................................................... 29
    5.4 Space-time Representations ................................................................................................. 33
    5.5 Summary ............................................................................................................................... 34
  6 Phase Extraction and Estimation ................................................................................................... 35
    6.1 Temporal Phase Shifting (TPS) .......................................................................................... 36
    6.2 Spatial Phase Shifting (SPS) ............................................................................................... 38
6.3 Direct Phase Estimation (DPE) ........................................................................................................ 42
   6.3.1 Monogenic Signal and Riesz Transform ........................................................................ 42
   6.3.2 Log-Gabor Bandpass Filter and its Monogenic Form .................................................... 44
   6.3.3 Riesz Transform as a Phase Estimator from Synthetic Images ..................................... 48
6.4 Summary ......................................................................................................................................... 57

7 Digital Image Processing ....................................................................................................................... 58
   7.1 Vibration-based Imaging Condition for Steady-state Vibration ............................................ 58
   7.2 Wave-based Imaging Condition for Transient Propagation .................................................. 59
   7.3 Vision-based and Statistical-based Imaging Condition for Randomness .............................. 60
   7.4 General Processing Procedures for Speckle Images .............................................................. 61
   7.5 The Insight of ADPNet – Advanced Damage Processing Network ..................................... 62
   7.6 What’s More – Machine Learning Framework ..................................................................... 63

8 Experiment Preparation for LSP/SG with DPE ................................................................................. 64
   8.1 Selection of Laser Source ...................................................................................................... 64
   8.2 Selection of Excitations ......................................................................................................... 66

9 Experiment Demo – Honeycomb Composite Panel .............................................................................. 67
   9.1 Baseline – X-ray CT and Ultrasonic C-scan .......................................................................... 67
   9.2 Baseline – Pulse Laser/LDV .................................................................................................. 69
   9.3 Thermal Excitation – Diode and High-power Diode Laser – LSP/SG .................................. 70
      9.3.1 Experimental Setup ........................................................................................................ 70
      9.3.2 Results and Comparisons ............................................................................................... 72
      9.3.3 Summary ........................................................................................................................ 76
   9.4 Thermal Excitation – HeNe Laser – SG ................................................................................ 78
      9.4.1 Experimental Setup ........................................................................................................ 78
      9.4.2 Results and Comparisons ............................................................................................... 79
      9.4.3 Summary ........................................................................................................................ 82
   9.5 Ultrasonic Excitation – HeNe Laser – SG ............................................................................. 83
      9.5.1 Experimental Setup ........................................................................................................ 83
      9.5.2 Results and Comparisons ............................................................................................... 84
      9.5.3 Summary ........................................................................................................................ 86

10 Experiment Demo – C-17 Globemaster II Composite Aileron .......................................................... 87
   10.1 Baseline – Ultrasonic C-scan ................................................................................................. 88
   10.2 Baseline – Pulse Laser/LDV .................................................................................................. 89
   10.3 Thermal Excitation – High-power Diode Laser – LSP/ADPNet ........................................... 90
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10.3.1</td>
<td>Experimental Setup</td>
<td>90</td>
</tr>
<tr>
<td>10.3.2</td>
<td>The Insight of ADPNet – Advanced Damage Processing Network</td>
<td>91</td>
</tr>
<tr>
<td>10.3.3</td>
<td>Results and Comparisons – LSP/ADPNet</td>
<td>92</td>
</tr>
<tr>
<td>10.3.4</td>
<td>Summary</td>
<td>94</td>
</tr>
<tr>
<td></td>
<td>III. Waves</td>
<td>95</td>
</tr>
<tr>
<td>11</td>
<td>Infinite Boundary Conditions</td>
<td>96</td>
</tr>
<tr>
<td>11.1</td>
<td>Pressure Waves – Longitudinal – P</td>
<td>96</td>
</tr>
<tr>
<td>11.2</td>
<td>Shear Waves – Transverse – SV and SH</td>
<td>96</td>
</tr>
<tr>
<td>11.3</td>
<td>Equation of Motion and Constitutive Relations</td>
<td>97</td>
</tr>
<tr>
<td>12</td>
<td>Semi-infinite Boundary Conditions</td>
<td>103</td>
</tr>
<tr>
<td>12.1</td>
<td>Rayleigh Waves – R</td>
<td>103</td>
</tr>
<tr>
<td>12.2</td>
<td>Love Waves – LO</td>
<td>106</td>
</tr>
<tr>
<td>12.3</td>
<td>Gravity Waves – GA</td>
<td>106</td>
</tr>
<tr>
<td>13</td>
<td>Finite Boundary Conditions</td>
<td>107</td>
</tr>
<tr>
<td>13.1</td>
<td>Lamb – PSV</td>
<td>107</td>
</tr>
<tr>
<td></td>
<td>IV. Planar and Cylindrical Waveguides in Elastic Solids</td>
<td>108</td>
</tr>
<tr>
<td>14</td>
<td>Planar Waveguides – PSV Waves</td>
<td>109</td>
</tr>
<tr>
<td>14.1</td>
<td>Formation of Lamb Waves</td>
<td>109</td>
</tr>
<tr>
<td>14.1.1</td>
<td>Symmetric Modes</td>
<td>114</td>
</tr>
<tr>
<td>14.1.2</td>
<td>Anti-symmetric Modes</td>
<td>115</td>
</tr>
<tr>
<td>14.2</td>
<td>Material Constants and Frequency Parameters Relation</td>
<td>116</td>
</tr>
<tr>
<td>14.3</td>
<td>Dispersion Relations</td>
<td>118</td>
</tr>
<tr>
<td>14.4</td>
<td>Representations of Dispersion Relations</td>
<td>119</td>
</tr>
<tr>
<td>14.5</td>
<td>Displacement Fields</td>
<td>123</td>
</tr>
<tr>
<td>14.6</td>
<td>Understanding of Spatial Frequencies – Wavenumbers</td>
<td>125</td>
</tr>
<tr>
<td>14.6.1</td>
<td>Transverse Wavenumbers and Slowness Curves</td>
<td>126</td>
</tr>
<tr>
<td>14.6.2</td>
<td>Longitudinal Wavenumber</td>
<td>129</td>
</tr>
<tr>
<td>14.6.3</td>
<td>Longitudinal Wavenumber – Real</td>
<td>130</td>
</tr>
<tr>
<td>14.6.4</td>
<td>Longitudinal Wavenumber – Real – Wavelength Point of View</td>
<td>133</td>
</tr>
<tr>
<td>14.6.5</td>
<td>Longitudinal Wavenumber – Pure Imaginary</td>
<td>136</td>
</tr>
<tr>
<td>14.6.6</td>
<td>Longitudinal Wavenumber – Complex</td>
<td>139</td>
</tr>
<tr>
<td>14.7</td>
<td>Mode Numbering &amp; Conversion</td>
<td>142</td>
</tr>
<tr>
<td>14.7.1</td>
<td>Mode Conversion</td>
<td>142</td>
</tr>
<tr>
<td>14.7.2</td>
<td>The Concept of First Derivative</td>
<td>144</td>
</tr>
<tr>
<td>Section</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>14.7.3</td>
<td>Analytic Solution for Group Velocity</td>
<td>146</td>
</tr>
<tr>
<td>14.7.4</td>
<td>Dispersion Bands</td>
<td>150</td>
</tr>
<tr>
<td>14.7.5</td>
<td>Mode Tracing and Filter Banks</td>
<td>151</td>
</tr>
<tr>
<td>14.8</td>
<td>Dispersion Curves Simulator</td>
<td>155</td>
</tr>
<tr>
<td>15</td>
<td>Planar Waveguide – SH Waves</td>
<td>157</td>
</tr>
<tr>
<td>16</td>
<td>Cylindrical Waveguides – CY Waves</td>
<td>161</td>
</tr>
<tr>
<td>16.1</td>
<td>Boundary Conditions</td>
<td>164</td>
</tr>
<tr>
<td>16.2</td>
<td>General Dispersion Relations</td>
<td>165</td>
</tr>
<tr>
<td>16.3</td>
<td>Cylindrical Wave – Longitudinal Modes</td>
<td>165</td>
</tr>
<tr>
<td>16.4</td>
<td>Cylindrical Wave – Flexural Modes</td>
<td>169</td>
</tr>
<tr>
<td>16.5</td>
<td>Cylindrical Wave – Torsional Modes</td>
<td>172</td>
</tr>
<tr>
<td>17</td>
<td>Analogy of Elastic Waves in Plane and in Rod</td>
<td>175</td>
</tr>
<tr>
<td>17.1</td>
<td>Comparison of PSV SYM Modes &amp; CY Longitudinal Modes</td>
<td>176</td>
</tr>
<tr>
<td>17.2</td>
<td>Comparison to PSV ASYM Modes &amp; CY Flexural Modes</td>
<td>177</td>
</tr>
<tr>
<td>17.3</td>
<td>Comparison to SH SYM Modes &amp; CY Torsional Modes</td>
<td>178</td>
</tr>
<tr>
<td>V.</td>
<td>Planar and Cylindrical Waveguides in Optical Dielectrics</td>
<td>179</td>
</tr>
<tr>
<td>18</td>
<td>Basic Optics</td>
<td>180</td>
</tr>
<tr>
<td>18.1</td>
<td>Wave Assumptions</td>
<td>180</td>
</tr>
<tr>
<td>18.2</td>
<td>The Definition of Polarization States</td>
<td>181</td>
</tr>
<tr>
<td>18.2.1</td>
<td>Linear Polarization</td>
<td>181</td>
</tr>
<tr>
<td>18.2.2</td>
<td>Circular Polarization</td>
<td>183</td>
</tr>
<tr>
<td>18.2.3</td>
<td>Elliptical Polarization</td>
<td>184</td>
</tr>
<tr>
<td>18.3</td>
<td>Production of Plane Polarized Waves</td>
<td>184</td>
</tr>
<tr>
<td>18.4</td>
<td>Light Propagation through a Quarter Wave Plate</td>
<td>185</td>
</tr>
<tr>
<td>19</td>
<td>EM Waves at Dielectric Interface</td>
<td>186</td>
</tr>
<tr>
<td>19.2</td>
<td>TE and TM Waves</td>
<td>186</td>
</tr>
<tr>
<td>19.3</td>
<td>Reflection and Deflection at a Plane Interface</td>
<td>188</td>
</tr>
<tr>
<td>19.3.1</td>
<td>Snell's Law</td>
<td>188</td>
</tr>
<tr>
<td>19.3.2</td>
<td>Waves in Medium with Refraction Index</td>
<td>190</td>
</tr>
<tr>
<td>19.4</td>
<td>External and Internal Reflection</td>
<td>191</td>
</tr>
<tr>
<td>19.4.1</td>
<td>External Reflection</td>
<td>194</td>
</tr>
<tr>
<td>19.4.2</td>
<td>Brewster Angle</td>
<td>195</td>
</tr>
<tr>
<td>19.4.3</td>
<td>Internal Reflection</td>
<td>197</td>
</tr>
</tbody>
</table>
19.4.4  Total Internal Reflection

20  Mode Theory and Ray Tracing
20.1 1D String Vibration in Free Space
20.2 2D Fields at Dielectric Interface
20.3 2D Fields in Two Parallel Boundaries
20.4 2D Fields in Two Parallel Boundaries – Ray Tracing Approach

21  Planar Waveguides
21.1 Uniform Plane Waves
21.2 Uniform Plane Waves – Summary
21.3 General Solutions
21.4 Coupled TEM Waves
21.5 Decoupled TE and TM Waves
21.6 Dispersion Curves
21.7 Analogy of Elastic and EM Waves in Planar Waveguides

22  Cylindrical Waveguides
22.1 Dispersion Curves and Energy Distributions

VI  Conclusions and Future Works
23  Conclusions
24  Future Works
25  Publications
26  Participated Project Summary
27  Declaration of Conflicting Interests
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.1</td>
<td>Applications of Bessel functions</td>
<td>291</td>
</tr>
<tr>
<td>32.2</td>
<td>Definitions</td>
<td>292</td>
</tr>
<tr>
<td>32.3</td>
<td>Bessel functions of the 1st kind: $J_{\alpha}$</td>
<td>292</td>
</tr>
<tr>
<td>32.4</td>
<td>Bessel functions of the 2nd kind: $Y_{\alpha}$</td>
<td>294</td>
</tr>
<tr>
<td>32.5</td>
<td>Bessel functions of the 3rd kind, Hankel function: $H^{(1)}<em>{\alpha}$ and $H^{(2)}</em>{\alpha}$</td>
<td>295</td>
</tr>
<tr>
<td>32.6</td>
<td>Modified Bessel functions: $I_{\alpha}$ and $K_{\alpha}$</td>
<td>297</td>
</tr>
<tr>
<td>32.7</td>
<td>Spherical Modified Bessel functions: $j_{n}$ and $y_{n}$</td>
<td>298</td>
</tr>
<tr>
<td>32.8</td>
<td>Spherical Hankel functions: $h_{n}^{(1)}$ and $h_{n}^{(2)}$</td>
<td>299</td>
</tr>
</tbody>
</table>

33 Maxwell's Equations ........................................................................................................300

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>33.1</td>
<td>Notation Conventions</td>
<td>301</td>
</tr>
<tr>
<td>33.1.1</td>
<td>Operator Notations</td>
<td>301</td>
</tr>
<tr>
<td>33.1.2</td>
<td>Electrical Notations</td>
<td>303</td>
</tr>
<tr>
<td>33.1.3</td>
<td>Magnetic Notations</td>
<td>304</td>
</tr>
<tr>
<td>33.2</td>
<td>Microscopic Formulation</td>
<td>305</td>
</tr>
<tr>
<td>33.2.1</td>
<td>Formulation in SI Unit Convention</td>
<td>306</td>
</tr>
<tr>
<td>33.2.2</td>
<td>Formulation in Gaussian Unit Convention</td>
<td>306</td>
</tr>
<tr>
<td>33.2.3</td>
<td>Relationship between Differential and Integral Formulations</td>
<td>308</td>
</tr>
<tr>
<td>33.2.4</td>
<td>Vacuum Equations, Electromagnetic Waves and Speed of Light</td>
<td>310</td>
</tr>
<tr>
<td>33.3</td>
<td>Macroscopic Formulation</td>
<td>311</td>
</tr>
<tr>
<td>33.3.1</td>
<td>Bound Charge and Current</td>
<td>312</td>
</tr>
<tr>
<td>33.3.2</td>
<td>Auxiliary Fields, Polarization and Magnetization</td>
<td>313</td>
</tr>
<tr>
<td>33.3.3</td>
<td>Constitutive Relations and Electromagnetism</td>
<td>314</td>
</tr>
<tr>
<td>33.4</td>
<td>Overdetermination of Maxwell's Equations</td>
<td>315</td>
</tr>
<tr>
<td>33.5</td>
<td>Limitations of the Maxwell Equations as A Theory of Electromagnetism</td>
<td>316</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 4.1  General characteristics of the four full-field optical NDI techniques ................................. 19
Table 4.2  The working principles of full-field optical NDI systems .................................................. 19
Table 6.1  Traditional phase extraction steps and proposed phase estimation procedures .................. 52
Table 6.2  Comparison of full-field optical NDI systems .................................................................... 52
Table 8.1  HeNe, diode, high-power diode laser beam profile characteristic comparison [112]........ 65
Table 8.2  Excitation and imaging condition selection guides ............................................................. 66
Table 11.1 The wave speeds and material properties for common materials ....................................... 100
Table 14.1 Mode numbering and cut-off frequencies ......................................................................... 142
Table 33.1 The definition of Bessel functions .................................................................................... 292
LIST OF FIGURES

Figure 1.1  Quantity mapping relations of speckle and vision-based metrology (a) three LSI techniques: DSPI, LSP and SG, (b) DIC with subset window ................................................. 2

Figure 1.2  The relation of a log-Gabor bandpass filter with Riesz transform to generate a monogenic filter in wavenumber and spatial domain ......................................................... 3

Figure 1.3  Laser ultrasonic composite inspection system (LUCIS) for rapid damage qualitative analysis and quantification – 2016 ................................................................................. 4

Figure 1.4  The analogy of Lamb waves in elastic solids and EM waves in dielectric materials: planar and cylindrical waveguides – 2018 ............................................................................. 5

Figure 1.5  Barely visible impact damage (BVID) inspection on C-17 Globemaster III composite aileron using laser speckle interferometry (LSI) – 2021 ............................................. 5

Figure 2.1  Scalable manufacturing-induces flaws/detects and in-service damage evolution .......... 6

Figure 2.2  Three categories of full-field non-contact optical non-destructive inspection (NDI) techniques ................................................................................................................. 8

Figure 2.3  Actuation bandwidth and signal processing dimensionality of P-NDI and V-NDI .......... 9

Figure 2.4  Sensor bandwidth and signal processing dimensionality of P-NDI and V-NDI .............. 11

Figure 2.5  The sensing field-of-interest (FOI) from nanometer to meter scale ............................. 11

Figure 4.1  Quantity mapping relations of speckle and vision-based metrology (a) three LSI techniques: DSPI, LSP and SG, (b) DIC with subset window ................................................. 13

Figure 5.1  DSPI with two light beams traveling from different optical paths .............................. 22

Figure 5.2  In-plane ($u$) and out-of-plane ($w$) displacement measurements from DSPI by two different optical-setup configurations ................................................................. 22

Figure 5.3  LSP working principles and ray tracing on a surface with arbitrary roughness ........... 24

Figure 5.4  LSP out-of-plane and in-plane measurement ................................................................ 24
Figure 5.5 Working principle of SG which is a common-path interferometer................................. 25
Figure 5.6 Phase difference between sheared images in the reference and deformed states.............. 26
Figure 5.7 Generalized SG EM waves and Lamb waves interrogation........................................... 30
Figure 5.8 Dispersion relation and analytic mode shapes for fundamental modes $A_0$ and $S_0$ on
5mm Aluminum plate at 200 kHz............................................................................................... 31
Figure 5.9 Comparison between optical and numerical operations................................................. 34
Figure 6.1 2D Fourier transform (global phase) and Riesz transform (local phase)......................... 41
Figure 6.2 Frequency response of a log-Gabor filter as an even function for $B = 1, \cdots, 5$ .............. 45
Figure 6.3 Impulse response of the log-Gabor filter as an even function for $B = 2$ ...................... 45
Figure 6.4 Frequency response of the log-Gabor filter quadrature for $B = 1, \cdots, 5$ ....................... 46
Figure 6.5 Impulse response of a log-Gabor filter as an odd function for $B = 2$ ............................ 46
Figure 6.6 The relation of a log-Gabor bandpass filter with Riesz transform................................. 46
Figure 6.7 The visualization of a log-Gabor bandpass filter with Riesz transform............................ 47
Figure 6.8 Choice between (left) a monogenic filter on a real-valued signal and (right) a real-valued
(unconditional even in wavenumber domain) filter on a monogenic signal .............................. 48
Figure 6.9 L-shape membrane function and its modulated form – 3D.............................................. 50
Figure 6.10 L-shape membrane function and its modulated form – 2D............................................. 50
Figure 6.11 Phase contribution of each 5-carrier modulated signal component .............................. 51
Figure 6.12 Phase extraction with 2D Fourier transform and DPE via Riesz transform...................... 51
Figure 6.13 Imaging processing procedures with SSIM and unwrapping algorithms ........................ 53
Figure 6.14 SG with TPS and Riesz DPE – L-shaped damaged membrane function – noise free ........ 54
Figure 6.15  SG with TPS and Riesz DPE – L-shaped damaged membrane function – with speckle noise ....................................................................................................................................... 54

Figure 6.16  SG with TPS and Riesz DPE – 5-carrier modulated L-shaped damaged membrane function – noise free .............................................................................................................. 55

Figure 6.17  SG with TPS and Riesz DPE – 5-carrier modulated L-shaped damaged membrane function – with speckle noise .............................................................................................. 55

Figure 6.18  SG with TPS and Riesz DPE – 15-carrier modulated L-shaped damaged membrane function – noise free ............................................................................................................ 56

Figure 6.19  SG with TPS and Riesz DPE – 15-carrier modulated L-shaped damaged membrane function – with speckle noise .............................................................................................. 56

Figure 7.1  General processing procedures for speckle images ................................................................................................................................. 61

Figure 8.1  (a) HeNe laser reflected from a rough surface, (b) collimated HeNe laser reflected from a rough surface, (c) HeNe laser with beam expander, and (d) high-power diode laser illumination ........................................................................................................................................ 65

Figure 8.2  Laser source comparison about on the same scale, (a) Thorlabs HNL210LB linear polarized HeNe laser 632.8nm 21mW, (b) Thorlabs PL202 diode laser 635nm 0.9mW, (c) Thorlabs L638P700M high power diode laser 638nm 700mW ............................................................................................................. 66

Figure 9.1  Honeycomb composite panel with low velocity impact BVID at the center ............................................................................................................................ 67

Figure 9.2  BVID characterization via ultrasonic C-scan and X-ray CT-scan ................................................................................................................................. 68

Figure 9.3  Ultrasonic C-scan through the layers ................................................................................................................................. 68

Figure 9.4  Different perspectives of X-ray CT-scan ................................................................................................................................. 69

Figure 9.5  Baseline measured by (a) X-ray CT scan and (b) ultrasonic C-scan ................................................................................................................................. 69

Figure 9.6  Thermal loading LSP/SG experimental setup ................................................................................................................................. 70

Figure 9.7  (a) LSP/SG setup for honeycomb composite sandwich panels with high power diode laser, (b) integrated cameras for LSP and SG ................................................................................................................................. 71
Figure 9.8 (a) Speckle cloud created by high-power diode laser, (b) a close look of front panel to monitor speckle cloud, (c) real-time imaging processing with LabVIEW FPGA .............. 71

Figure 9.9 High-power diode laser drive and assembly ................................................................. 72

Figure 9.10 LSP results (A, LE, LP) with high-power diode (1st row) and diode (2nd row) laser .......... 73

Figure 9.11 LSP results (A, LE, LP) with high-power diode (1st row) and diode (2nd row) laser .......... 73

Figure 9.12 SG results (A, LE, LP) with (1st row) and without (2nd row) spectrum filter .................. 74

Figure 9.13 SG results (A, LE, LP) with (1st row) and without (2nd row) spectrum filter .................. 74

Figure 9.14 Shear distance (SD) comparison – panel front side .................................................... 75

Figure 9.15 Shear distance (SD) comparison – panel back side ..................................................... 75

Figure 9.16 LSP/SG – LE comparison with and without spectrum filter ........................................... 76

Figure 9.17 LSP/SG – LP comparison with and without spectrum filter .......................................... 76

Figure 9.18 Shear distance (SD) long duration study – front side panel – LE .................................... 77

Figure 9.19 Shear distance (SD) long duration study – front side panel – LP .................................... 77

Figure 9.20 Thermal loading SG experimental setup – version one .................................................. 78

Figure 9.21 Thermal loading SG experimental setup – version two ................................................. 78

Figure 9.22 Speckle clouds recorded in time sequence during cool down – small SD ....................... 79

Figure 9.23 Speckle clouds recorded in time sequence during cool down – large SD ...................... 79

Figure 9.24 Local phase (LP) extracted via Riesz transform – small SD ......................................... 80

Figure 9.25 Local phase (LP) extracted via Riesz transform – large SD ......................................... 80

Figure 9.26 Envelope extracted via Riesz bp filter to remove global fringes – small SD .................... 80

Figure 9.27 Envelope extracted via Riesz bp filter to remove global fringes – large SD ................. 80
Figure 9.28  Image projection on the 1st, 2nd and 3rd principal component directions – small SD ........ 81
Figure 9.29  Image projection on the 1st, 2nd and 3rd principal component directions – large SD ........ 81
Figure 9.30  Pulse laser/LDV baseline images with different vibration modes of the delaminated area appear at resonance frequencies ................................................................................................. 81
Figure 9.31  Setup of (a) integrated SG with HeNe laser, (b) piezo shaker attached on the back of the panel, and (c) function generator, amplify and oscilloscope ................................................................. 83
Figure 9.32  (Left) piezo shaker attachment and (right) expected wavefield under steady-state vibration with Hygiene wavefront ........................................................................................................ 84
Figure 9.33  (Left) measured wavefield at 20 kHz and (right) simulated mode shapes around 20 kHz ............................................................................................................................................. 84
Figure 9.34  The hidden damage at the right cross position (as reference left cross is intact position) was detected at 15 kHz and mode shape at 20 kHz .............................................................................. 85
Figure 9.35  Measured mode shapes at every 5 kHz incremental frequency ........................................ 85
Figure 9.36  Measured mode shapes with incremental frequencies to identify the local resonance ...... 86
Figure 9.37  Recall the pulse/laser LDV ........................................................................................ 86
Figure 10.1  Baseline measured by ultrasonic C-scan setup ............................................................ 87
Figure 10.2  Baseline measured by ultrasonic C-scan on impact location one ................................. 88
Figure 10.3  Baseline measured by pulse laser/LDV ........................................................................ 89
Figure 10.4  Wavenumber index (WI) and zero-lag cross-correlation (ZLCC) comparison .......... 89
Figure 10.5  (a) High-power diode laser controller with polarizer, cylindrical lens and diffuser set, (b) LSP setup with high-power diode laser for large-area inspection ......................................................... 90
Figure 10.6  Cylindrical lens for one side collimation ........................................................................ 91
Figure 10.7  Speckle clouds recorded in time sequence during cool down ........................................ 93
Figure 10.8  Instantaneous phase extracted via Riesz bp transform to create a phase map for each image so that they can be correlated to each other in time sequence........................................ 93

Figure 10.9  Image projection on the largest and the smallest principal directions by PCA to highlight the large variation regions and keep geometric features for imaging fusion........ 93

Figure 10.10  Post-processed image with different types of non-linear filters for segmentation and feature separation................................................................................................................ 93

Figure 10.11  Proposed thermal LSP/ADPNet (middle), baseline measured by pulse laser/LDV (left) and ultrasonic C-scan (right)...................................................................................... 94

Figure 12.1  Rayleigh surface wave particle motion................................................................. 103

Figure 13.1  Lamb waves (PSV) with(a) symmetric and (b) anti-symmetric modes................. 107

Figure 14.1  Elastic waves in planar waveguides........................................................................ 109

Figure 14.2  Lamb wave formation in an isotropic plate with free boundaries......................... 109

Figure 14.3  Lamb waves in (a) symmetric modes and (b) anti-symmetric......................... 113

Figure 14.4  Mode shapes in region III (1st row), II (2nd row) and I (3rd row)............................ 113

Figure 14.5  The fundamental modes, $A_o$ and $S_o$, and their particle motions..................... 125

Figure 14.6  Three regions for the propagating Lamb waves in the real (a) dimensional $(\omega,k)$ and (b) non-dimensional $(\Omega,K)$ planes: (1) $c_p > c_L$, $p$ and $q$ are real; (2) $c_L > c_p > c_T$, $p$ is pure imaginary and $q$ is real; and (3) $c_T > c_p$, $p$ and $q$ are pure imaginary .................................................................................................................. 126

Figure 14.7  Lamb wave dispersion relations with plane divisions.............................................. 127

Figure 14.8  Slowness curve Lamb waves in an isotropic plate of thickness $h$............................. 128

Figure 14.9  Lamb wave propagating mode dispersion curves.................................................. 131

Figure 14.10  Lamb wave propagating mode dispersion curves – phase velocity.................... 132
Figure 14.11  Lamb wave propagating mode dispersion curves – group velocity ................................. 133
Figure 14.12  Wavenumber and wavelength dispersion curves – SYM .................................................. 134
Figure 14.13  Wavenumber and wavelength dispersion curves – SYM – phase velocity .................. 134
Figure 14.14  Wavenumber and wavelength dispersion curves – SYM – group velocity ................. 134
Figure 14.15  Wavenumber and wavelength dispersion curves – ASYM ............................................. 135
Figure 14.16  Wavenumber and wavelength dispersion curves – ASYM – phase velocity ............... 135
Figure 14.17  Wavenumber and wavelength dispersion curves – ASYM – group velocity .............. 135
Figure 14.18  Asymptotes on pure imaginary (evanescent) & real (propagating) planes .............. 136
Figure 14.19  Dispersion curves on pure imaginary (evanescent) & real (propagating) planes .......... 137
Figure 14.20  Dispersion curves on pure imaginary (evanescent) & real (propagating) planes with asymptotes .................................................................................................................... 138
Figure 14.21  Complex roots projected on (a) real and (b) imaginary planes ...................................... 139
Figure 14.22  Real, pure imaginary and complex roots in different 3D perspectives ....................... 140
Figure 14.23  Real, pure imaginary and complex roots in 3D perspective with symmetry plane .... 141
Figure 14.24  Complex roots and up-conversion modes ..................................................................... 143
Figure 14.25  Complex roots and down-conversion modes ............................................................. 143
Figure 14.26  ZGV points for $\nu = 0.265$ ................................................................................................. 145
Figure 14.27  ZGV points for $\nu = 0.3$ ........................................................................................................ 145
Figure 14.28  ZGV points for $\nu = 0.34$ .................................................................................................... 145
Figure 14.29  For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands ........................................................... 150
Figure 14.30  For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands – phase velocity ............................. 150
Figure 14.31 For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands – group velocity ................................................ 150

Figure 14.32 Example of how to assign five filters as a filter bank at 1.25 MHz excitation..................... 151

Figure 14.33 SG simulated strain field and measured local phase (LP).................................................. 152

Figure 14.34 Riesz bp local phase (LP) of measured local phase (LP).................................................. 152

Figure 14.35 Riesz bp local energy (LE) of measured local phase (LP)................................................ 152

Figure 14.36 Excitation conditions for plate with three dummy damages............................................. 153

Figure 14.37 Riesz bp can be applied in either spatial or frequency domain to separate the wavefields ......................................................................................................................... 153

Figure 14.38 Wavefield separation into ASYM and SYM modes – initiating stage ............................ 154

Figure 14.39 Wavefield separation into ASYM and SYM modes – first reflection stage ................. 154

Figure 14.40 Wavefield separation into ASYM and SYM modes – multiple reflection stage ........... 154

Figure 14.41 Dispersion curves simulator – GUI front panel ................................................................. 155

Figure 14.42 Dispersion curves simulator – block diagram................................................................. 156

Figure 15.1 Elastic waves in planar waveguides.................................................................................. 157

Figure 15.2 SH wave dispersion curves for all propagating modes..................................................... 159

Figure 15.3 SH wave dispersion curves for all propagating modes – phase velocity ................. 160

Figure 15.4 SH wave dispersion curves for all propagating modes – group velocity .................. 160

Figure 16.1 Elastic waves in cylindrical waveguides.......................................................................... 161

Figure 16.2 Cylindrical coordinates $(r, \theta, z)$ and the radius of the cylinder is $a$ .......................... 161

Figure 16.3 Cylindrical wave longitudinal mode dispersion curves in rod.......................................... 167

Figure 16.4 Cylindrical longitudinal wave dispersion curves in rod – phase velocity............... 168
Figure 16.5  Cylindrical longitudinal wave dispersion curves in rod – group velocity ......................... 168
Figure 16.6  Cylindrical wave flexural mode dispersion curves in rod ................................................. 170
Figure 16.7  Cylindrical wave flexural mode dispersion curves in rod – phase velocity ...................... 171
Figure 16.8  Cylindrical wave flexural mode dispersion curves in rod – group velocity ...................... 171
Figure 16.9  Cylindrical wave torsional mode dispersion curves in rod ............................................... 173
Figure 16.10  Cylindrical wave torsional mode dispersion curves in rod – phase velocity .................... 174
Figure 16.11  Cylindrical wave torsional mode dispersion curves in rod – group velocity .................... 174
Figure 17.1  Analogy of elastic waves in plane and in rod ................................................................. 175
Figure 17.2  Dispersion curves analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod ........................................................................................................ 176
Figure 17.3  Phase velocity analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod ........................................................................................................ 176
Figure 17.4  Group velocity analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod ........................................................................................................ 176
Figure 17.5  Dispersion curves analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod ........................................................................................................ 177
Figure 17.6  Phase velocity analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod ........................................................................................................ 177
Figure 17.7  Group velocity analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod ........................................................................................................ 177
Figure 17.8  Dispersion curves analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod ........................................................................................................ 178
Figure 17.9  Phase velocity analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod ........................................................................................................ 178
Figure 17.10  Group velocity analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod
............................................................................................................................................. 178

Figure 18.1  Polarization states of EM waves (wiki) ................................................................. 181

Figure 18.2  A linear polarized EM wave propagating along the z direction (wiki) ............. 182

Figure 18.3  Circular polarized EM wave (wiki) ................................................................. 183

Figure 18.4  Elliptical polarized EM wave (wiki) ................................................................. 184

Figure 18.5  Production of plane polarized light from an unpolarized light source using a polarizer. The vertical lines in the polarizer represent the pass axis as a polarizing filter ............... 185

Figure 18.6  The transition from an unpolarized wave to a linearly polarized wave and then to a circularly polarized wave ........................................................................................................ 185

Figure 19.1  TE (left) and TM (right) polarization waves ........................................................... 187

Figure 19.2  Reflection and transmission of a plane wave incident at an interface between media of refractive indices \( n_i = n_r \) and \( n_s \), (a) plane of incident, (b) p-polarization, (c) s-polarization. ................................................................................................................................. 188

Figure 19.3  Incident and transmitted wave ............................................................................. 189

Figure 19.4  External and internal reflection ............................................................................ 192

Figure 19.5  External reflection ............................................................................................. 194

Figure 19.6  Coefficients for external reflection .................................................................... 195

Figure 19.7  Incident at Brewster angle and partially polarized refraction light ..................... 196

Figure 19.8  Water surface reflection without and with polarizer ......................................... 197

Figure 19.9  Internal reflection ............................................................................................. 197

Figure 19.10 Coefficients for internal reflection ..................................................................... 198

Figure 19.11 Refraction of light at the interface between two media. .................................... 199
Figure 20.1  (a) Out-going (blue) and in-coming (red) waves; and (b) the formation of standing wave with two-sided boundary conditions

Figure 20.2  Single boundary

Figure 20.3  Two parallel boundaries

Figure 20.4  Wavefront geometry

Figure 20.5  The ray tracing procedure and path phase compatibility

Figure 21.1  EM waves in planar waveguides

Figure 21.2  Three layers of medium for EM waves

Figure 21.3  Dispersion curves of planar waveguides

Figure 21.4  Lamb and EM waves analogy – frequency parameters

Figure 21.5  Lamb and EM waves analogy – wavevectors

Figure 21.6  Lamb and EM waves analogy – dispersion curves – Lamb convention

Figure 21.7  Lamb and EM waves analogy – dispersion curves – EM convention

Figure 22.1  Dispersion relation of EM waves in planar waveguides

Figure 22.2  Dispersion relation of EM waves in cylindrical waveguides

Figure 22.3  Dispersion curves of cylindrical waveguides – TE, TM, HE, EH

Figure 22.4  Dispersion curves of cylindrical waveguides – LP

Figure 22.5  Dispersion curves of cylindrical waveguides – TE, TM, HE, EH – LP

Figure 22.6  LP modes and their mode shapes

Figure 23.1  Bessel functions of (a) the 1st kind, $J_\alpha(x)$ for integer orders $\alpha = 0, 1, 2$ and (b) the 2nd kind, $Y_\alpha(x)$ for integer orders $\alpha = 0, 1, 2$
Figure 33.2  The plots above show the structure of $H_o^{(1)}(z)$ in the complex plane. ....................... 296

Figure 33.3  The plots above show the structure of $H_o^{(2)}(z)$ in the complex plane. ....................... 296

Figure 33.4  Modified Bessel function of (a) the 1st kind, $I_\alpha(x)$ for integer orders $\alpha = 0, 1, 2, 3$
and (b) the 2nd kind, $K_\alpha(x)$ for integer orders $\alpha = 0, 1, 2, 3$ ............................................. 297

Figure 33.5  The real and imaginary parts $h_o^{(1)}$ in the complex plane........................................... 299

Figure 33.6  The real and imaginary parts $h_o^{(2)}$ in the complex plane ........................................... 299

Figure 34.1  Volume $\Omega$ and its closed boundary $\partial\Omega$, containing (respectively enclosing) a source
(+ and sink (−) of a vector field $F$ .......................................................... 308

Figure 34.2  Surface $\Sigma$ with closed boundary $\partial\Sigma$ where $F$ could be the $E$ or $B$ fields ............. 309

Figure 34.3  A schematic view of how an assembly of microscopic dipoles produces opposite
surface charges as shown at top and bottom (left). ................................................................. 313
I. Preface

The dissertation consists of the following units where Unit I provided an overview of sensing technologies, state-of-the-art NDI/SHM techniques and motivations; Unit II explored vision-based measurement method including three types of laser speckle interferometry (LSI) and digital image correlation (DIC) with several experimental demonstrations wither in large-area or ultrasonic range; Unit III summarized the wave formation in different types of boundary conditions; Unit IV & V derived and discussed the wave properties in bounded elastic solids and optical dielectrics; Unit VI concluded above works to accomplish this journey.

I. Preface ......................................................................................................................................................... 1
II. Vision-based Large-area Damage Inspection ......................................................................................... 12
III. Waves ...................................................................................................................................................... 95
IV. Planar and Cylindrical Waveguides in Elastic Solids ......................................................................... 108
V. Planar and Cylindrical Waveguides in Optical Dielectrics .................................................................. 179
VI. Conclusions and Future Works ........................................................................................................... 238
1 Technology Brief

Physical quantities through phase are interpreted via Riesz transform in a higher dimensional hypercomplex domain with versatile choice of variables such as quadratures, quadrature vector norm, orientation and phase between variables, phase congruency (PC), cross-entropy (CE) or any user-defined variables. Dimensional expansion, projection, reduction are the mathematical manipulations behind this scheme. The limitation of phase stepping procedures such as TPS and SPS is that they never interpret the image in higher dimensions, for phase sensitive damage they work well but lack of choice for other variables to look the images in different perspective.

DPE is a universal higher dimensional phase estimator and accommodate all types of stress loading conditions in terms of frequency ranging from quasi-static, steady-state vibration, or ultrasonic transient state with correspondent image conditions. Commonly used excitation such as transient thermal seeks for the thermal induced strain gradient around damages, ultrasonic steady-state vibration is based on local resonance and the summation of modal contribution to highlight the damage features, and transient wave propagation is based on observing the wavefront variation or energy concentration around damage locations.

![Quantity mapping relations of speckle and vision-based metrology](image)

Figure 1.1 Quantity mapping relations of speckle and vision-based metrology (a) three LSI techniques: DSPI, LSP and SG, (b) DIC with subset window

Recently vision-based metrologies are developed for rapid quality assurance of mass manufactured parts and has the potential to inspect/monitor the damage of large-scale structures in real-time. LSI and DIC are the two speckle-based categories but working on very different principles. Their quantity mapping relations
are shown as (Figure 1.1) for comparison. These vision-based metrologies had drawn significant attention and interest in recent years in a broad range of commercial applications due to its unique and superior properties. The image from the camera acquired from the LSI speckles contain rich information from the phase, yet intrinsically noisy. In addition, the high-power diode laser can cause serious phase drifting and lower coherence. Correlation of phase signatures together with the bandpass filter was applied to supplant the phase-difference algorithm to localize damage regions.

To employ numerical DPE implementation via the Riesz transform shown as (Figure 1.2) with a log-Gabor bandpass filter enables a very compact, portable, and cost-effective system to evaluate the phase instantly and envelop the monogenic signal where the phase unwrapping is often needed in TPS/SPS as a subsequent step for featuring. Consequently, the phase-correlation operation is a more robust means than phase-difference operation to cope with interferometry speckles yielding the damage image with more assurance.

![Equation](equation.png)

**Figure 1.2** The relation of a log-Gabor bandpass filter with Riesz transform to generate a monogenic filter in wavenumber and spatial domain

For wave measurement above ultrasonic range, Nyquist criteria has to be satisfied so that the wavefield can be measured without aliasing. DSPI and LSP are pixelwise interferometry that is mostly over-sampled the wavefields in space whereas SG has one more restriction with its shear distance. Increasing shear distance can increase sensitivity as well but the strain measurement will be compromised since shear distance is like the gage length of a strain gage to measure the deformation difference.

Larger shear distance incorporates larger phase difference between the original and sheared images, but images may lose the correlation of speckle clouds corresponding to each other when shear distance is too large. Smaller shear distance on the other hand has a more precise strain measurement but the sensitivity is also smaller with fewer fringes and may be harder to quantify when deformation is comparative small.

Shear direction is also important for measurement when the direction of structural discontinuity caused by hidden damages is unknown. Rotating either shear device or camera is equivalent to take displacement gradient along different directions and look for damages subjected to a certain direction. Image fusion is
often applied at post-processing to fuse images from different directions when structure is subjected to static load or steady-state vibration or transient wave propagation with repeated excitations.

When considering dynamic measurement for steady-state vibration or transient wave propagation problem, the intrinsic wavelength within the structure must be at least twice larger than the shear distance to avoid spatial aliasing. Structure wavelength itself acts like the upper bound of shear distance and experimentally would be selected at least five times larger than shear distance to have a better description of the wavefields.

For ultrasonic measurement, since wavelength is getting smaller at higher frequency range, subtracting the phase between different stages will be difficult when shear distance is shrinking, and sensitivity is decreasing as well, and eventually it works like an LSP where the phase information only reflects the displacement at current state. At this scenario, quantifying strain field is getting harder whereas correlating speckle clouds for qualitative analysis would be one of the alternative solutions to accommodate this shear distance/structural wavelength or quantitative/qualitative dilemma.

The key milestones of the projects are summarized as below: understanding of fiber optics and damage inspection in 2016, unifying the representation of waveguides for elastic and EM waves in 2018, and making a real-time vision-based interferometry system for C-17 aileron damage inspection in 2021.

![Figure 1.3](image.png)

Laser ultrasonic composite inspection system (LUCIS) for rapid damage qualitative analysis and quantification – 2016
Figure 1.4  The analogy of Lamb waves in elastic solids and EM waves in dielectric materials: planar and cylindrical waveguides – 2018

Figure 1.5  Barely visible impact damage (BVID) inspection on C-17 Globemaster III composite aileron using laser speckle interferometry (LSI) – 2021
2 Introduction

High performance fiber reinforced composite and composite sandwich honeycomb sandwich structures were widely used in aerospace applications with its high specific stiffness and strength, excellent fatigue and corrosion resistance. Due to its layup structures and tailorability, shape, topology, topography and parametric optimization can be applied to meet the load requirements with objective weight, stiffness and strength. A critical safety issue for the design of primary composite structures is vulnerability and damage tolerance due to sudden and unpredictable loading such as foreign object impacts from bird strikes, hail, metal fragments, dropped tools, and small runway debris [1].

Composite structures are vulnerable to impact damage due to the brittle behavior of the matrix and to the low through-the-thickness strength. Specifically, the honeycomb composite panels which this paper will focus on are also susceptible to low velocity impact damage in which surface indentations are too small to be observed in routine inspection that can have significant internal macro-damage such as disbond, delamination, and core. These panels must be designed to sustain ultimate load with barely visible impact damage (BVID). While BVID is subjective by nature, it is defined [2] as small damages which may not be visible during heavy-maintenance, general visual surface inspections under typical lighting conditions from a distance of 1.5 m or the damage causing a permanent indentation depth of 0.25 mm to 0.51 mm at the outer skin surface. The damage state establishes the design strength values to be used in analyses.

Figure 2.1 Scalable manufacturing-induces flaws/detects and in-service damage evolution

These values are in compliance with the regulatory ultimate load strength requirements of FAR 25.305 [3]. The BVID which can be considered an evaluation of damage tolerance of the structures degrades the aspect
of structural integrity which should be addressed during the design and certification phases. Structural health monitoring (SHM) and nondestructive inspection (NDI) can be a viable solution to this problem. In order to select an appropriate technique for structure integrity evaluation, ASTM E2533 [4] serves as a practical guide on polymer matrix composites for aerospace applications.

2.1 NDI/SHM

Nondestructive testing (NDT), nondestructive evaluation (NDE) and nondestructive inspection (NDI) are similar concepts where NDT is a more general term of testing methods, NDE focuses on the process of evaluation and NDI emphasizes the techniques of inspection. NDI in the field of aviation was developed since 1940s and represented a major innovation in aircraft maintenance and inspection technology. Nondestructive inspection implies a method that will detect "defective" parts and assemblies while avoiding any damage to serviceable items during the course of the inspection process. NDI has grown exponentially over the past 25 years in both application and innovation. It has done more to improve safety than just about any other existing technology.

"The Big Five" of NDI are the core disciplines that represent the bulwark of the aircraft maintenance and manufacturing industries including 1) fluorescent penetrant testing (FPT), 2) magnetic particle testing (MPT), 3) x-ray/neutron radiography/tomography (RT), 4) eddy current testing (ECT) and 5) ultrasonic testing (UT). Other techniques [5]–[7] such as visual inspection (VI), liquid penetration testing (LPT), acoustic emission (AE), infrared thermography (IRT), neutron radiography (NR), fiber Bragg grating (FBG), terahertz (THz) microwave testing, holography (HL), electronic speckle pattern interferometry (ESPI), laser speckle photography/photometry (LSP), shearography (SG), and digital image correlation (DIC) are also widely used in industrial applications to observe damage sensitive parameters (e.g., intensity, frequency, phase, polarization state, dielectric properties, or other observable physical quantities) in specific domains (e.g., real, complex, hypercomplex or even higher dimension domain) for various damage detection applications.

Ultrasonic guided wave based techniques have shown great potential for practical applications in nondestructive inspection (NDI) and structural health monitoring (SHM) for reducing the risks of catastrophic failures in critical structures like aircraft [8]–[12]. By investigating the scattered ultrasonic guided waves in a structure using appropriate signal/image processing algorithms, a wealth of information about the hidden details of the structure can be unearthed, including information about the location of the damage, if present, and its characteristics. NDI/SHM techniques can be broadly classified as active or passive: while in active sensing, a well-controlled artificial actuation source with a dedicated power supply
is required to excite ultrasonic guided waves in the structure; in passive sensing, these can emanate due to an impact source [13] or acoustic emission [14], [15] or from ambient disturbances generated naturally from turbulent flows [16]. In each of the two cases, scattered waves can be sensed using either contact (e.g., a network of piezoelectric (PZT) sensor arrays [17], [18]) or fully non-contact means (e.g., air-coupled transducers, LDV) [19]–[24].

Nowadays the advances of computer vision, machine learning and parallel computing architectures (from CUDA to Tensor Cores) with advanced image processing techniques push the technologies to the next stage with full-field non-contact and real-time inspection capabilities.

![Figure 2.2](image)

**Figure 2.2**  Three categories of full-field non-contact optical non-destructive inspection (NDI) techniques

### 2.2 Point-based Laser NDI

Point-based non-destructive inspection (P-NDI) system employing laser Doppler vibrometer (LDV) was widely used as baseline due to its extremely high spatial sensitivity (sub-nano to picometer scale) and very high dynamic range (DC to GHz) under modern signal processing scheme (quantization and modulation/demodulation scheme of AM, FM and PM signals) so that discrete wavefield can be captured by pointwise scanning and reconstructed after a series of acquisition.
The scanning LDV (SLDV) has been widely used recently and has attracted a lot of interest in investigating full-field ultrasonic guided wave propagation in a non-contact manner; however, it is limited to point-by-point measurement/sensing (i.e., one point at a time). In order to sense all the data available from the scan region for reconstructing the full-field, this point-by-point technique requires the excitation to be repeated for each sensing event. This makes the process not only time-consuming but also inconsistent, as the physical excitation and/or the system itself can change over the repeated sensing cycles.

2.3 Vision-based Laser NDI

Vision-based laser sensing methods such as interferometric techniques including moiré interferometry, holographic interferometry, electronic/digital speckle pattern interferometry (holography/ESPI/DSPI), laser speckle photography/photometry (LSP) and shearography (SG) allow full-field non-contact measurement of surface deformations or deformation gradients. They generally require a coherent light source and have stringent requirements of experimental conditions that limit their practicality in NDI/SHM.

In order to overcome these limitations, a sensing paradigm capable of taking qualitative and quantitative measurements simultaneously is desirable. The vision-based non-destructive inspection (V-NDI) system captures the wavefield of at least an order higher spatial resolution and nearly continuously by a CMOS...
camera with a global shutter for every time instance without repeating the experiment tens of thousand times to reassemble the wavefield where P-NDI does. The development of V-NDI has two routes, interference and correlation. Interference stands for the interrogation between two electromagnetic (EM) waves and interference pattern was generated instantly whereas correlation stands for image correlation as an image conditioning algorithm and always suffers a certain delay.

Traditional holography/ESPI/DSPI relies on the observation of interference pattern and digital image correlation (DIC) simply correlates the images under different deformation stages. LSP/SG is the combination of interference and correlation (fetches the interference speckles firstly and then correlates speckles in time sequence or different states) and takes the advantages of both routes.

The selection of speckle types depends on the measurement environment and scale of measurement. Laser speckles are designed for precise measurement with interferometry and rigid body motion always has to be taken care with iterative or asymptotic estimator (e.g. Bayesian update or inference) and painted speckles at the other side is more suitable for large deformation measurement and is implicitly rigid body motion free with its feature matching and speckle tracing algorithm. Another major difference of laser speckle and painted speckle is that the first one was defined in Eulerian coordinates and the second one was in Lagrangian coordinate.

2.4 The Transition of Sensing Technologies

When calculating the data throughput for PD and CMOS, they are exactly having the same level of processing throughput provided by current state-of-the-art microprocessor chips (MPU), for example, a PD generally takes single point of data with high resolution digitizer (12 to 16 bits) at hundreds of MHz to GHz range (e.g. Thorlabs InGaAs Avalanche PD, APD450C has 3dB bandwidth 0.3 to 1600 MHz and APD310 has 3dB bandwidth 5 to 1000 MHz) and that makes the total data throughput to be 12 to 16 Gbits/s; a 10 Mega pixels 1000 frames per second monochrome CMOS camera (pixel depth 8 to 14 bits) has about the same level of data throughput from 8 to 14 Gbits/s. As the computing power is getting cheaper and faster, the computational cost/throughout isoline (Figure 2.4) will move toward the right upper direction.

This means no matter which technique we select, the maximum data throughput shown as (Figure 2.4) is always the bottleneck and it also limits the types of measurement to be a point-based (we called P-NDI in this paper) or a vision-based (V-NDI) inspection system. The transition of technologies from P-NDI to V-NDI is quite promising, from single point to multi-point to mega-pixels resolution, from early-stage hologram film plate to TV-holography also called electronic speckle pattern interferometry (ESPI) to laser
speckle contrast imaging (LSCI) to laser Doppler imaging (LDI). The faster we can process the image, the larger bandwidth and the higher frequency of data we can resolve.

![Image](image_url)

**Figure 2.4** Sensor bandwidth and signal processing dimensionality of P-NDI and V-NDI

![Image](image_url)

**Figure 2.5** The sensing field-of-interest (FOI) from nanometer to meter scale

In summary, the technologies migrate from point detection to vision (2D image) to video (3D volumetric data) to higher dimension data, from analog element to digital sensor, from baseband to passband, from AM to FM to PM, from small bit rate to high bit rate, from small bandwidth to high bandwidth, from sequential to parallel processing, from software to onboard hardware processing, etc.

### 3 Motivation and Plan of Works

Due to the slow (a few hours) and small area (inch by inch) inspection of traditional NDI/NDT techniques and complex phase stepping hardware setup for laser speckle interferometry (LSI), a vision-based full-field LSI with numerical phase estimation capability was proposed for large area (meter by meter) and fast (a few seconds) inspection. Several experiments are performed and compared with traditional baselines measured by pulse laser/LDV, X-ray CT scan and ultrasonic C-scan. In order to understand the how optical metrology works, the interrogation of elastic and EM waves are studied with analogies in many perspectives.
## II. Vision-based Large-area Damage Inspection

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-field Optical NDI Techniques</td>
<td>13</td>
</tr>
<tr>
<td>Working Principles of Laser Speckle Interferometry</td>
<td>20</td>
</tr>
<tr>
<td>Phase Extraction and Estimation</td>
<td>35</td>
</tr>
<tr>
<td>Digital Image Processing</td>
<td>58</td>
</tr>
<tr>
<td>Experiment Preparation for LSP/SG with DPE</td>
<td>64</td>
</tr>
<tr>
<td>Experiment Demo – Honeycomb Composite Panel</td>
<td>67</td>
</tr>
<tr>
<td>Experiment Demo – C-17 Globemaster II Composite Aileron</td>
<td>87</td>
</tr>
</tbody>
</table>
4 Full-field Optical NDI Techniques

Recently vision-based metrologies are developed for rapid quality assurance of mass manufactured parts and has the potential to inspect/monitor the damage of large-scale structures in real-time. LSI and DIC are the two speckle-based categories but working on very different principles. Their quantity mapping relations are shown as below for comparison. These vision-based metrologies had drawn significant attention and interest in recent years in a broad range of commercial applications due to its unique and superior properties.

![Quantity mapping relations of speckle and vision-based metrology](image)

(a) (b)

Figure 4.1 Quantity mapping relations of speckle and vision-based metrology (a) three LSI techniques: DSPI, LSP and SG, (b) DIC with subset window

In this section four full-field optical NDI techniques: DSPI, LSP, SG, and DIC are described in detail. Regarding the full-field measurement methods, two groups shown as (Figure 4.1) can be categorized: laser speckle (DSPI, SLP and SG) and painted speckle (DIC). [25]–[29]

4.1 Digital Speckle Pattern Interferometry (DSPI)

A stricter definition of laser speckle related techniques usually quite confused and the terminologies are used interchangeable. To be clear, the term “interferometry” should only be used when the device itself can provide a certain level of interferometry or another name should be used, but there is still an exception with ambiguous terminology like holography which is a Mach–Zehnder interferometer with hologram film plates as sensing elements before CMOS was introduced, and after so called TV-holography which is known as ESPI/DSPI. DSPI measure surface displacement but suffer with ambient noise a lot because of

13
the reflection beam and reference arm are from different source where shearography has been proven to be more practical with a simpler setup and does not require special vibration isolation [7], [30], [31].

Interferometry is a measurement method using the phenomenon of interference of waves. Interference occurs when two or more waves overlap each other in space. Interferometers separate illuminated light so that it travels two independent paths, one of which follows a reference surface and the other the object (structural) surface. The separated light beam then recombines and are directed to a digital camera that measures the resultant light intensity over the FOV simultaneously and interferograms are formed.

Modification of the classical interferometry such as Michelson interferometer is digital speckle pattern interferometry (DSPI) which is a technique which uses laser light, together with video detection, recording and electronic processing and imaging processing to visualize static and dynamic displacements of components with optically rough surfaces. DSPI has evolved to become one of the most accurate optical technique that is widely used for measurements to study object deformations, vibrations, temperature gradients, material properties, defects, damage assessment, among many other applications. The main challenges in speckle interferometry manifest on phase distribution extraction leading to the direct determination of surface displacements. Furthermore, speckle fringe patterns are characterized by a strong speckle noise and thus are intrinsically noisy. This undesired speckle noise negates the phase distribution estimation accuracy, and therefore, a denoising scheme is necessary to reduce the speckle noise from speckle fringe patterns before image analysis.

Speckle pattern images are formed from the illumination of the laser on optically rough surfaces. As the object is subject to a variety of loading conditions, the speckle pattern changes. Subtraction of the different states speckle patterns resulting fringes, which depicts deformation of the tested object. DSPI measures surface displacements but suffer from ambient noise because of the reflective beam and reference arm arise from different optical paths where shearography [8], [31], [32] has been proven to be more practical in industrial environments with a simpler setup and does not require special vibration isolation.

4.2 Laser Speckle Photometry/Photography (LSP)

Laser speckle photography/photometry (LSP) was proposed by Shin et al. [33] at MIT CSMIL with advanced correlation function called Structural Similarity Index Measure (SSIM) combining several parameters to outperform traditional correlation functions [34]–[37] including normalized cross-correlation (NCCR), normalized cross-covariance (NCCV), etc. Other correlation functions [38], [39] such as sum of squared differences (SSD), normalized sum of squared differences (NSSD), zero-normalized sum of
squared differences (ZNSSD) are discussed to quantify speckle variation and evaluate surface displacement in time sequence for most of the DIC applications.

Although DIC can have accurate evaluation of displacement and strain fields from painted speckles, its spatial resolution is proportional to its field of view (FOV) and in-plane components are at least an order smaller than out-of-plane components for most of the structural dynamic problems. For guided wave excitation in ultrasonic range, out-of-plane displacement is generally in sub-nm range and in-plane is even smaller in just a tens of pm that makes the FOV shrink to a few mm by mm so that this minute displacement can be resolved with pixels. Painting speckles in this very small area is not practical and tons of post-processing work has to be done to stitch the images together. Otherwise, the processing time to trace speckle relative displacement in global coordinate is very time consuming which makes DIC almost impossible be an in-situ real-time tool for large area inspection, say one meter by one meter.

As long as the processor is more powerful especially with field-programmable gate array (FPGA) image volumetric data can be processed in real-time in nowadays technologies. Unlike DSPI having instantaneous fringes, laser speckle (LS) techniques rely on correlating images from different loading stages in time sequence. This makes LS technique more flexible than DSPI since there are tons of correlation functions to apply other than just looking for the intrinsic phase variation in the fringes.

Unlike DIC with painted speckle which cannot capture out-of-plane displacement with a single camera, LSP itself deploys random distributed laser speckles and intrinsically has spatial resolution ranging in nm to pm scale. Other advantages of LSP are its FOV has nothing to do with its spatial resolution, the larger power of laser the larger FOV, and the correlation can be done in real-time for in-situ inspection, say 30 frames per second where DIC has to spend few minutes to process a single frame. In contrast to DSPI relies on interferometry with reference arm, LSP and shearography \[7\], \[30\], \[40\], \[41\] self-interfere either from rough surface scattering field or from duplicated image created by shear device (e.g. shear plate, wedge lens, tilt mirror, Bragg cell, etc.) and are very insensitive to ambient noise.

4.3 Shearography (SG)

Digital Speckle Pattern Shearing Interferometry (DSPSI), more commonly known as shearography (SG), introduced by Hung and Taylor \[42\] in 1973, was developed from DSPI. SG can measure the surface deformation gradients, from which the surface strain components can be determined. The interference is usually presented as a fringe pattern, or called shearogram, although both DSPI and SG are full-field optical techniques which generate interference patterns by combining a reference beam with the reflected light from an optically rough surface, a significant difference exists between the manner in which these two
methods form their resulting fringe patterns. SG is a common-path interferometer. The reference beam is derived from the same laser beam as that used to probe the structure, thus it represents the interference (two coherently superimposed) of two sheared displayed interferometric speckle patterns of the same wavefields. This distinction has several advantages including tolerance to rigid body motions, reduced laser coherence requirements, compact design, convenient sensitivity adjustment, and direct measurement of differential displacement. The advent of modern video cameras has allowed designers to build upon these intrinsic advantages producing a portable low-cost sensor. In this configuration, albeit the limitation of low contrast fringes, SG has gained acceptance as a nondestructive inspection technique [8], [31], [32], which is commonly used on aerospace structures and materials [8], [9], [43]–[45].

In any SG system, the field of view (FOV) is illuminated by the laser (light source) via an expander and/or a diffuser. The light, physically represented by electromagnetic wavefields, scattered from the rough surface of the FOV forms a laser speckle pattern which is imaged through a shearing device onto a CCD or CMOS camera (image sensor). The shearing device serves to divide the scattered wavefields into two identical paths but laterally displaced images sensed by the camera. Since both wavefields travel the same common wave path, shear interferometry makes low demands on the light coherence and at the same time provides enhanced stability and robustness compared with conventional interferometric methods. Most importantly since SG employs a common-path interferometer configuration, it is resilient to the environmental disturbances and vibrations which can be applied to industrial environments.

The two laterally sheared (displaced) scattered wavefields are recombined coherently into images to produce interferometric speckle patterns at the sensor of the camera. The light contribution to each speckle in the pattern is scattered from points on the rough structural surface separated by the shear distance. Any subsequent deformation on the surface will result in the phase difference between the light scattered from those points resulting in a change in intensity of each speckle. A comparison of images recorded before and after some loading event, commonly by digital subtraction and rectification, results in a fringe pattern where the fringes represent a locus of points with the same magnitude (or intensity) of deformation gradient. Although both systems share the noisy character from speckle noise, SG is not as prone as to environmental disturbances due to SG’s common-path nature.

4.4 Digital Image Correlation (DIC)

Digital image correlation (DIC), which is a non-interferometry optical technique, extracts the surface deformations by comparing digital images of the structure’s surface captured before and after deformation. DIC does not require a complicated optical platform for measurements, is less sensitive to ambient
environmental factors, and white light can be used for illumination instead of a coherent laser source making it suitable for in-field NDI/SHM applications. Also, unlike other interferometry techniques (which often present results in the form of fringe patterns), phase analysis of the fringe patterns and subsequent phase unwrapping processes are not required. These factors have led to the widespread adoption of DIC as a practical, effective, and reliable tool for full-field measurement of surface deformations in the field.

Even though the first known use of image processing dates back to the 1960s, the determination of surface deformations using a single camera with DIC image analysis started in the early 1980s. Peters and Ranson [46] in 1982 established a foundation for converting digitized 2D images of a planar specimen undergoing in-plane deformations into full-field measurement of displacements. This marked the first use of computer vision for measuring 2D deformations and deformation gradients. Computer vision is a branch of computer science that develops computational methods to extract information from visual images. In this context, it refers to computer-based, non-contact, surface deformation measurement methods used for studying solid media. Chu et al. [47] developed the mathematical theory of DIC and performed several validation experiments to demonstrate the viability of the proposed methodology.

Due to the limitations of 2D DIC vis-à-vis practical engineering measurements, Luo et al. [48], [49] extended 2D DIC to StereoDIC or 3D DIC for making full-field 3D displacement and strain measurements in 1993. Advances in 2D and StereoDIC up to the early 2000s have been summarized in [50], a review on the 2D DIC was reported in [38], and recent advances/applications of DIC image analysis have been reviewed in [39], [51]. Advances in computers and digital imaging technology have enabled 2D/3D DIC to extend from static or quasi-static to dynamic applications, such as low-frequency vibration (in the range of kHz) and full-field modal analysis for characterizing the behavior of structures[52]–[56].

Digital image correlation (DIC), which is a processing software, extracts the surface deformations or surface deformation gradients by comparing digital images of the structure’s surface captured before and after deformation. DIC does not require a complicated optical platform for measurements, is less sensitive to ambient environmental factors, and white light can be used for illumination instead of a coherent laser source making it suitable for in-field NDI/SHM applications. Also, unlike other interferometry techniques (which often present results in the form of fringe patterns), phase analysis of the fringe patterns and subsequent phase unwrapping processes are not required. These factors have led to the widespread adoption of DIC as a practical, effective, and reliable tool for full-field measurement of surface deformations.

Chu et al. [47] developed the mathematical theory of DIC and performed several validation experiments to demonstrate the viability of the proposed methodology. Due to the limitations of 2D DIC vis-à-vis practical
engineering measurements, Luo et al. [48], [49] extended 2D DIC to StereoDIC or 3D DIC for making full-field 3D displacement and strain measurements in 1993. Advances in 2D and StereoDIC up to the early 2000s have been summarized in [50], a review on the 2D DIC was reported in [57], and recent advances/applications of DIC image analysis have been reviewed in [39], [51]. Advances in computers and digital imaging technology have enabled 2D/3D DIC to extend from static or quasi-static to dynamic applications, such as low-frequency vibration (in the range of kHz) and full-field modal analysis for characterizing the behavior of structures [52]–[56].
4.5 Summary

All the interferometry techniques are based on the observation of optical path length represented as displacement where displacement = rigid body motion (translation + rotation) + deformation. The gross working principles and essential characteristics of full-field optical NDI systems are summarized in (Table 4.1) and (Table 4.2) shown below and will be discussed in details in the next section.

Table 4.1 General characteristics of the four full-field optical NDI techniques

<table>
<thead>
<tr>
<th>Technique</th>
<th>Light source</th>
<th>Speckle</th>
<th>Sensing device</th>
<th>Measurand *</th>
<th>Generic measurand component **</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSPI</td>
<td>Narrowband laser, high coherence</td>
<td>Laser induced</td>
<td>Can be implemented without camera</td>
<td>Displacement</td>
<td>Out-of-plane</td>
<td>nm</td>
</tr>
<tr>
<td>LSP</td>
<td>Narrowband laser, high coherence</td>
<td>Laser induced</td>
<td>Camera (monochrome preferable)</td>
<td>Displacement</td>
<td>Out-of-plane</td>
<td>nm</td>
</tr>
<tr>
<td>SG</td>
<td>Wideband white light, low coherence</td>
<td>Paint digital speckle or random pattern</td>
<td>Camera (monochrome or color)</td>
<td>Deformation gradient ***</td>
<td>Out-of-plane</td>
<td>nm/m</td>
</tr>
<tr>
<td>DIC</td>
<td>Wideband white light, low coherence</td>
<td>Paint digital speckle or random pattern</td>
<td>Camera (monochrome or color)</td>
<td>Displacement and strain</td>
<td>In-plane</td>
<td>Pixel in full-scale</td>
</tr>
</tbody>
</table>

* Homodyne interferometry for displacement and heterodyne interferometry for velocity.

** Generic out-of-plane and can be in-plane displacement or deformation gradient when applying symmetric light source for LSI or with dual cameras setup for DIC.

*** When taking the gradient of displacement with rigid body motion, it turns out to be the deformation gradient

Table 4.2 The working principles of full-field optical NDI systems

<table>
<thead>
<tr>
<th>Technique</th>
<th>Reference arm</th>
<th>Noise susceptible</th>
<th>Coordinate system</th>
<th>Instantaneous pattern</th>
<th>Correlation treatment</th>
<th>Correlation output</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSPI</td>
<td>Yes</td>
<td>Yes</td>
<td>Eulerian</td>
<td>Interference fringes</td>
<td>Pixel tracing</td>
<td>No correlation Already have fringe</td>
</tr>
<tr>
<td>LSP</td>
<td>No</td>
<td>No</td>
<td>Eulerian</td>
<td>Interference* speckles</td>
<td>Pixel tracing</td>
<td>Fringe or other index</td>
</tr>
<tr>
<td>SG</td>
<td>No</td>
<td>No</td>
<td>Lagrangian</td>
<td>Deformed speckles</td>
<td>Speckle tracing</td>
<td>Nodal displacement</td>
</tr>
</tbody>
</table>

* Low contrast fringe pattern
5 Working Principles of Laser Speckle Interferometry

Among the three laser speckle interferometry (LSI) techniques, this section focuses on the working principle of DSPI, LSP and SG, and the working principle of DIC can be seen in the review paper [51] and thus, is not discussed here. Laser is a highly coherent, convergent, and narrowband (ideally monochromatic) beam of electromagnetic radiation. At far field, the electromagnetic radiation phenomenon can be described by a plane electromagnetic (EM) wave characterized by time-varying electric (E) and magnetic (H) fields in space. The E and H wavefields are orthogonal to each other and to the direction of propagation. These two time-varying fields are mutually dependent through Maxwell’s equations; thus hereafter only electric fields are selected for discussion. An EM wave $E_1$ can be mathematically described in electric field as

$$E_1 = A \exp i \left( k_{EM} \cdot r - \omega t + \phi \right)$$ Eq. (5.1)

Physically the EM wave propagates along direction $k_{EM}$ in optical path direction $r = (x, y, z)^T$ with a given frequency $\omega$ and oscillating with a real-valued amplitude $A$ and the initial phase $\phi$. Another EM wave $E_2$ oscillating in the same direction with amplitude $B$ and phase difference $\varphi$ with respect to $E_1$ can be described as

$$E_2 = B \exp i \left( k_{EM} \cdot r - \omega t + \phi + \varphi \right)$$ Eq. (5.2)

When two waves $E_1$ and $E_2$ interfere each other, their fields are superimposed as $E$

$$E = E_1 + E_2 = [A + B \exp i \varphi] \exp i \left( k_{EM} \cdot r - \omega t + \phi \right)$$ Eq. (5.3)

Then the intensity can be calculated by taking the inner dot product of its complex conjugate:

$$I = E \cdot E^* = A^2 + B^2 + 2AB \cos \varphi = I_{DC}(1 + m \cos \varphi)$$ Eq. (5.4)

where $I_{DC} = A^2 + B^2$ is the DC offset, $m = 2AB / I_{DC}$ can be viewed as contrast (or visibility), or as an intensity modulation function. Typically, in the fringe pattern, the DC offset and the modulation respectively are slowly and smoothly varying functions. The phase function is also smoothly varying. As a result, the intensity is a rapidly oscillating function. Note that all the variables in Eq. (5.4), $I_{DC}$, $m$, and $\varphi$ are functions of $(x, y, t)$. 

20
Regardless which type of interferometry is employed, a primary purpose is to evaluate the phase difference \( \phi(x, y, t) \) from \( I(x, y, t) \) where important information is encoded in the phase projected to \( z \) plane, say from naked eye or a camera. In DSPI, the phase difference \( \phi \) arises from the optical path between the reflected beam from the structure and the reference beam whereas in SG the phase difference is from the reflective beam from the location and laterally displaced location.

### 5.1 Digital Speckle Pattern Interferometry (DSPI)

DSPI [45] is typically composed of a Michelson interferometer with a reference arm shown in (Figure 5.1), thus it is very sensitive to ambient noise and minute optical path length change caused by deformation and rigid body motion (translation and/or rotation). Several variants of DSPI have been proposed to mitigate its susceptibility to ambient noise. For example, by converting from homodyne (AM) to heterodyne (FM and PM) under the framework of Mach–Zehnder interferometer through a signal modulation scheme, the detection sensitivity is no longer stationary in the frequency band. Another means is to add a resonance cavity to build a scanning Fabry-Perot interferometer to detect frequency variations and then process the signal solely in the frequency domain.

The DSPI almost follows the same working principle as the LDV where the difference lies in the sensing element; one uses CMOS camera and the other uses photo detector (PD). From data throughput and sensor point of view, DSPI uses high resolution camera (few mega pixels) at hundreds of frames per second and LDV with displacement (homodyne) or velocity (heterodyne) decoder acquires samples with high resolution digitizer (12 to 16 bits) at GHz range. The DSPI itself can be treated as the 2D version of LDV with displacement decoder and both trace the variation of speckle intensities in the same manner. At present most of the cameras are still working on baseband unless further modulation and image processing techniques can be applied to push the image into passband.

Two illuminated light beams (Figure 5.1) travel two independent paths from beam splitter. One follows a reference surface, and the other is incident on the FOV of the structural surface and reflected. The separated light beam then recombines forming fringes instantaneously and are directed to a digital camera that measures the resultant light intensity over the FOV. Since the two light beams travel in different paths, one from the device itself and another from the structure, this configuration becomes very susceptible to ambient noise and need additional isolation of vibration. However, there are still several advantages to apply DSPI, the fringes are instantaneous, and the sensing element is not limited to digital camera, say using a hologram film to store the image on an analog agent or simply projecting the image on a wall for direct observation.
By arranging the optical set-up configuration, the out-of-plane and in-plane displacements can be measured separately shown in (Figure 5.1) [58]. Without loss of generality, a two-dimensional x-z plane with the x-direction along the surface and the z-direction parallel the surface normal is illustrated. For measuring out-of-plane displacement, a single illumination (or emission) is incident on the structural surface and observations from two cameras; one aligned with the reflected beam direction gives the out-of-plane displacement ($w$) and the other gives the combination of the in-plane displacement ($u$) and out-of-plane displacement ($w$), after interfering with the reference beam are shown in the left side of (Figure 5.2).

Figure 5.1  DSPI with two light beams traveling from different optical paths

Figure 5.2  In-plane ($u$) and out-of-plane ($w$) displacement measurements from DSPI by two different optical-setup configurations.
In contrast, for measuring in-plane displacement \( (u) \), two laser beams symmetrically illuminate the structural surface and a single observation direction from the camera is normal to the surface shown in the right side of (Figure 5.2). Light beams are incident on the structural surface or split from the beam splitter are marked in red; the light beams reflected from the structural surface are marked in blue.

It is convenient to define \( \mathbf{k}_{EM} = \mathbf{k}_s - \mathbf{k}_o \) as the sensitivity vector that is directly related to the displacements in DSPI and deformation gradients later in SG. The sensitivity vector subtracts the observation or camera direction, wavevector \( \mathbf{k}_o \), from the laser emission (or incident) direction toward structural surface, wavevector \( \mathbf{k}_s \).

Due to its inherent optical setup in DSPI with the reference beam directly emitting from the laser with no need to go through the optically rough surface, the interferometric fringes can be observed instantly instead of post-processed. In contrast, SG requires random speckles as information carrier on the surface (diffuser is required if the surface is smooth) for each state then processes the images from the two states to perform fringe patterns. The phase-difference operation by subtraction in space-time domain for any time instance can be represented as

\[
\varphi_{\text{ref}}(x, y, t) = \mathbf{k}_{EM} \cdot \mathbf{r} \\
\varphi_{\text{deform}}(x, y, t) = \mathbf{k}_{EM} \cdot \mathbf{r}'
\]  

\[
\rightarrow \varphi_{\text{sub}}(x, y, t) = \varphi_{\text{deform}}(x, y, t) - \varphi_{\text{ref}}(x, y, t) = \mathbf{k}_{EM} \cdot \Delta \mathbf{r} \quad \text{Eq. (5.5)}
\]

where \( \mathbf{r} = (x, y, z)^T \) is position in Cartesian coordinate, \( \Delta \mathbf{r} = (u, v, w)^T \) is displacement vector. The two images at two different time instances can be cross-correlated in the proposal DPE technique, called phase-correlation, so that the location with damage induced fringes can be enhanced and separated.

\[
\varphi_{\text{sub}}(x, y, t) \rightarrow \varphi_{\text{sub}}(x, y, t_1) \rightarrow \varphi_{\text{corr}}(x, y, \Delta t) = \varphi_{\text{sub}}(x, y, t_2) \otimes \varphi_{\text{sub}}(x, y, t_1) \quad \text{Eq. (5.6)}
\]

where \( \otimes \) denotes cross-correlation. Note that the image correlation is not an essential step for DSPI and is thus demonstrated as blurred at the right-hand side in contrast to (Figure 5.5).

To trace the phase differences from the incident and reflected laser beams, the difference of the directional wavevectors denoted as \( \mathbf{k}_o \) and \( \mathbf{k}_s \) are multiplied with displacement vector \( \mathbf{d} \). The phase of the other beam split by the beam splitter does not change and can be treated as a constant which is omitted in the phase equation.
5.2 Laser Speckle Photometry/Photography (LSP)

LSP needs laser speckles created either by the scattered field from object rough surface or a diffuser to deploy speckle cloud so that CMOS camera can extract speckle intensity for every loading stage. One major difference between DSPI and LSP is the type of instantaneous pattern where the previous one is interference fringes and later one is interference speckle cloud. This makes LSP need additional works such as correlating speckle clouds in different time sequence which is not a necessary step for DSPI.

![LSP Working Principle](image)

Figure 5.3 LSP working principles and ray tracing on a surface with arbitrary roughness

When operating LSP shown as (Figure 5.3), speckle clouds was recorded in time sequence and then correlating (a more general operation not limited to subtraction but with more advanced correlation functions) in sequence with specified reference frame as the zero-return key, say the first frame.

![LSP out-of-plane and in-plane measurement](image)

Figure 5.4 LSP out-of-plane and in-plane measurement
Instead of having instantaneous fringes, LSP records and correlates speckle clouds induced either by out-of-plane or in-plane deformation under different configurations shows as (Figure 5.4) which is exactly the same as DSPI without beam splitter for out-of-plane detection. Without instantaneous phase difference in space-time like DSPI, phase difference with specified reference frame can be represented as below and it’s no longer a function of $t$ but $\Delta t = t_2 - t_1$

\[
\begin{align*}
\varphi_{\text{ref}}(x, y, t_1) &= k_{EM} \cdot r \\
\varphi_{\text{deform}}(x, y, t_2) &= k_{EM} \cdot r' \\
\rightarrow \quad \varphi_{\text{sub}}(x, y, \Delta t) &= \varphi_{\text{deform}}(x, y, t_2) - \varphi_{\text{ref}}(x, y, t_1) = k_{EM} \cdot \Delta r
\end{align*}
\]

Eq. (5.7)

or correlated as

\[
\varphi_{\text{corr}}(x, y, \Delta t) = \varphi_{\text{deform}}(x, y, t_2) \otimes \varphi_{\text{ref}}(x, y, t_1) = \text{func}(\Delta r)
\]

Eq. (5.8)

where $\otimes$ can be any form of correlation functions such as normalized cross-correlation (NCCR), normalized cross-covariance (NCCV), structural similarity index measure (SSIM), cross-entropy (CE), or any used defined function, etc.

### 5.3 Shearography (SG)

Shearography extends one more step by sensing deformation gradients instead of displacements, for which the effects of rigid body motion can be minimized. Due to the self-interference at every time instant, shearography does not need a diffuser and can be applied on any surface even on a smooth surface. On an arbitrary roughness smooth shown as (Figure 5.5) the reflected light interferes with an overlapped image created by a shear device which can be a wedge, parallel plate, or beam splitter with a tilt mirror.

![Shearography Working Principle](image)

Figure 5.5 Working principle of SG which is a common-path interferometer.
The reference beam is derived from the same laser beam as that used to probe the structure; it represents the interference (two coherently superimposed) of two sheared displayed interferometric speckle patterns of the same wavefields. The purpose of the shear device is to image two points on the object, separated by $\Delta r$, as a single point in the image plane. A diffuser can be added when the surface cannot create the random speckle with sufficient contrast.

The reference beam is derived from the same laser beam as that used to probe the structure; it represents the interference (two coherently superimposed) of two sheared displayed interferometric speckle patterns of the same wavefields. At each state, speckle patterns are already self-interfered by the displaced sheared images with a certain of phase difference which is a function of shear distance $\Delta r = d$ and sensitivity vector $k_{EM} = k_s - k_o$. The phase difference $\Delta \varphi_{ref}(x, y, t_1)$ between duplicated sheared images at state $t = t_1$ and $t = t_2$ are shown as

$$
\Delta \varphi_{ref}(x, y, t_1) = k_{EM} \cdot \Delta r \equiv (k_{EM-x}, k_{EM-y}, k_{EM-z})^T \cdot d
$$

Eq. (5.9)

$$
\Delta \varphi_{deform}(x, y, t_2) = k_{EM} \cdot \Delta r' \equiv (k_{EM-x}, k_{EM-y}, k_{EM-z})^T \cdot d'
$$

Eq. (5.10)

The red and blue shading rays are FOVs for emission and sheared emission respectively; solid lines are ray emission and reflections; red and blue dots are interfered shown in (Figure 5.6) with phase difference for each time instant; shear distance before and after deformation are $\Delta r = d$ and $\Delta r' = d'$ respectively. The two displaced sheared laser beams (red and blue) are exaggerated separated for clarity.

![Figure 5.6](image)

**Figure 5.6** Phase difference between sheared images in the reference and deformed states.
The fringe patterns will be revealed by seeking for the phase-difference between two states where the change of shear distance is defined as $\Delta d = \Delta r' - \Delta r = d' - d$

$$
\Delta \phi_{sub}(x, y, \Delta t) = \Delta \phi_{deform}(x, y, t_2) - \Delta \phi_{ref}(x, y, t_1) = k_{EM} \cdot \Delta d
$$

Eq. (5.11)

The instantaneous speckle patterns at different states can be correlated just as cross-correlation, phase-correlation, previously described in DSPI.

$$
\Delta \phi_{corr}(x, y, \Delta t) = \Delta \phi_{deform}(x, y, t_2) \otimes \Delta \phi_{ref}(x, y, t_1) = \text{func}(\Delta d)
$$

Eq. (5.12)

When the measurement is noisy with low coherent light source such as high-power diode laser, correlating the phase from different loading conditions can provide better results than compared to phase-difference $\Delta \phi_{sub}(x, y, \Delta t)$ algorithm. Correlating speckle patterns $\Delta \phi_{corr}(x, y, \Delta t)$ under different loading states can result in intensity-based fringe patterns but will lose connection of intrinsic phase information to strain components within the speckle patterns. The output of correlation function in general is a metric (additional interpretation of physical meaning is required) which represents the similarity of two speckle patterns, a qualitative analysis.

When the shear device is tilted along the $x$ direction $\Delta r = \Delta x$ then the change of shear distance $\Delta d$ can be approximated by the first-order Taylor series expansion with respect to $\Delta x$ so that

$$
\Delta d = \Delta r' - \Delta r = \frac{\partial d}{\partial x} \Delta x + \frac{1}{2} \frac{\partial^2 d}{\partial x^2} (\Delta x)^2 + \frac{1}{6} \frac{\partial^3 d}{\partial x^3} (\Delta x)^3 + \cdots
$$

Eq. (5.13)

then the phase difference between states appears to be the summation of deformation gradient components

$$
\Delta \phi_{sub}(x, y, \Delta t) = k_{EM} \cdot \Delta d = k_{EM} \cdot \frac{\partial d}{\partial x} \Delta x = \left( k_{EM-x} \frac{\partial u}{\partial x} + k_{EM-y} \frac{\partial v}{\partial x} + k_{EM-z} \frac{\partial w}{\partial x} \right) \Delta x
$$

Eq. (5.14)

If the shear device is tilted to another direction, say $\Delta y$, it leads to

$$
\Delta \phi_{sub}(x, y, \Delta t) = k_{EM} \cdot \Delta d = k_{EM} \cdot \frac{\partial d}{\partial y} \Delta y = \left( k_{EM-x} \frac{\partial u}{\partial y} + k_{EM-y} \frac{\partial v}{\partial y} + k_{EM-z} \frac{\partial w}{\partial y} \right) \Delta y
$$

Eq. (5.15)

Each phase measurement consists of three deformation gradient components with respect to the sensitivity vector $k_{EM}$ shown in each phase equation. Due to the limitation of surface measurement, any component
with respect to through thickness direction $z$, say $\partial u/\partial z, \partial v/\partial z, \partial w/\partial z$ cannot be taken. Only components marked in red are measurable shown as below when strain tensor is given by $\varepsilon_{ij} = (u_{ij} + u_{ji})/2 = \gamma_{ij}/2$

$$\varepsilon_{ij} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \text{SYM} & \text{SYM} & \varepsilon_{zz} \end{pmatrix} = \begin{pmatrix} \partial u/\partial x & (\partial u/\partial y + \partial v/\partial x)/2 & (\partial u/\partial z + \partial w/\partial x)/2 \\ (\partial v/\partial x) & \partial v/\partial y & (\partial v/\partial z + \partial w/\partial y)/2 \\ \partial w/\partial x & \partial w/\partial y & \partial w/\partial z \end{pmatrix}$$  \hspace{1cm} \text{Eq. (5.16)}

### 5.3.1 Strain Components Isolation

Several methods can be applied to isolate strain components from each SG measurement.

Observing in $z$ direction

$$k_o = (0, 0, k_{oz})$$  \hspace{1cm} \text{Eq. (5.17)}

When light sources in symmetrical directions

$$k_{s1} = (-k_{sx}, -k_{sy}, -k_{sz})$$
$$k_{s2} = (k_{sx}, -k_{sy}, -k_{sz})$$  \hspace{1cm} \text{Eq. (5.18)}
$$k_{s3} = (k_{sx}, k_{sy}, -k_{sz})$$

Sensitivity vector is defined as $k_{EM-i} = k_{s1} - k_o$ and let $k_{EM-x} = -k_{sx}, k_{EM-y} = -k_{sy}, k_{EM-z} = -k_{sz} - k_{oz}$ for each light source so that we have

$$k_{EM-1} = (-k_{sx}, -k_{sy}, -k_{sz} - k_{oz}) = (k_{EM-x}, k_{EM-y}, k_{EM-z})$$
$$k_{EM-2} = (k_{sx}, -k_{sy}, -k_{sz} - k_{oz}) = (-k_{EM-x}, k_{EM-y}, k_{EM-z})$$  \hspace{1cm} \text{Eq. (5.19)}
$$k_{EM-3} = (k_{sx}, k_{sy}, -k_{sz} - k_{oz}) = (-k_{EM-x}, -k_{EM-y}, k_{EM-z})$$

The phase differences for each light source setup are

$$\Delta \varphi_{1}(x, y, \Delta t) = \left( k_{EM-x} \frac{\partial u}{\partial x} + k_{EM-y} \frac{\partial v}{\partial x} + k_{EM-z} \frac{\partial w}{\partial x} \right) \Delta x$$
$$\Delta \varphi_{2}(x, y, \Delta t) = \left( -k_{EM-x} \frac{\partial u}{\partial x} + k_{EM-y} \frac{\partial v}{\partial x} + k_{EM-z} \frac{\partial w}{\partial x} \right) \Delta x$$  \hspace{1cm} \text{Eq. (5.20)}
$$\Delta \varphi_{3}(x, y, \Delta t) = \left( -k_{EM-x} \frac{\partial u}{\partial x} - k_{EM-y} \frac{\partial v}{\partial x} + k_{EM-z} \frac{\partial w}{\partial x} \right) \Delta x$$
Strain components can be separate by taking the phase differences in between light source setups

\[
\Delta \varphi_{s_1}(x, y, \Delta t) - \Delta \varphi_{s_2}(x, y, \Delta t) = 2k_{EM-x} \cdot \frac{\partial u}{\partial x} \cdot \Delta x
\]

\[
\Delta \varphi_{s_2}(x, y, \Delta t) - \Delta \varphi_{s_3}(x, y, \Delta t) = 2k_{EM-y} \cdot \frac{\partial v}{\partial x} \cdot \Delta x
\]

\[
\Delta \varphi_{s_1}(x, y, \Delta t) + \Delta \varphi_{s_3}(x, y, \Delta t) = 2k_{EM-z} \cdot \frac{\partial w}{\partial x} \cdot \Delta x
\]

\text{Eq. (5.21)}

5.3.2 Interaction of EM and Lamb Waves

The wavelength \( \lambda = c / f \) of light source, sensitivity vector \( \mathbf{k}_{EM} = \mathbf{k}_s - \mathbf{k}_o \) (relative direction between emission and observation) and shear distance \( \Delta r = \mathbf{d} \) are key parameters to setup SG. The wavelength of light source is decided at the beginning of experiment setup and is not tunable during the experiment. Sensitive vector is something we can play around to isolate in-plane or out-of-plane components and will dramatically affect measurement sensitivity. Once the sensitivity vector was decided, the only tunable parameter is shear distance.

Increasing shear distance can increase sensitivity as well but the strain measurement will be averaged out since shear distance is like the gage length of a strain gage to measure deformation variation. Larger shear distance incorporates larger phase difference between the original and sheared images, but images may lose the correlation of speckle clouds corresponding to each other when shear distance is too large. Smaller shear distance on the other hand has a more precise strain measurement but the sensitivity is also smaller with fewer fringes and may be harder to quantify when deformation is comparative small.

Shear direction is also important for measurement when the direction of structural discontinuity caused by hidden damages is unknown. Rotating either shear device or camera is equivalent to take displacement gradient along different directions and look for damages subjected to a certain direction. Image fusion is often applied at post-processing to fuse images from different directions when structure is subjected to static load or steady-state vibration or transient wave propagation with repeated excitations.

When considering dynamic measurement for steady-state vibration or transient wave propagation problem, the intrinsic wavelength within the structure has to be at least twice larger than the shear distance to avoid spatial aliasing. Structure wavelength itself acts like the upper bound of shear distance and experimentally would be selected at least five times larger than shear distance so that a better description of displacement gradient can be performed.
For ultrasonic measurement, since wavelength is getting smaller at higher frequency range, subtracting the phase between different stages will be difficult when shear distance is shrinking and sensitivity is decreasing as well. At this scenario, quantifying strain field is getting harder whereas correlating speckle clouds for qualitative analysis would be one of the alternative solutions to accommodate this shear distance/structural wavelength or quantitative/qualitative dilemma.

In order to avoid double jeopardy, the wavenumber of light is denoted as $k_{EM} = k_x - k_o$ in 3D space, $k = (k_x, k_y)$ for structure with planar waveguide in 2D arbitrary coordinate. The relation of wavevectors is shown as (Figure 5.7) along $xz$ cross-section with generalized notation.

![Diagram of Lamb wave in planar waveguide and SG EM waves in free space](image)

The formation of Lamb waves consists longitudinal wave $k_p$ and transverse shear wave $k_{SV}$ with longitudinal/propagation wavevector $k$ and transverse/oscillation wavenumber $p$ and $q$. For a given excitation frequency $\omega$, Lamb wave will have several symmetric and anti-symmetric modes according to its excitation frequency, and each mode has its own wavevector, $k_{sym}$ and $k_{anti}$, with dispersion relation of wavevector norm $\|k\|$ and its excitation frequency $\omega$ shown as below.

\[
(\omega, \|k\|) \xrightarrow{form from Bulk} k_p = (k, p) \xrightarrow{interfere into Lamb} k = (k_x, k_y) \xrightarrow{separate into} \begin{cases} k_{sym} = (k_x, k_y) & \text{if } \omega = \omega_{sym} \\ k_{anti} = (k_x, k_y) & \text{if } \omega = \omega_{anti} \end{cases}
\]  

Eq. (5.22)
where \( \| \mathbf{k}_P \| = \sqrt{k_x^2 + p^2} = \sqrt{k_x^2 + k_y^2 + p^2} = \frac{\omega}{c_L} \) and \( \| \mathbf{k}_{SV} \| = \sqrt{k_x^2 + q^2} = \sqrt{k_x^2 + k_y^2 + q^2} = \frac{\omega}{c_T} \).

When the wavefront is aligned with coordinate axis on a plane strain cross section, the wavevector \( \mathbf{k} \) will degenerate into a wavenumber \( k \) so that the dispersion relation can be discussed in a more general way in complex domain.

\[
(\omega, k) \xrightarrow{\text{form from Balks}} \mathbf{k}_P = (k, p) \quad \text{interfere into Lamb} \quad \mathbf{k}_S = (k, q) \quad \text{separate into} \quad k_{\text{anti}} \quad k_{\text{sym}} \quad \text{Eq. (5.23)}
\]

where \( \| \mathbf{k}_P \| = \sqrt{k_x^2 + p^2} = \frac{\omega}{c_L} \) and \( \| \mathbf{k}_{SV} \| = \sqrt{k_x^2 + q^2} = \frac{\omega}{c_T} \).

Dispersion relation and analytic mode shapes for fundamental modes \( A_o \) and \( S_o \) on 5mm Aluminum plate at 200 kHz; in-plane and out of plane displacement direction separation for anti-symmetric and symmetric modes are shown as below.

Recall the symmetric and anti-symmetric modes and then compare with what can be measured by SG, only shear components will be left shown as below. This means even though SG can measure extensional components such as \( \varepsilon_{xx} \) and \( \varepsilon_{yy} \), for wave propagation problem these components are not existed.
Under plane strain assumption, strain and stress tensor for symmetric modes are shown as

\[
\varepsilon_{ij} = \begin{pmatrix}
\varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\
\varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\
SYM & SYM & SYM
\end{pmatrix} = \begin{pmatrix}
0 & 0 & \left[-ikpA_2 \sin(pz) + q^2 B_1 \sin(qz)\right]/2 \\
0 & 0 & 0 \\
SYM & -p^2 A_2 \cos(pz) + ikqB_1 \cos(qz)
\end{pmatrix}
\text{ Eq. (5.24)}
\]

For anti-symmetric modes is

\[
\varepsilon_{ij} = \begin{pmatrix}
\varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\
\varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\
SYM & SYM & SYM
\end{pmatrix} = \begin{pmatrix}
0 & 0 & \left[-ikpA_2 \cos(pz) + q^2 B_2 \cos(qz)\right]/2 \\
0 & 0 & 0 \\
SYM & -p^2 A_2 \sin(pz) - ikqB_2 \sin(qz)
\end{pmatrix}
\text{ Eq. (5.25)}
\]

The coefficients \(A_1, A_2, B_1\) and \(B_2\) can be solved by excitation conditions which will decide the domination of strain components. Comparing the strain components measured by SG and generated by Lamb waves, the strain tensor will be degenerated as below

\[
\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{xy} = \varepsilon_{yz} = 0
\]

\[
\varepsilon_{xz} = \frac{\partial u/\partial z + \partial w/\partial x}{2} = \frac{-ikpA_2 \sin(pz) + q^2 B_1 \sin(qz)}{2} + \frac{\left[-ikpA_2 \cos(pz) + q^2 B_2 \cos(qz)\right]/2}{2}
\text{ Eq. (5.26)}
\]

\[
\varepsilon_{zz} = \frac{\partial w/\partial z}{2} = -p^2 A_2 \cos(pz) + ikqB_1 \cos(qz) - p^2 A_1 \sin(pz) - ikqB_2 \sin(qz)
\]

Red words are components measured by SG and components that do not exist for wave propagation are strikethrough. The only measurable component will be \(\partial w/\partial x\) when shear plate is tilting toward x direction no matter the light source setup is in-plane or out-of-plane.
5.4 Space-time Representations

The three measurement techniques are quite similar with only slight difference and the representation of phase was unified for comparison. The concept of quantitative analysis with phase difference and qualitative analysis with phase correlation are also listed out as below.

**DSPI – phase difference in space-time**

Quantitative analysis – instantaneous fringes with phase difference $\varphi_{\text{sub}}(x, y, t)$

$$\varphi_{\text{ref}}(x, y, t) = k_{EM} \cdot r$$
$$\varphi_{\text{deform}}(x, y, t) = k_{EM} \cdot r' \quad \rightarrow \quad \varphi_{\text{sub}}(x, y, t) = \varphi_{\text{deform}}(x, y, t) - \varphi_{\text{ref}}(x, y, t) = k_{EM} \cdot \Delta r = k_{EM} \cdot d$$

Qualitative analysis – global phase decoupling in time sequence $\varphi_{\text{corr}}(x, y, \Delta t)$ if necessary

$$\varphi_{\text{sub}}(x, y, t) \quad \rightarrow \quad \varphi_{\text{sub}}(x, y, t_1) \quad \rightarrow \quad \varphi_{\text{corr}}(x, y, \Delta t) = \varphi_{\text{sub}}(x, y, t_2) \otimes \varphi_{\text{sub}}(x, y, t_1) = \text{func}(d)$$

**LSP – phase difference in space-time**

Quantitative analysis – instantaneous speckle clouds with phase difference $\varphi_{\text{sub}}(x, y, \Delta t)$

$$\varphi_{\text{ref}}(x, y, t_1) = k_{EM} \cdot r$$
$$\varphi_{\text{deform}}(x, y, t_2) = k_{EM} \cdot r' \quad \rightarrow \quad \varphi_{\text{sub}}(x, y, \Delta t) = \varphi_{\text{deform}}(x, y, t_2) - \varphi_{\text{ref}}(x, y, t_1) = k_{EM} \cdot \Delta r = k_{EM} \cdot d$$

Qualitative analysis – correlating speckle clouds in time sequence

$$\varphi_{\text{corr}}(x, y, \Delta t) = \varphi_{\text{deform}}(x, y, t_2) \otimes \varphi_{\text{ref}}(x, y, t_1) = \text{func}(d)$$

**SG – phase difference in space-time**

Quantitative analysis – instantaneous speckle clouds with phase difference $\Delta \varphi_{\text{sub}}(x, y, \Delta t)$

$$\Delta \varphi_{\text{ref}}(x, y, t_1) = k_{EM} \cdot \Delta r = k_{EM} \cdot d$$
$$\Delta \varphi_{\text{deform}}(x, y, t_2) = k_{EM} \cdot \Delta r' = k_{EM} \cdot d' \quad \rightarrow \quad \Delta \varphi_{\text{sub}}(x, y, \Delta t) = \Delta \varphi_{\text{deform}}(x, y, t_2) - \Delta \varphi_{\text{ref}}(x, y, t_1) = k_{EM} \cdot \Delta d$$

Qualitative analysis – correlating speckle clouds in time sequence

$$\Delta \varphi_{\text{corr}}(x, y, \Delta t) = \Delta \varphi_{\text{deform}}(x, y, t_2) \otimes \Delta \varphi_{\text{ref}}(x, y, t_1) = \text{func}(\Delta d)$$
5.5 Summary

All the interferometry techniques are based on the observation of path length represented as displacement where displacement = rigid body motion (translation + rotation) + deformation. Subtraction and correlation are the two most used operations, each of which can be performed either optically or numerically as shown in (Figure 5.9) depending on the optical design framework. Common optical operations can include addition, subtraction, vector-matrix multiplication, and even Fourier transform by Fraunhofer diffraction. Numerical operations can perform not only all above but also any operations such as division, correlation, differentiation, integration, etc.

![Figure 5.9](image.png)

**Figure 5.9** Comparison between optical and numerical operations

DSPI performs interferometry optically for every time instant so that the phase difference $\varphi_{\text{sub}}$ is a function of space-time $(x, y, t)$ but can be extended as a function of space-time-difference $(x, y, \Delta t)$ if correlation $\varphi_{\text{corr}}$ was performed. LSP no matter performing in which ways, subtraction $\varphi_{\text{sub}}$ or correlation $\varphi_{\text{corr}}$, is a function of space-time-difference $(x, y, \Delta t)$. SG on the other hand, firstly performs speckle interferometry optically and then numerically subtracts or correlates the speckles. At each time instance, the phase is a phase difference denoted as $\Delta \varphi$ instead of $\varphi$ like DSPI and LSP. In this way, its subtraction $\Delta \varphi_{\text{sub}}$ or correlation $\Delta \varphi_{\text{corr}}$ is a function of space-time-difference $(x, y, \Delta t)$ like LSP.
6 Phase Extraction and Estimation

The correlation fringe which is composed of random intensity variations; namely, well known as speckle noise, is highly sensitive to noise in the speckle and deteriorates the fringe quality. It makes it difficult to analyze the correlation fringes and to obtain phase distribution with high accuracy. Thus, conventional speckle-interferometry techniques have not produced quantitative results with high accuracy, although they have some practical advantages, such as validity for rough surface objects, and are good for qualitative measurements. One of the methods used to achieve accurate quantitative measurements is phase shifting.

Since the phase contains information of the measurand which is the displacement for DSPI and deformation gradient for SG, reconstruction (or demodulation) of the phase changes due to the deformation is an essential step in the quantification of the displacement with DSPI or deformation gradient with SG. Two phase shifting methods have been proposed and implemented in the optical setup to unravel the phase information: temporal phase shifting (TPS) and spatial phase shifting (SPS). The TPS involves recording a series of images sequentially with a known phase step between them. The images are combined to produce a wrapped phase map with a phase modulo 2. These wrapped phase fringe discontinuities can then be removed so that a continuous measurement is obtained. This process is known as phase unwrapping.

The aim to apply phase shifting algorithms is to evaluate the phase from intensity images captured from cameras (image sensors). For the homodyne interferometry like DSPI, the phase itself is simply proportional to displacement whereas the heterodyne interferometry is proportional to velocity. In contrast, the SG follows the homodyne architecture, and the phase measurement is proportional to deformation gradient. TPS and SPS are techniques for quantitative analysis to evaluate the phase state with respect to deformation gradient components from intensity map captured by cameras and then convert the phase distribution to deformation gradient fields. For static or quasi-static problems, TPS was applied for every loading state with multiple images to evaluate the phase state whereas SPS determines the phase state for every time instant and was designed for transient dynamic problems when loading conditions were time-varying or uncertain. A modulated intensity map for every pixel recorded by a camera can be described from Eq. (5.4) shown as below.

\[ I(x, y) = I_{DC}(x, y)[1 + m(x, y) \cos \Delta \phi(x, y)] \]

Eq. (6.1)

where \( I_{DC} \) is DC offset, \( 0 \leq m \leq 1 \) is contrast or intensity modulation function, and \( \Delta \phi \) is phase difference and denoted as a phase state under a loading state.
In order to evaluate the phase state, simply correlating speckle clouds from intensity maps cannot extract the intrinsic phase information so that phase shifting techniques are introduced. With above three unknowns, $I_{DC}$, $m$ and $\Delta \varphi$, at least three equations are required to determine the phase state. When the loading condition was unchanged, these three unknowns remain the same. A unified representation of the intensity map $I_i$ was shown below with a provided phase step $2\alpha_i$

$$I_i(x, y) = I_{DC}(x, y)[1 + m(x, y)\cos(\Delta \varphi(x, y) + 2\alpha_i)]$$

Eq. (6.2)

After the phase state was determined by applying phase shifting algorithm, phase maps can be obtained. The phase difference between phase states then can be subtracted for quantitative or correlated for qualitative analysis. No matter in which way, fringe patterns can always be obtained so that the consecutive phase unwrapping algorithm can be applied for a continuous output.

### 6.1 Temporal Phase Shifting (TPS)

If the interferometer is insensitive to vibration many measurements can be averaged to reduce the incoherent noise. An interferometer using temporal phase-shifting (TPS) is very sensitive to vibration because the various phase shifted frames of interferometric data are taken at different times and vibration causes the phase shifts between the data frames to be different from what is desired. Vibration effects can be reduced by taking all the phase shifted frames simultaneously. There are several techniques for simultaneously obtaining three or more phase-shifted interferograms.

For the SG with Michelson interferometry architecture, the phase shifting device is often incorporated with a PZT mirror to control the relative path length and generate multiple phase differences precisely during the phase reconstruction procedures. Commonly used TPS techniques require deterministic phase steps although there are some more advanced reconstruction algorithms for blind phase detection but not discussed here.

Since there are three unknowns in Eq. (6.1), a minimum of three fringe patterns is required to determine the phase. These fringe patterns can be combined using a three-step algorithm [59] requires three images $i = 1, 2, 3$ and is the fastest algorithm with symmetric non-quadrature phase steps $2\alpha_i = (i - 2) \cdot 2\pi / 3$

$$\Delta \varphi_{3\text{-step}} = \tan^{-1}\left(\sqrt{3} \frac{I_3 - I_1}{2I_2 - I_1 - I_3}\right)$$

Eq. (6.3)
To yield a wrapped phase map. Other commonly used phase-stepping algorithms are the four-step algorithm, the five-step algorithm and the Carre algorithm, which is also four steps, each of which requires images with a relative phase of step of \(\pi/2\).

A four-step algorithm [60] requires four images \(i = 1, 2, 3, 4\) with quadrature phase steps \(2\alpha_i = (i-1) \cdot \pi\) and can be realized by a PZT mirror or the combination of wave plates to alter the polarization direction of incoming light:

\[
\Delta \varphi_{4\text{-step}} = \tan^{-1}\left(\frac{I_4 - I_2}{I_1 - I_3}\right) \tag{6.4}
\]

A five-step algorithm [61] compensates the error introduced by the phase shifter with one additional image so that five images \(i = 1, 2, 3, 4, 5\) are required with quadrature phase steps \(2\alpha_i = (i-1) \cdot \pi\) so that

\[
\Delta \varphi_{5\text{-step}} = \tan^{-1}\left(\frac{2(I_2 - I_4)}{2I_3 - I_5 - I_1}\right) \tag{6.5}
\]

A Carre four-step algorithm [62] is the optimal estimator to evaluate the phase state with four images \(i = 1, 2, 3, 4\) when phase shifter in general is hardly to provide exactly \(\pi/2\) phase shift especially if not well calibrated. Every phase steps \(2\alpha_i \neq i\pi\) can be applied to evaluate the phase state and is more robust than traditional four-step algorithm.

\[
\Delta \varphi_{\text{Carre-4\text{-step}}} = \tan^{-1}\sqrt[4]{\frac{(I_1 - I_4)(I_2 - I_3)}{(I_2 - I_5)(I_1 - I_4)}} \tag{6.6}
\]

There are techniques to evaluate phase state with several unknown steps so called blind evaluation which is not discussed here as it does not provide better performance of estimation error with measurement uncertainties. This technique is more likely a contingency plan when shifting conditions cannot be applied precisely.
6.2 Spatial Phase Shifting (SPS)

In most of SG applications, the TPS is a preferred choice since it is straightforward to implement and generates wrapped phase maps with good spatial resolution. However, the measurand may vary during the time taken to record the phase-stepped images, resulting in an inaccurate measurement. This is the case when dynamic or transient situations occur. One solution is to record a series of phase stepped images simultaneously, a technique is known as spatial phase shifting.

As stated before, the TPS technique is only suitable for static or quasi-static problems and the reference state was assumed unchanged when recording phase shifting images which is not a case for dynamic problems. In practice, most of the loading conditions are time-varying generating dynamic response such as steady-state vibration and transient wave propagation. Strictly speaking, even the heat transfer process can be categorized as a dynamic problem. There is only one image recorded for every time instance and the phase state must be evaluated with this single image. This technique was known as spatial phase shifting.

The SPS can be conducted by recorded multiple images by using synchronized cameras [63] with well-designed optical arrangement of sensitivity vectors. Another approach is to use a single camera with diffractive optical elements [64] and generate different orders of diffractive images with respect to different sensitivity vectors. Each diffractive pattern occupies only a small portion of camera sensing element and spatial resolution was reduced, e.g., a commercial system (4D Technology Corporation) with a pixelated phase mask [65].

Polarization-based shearing interferometer [66] divides a single image into two images with 180-degree relative phase difference. This setup does not need to calibrate the relative position of cameras and full spatial resolution can be retrieved. Otherwise, the treatment of unfavorable polarization states was discussed as well. In order to maintain accurate measurements, sub-pixel image registration [67] was an essential step to analyze the image.

In applications, Gabor filters that yield an approximation of the complex signal of a bandpass filtered version of the input signal are often used. One of the well-known methods called windowed Fourier ridge (WFR) [67]–[69] is to interrogate local fringe patterns and compare its spatial frequency with windowed transform spectrum. The frequency with maximum resemblance to windowed transform spectrum was taken as the ridge with local phase information. Then the local phase can be calculated from its complex representation in frequency domain. Other methods such as normalized Hilbert transform [70] and continuous wavelet transform (CWT) also can be used to determine the phase from a single fringe pattern [71], [72] since the intensity was modulated with a carrier frequency.
In conjunction with the spatial carrier technique, a carrier frequency is introduced into the recorded intensity distribution which serves to separate the phase information from the background information in frequency domain. The concept of spatial carrier technique \[73\]–\[78\] with Fourier transform is to send the effective information to a higher frequency band, say passband, so that it can be distinguished from the ground noise which is at lower frequency band, called as baseband. If the intensity was modulated with carrier wavenumber \(k_o\) along the \(x\) direction, then it can be denoted as

\[
I(x, y) = I_{DC}(x, y) \left[1 + m(x, y) \cos(\Delta \phi(x, y) + k_o x)\right] = I_{DC}(x, y) \left[1 + c(x, y) + c^*(x, y)\right] \quad \text{Eq. (6.7)}
\]

where superscript * is complex conjugation and

\[
c(x, y) = \frac{m(x, y)}{2} . \exp i (\Delta \phi(x, y) + k_o x) \quad \text{Eq. (6.8)}
\]

Converting a real-valued function to its complex representation is equivalent to folding a double-side band (DSB) Fourier spectrum into a single-sided band (SSB) so that the phase information can be extracted from its frequency domain as a global phase or from its original domain as a local phase when inversely converting its SSB back to its original domain.

The analytic signal calculated from inversed SSB 1D Fourier transform is equivalent to the one calculated from Hilbert transform which is a 1D transform as well.

\[
I_{x-a}(x, y) = F_{x-SSB}^{-1} \left[ F_x \{I(x, y)\}\right] = I(x, y) + jH_x(I(x, y)) \quad \text{Eq. (6.9)}
\]

where \(I_{x-a}(x, y)\) is the analytic signal of an intensity modulated signal \(I(x, y)\) along \(x\) direction, \(F_x(\ )\) is 1D Fourier transform along \(x\) direction, \(F_{x-SSB}^{-1}(\ )\) is inversed folded spectrum SSB Fourier transform along \(x\) direction, \(H_x(\ )\) is Hilbert transform along \(x\) direction. The local phase of the original signal can be calculated by taking the phase of its lowpass equivalent just like in 1D signal after excluding the ground noise in the frequency domain

\[
\Delta \phi(x, y) = \tan^{-1} \frac{\text{Im}[c_i(x, y)]}{\text{Re}[c_i(x, y)]} \quad \text{Eq. (6.10)}
\]

where \(c_i(x, y) = c(x, y) \exp i(-k_o x)\) is a lowpass equivalent signal. Many demodulation methods based on Fourier transform \[73\], windowed Fourier transform \[69\], Gabor transform \[79\], and wavelet transform \[80\]–\[84\] has been developed for extracting the phase from the carrier fringes.
Since the Hilbert transform kernel $k/\|k\|$ is a linear operator and so as the analytic operator kernel $1+k/\|k\|$, it’s by all mean following exactly the same principles for a 2D analytic signal which was defined as the inversed double-folded spectrum from 2D Fourier transform and is equivalent to the original signal plus its 1D quadrature calculated by Hilbert transform along two directions

$$I_a(x, y) = \mathcal{F}^{-1}_{SSB} \left[ \mathcal{F} \left\{ I(x, y) \right\} \right] = I(x, y) + i\mathcal{H}_x(I)(x, y) + j\mathcal{H}_y(I)(x, y) \quad \text{Eq. (6.11)}$$

where $I_a(x, y)$ is the 2D analytic signal of an intensity modulated signal $I(x, y)$, $\mathcal{F}(\ )$ is 2D Fourier transform, $\mathcal{F}^{-1}_{SSB}(\ )$ is inversed double-folded spectrum SSB 2D Fourier transform, $\mathcal{H}_x(\ )$ is Hilbert transform along $x$ direction, $\mathcal{H}_y(\ )$ is Hilbert transform along $y$ direction.

The technique can be extended to two-dimensional fringe patterns but it requires that the carrier frequency is constant across the image, i.e. the carrier fringes need to be perfectly straight. This can be difficult to achieve in a practical SG system. Any curvature in the fringes will be unrecognizable from the distortion in the fringes due to the phase modulation and therefore result in an erroneous measurement when Hilbert transform can only extract phase along one direction with linear kernel.

The selection of Fourier transform direction aligned with shearing and carrier fringes direction may result in unfavorable error since there are multiple axes to align. When considering the coupling of two axes with a monogenic operator from Riesz transform to replace the linear operator in Hilbert transform which considers the two axes as independent directions shown as below.

$$1 \xrightarrow{1D_{\text{linear}}} 1 + \frac{k}{\|k\|} \xrightarrow{2D_{\text{linear}}} 1 + \frac{k_x}{\|k_x\|} + \frac{k_y}{\|k_y\|} \xrightarrow{2D_{\text{coupling}}} 1 + \frac{k_x}{\|k\|} + \frac{k_y}{\|k\|} \quad \text{Eq. (6.12)}$$

The logic flow (Figure 6.1) of phase extradition $\Delta\varphi_{\text{ref}}(x, y)$ and $\Delta\varphi_{\text{deform}}(x, y)$ from reference state intensity $I_{\text{ref}}(x, y)$ and deformed state intensity $I_{\text{deform}}(x, y)$ is quite similar for 2D Fourier transform and Riesz transform shown as below. The lowpass equivalence $c_{I_{\text{ref}}}(x, y)$ and $c_{I_{\text{deform}}}(x, y)$ are calculated by the inversed SSB 2D Fourier transform whereas Riesz transform outputs the quadrature vector $\mathcal{R}(I_{\text{ref}})(x, y)$ and $\mathcal{R}(I_{\text{deform}})(x, y)$. Once the phase map for each state, say reference state and deformed state, then phase difference can be found by subtraction or correlation for further analysis.
Figure 6.1 2D Fourier transform (global phase) and Riesz transform (local phase)

In a 2D scheme, the analytic signal $I_a(x, y)$ differs from the monogenic signal $I_m(x, y)$ and is more preferred to be denoted as a displacement function $u_m(x, y)$ to be more specific instead of the intensity function $I_m(x, y)$, which is more general.
6.3 Direct Phase Estimation (DPE)

In order to achieve the goal for large area inspection with high power diode laser based on correlation speckles, being random distributed without clear spatial carriers. Direct phase estimation (DPE) from image intensity distribution via Riesz transform appears to be a viable approach than traditional TPS and SPS.

A complex signal of a real-valued 1D time-series signal, called analytic signal, was introduced to signal theory for digital communication by Gabor in 1946. The analytic signal is constructed from the real-valued (base) signal by adding its Hilbert transform – which is a phase-shifted version of the signal by \(-\pi / 2\) – as an imaginary (or quadrature) part to the signal. Its real and imaginary part form a quadrature pair. The analytical signal suppresses all negative frequency components of a real-valued signal.

This analytic signal representation concerning the analysis of the 1D signal has become an important tool in 1D signal processing for amplitude or frequency modulation (AM or FM) and demodulation in narrowband communication. Using the complex-valued signal to represents the real-valued signal make various processes conceptually simpler. This analytic signal enables the analysis of instantaneous (or local) amplitude and phase of a signal. From the complex signal, the local amplitude (the envelope) and the local phase of the original signal can be derived as modulus and angular argument, respectively.

6.3.1 Monogenic Signal and Riesz Transform

Monogenic signal is a generalization of the analytic signal in 1D to multi-dimensional signals such as 2D image signals, simply called 2D images in the spatial domain \((x, y)\). The monogenic signal introduced by Felsberg and Sommer in 2001 [85], [86] has been employed in many domains such as image analysis and computer vision. Oppenheim and Lim [87] showed that the main information content of an image intensity \(I(x, y)\) lies in its phase components.

For SPS with spatial carrier, due to the difficulty of generating sufficiently high carrier frequencies to discern the signal from DC to higher frequency band, the Riesz transform provides a viable tool to evaluate the phase with its hypercomplex monogenic signal and agile choice of output phase signatures, e.g., local energy, local phase, phase between quadratures, quadrature vector norm [88], [89], phase congruency [90], [91], invariance [92] and etc. [93], [94]. This unfolds many possibilities to interpret or further classify the signal via image processing. As the Riesz transform converts an image into its hypercomplex monogenic signal, there are several phase signatures that can be selected as the candidate to feed into correlation function, say amplitude (also called local energy, LE), phase (local phase, LP) and orientation (local orientation, LO).
where the Riesz transform $\mathcal{R}$ is a scalar-to-vector signal transformation specified by the mapping from $I(x, y)$ to $\mathcal{R}(I)(x, y)$ in spatial domain $(x, y) = x$ shown as follows:

$$\mathcal{R}(I)(x, y) = \left( \frac{R_x(I(x, y))}{R_y(I(x, y))} \right) = \frac{x}{2\pi \|x\|^3} \otimes I(x, y)$$ \hspace{1cm} \text{Eq. (6.14)}

where $\| \|$ denote the norm of the functions in the Hilbert space in which the functions are finite in energy and compactly supported. $\otimes$ denotes the convolution operator. The hypercomplex variables $i, j, k$ in quaternion algebras are defined as $i^2 = j^2 = k^2 = -1$ and $ijk = -1$.

Local parameters associated with the quadrature vector are defined as

- **Amplitude – local energy (LE)** $A(x, y) = \| I_m(x, y) \|^2 = \sqrt{I^2(x, y) + \|\mathcal{R}(I)(x, y)\|^2}$ \hspace{1cm} \text{Eq. (6.15)}

- **Phase – local phase (LP)** $\varphi(x, y) = \angle I_m(x, y) = \tan^{-1} \left( \frac{\|\mathcal{R}(I)(x, y)\|}{I(x, y)} \right)$ \hspace{1cm} \text{Eq. (6.16)}

- **Orientation – local orientation (LO)** $\theta(x, y) = \angle \mathcal{R}(I)(x, y) = \tan^{-1} \left( \frac{R_y(I(x, y))}{R_x(I(x, y))} \right)$ \hspace{1cm} \text{Eq. (6.17)}

In order to find the monogenic signal we have two routes to go, one is to calculate the quadrature components $R_x$ and $R_y$ with the quadrature operator to convert a real-valued signal into its quadrature so that we can form the monogenic signal from these components. The vector form is defined as below to generate the quadrature vector

$$- \left( \begin{array}{l} i \\ j \end{array} \right) \circ \frac{\mathbf{k}}{\|\mathbf{k}\|} \cdot \mathcal{F}(I(x)) = \mathcal{F}(\mathcal{R}(I)(x)) \iff \frac{x}{2\pi \|x\|^3} \otimes I(x) = \mathcal{R}(I)(x)$$ \hspace{1cm} \text{Eq. (6.18)}

where $\circ$ is element wise multiplication, $\cdot$ is dot product and $\otimes$ is convolution. Another route directly goes for monogenic operator to convert the real-valued signal into monogenic showing as

$$\left( 1 + I^T \cdot \frac{\mathbf{k}}{\|\mathbf{k}\|} \right) \cdot \mathcal{F}(I(x)) = \mathcal{F}(I_m(x)) \iff \left( \delta(x) + \left( \begin{array}{l} i \\ j \end{array} \right)^T \cdot \frac{x}{2\pi \|x\|^3} \right) \otimes I(x) = I_m(x)$$ \hspace{1cm} \text{Eq. (6.19)}
where \( \mathbf{1} = [1, 1, \ldots]^T \) is unity column vector, \( \cdot \) is dot product. \( \mathbf{x} \) and \( \mathbf{k} \) are position and wavevector respectively.

### 6.3.2 Log-Gabor Bandpass Filter and its Monogenic Form

Local signatures derived from Eq. (6.15) to Eq. (6.17) are defined over the entire the spatial domain of the image. For feature detection, localization in spatial domain is highly desirable. To extract localized image features, filters such as Gaussian, gradient of Gaussian (GoG) and Laplacian of Gaussian (LoG) are commonly chosen in accompany with Riesz transform. Although the Gaussian filter is quite useful for denoising and smoothing, flexibility of choice of the center frequency and bandwidth based on the prior knowledge of excitation frequency and structural wavelength renders the log-Gabor bandpass filter as a preferable choice due to its wave packet-like impulse response and tunable bandwidth.

The log-Gabor filter in the frequency domain is a Gaussian function on a logarithmic scale.

\[
G(\omega) = \exp \left( -\frac{1}{2} \left( \frac{\log(\|\omega\|/\omega_o)}{\log(\sigma_o)} \right)^2 \right) \tag{6.20}
\]

The log-Gabor filter consists of two independent variables. First, \( \omega_o \) is passband center frequency and governs what dominant frequency is tuned by the filter; second, \( \sigma_o \) is a shape parameter governing how wide the filter is in the frequency domain. The shape parameter is related to the bandwidth \( B \) of the filter by the following relation

\[
B = 2\sqrt{2/\log(2)} \cdot \|\log(\sigma_o)\| \tag{6.21}
\]

One of the advantages of log-Gabor filter is that it can be assigned with an arbitrarily bandwidth by \( \sigma_o \) while maintaining a zero gain at \( \omega = 0 \) (i.e., DC filter). It is an even symmetric function with respect to \( \omega \) and its frequency response is demonstrated as below with center frequency \( \omega_o = 0.2\omega_s \) where \( \omega_s \) is the sampling frequency and shape parameters \( \sigma_o \) with respect to times octave band as a normalized coefficient.

It is convenient to define a filter with symmetric spectrum, \( G_{\text{even}}(\omega) = G(\omega) \) in Eq. (6.20), so that it can be rather fit into the scheme of analytic filter. (Figure 6.2) shows \( G(\omega) \) for five bandwidths \( B = 1, \ldots, 5 \); the corresponding impulse response in time domain \( g_{\text{even}}(t) = g(t) \) for \( B = 2 \) shown in (Figure 6.3).
It is worth noting that the log-Gabor filter is usually expressed in the frequency domain since no closed form solution is obtainable in the time domain (i.e., impulse response); it can be obtained numerically by taking the inverse Fourier transform.

In theory, the imaginary part of its time domain impulse response shown in (Figure 6.3) should be zero because of its symmetrical spectrum. It can be verified by taking the inverse Fourier transform of its spectrum and find out comparative small value of its imaginary part (theoretically will be aero for infinite length series).

For multi-dimensional signals using monogenic signals, the corresponding quadratic filters in the frequency domain need to be generated. Quadratic filters consist of two filters: an even and an odd bandpass filter. To generate its quadrature which is an odd filter, it can readily apply quadrature operator just like in 1D Hilbert transform showing as

$$G_{\text{odd}}(\omega) = \left(-j\frac{\omega}{\|\omega\|}\right) \cdot G_{\text{even}}(\omega)$$

Eq. (6.22)

Then its phase shift spectrum (Figure 6.4) and impulse response (Figure 6.5) can be obtained.
The 2D log-Gabor filter itself is also an even function expanded from Eq. (5.8) as shown below

$$G(k) = \exp\left(-\frac{1}{2} \log\left(\frac{k}{k_o}\right)^T \log\left(\Sigma_o\right)^{-1} \log\left(\frac{k}{k_o}\right)\right)$$  \hspace{1cm} \text{Eq. (6.23)}$$

where $k$ is wavevector, $k_o$ is center wavevector, and $\Sigma_o$ is shape parameter matrix.

The Riesz transform can generate the quadratures of the 2D log-Gabor filter shown as (Figure 6.6) and (Figure 6.7) to form a monogenic signal as a monogenic filter consisting three parts: one even function $G_{even}(k)$, two odd functions $G_{odd-x}(k)$ and $G_{odd-y}(k)$.

The visualization of log-Gabor bandpass filters with Riesz transform and their impulse response of kernel functions are shown in (Figure 6.6) and (Figure 6.7) to better understand the formation of a monogenic filter for damage detection.
After applying the Riesz transform on the log-Gabor bandpass filter, it becomes a monogenic filter just like other real-valued signals being converted into monogenic signals. Their frequency and impulse responses are shown in (Figure 6.6) and visualized in (Figure 6.7) to have better understanding their relations.

Another representation of this monogenic filter is to take its quadratures $R_x$ and $R_y$ with hypercomplex variables $i$ and $j$ shown below

$$i \cdot g_{odd-x}(x) + j \cdot g_{odd-y}(x) = i \left[ g_{odd-x}(x) - k \cdot g_{odd-y}(x) \right]$$  \hspace{1cm} Eq. (6.24)

and then turn them into a complex filter by substituting $k$ with $i$ in spatial domain shown as Eq. (6.25) where the hypercomplex variables follow the rules $ijk = -1$ and $i^2 = j^2 = k^2 = -1$ in quaternion algebra

$$g_{complex}(x) = g_{odd-x}(x) - i \cdot g_{odd-y}(x)$$  \hspace{1cm} Eq. (6.25)

The complex filter $g_{complex}(x)$ has a 2D analytic form like a 1D analytic signal. Here to avoid the confusion “complex” is used instead of “analytic”.

Figure 6.7  The visualization of a log-Gabor bandpass filter with Riesz transform
So far, we have a monogenic filter $G_m(k)$ to apply on the image $I(x)$ for feature extraction which can either be performed in spatial domain or wavenumber domain shown as (Figure 6.8)

$$G_m(k) \cdot \mathcal{F}(I(x,y)) \iff g_m(x) \otimes I(x,y) \quad \text{or} \quad G_{even}(k) \cdot \mathcal{F}(I(x,y)) \iff g_{even}(x) \otimes I_m(x,y)$$

**Figure 6.8** Choice between (left) a monogenic filter on a real-valued signal and (right) a real-valued (unconditional even in wavenumber domain) filter on a monogenic signal

The interpretation of the entire phase extraction process could be either applying a monogenic filter on either the real-valued signal or applying a real-valued filter on a monogenic signal shown in (Figure 6.8). The former is preferable because of the symmetrical spectrum and analytical derivative forms, the filter as a part of the exponential family with natural parameters.

The process to apply the filter is pretty much like estimating parameters in statistics by taking samples. The only difference is that we don’t take statistic testing to form the asymptotic normality to evaluate the confidence of our hypothesis, here we form the estimator with pre-defined parameters from our prior knowledge and then evaluate the confidence by the variation of samples.

### 6.3.3 Riesz Transform as a Phase Estimator from Synthetic Images

To demonstrate the Riesz transform capable of directly estimating the phase from an image, a synthetic image from an L-shaped membrane is selected for several reasons. Being geometrically simple, yet solutions to the wave equation cannot be expressed analytically. Thus, numerical computation is necessary. Second, the 270º nonconvex corner induces a singularity in the solution.

Mathematically, the gradient of the first eigenfunction is unbounded near the corner. Lastly, physically a membrane stretched over such a region would cause wrinkling at the corner. This singularity limits the accuracy of any numerical methods with uniform grids and yet appears to be a particularly good trial function for imaging processing algorithms.

The wave equation governing the membrane is most basic dynamic equation that describes how a wave disturbance travels in the membrane. For simplicity, the L-shaped membrane has the long side with a unit length. The amplitude of a wave with the wave speed being unity is governing by
\[ \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial^2 u}{\partial t^2} \]  
Eq. (6.26)

where \( u \) is the transverse displacement \( x \) and \( y \) are spatial coordinates and \( t \) is time. The steady-state wave solutions at resonance are described as

\[ u(x, y, t) = v(x, y) e^{-i\omega t} \]  
Eq. (6.27)

leading Eq. (6.26) to

\[ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \omega^2 v = 0 \]  
Eq. (6.28)

The periodic oscillation in time gives solutions of the form to separate the space and time

\[ u(x, y, t) = v(x, y) \sin(y^{1/2} t) \]  
Eq. (6.29)

So that

\[ \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \lambda v = 0 \]  
Eq. (6.30)

The quantities \( \lambda \) are the eigenvalues and the corresponding functions \( v(x, y) \) are the eigenfunctions or modes of vibration. They are determined by the physical properties, the geometry, and the boundary conditions. Any solution to the wave equation can be expressed as a linear combination of these eigenfunctions. The square roots of the eigenvalues are resonant frequencies. A periodic external driving force at one of these frequencies will generate an unboundedly strong response in the medium. The membrane mode shape \( v(x, y) \), mathematically treated as an envelope function can be modulated with \( \exp(-ikr) \) in space which can be represented as

\[ v(x, y) \exp(-ikr) \quad \text{and} \quad r = \sqrt{x^2 + y^2} \]  
Eq. (6.31)

where \( k \) is the fast-varying wavenumber (spatial carrier frequency). The first mode of the eigenfunction and its modulated form are shown as (Figure 6.9) in 2D and (Figure 6.10) in 3D.
Then the constitutive equation Eq. (6.30) turns out to be

\[
\left( k \frac{dr}{dx} \right)^2 + \left( k \frac{dr}{dy} \right)^2 - ik \left( \frac{d^2 r}{dx^2} + \frac{d^2 r}{dy^2} \right) + \lambda = 0
\]

We’re not going to solve the PDE even we’d like to there’s no analytical solutions. In this example, \( k = 5, 10, 15 \) represents the number of carriers. Here the intensity of \( \nu(x, y) \exp{-ikr} \) is defined as the image \( I(x, y) \) to test the capabilities Riesz transform. Since any input image \( I(x, y) \) can be interpreted as the general form with a real-valued \( I \) and complex \( Q \) where local energy (LE) denoted as \( A \) and local phase (LP) denoted as \( \phi \).

The general form of the monogenic signal for the ease of explanation can be shown as

\[
I_m(x, y) = I + i \cdot Q = A(\cos \phi + i \sin \phi) = A \exp{i\phi}
\]

Till now, we have to figure out the phase contribution from each of above components, say a 5-carrier modulated membrane function for example shown as below
Traditional interferometry relies on either temporal (TPS) or spatial phase shifting (SPS) to back calculate the intrinsic phase information from image intensity pattern taken by the camera. The TPS is only suited to static, quasi-static or steady-state problems with given recursive phase shifts whereas the SPS can reconstruct the instantaneous phase information at every time instant.

For comparison, three shifting techniques: TPS, SPS and SPS with spatial carrier require four, two and one images/image respectively at each state $I_{\text{ref}}$ or $I_{\text{deform}}$ consecutively denoted as (4, 2, 1) shown in the top row of (Table 6.1) whereas the DPE only requires a single image to evaluate the phase.
Table 6.1  Traditional phase extraction steps and proposed phase estimation procedures

<table>
<thead>
<tr>
<th>Technique</th>
<th>Current image processing techniques</th>
<th>Proposed image processing algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPS</td>
<td>1. Correlate speckle images</td>
<td></td>
</tr>
<tr>
<td>SG</td>
<td>1. Evaluate speckle phase by shifting 2. Unwrap phase for deformation gradients</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.2  Comparison of full-field optical NDI systems

In order to have minimal hardware setup, an advanced image processing sequence was proposed for direct phase estimation (DPE). The consecutive steps to obtain fringe patterns and phase map are also replaced with correlation instead of subtraction and enveloping by Riesz transform as a replacement of phase unwrapping algorithm. The enveloping images are further processed with PCA for segmentation and feature separation so that damage features can be highlighted with significance index, eigenvalue, etc.

However, when the measurement is noisy with low coherent light source, correlating the phase from different loading states can provide better results compared to subtracting. Correlating speckle clouds under different loading states can result in intensity-based fringe patterns but will lose connection of intrinsic phase information to strain components within the speckle clouds. The output of correlation function in general is an index (additional interpretation of physical meaning is required) which represents the similarity of speckle patterns from different states, and it turns out to be a qualitative analysis.

Advances of signal processing techniques permit to interpret the signal in higher dimensions with hypercomplex representation. Traditional phase shifting techniques can be replaced by Riesz transform with its instantaneous phase representation so that the phase state can be extracted at every time instant.
A simulated indentation damage induced phase change $\Delta \varphi_{\text{damage}}$ was described as a Gaussian distribution and assigned at the corner of the L-shape membrane so that the phase at deformed state will be

$$\Delta \varphi_{\text{damage}} = \Delta \varphi_{\text{deform}} - \Delta \varphi_{\text{ref}}$$  \hspace{1cm} \text{Eq. (6.34)}

where $\Delta \varphi_{\text{damage}}$ is the true value from measured intensities $I_{\text{ref}}$ and $I_{\text{deform}}$. In order to estimate the phase from intensity, phase estimators can be extracted in several ways: the TPS estimator with phase-difference algorithm executing subtraction followed by unwrapping algorithm is represented as

$$\Delta \hat{\varphi}_{\text{damage}} = \text{unwrap}(\Delta \varphi_{\text{sub}}) = \text{unwrap}(\Delta \varphi_{\text{deform}} - \Delta \varphi_{\text{ref}})$$  \hspace{1cm} \text{Eq. (6.35)}

whereas the Riesz bp estimator employing the hyperparameters including local energy (LO), local phase (LP), and quadrature norm $\| \mathcal{R} \|$ is represented as

$$\Delta \hat{\varphi}_{\text{damage}} = \text{LE}_{\text{corr}} \text{ or } \text{LP}_{\text{corr}} \text{ or } \| \mathcal{R} \|_{\text{corr}}$$  \hspace{1cm} \text{Eq. (6.36)}

There are several choices of correlation functions to correlate speckle clouds. Here we select structural similarity index measure (SSIM) as the correlation function which is a function of luminance, contrast and structure with power coefficients. The phase estimation process is shown as (Figure 6.13).

![Figure 6.13](image)

Figure 6.13 Imaging processing procedures with SSIM and unwrapping algorithms

To compare the performance of traditional TPS and Riesz DPE, the reference state phase $\Delta \varphi_{\text{ref}}$ was selected as a membrane function with the predefined circular damage phase $\Delta \varphi_{\text{damage}}$. The phases were converted into intensity $I_{\text{ref}}$, $I_{\text{damage}}$ and $I_{\text{def}}$ where only $I_{\text{ref}}$ and $I_{\text{def}}$ are measurable, and the aim is to evaluate the phase $\Delta \varphi_{\text{damage}}$ under $I_{\text{damage}}$. Different numbers of carriers are compared to test whether DPE still can estimate the phase correctly. Results are shown as from (Figure 6.14) to (Figure 6.19).
Figure 6.14  SG with TPS and Riesz DPE – L-shaped damaged membrane function – noise free

Figure 6.15  SG with TPS and Riesz DPE – L-shaped damaged membrane function – with speckle noise
Figure 6.16 SG with TPS and Riesz DPE – 5-carrier modulated L-shaped damaged membrane function – noise free

Figure 6.17 SG with TPS and Riesz DPE – 5-carrier modulated L-shaped damaged membrane function – with speckle noise
Figure 6.18  SG with TPS and Riesz DPE – 15-carrier modulated L-shaped damaged membrane function – noise free

Figure 6.19  SG with TPS and Riesz DPE – 15-carrier modulated L-shaped damaged membrane function – with speckle noise
6.4 Summary

It’s no doubt that Riesz DPE is almost impossible to outperform TPS when TPS takes several speckle images to evaluate the phase. Since TPS evaluates the phase from an inverse function by back calculating the phase from the intensity, the extracted phase is very deterministic, in contrast, Riesz DPE relies on evaluating the intensity by the kernel function to estimate the phase. The selection of the kernel function, the feature of the intensity map, the hypercomplex variables for correlation, and the selection of correlation functions, all of them bring the uncertain factors for phase estimation.

Since there are several hyperparameters and it is possible to optimize and select a set of hyperparameters for a learning algorithm. Hyperparameter optimization finds a tuple of hyperparameters that yields an optimal model which minimizes a predefined loss function on given independent data. The objective function takes a tuple of hyperparameters and returns the associated loss. Cross-validation is often used to estimate this generalization performance. Possible choice of cost functions such as PSNR evaluated in dB scale, cross-entropy (CE) in between, accumulated principal values, or else more.

Riesz DPE has its value on single image phase estimation, no additional phase stepping hardware setup, and can be applied to any kind of measurement including transient dynamic problems. Also, aim to the correlation function, it has much larger tolerance to the coherence of the illumination source and much more capable to realize large-area inspection with high-power low coherence laser like diode laser. Since the entire Riesz DPE process is numerical computation, it can be fit into machine learning framework for feature engineering such as segmentation, classification, and identification. This once more makes the Riesz DPE a practical approach for industrial applications.
7 Digital Image Processing

Digital image processing is the use of a digital computer to process digital images through an algorithm. As a subcategory or field of digital signal processing, digital image processing has many advantages over analog image processing. It allows a much wider range of algorithms to be applied to the input data and can avoid problems such as the build-up of noise and distortion during processing. Since images are defined over two dimensions (perhaps more) digital image processing may be modeled in the form of multidimensional systems. The generation and development of digital image processing are mainly affected by three factors: first, the development of computers; second, the development of mathematics (especially the creation and improvement of discrete mathematics theory); third, the demand for a wide range of applications in environment, agriculture, military, industry and medical science has increased. [95]

For damage inspection applications, the full-field measurement has to be applied with specified imaging condition to analyze the wavefield and reveal the damage information. Detection, localization, feature extraction, and severity evaluation are general procedures to analyze structural hidden damages. According to the types of excitations, imaging conditions can be categorized into vibration-base, wave-based, vision-based and statistical-based methods discussed in this chapter.

7.1 Vibration-based Imaging Condition for Steady-state Vibration

Local wavenumber (LW) was developed by Rogge et al. [96], Flynn et al. [24], [97], [98], Goodman et al. [99] with the extension of short-time Fourier transform (STFT) into higher dimensions in space so called short-space Fourier transform (SSFT). It estimates the local wavenumber with SSFT with a sliding window at each spatial location but have to select a proper window function to suppress sidebands and negotiate with the trade-off between spatial and wavenumber resolution just like other Fourier-based time frequency analysis. With the knowledge of the material properties and the layup of composites, the wavenumber map can then be converted to a delamination depth map based on the dispersion relations.

Instantaneous wavenumber (IW) was conducted by Mesnil et al. [100]–[103] and derived by Riesz transform which is the 2D version if Hilbert transform so that the instantaneous phase and other hypercomplex parameters can be obtained in a wavefield. For a propagating wave, traveling in space is equivalent to phase shifting and it gradient norm is wavenumber (wavenumber vector norm) so that wavenumber can be evaluated at any position without losing any resolution in spatial and wavenumber domain.
Wavenumber index (WI) was a normalized version of IW with time average and proposed by Chang et al. [104], [105] to suppress the uncertainty level as the wavefield was taken in time sequence.

### 7.2 Wave-based Imaging Condition for Transient Propagation

Since the entire wavefield can be attained from the use of point by-point scanning, for example using LDV, Ruzzene [106] employed a signal filtering technique in the wavenumber-frequency domain to separate the reflected wavefield from the incident wavefield in an aluminum plate, each of them propagating in different directions which can be readily dictated by their corresponding wavevectors. This approach clears a way to attain the reflected waves whose magnitudes are often an order of magnitude smaller than the incident wavefield, thus they are largely masked in the entire wavefield.

As a result, most of the damage detection methods relying on baseline subtraction; that is, subtracting the wavefield from pristine structures, can be directly applied to the acquired reflected wavefield for locating the damages. Different damage imaging techniques have been used to isolate and quantify the damage area within the guided wavefield.

The energy-based imaging condition RMS which is the square root of the averaged sum of the sensing signals squared for damage imaging after wavefield filtering was suggested by Ruzzene et al. [107], Zak et al. [108] and Kudela et al. [109].

An et al. [110] and Park et al. [20] developed an algorithm to isolate standing waves trapped within delaminations and disbonds from the propagating waves in the wavenumber-frequency domain. After inverting the standing wave signal from the wavenumber-frequency back into the space-time domain, a cumulative standing wave energy (CSWE) is calculated in the entire scanned domain to image the damage.

Chia et al. [111] developed an anomalous wave propagation imaging (AWPI) method with adjacent wave subtraction to locate artificial defects introduced in a composite wing. Rogge and Leckey [96] developed a local wavenumber domain analysis to determine the depth of the delamination in composite structures, while Tian et al. [6] employed numerical and experimental Lamb wavefield for delamination detection in composite structures. They used wavenumber-frequency ($k$–$f$) analysis and spatial-wavenumber ($x$–$k$) analysis to determine both the dimension of the delaminated region and the plies between which the delamination occurred.
7.3 Vision-based and Statistical-based Imaging Condition for Randomness

One of the most prolific algorithms for image comparison and template matching, NCC, often refers to one of two algorithms, normalized cross-correlation and normalized cross-covariance. Many previous works have focused on only one of the two, trying to optimize the algorithm as much as possible. Both normalized cross-correlation (NCCR) and normalized cross-covariance (NCCV) are presented in various forms covering both the one-dimensional and two-dimensional cases as well as the applications of image comparison and template matching. However, they did not perform well for speckle-based interferometry since noise and lighting condition uncertainties also contribute to speckle intensity.

Another reason is speckle clouds are random distributed and do not have any meaningful feature for NCC-based algorithm to distinct the speckle clouds before and after loading unless the statistics-base algorithm are applied with enough samples to satisfy the law of large numbers.

Structural similarity index measure (SSIM) augmented traditional NCC-based algorithm with image luminance, contrast and structure with power coefficients to adapt with environmental variations. This makes SSIM a more robust and practical choice over other correlation function.

\[
SSIM(x, y) = \left[ l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma \right]
\]

Eq. (7.1)

where

- \( l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \) is luminance
- \( c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \) is contrast
- \( s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3} \) is structure.

When \( \alpha = \beta = 1 \) and \( \gamma = 0 \), SSIM degenerates to

\[
SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}
\]

Eq. (7.2)
7.4 **General Processing Procedures for Speckle Images**

There are several ways to process speckle images but here we propose a universal framework shown as (Figure 7.1) that can be applied to any kinds of speckle interferometry measurements. We named this framework as ADPNet – advanced damage processing network. The framework is described with several blocks and each block can be treated as a processing kernel.

**Green block** is for measurement, and the hardware setup can vary according to the desired measurement quantities; **blue block** is an interpreter or hypercomplex parameter generator according to the measured quantities from the previous block; **orange block** is for correlation and DIC (nothing to do with its hardware setup) is just one of the selections of correlation functions; **purple block** is for feature extraction and most of the machine learning processing are in this block.

![Figure 7.1 General processing procedures for speckle images](image)

The following chapters demonstrate this vision-based and statistical-based image processing under this framework with LSP and SG measurement under different excitations. Some preparations before and after the ADPNet are described as below.

**Pre-processing**

Feature matching with RANdom SAmples Consensus (RANSAC) algorithm deals with image alignment so that rigid body motion (translation and rotation) can be suppressed, and distorted images (caused by lens and aperture) can be recovered by blind de-convolution algorithm.
Riesz transform acts as a *quadrature operator* to convert real-valued signal into a monogenic signal in with hypercomplex representation so that implicit information such as phase, orientation, congruency, etc. can be further examined.

Motion magnification is the following step with Riesz pyramid (multi-scale Riesz transform) to process the image with a filter bank so that image contrast can be enhanced with very limited noise amplification.

**Post-processing**

Feature extraction, image registration, unsupervised image/semantic segmentation, pattern recognition, etc. are not limited to analytical methods. Machine learning and deep learning framework are always good to have. For example, we can do feature extraction by PCA, kernel-PCA, ICA, NMF, auto-encoder, etc. but machine learning with neural network kind of provides a unified method to get the work done under its framework without doing parametric studies and making assumptions. Machine learning somehow wipes out the selection bias between analytical methods.

**Add-on**

Beyond the machine learning framework, some study can be made such as

- Optimization of hyperparameters
- Choice of imaging conditions, correlation and cost functions
- Multi-sensor data fusion at the beginning or the final stage

### 7.5 The Insight of ADPNet – Advanced Damage Processing Network

Under the machine learning framework, the majority of feature related parameters are learned iteratively. Since there are several hyperparameters, it’s possible for a learning algorithm to optimize the selection of them. The optimization procedure yields an optimal solution to minimize the predefined loss function such as PSNR to evaluate the contrast, cross-entropy to observe the additive information, log-likelihood and KL divergence to estimate the similarities, etc.

The proposed ADPNet includes the following kernels: (1) several pre-processors such as RANSAC (matching geometric features to remove translation and rotation by RANdom SAmple Consensus), Riesz bp transform (converting image into hypercomplex domain); (2) an image correlation function with SSIM (comparing image similarities); and (3) several post-processors such as PCA (extracting and projecting image into maximum variance direction by Principal Component Analysis), image fusion (de-correlating...
geometry features from damage imaging), non-linear filter (de-noising and removing speckles) and unsupervised image segmentation (clustering image features and highlighting damage profiles)

7.6 What’s More – Machine Learning Framework

Although NI tool provides a unified platform and easy customized GUI for prototyping, but the block diagram architecture is pretty hard to manage for collaborating and version control as the project iterates. Otherwise, FPGA itself has lots of limitations for compiling and LabVIEW itself does not support open-source library to expand its capabilities. The migration to the open-source world with TensorFlow, Caffe or Pytorch on a CUDA platform such as Jetson Nano, Xavier or Tesla, etc. will be the next affordable solution to accelerate the development for the consultative projects.
8 Experiment Preparation for LSP/SG with DPE

Experiments are applied with thermal loading and ultrasonic vibration for large-area inspection. Through each stage, different light sources are applied and compared on several dimensions.

Honeycomb composite panel

- For small scale inspection and hyperparameter study
- Providing insights (power, coherence, portability, etc) for laser source selection

C-17 Globemaster III composite aileron

- Realization of large-area inspection
- Demonstrating the capabilities of blind inspection
- Compromising the power-coherence trade-off

Evaluation dimensions

- Qualitative analysis for localization and defect detection
- Quantitative analysis for size estimation and severity evaluation
- Image fusion, segmentation and, classification.
- Clustering with PCA, ICA, etc

8.1 Selection of Laser Source

All the laser speckle interferometry (LSI) techniques are based around the same working principles to measure the change of physical quantities (intensity, frequency or phase) in the laser speckle caused by rigid body motion (translation and rotation) or deformation on the object surface. It could be a single point laser speckle for high-speed measurement with a photo detector (PD) at GHz range like LDV as one of the P-NDI techniques or millions of randomly distributed laser speckles as a speckle cloud with a CMOS camera (few mega pixels spatial resolution and hundreds of frames per second) as one of the mainstream V-NDI techniques.

Laser speckles created by different lasers are shown in (Figure 8.1); (a) even the reflection light from a HeNe laser on a rough surface creates speckles and (b) needs further collimation to bring it into PD sensing area. For LSP large area applications using a HeNe laser, we need to (c) expand the beam size so that it can cover a larger area of interest (AOI) or use (d) a diode laser which has a larger divergence angle, smaller
waist distance, smaller coherence length but much larger power to cover the entire AOI. According to their minute working principles, LSI techniques can be categorized into DSPI, LSP and SG for different applications.

![Image](image.png)

Figure 8.1 (a) HeNe laser reflected from a rough surface, (b) collimated HeNe laser reflected from a rough surface, (c) HeNe laser with beam expander, and (d) high-power diode laser illumination

Traditional interferometry has very high demands on laser coherence length and beam profile where HeNe laser turns out to be one of the candidates to satisfy these highly sensitive applications for single-base detection. However, the advancing of laser speckle and image processing techniques open a new paradigm to lower down the requirement of coherence length and make low coherence interferometry/correlation possible. The comparison of beam profiles and characteristics are shown as (Table 8.1) for HeNe, diode, and high-power diode laser.

<table>
<thead>
<tr>
<th></th>
<th>HeNe</th>
<th>Diode</th>
<th>High power diode</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HeNe</strong></td>
<td>- Tens of mW and high coherence</td>
<td>- Tens of mW and low coherence (better than high power diode)</td>
<td>- Hundreds of mW to few W and low coherence</td>
</tr>
<tr>
<td></td>
<td>- Gaussian beam with low divergence angle &lt; 0.1 degree</td>
<td>- Close to Gaussian beam with low divergence angle &lt; 1 degree</td>
<td>- Not even close to Gaussian beam</td>
</tr>
<tr>
<td></td>
<td>- Bulky, fragile and expensive</td>
<td>- Light and more cost effective to build an array</td>
<td>- Very large divergence angle &gt; 20 degree</td>
</tr>
<tr>
<td></td>
<td>- Good for remote sensing and interferometry applications</td>
<td>- Compromising between HeNe and high power diode</td>
<td>- Need controller for active cooling and current stabilization</td>
</tr>
</tbody>
</table>

Table 8.1 HeNe, diode, high-power diode laser beam profile characteristic comparison [112]
For experiment, different types of lasers are shown as (Figure 8.2) are selected according the size of testing objects and the study of how much coherence length affects the results. On honeycomb composite sandwich panels with few inches by few inches area, diode and high-power diode laser are selected; and on C-17 Globemaster III composite aileron with two feet by three feet area (about one by 1.5 meters) high-power diode turns out to be the only option.

(a) (b) (c)

Figure 8.2 Laser source comparison about on the same scale, (a) Thorlabs HNL210LB linear polarized HeNe laser 632.8nm 21mW, (b) Thorlabs PL202 diode laser 635nm 0.9mW, (c) Thorlabs L638P700M high power diode laser 638nm 700mW

8.2 Selection of Excitations

In short, inspection is an interrogation process where the imaging condition has to be selected according to the type of excitations. Their characteristics are summarized as below.

<table>
<thead>
<tr>
<th>Excitations</th>
<th>Bandwidth</th>
<th>Power or energy output</th>
<th>Coupling</th>
<th>Problem types</th>
<th>Imaging condition category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermal</td>
<td>&lt; 10 Hz</td>
<td>Hundreds of W</td>
<td>Non-contact</td>
<td>Quasi-steady-state</td>
<td>State correlation</td>
</tr>
<tr>
<td>Impact hammer</td>
<td>&lt; 10 kHz</td>
<td>A few J</td>
<td>Contact</td>
<td>Transient dynamics</td>
<td>Wave-based correlation</td>
</tr>
<tr>
<td>Mechanical shaker</td>
<td>&lt; 15 kHz</td>
<td>Tens of W</td>
<td>Contact</td>
<td>Steady-state dynamics</td>
<td>Vibration-based correlation</td>
</tr>
<tr>
<td>PZT shaker</td>
<td>&lt; 100 kHz</td>
<td>Tens of W</td>
<td>Contact</td>
<td>Steady-state / transient dynamics</td>
<td>Vibration / wave-based correlation</td>
</tr>
<tr>
<td>Speaker</td>
<td>&lt; 20 kHz</td>
<td>Hundreds of W</td>
<td>Non-contact</td>
<td>Steady-state / transient dynamics</td>
<td>Vibration / wave-based correlation</td>
</tr>
<tr>
<td>ACT</td>
<td>&lt; 300 kHz</td>
<td>Tens of W</td>
<td>Non-contact</td>
<td>Steady-state / transient dynamics</td>
<td>Vibration / wave-based correlation</td>
</tr>
<tr>
<td>PZT wafer</td>
<td>&lt; 300 kHz</td>
<td>A few W</td>
<td>Contact</td>
<td>Steady-state / transient dynamics</td>
<td>Vibration / wave-based correlation</td>
</tr>
<tr>
<td>Pulse laser</td>
<td>&gt; 200 kHz</td>
<td>mJ</td>
<td>Non-contact</td>
<td>Steady-state / transient dynamics</td>
<td>Vibration / wave-based correlation</td>
</tr>
</tbody>
</table>
9  Experiment Demo – Honeycomb Composite Panel

The structural component to be tested was a honeycomb composite panel (254 mm × 178 mm) shown in (Figure 9.1). The face-sheet is a 6-ply of TE-1 grade 190 type 35 carbon/epoxy with a stacking sequence of [60/0/-60]s, with nominal thickness 1.14 mm. The core is a Hexcel CR III 1/8-5052-.0007 lightweight aluminum honeycomb core with 25.4 mm thick. The material properties can be found in [113]. The panel was impacted at the center and normal to the surface with a 25.4 mm diameter blunt hemispherical impactor made of hardened steel and impact energy of 2 J. The damage is about at the center of the specimen with two indentations.

![Figure 9.1] Honeycomb composite panel with low velocity impact BVID at the center

The impacted panel was bent convex slightly and the maximum dent was measured to be 0.53 mm deep. The resulting barely visible impact damage (BVID) may encounter delamination in the impacted face-sheet, disbond between face-sheet and core, and crushed honeycomb core itself. The scan area of 50 mm × 50 mm covering the impacted damage area in the panel, as shown in (Figure 9.1), was first evaluated through an ultrasonic C-scan to roughly identify the dent region and show the existence of delaminations; then a detail information for the delaminated area was obtained via a 27.5 μm resolution X-ray computed tomographic (X-ray CT).

9.1  Baseline – X-ray CT and Ultrasonic C-scan

Ultrasonic C-scan merely showed the existence of delamination in the face-sheet, but without information of the delaminated area. The panel was then interrogated via X-ray CT scan for further evaluation. The images obtained by X-ray CT scan focusing on a plane 0.15 mm below from the core/facesheet interface where the maximum projected area of delamination was reached.
The horizontal AA and BB sections, respectively shown at $x = 23 \, mm$ and $x = 17 \, mm$, display the crushed core area and show the depth of delamination. The vertical cross-section CC at $x = 22 \, mm$ also displays a similar damage pattern. From the three cross-sectional images, the shape of main delamination area can be approximated as an ellipse with major and minor axis lengths, $17 \, mm$ and $10 \, mm$ respectively and oriented 105 degrees from the positive $x$ axis. More details referred to [105], [114], [115].

The CT scan area $30 \, mm \times 35 \, mm$ was imaged at the impact location instead of the entire laser interrogating area $50 \, mm \times 50 \, mm$ using LDV. Three cross-sections side view (AA, BB, and CC) near the delamination location were taken from the X-ray CT scan. The top view reveals two dark projected delaminated regions in the face-sheet, where the main delamination is centered around $x = 23mm$, $y = 20mm$ and the secondary delamination is centered around $x = 17mm$, $y = 17mm$. 

Figure 9.2  BVID characterization via ultrasonic C-scan and X-ray CT-scan

Figure 9.3  Ultrasonic C-scan through the layers
9.2 Baseline – Pulse Laser/LDV

The zero-lag cross-correlation (ZLCC) result shows the damage region at several resonant frequencies which are consistent with the X-ray CT-scan and ultrasonic C-scan results. Three frequencies yielding higher ZLCC image intensities and highlights the corresponding first two resonance modes of the main delaminated area. The lowest fundamental mode is shown at a frequency of about $25 \text{ kHz}$, while the second mode is appeared at about $65 \text{ kHz}$ and then the third at about $90 \text{ kHz}$ shown as (Figure 9.5). More details referred to [116].
9.3 Thermal Excitation – Diode and High-power Diode Laser – LSP/SG

The preliminary study of LSP and SG for large-area damage inspection. These techniques should be extended readily to visualize steady-state vibration or transient wavefield by stroboscopic if desirable. Original amplitude (A), local energy (LE) and local phase (LP) are the parameters we are interested in, and we also study the effectiveness of a bandpass spectrum filter in front of the camera to only allow a selected band of light passing through.

9.3.1 Experimental Setup

Thermal loading was applied on the structural surface by a 1500W heat lamp (110V ceramic infrared heat emitter) and monitored by a thermal camera (FLIR A310, radiometric thermal imaging IR camera) with resolution 320×320 pixels to ensure the composite specimen was not overheated. The heat lamp was placed with stand-off distance 10 inch to gradually heat the composite panel to 60 degree and the cool down in room temperature. Speckle patterns were recorded every second during the cool down process with duration 30 seconds, say total 30 images per loading cycle. As described before, correlating speckle patterns in time sequence has a larger tolerance speckle pattern fluctuation during measurement and is more preferable for qualitative analysis. The effect of light source coherence was also studied. NDI procedures include (1) qualitative analysis for localization and damage detection, and (2) quantitative, damage size estimation and severity evaluation.

![Figure 9.6 Thermal loading LSP/SG experimental setup](image)

The output of a laser source is expanded to illuminate the field of view (FOV). Two light sources are selected for comparison, a low-power diode laser (Thorlabs PL202, 635 nm, 0.9 mW) and a high-power diode laser (Thorlabs L638P700M, 638 nm, 700 mW). The coherent length of the laser was approximately inversely proportional to its power. Although both are diode lasers the PL202 has much higher coherent...
length than L638P700M because PL202 only has 1/800 power of L638P700M. The PL202 is however a compact diode laser, and it is rather easy to build an array when a larger illumination area is required for inspection. By contrast, the L638P700M requires a diode mount (Thorlabs LTC56B) with lens adapter, a driver (Thorlabs ITC5102 or LDC240C) for current control and a temperature controller (Thorlabs ITC5102 or TED200C) for active cooling due to its high power throughout. The entire system can be quite bulky with high power diode laser, but it can illuminate an area larger than 1 $m^2$ for fast inspection.

![Figure 9.7](image1.png)

**Figure 9.7** (a) LSP/SG setup for honeycomb composite sandwich panels with high power diode laser, (b) integrated cameras for LSP and SG

The light reflected from the structural surface is focused on the image plane of a camera. A high-resolution CMOS camera (Basler, acA3800-10gm, 3840×2748 pixels, 10 fps, monochrome) was selected with 12 bits (0 to 4095) pixel depth so that any minute intensity variation can be detected where traditional CMOS camera with 8 bits (0 to 255) pixel depth cannot. Monochrome is not just a software option, and it has a native monochrome sensor with no color filters. These results in twice the light sensitivity, and higher effective resolution due to the lack of color filters and the subsequent Bayer pattern demosaic algorithm. Monochrome sensors commonly used for scientific research and experiments.

![Figure 9.8](image2.png)

**Figure 9.8** (a) Speckle cloud created by high-power diode laser, (b) a close look of front panel to monitor speckle cloud, (c) real-time imaging processing with LabVIEW FPGA
Images are acquired by a dual channel frame grabber (NI PXIe-8234) through Ethernet with vision acquisition software in LabVIEW. With the customized front panel GUI, speckle patterns can be monitored and recorded on a controller (PXIe-8133) integrated in a high bandwidth PXI chassis (NI PXIe-1062Q). Real-time image processing was done by LabVIEW FPGA (Figure 9.7 and Figure 9.8).

An interferometry system highly relies on illumination light source which will affect image contrast and the choice of algorithms for image processing. The higher the coherent length, the higher contrast of the image but the more noise will be picked up during the measurement. For quantitative analysis, high coherent light source is preferable with higher sensitivity to hidden damages.

### 9.3.2 Results and Comparisons

Results are discussed for LSP and SG with different dimensions. Since LSP and the degenerated SG, we would like to know more about the how the light source effects the results. Then we can more to SG to see the efficiency of a spectrum filter to reject noise from ambient illumination.

#### Case 1 – LSP with high-power diode and diode for coherence study

We firstly use high-power laser to cover the entire specimen shown as (Figure 9.10 the 1st row). The results of A, LE and LP are similar. LE has higher contrast and A where LP reveals more information such as fiber waviness textures than A and LE. This makes sense that LP is a more sensitive but also a more noisy parameter as expected.

Then replace with diode laser with higher coherence laser but can only cover a circular region shown as (Figure 9.10 the 2nd row). With this higher coherence laser, A, LE and LP can see the in-phase pattern which
is an important evidence of interferometry where low coherence interferometry does not have. The confined damage area from all the variables is about the same.

Figure 9.10  LSP results (A, LE, LP) with high-power diode (1\textsuperscript{st} row) and diode (2\textsuperscript{nd} row) laser

For visualization, gray scale provides a continuous map which is a good candidate for phase such as LP observation where rainbow map is feature-oriented choice for LE.

Figure 9.11  LSP results (A, LE, LP) with high-power diode (1\textsuperscript{st} row) and diode (2\textsuperscript{nd} row) laser
Case 2 – SG with diode laser for spectrum filter effectiveness study

Use diode laser with spectrum filter for SG shown as (Figure 9.12 the 1st row) to see out-of-phase fringe patterns where the patterns in LSP is in-phase. This provides another evidence that the SG setup is correct. A and LE have similar results where LP is noisy in this case.

For visualization, gray scale provides a continuous map which is a good candidate for phase such as LP observation where rainbow map is feature-oriented choice for LE. Use diode laser without spectrum filter for SG shown as (Figure 9.12 the 2nd row). Without spectrum filter, all of the images are quite blurred. This confirms the requirement of spectrum filter in an open space or industrial environment.
Case 3 – SG with high-power diode laser for shear distance and sensitivity study

Since shear distance (SD) dominates the measurement sensitivity, we’d lie to know more about it. Following the same processing procedures as previous case, LE and LP are generated by Riesz bp for correlation. Thermal loading is applied on front (Figure 9.14) and back (Figure 9.15) sides of the panel. When SD is larger, the damage feature will be more separated and like two adjacent out-of-phase damages. This is the key to judge it is a damage with large SD or two small damages. Although LE and LP have very similar results, LP has more smooth features than LE.

Figure 9.14  Shear distance (SD) comparison – panel front side

Figure 9.15  Shear distance (SD) comparison – panel back side
9.3.3 Summary

After replacing the high-power 800mW diode laser with regular 5mW diode laser, the 1st order in-phase fringes are revealed but the inspection region is quite limited due to the limited power throughput. The circular region was created by beam expander to collimate the laser beam to the region of interest (ROI) to those hidden damages so that they can be further analyzed according to the location enlightened by the 1st stage inspection. Summary are made for the following cases.

- Case 1 – LSP with high-power diode and diode for coherence study
- Case 2 – SG with diode laser for spectrum filter effectiveness study
- Case 3 – SG with high-power diode laser for shear distance and sensitivity study

Case 1 – LSP with high-power diode and diode for coherence study & Case 2 – SG with diode laser for spectrum filter effectiveness study

Side-by-side comparisons are shown as (Figure 9.16) for LE and (Figure 9.17) for LP.

High-power low coherence light source illuminates the entire specimen with good position agreement of baseline although damage boundary contour is smaller than it should be. LSP takes the benefit of fast
evaluation to localize the hidden damages, but it’s limited to qualitative analysis when its phase is a function displacement.

Furthermore, SG has out-of-phase fringe patterns where the fringe patterns in LSP is in-phase. Mathematically, the out-of-phase pattern is derived from taking the derivative optically of the in-phase pattern so that SG intrinsically has higher sensitive inspection capabilities for small features such as fiber woven textures. Shear distance (SD) was carefully tuned until the 1st order fringes can be seen. With spectrum filter, ambient light was filtered out and the image has much higher contrast than the one without spectrum filter.

**Case 3 - SG with high-power diode laser for shear distance and sensitivity study**

(Figure 9.14) is the results of shot duration correlation, if we correlate the images with a longer duration, local and global fringes will be revealed (Figure 9.18 and Figure 9.19) where the global fringe may overlay the local fringe and make it difficult to be saw. In most of the cases, a marginal SD is good for damage detection, unless the damage induced phase change is very small. A good value of SD is large enough to reveal the damage without seeing the global fringe so that the maximum image contrast can be obtained.

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 9.18**  Shear distance (SD) long duration study – front side panel – LE

<table>
<thead>
<tr>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 9.19**  Shear distance (SD) long duration study – front side panel – LP
9.4 Thermal Excitation – HeNe Laser – SG

To demonstrate the large area inspection in real-time using the Riesz bp SG, a honeycomb panel with a barely visible impact damage (BVID) was excited by a heating lamp through the thickness of the panel. Speckle pattern variation was introduced by inhomogeneous displacement distribution caused by thermal gradient induced by the BVID. This loading can also be substituted by other devices such as mechanical shaker or speaker for different applications.

9.4.1 Experimental Setup

An improved version two LSP/SG was released to with dual high-power diode laser and HeNe laser so that this system can be adaptive to different measurement scenarios. Also, another independent LSP is included with the integrated LSP/SG shown as below.

Figure 9.20 Thermal loading SG experimental setup – version one

Figure 9.21 Thermal loading SG experimental setup – version two
9.4.2 Results and Comparisons

SG with thermal excitation is a gradient-based damage method where any damage-induced discontinuities can be revealed effectively. In order to have a minimal hardware setup, an advanced image processing sequence was proposed for SG. The proposed Riesz bp transform with SG directly assesses phase information for each image using hypercomplex monogenic signal, derives a number of phase signatures for different image features.

Riesz bp here plays an important role to extract the local phase (LP) for correlation and local energy (LE) for local feature enveloping. The enveloping images are further processed with PCA for segmentation and feature separation. Different shear distance (SD) results are also studied.

Step 1. Speckle clouds are recorded in time sequence during cool down since the intensity is the only quantity can be captured by the camera. Any deformation will cause the variation of intensity and the key is to highlight the large variation regions which are potential damage locations.

![Figure 9.22 Speckle clouds recorded in time sequence during cool down – small SD](image)

![Figure 9.23 Speckle clouds recorded in time sequence during cool down – large SD](image)

Step 2. Before correlating images in time sequence, instantaneous phase (also called local phase, LP) has to be extracted via Riesz transform from their intensity map. Once the phase maps were correlated, fringes are revealed for quantitative analysis if necessary. However, potential damage locations are tangled with fringes especially there is global deformation causing global fringes.
Step 3. Speckle images are correlated to the selected reference image, here selecting the first image. Fringes itself can be interpreted as the information carrier but what its envelope is sought to reveal its features. In spatial representation, global fringes have a larger wavelength and can be easily filtered by a spatial lowpass or bandpass filter. The Riesz bp filter with selected wavelengths was applied on top of the fringes to extract the envelopes of damage features.
Step 4. The representation of the envelope in high dimensions is the also called local energy. Riesz transform can unfold the local energy for each of the image in time sequence but has nothing to do with image identification and segmentation where PCA appears to be a useful tool to find the direction with largest variance of the local energy, so called the 1st principal direction.

The secondary large one is the 2nd principal direction and so on. The eigenvalue here can be interpreted as how significant of each direction. Here the first three principal directions were extracted and each one has its own characteristic.

![Image projection on the 1st, 2nd and 3rd principal component directions – small SD](image1)

![Image projection on the 1st, 2nd and 3rd principal component directions – large SD](image2)

Step 5. Recall the baseline of a pulse laser/LDV system with selected resonance frequencies, 25 kHz, 65 kHz and 90 kHz as references.

![Pulse laser/LDV baseline images with different vibration modes of the delaminated area appear at resonance frequencies](image3)
9.4.3 Summary

A full-field optical non-destructive inspection (NDI) system based on shearography was developed. The system comprises of four major components: (1) a laser used to illuminate the structural surface providing the field of view (FOV); (2) a heating lamp for providing thermal stress loading on the structure; (3) an image acquisition subsystem including a CMOS camera (monochrome sharp contrast image sensor is preferable) with a shear device to capture the interference speckle patterns from time-varying temperature distribution (intensity) on the laser-illuminated surface; and (4) a direct phase estimation (DPE) imaging processing algorithm, Riesz transform with log-Gabor bandpass (bp) filter, called Riesz bp transform in short, for damage visualization.

The Riesz bp transform with SG utilizes interference speckle patterns in time sequence without reference arm and thus are insensitive to external disturbances, in which other optical methods such as DSPI techniques suffers greatly. The traditional SG system which relies on the optical setup for phase shifting to evaluate the phase difference pixel by pixel via subtraction.

The proposed Riesz bp transform with SG directly assesses phase information for each image using hypercomplex monogenic signal, derives a number of phase signatures for different image features. The log-Gabor bandpass filter permits the correlation between localized windows for phase change. As a result, the system employing the numerical DPE can be much more compact and robust to successively monitor the phase state changes when more images are taken.

The proposed imaging algorithm can be used to supplant the optical setup for estimating phase shifting in SG and can be applied on other laser speckle interferometry (LSI) such as laser speckle photometry (LSP) and electronic speckle pattern interferometry (DSPI). Direct phase estimation (DPE) is a generalized method to evaluate the intrinsic phase from intensity image so that it can be implemented not only to quasistatic or steady state but also to transient states with the use of stroboscopic setup or high-speed camera.

Comparing to temporal phase stepping (TPS) and spatial phase stepping (SPS) which rely on the hardware (optical setup) mostly, DPE brings the entire process into image post-processing (numerically) and takes the benefits of using consumer level cameras or even the smart phone in the future.
9.5 Ultrasonic Excitation – HeNe Laser – SG

Instead of applying thermal loading, piezo shaker from isi-sys with suction cup was attached on the back of honey composite sandwich panel as an excitation source. The goal is looking for local resonance of the hidden damage at the front side of the panel. Time-average was taken for each incremental frequency with image frame rate 10 fps for two seconds – total 20 images per frequency.

9.5.1 Experimental Setup

The setup (Figure 9.31) is pretty much the same as the thermal loading experiment in the previous section. The PS-X-03-Series piezo shaker is designed for non-destructive testing purpose in combination with laser speckle interferometry (LSI). It is fixed on the object surface by vacuum via a suction cup. The pressing force is adjustable by the vacuum level. Different actuator types are available for different frequency range and forces in combination with different counterweight mass. Coupler or ultrasonic gel is recommended for better power transmission efficiency.

(a) (b) (c)

Figure 9.31 Setup of (a) integrated SG with HeNe laser, (b) piezo shaker attached on the back of the panel, and (c) function generator, amplify and oscilloscope

Since the piezo shakers are consumables and can be destroyed by high loads through the amplifier as well as at high operation temperatures, we would like to know the excitation level of deformation and the correlation performance under the steady-state vibration in ultrasonic range. The ideal setup is the amplifier gain is large enough to drive the piezo shaker, but not too large to burn it out, and shear distance (SD) is not too large to avoid spatial aliasing. The spatial aliasing only happens under steady-state vibration when generating a full-field deformation represented as mode shapes. As mentioned in the previous chapter, the high frequency excitation in ultrasonic range limits the upper bound of SD and has to be carefully checked when boosting up the excitation frequencies.
9.5.2 Results and Comparisons

Two experiments are performed shown as below. Firstly, looking for the an expected wavefield to tune experimental parameters and then adding the frequency incrementally for local resonance.

Case 1 – excitation trial test to capture the mode shape – panel back side

The piezo shaker was attached on the front side right bottom corner of the panel (Figure 9.32) and measured on the back side of the panel. There is an impact damage at the center of the back side of the panel and another one at the right cross of the back side. When the mode shape was firstly measured (Figure 9.33 left), we’re not quite sure how does it should look like at 20 kHz, so the simulation was performed looking for resonant frequencies around 20 kHz. A wavefront towards right-up corner can be observed in both measured wavefield and the simulation (Figure 9.33 right) mode shape at its 82th mode – 19.338 kHz.

Figure 9.32 (Left) piezo shaker attachment and (right) expected wavefield under steady-state vibration with Hygiene wavefront

- Try to find the domination modes around 20kHz
- Haven’t run force excitation SSD yet
- Wavelength and frequency should be correct although accurate material properties are unknown
  - Lumped macro model – equivalent stiffness – isotropic

Figure 9.33 (Left) measured wavefield at 20 kHz and (right) simulated mode shapes around 20 kHz
At resonance frequency demonstrated as below (Figure 9.34) at 15 kHz, the hidden damage can be revealed clearly which is not a case at 20 kHz but the mode shape can be observed under the continuous excitation as a steady-state vibration.

![Figure 9.34](image)

**Figure 9.34** The hidden damage at the right cross position (as reference left cross is intact position) was detected at 15 kHz and mode shape at 20 kHz

One of the noticeable signs of losing correlation shown as (Figure 9.35) can be observed in jet color scale but very hard to notice in gray scale. When the domination color of the color map suddenly changes, for example the map at 10 and 15 kHz are mostly red but at 20, 25 and 30 kHz are blue-green dominated, indicating something is wrong during the correlation. That’s why in this case at 15 kHz (Figure 9.35) not revealing any damage where (Figure 9.34) is correctly correlated.

![Figure 9.35](image)

**Figure 9.35** Measured mode shapes at every 5 kHz incremental frequency

**Case 2 – looking for local resonance by incremental frequencies – panel front side**

Since we have confidence to measure the mode shape, we can incrementally increase the frequencies to identify the resonance frequency. According to the baseline (Figure 9.37) measurement, the 1st and 2nd resonant frequency should be 25 and 65 kHz respectively.

Due to the limitation of piezo shaker that 65 kHz cannot be reached, then is pretty much high enough to excite up to 40 kHz. Then, our goal is to duplicate the damage image with proposed vision-based full-field ultrasonic SG. The excitation starts from 1 kHz to 35 kHz and the correlated images (DC image as the reference) are shown as (Figure 9.36). Every frequency takes 20 images with a 10-fps camera so it take
about two seconds measurement at each frequency, total about 30 seconds for the demonstrated 10 frequencies. The white square is indicated as the LDV/pulse laser measurement region for comparison.

Figure 9.36  Measured mode shapes with incremental frequencies to identify the local resonance

The correlation images (Figure 9.36) are the results of mode shape based on the SG measurement without adding any imaging condition yet so the damage region so far looks larger than the baseline (Figure 9.37) as expected. When the excitation is close to the resonance, the local deformation (so as its gradient, strain) is much larger than the global mode shape so that the damage location can be identified.

Figure 9.37  Recall the pulse/laser LDV

9.5.3 Summary

Imaging conditions are not applied in this experiment since our goal is to identify the damage location and prove the concept of this vision-based ultrasonic SG with full-field measurement. Imaging conditions such as 2D wavelet transform, wavefield RMS, wavenumber index (WI), instantaneous wavenumber (IW) filter, etc. can be the consecutive step to lump these images (Figure 9.36) at each frequency to a single damage image just like the baseline (Figure 9.37). In short, the vision-based inspection accomplished the full-field measurement in 30 seconds, in contrast the pulse laser/LDV system takes several hours for a 2 by 2 inch measurement. There’s no doubt the vision-based interferometry will be the future for fast inspection.
### 10 Experiment Demo – C-17 Globemaster II Composite Aileron

After successful demonstrated the inspection capabilities on a sandwich composite honeycomb, here we extend the inspection area from inch-by-inch to meter-by-meter scale. LSP and LSP/SG experiments are performed with thermal loading on a C-17 Globemaster III composite aileron (Figure 10.1 right) to realize *large-area damage blind detection* without prior knowledge of the structure.

The complex geometry with pivot joints and material/geometry discontinuities increases the inspection difficulty. The panel was impacted at NASA Langley Research Center with a $\phi = 25.4 \text{ mm}$ hemispherical impactor with impact energy of $15 J$ ($11 \text{ ft} \cdot \text{lbs}$). The aileron geometry and impact locations (# 1 and # 2) are shown in (Figure 10.1 left top).

![Figure 10.1 Baseline measured by ultrasonic C-scan setup](image)

The T-joint stiffeners are also made of composites, and the layup is yet unknown. The joint region, stiffener, and damage location is illustrated in as a side view of cross-section in (Figure 10.1 center top). The joint region is about $42 \text{ mm}$ wide with the $3 \text{ mm}$ thick stiffener in the center and the impact damage is located only about $7 \text{ mm}$ from the edge of the joint. Even though the BVID created on the face laminate only shows as a small superficial dent on the surface, it may reduce the strength of the composite structure significantly by 20 to 30%. More details referred to [105], [114], [115].
10.1 Baseline – Ultrasonic C-scan

For comparison, the impact damage was imaged by ultrasonic C-scan as the baseline. The ultrasonic C-scan system consists of an Olympus submersion transducer and a Panametrics Model 5055 PR-101 pulser receiver operating at 10 MHz. The sampling resolution was 0.254 mm, and the sampling area was a 38.3×38.3 mm (1.5×1.5 inch) covering the damage area with 151×151 pixels image resolution.

The images were interpolated in MATLAB to produce higher-definition images. The ultrasonic C-scan system setup is shown in (Figure 10.1) and the layer-by-layer results are shown in (Figure 10.2). An RMS algorithm embedded in the C-scan system was employed to generate the damage image throughout the thickness to highlight all of the delaminations underneath the composite laminate. More details referred to [105], [114], [115].
10.2 Baseline – Pulse Laser/LDV

The pulse laser/LDV setup (Figure 10.3 left) and ZLCC results (Figure 10.3 right) are shown below. The joint region is marked as a white shaded area on the left of the wave-field image, and the stiffener is marked as the shaded area with white dotted line. The damage boundaries are marked as the red dashed line with the rectangular ultrasonic C-scan region marked as the red solid line.

The damage imaging at location #1 with guided waves propagating toward right is shown as (Figure 10.3a) where the T-joint stiffener is detected at the left. At location #2 with left propagating shown in (Figure 10.3b), at location #3 shown as (Figure 10.3c) and at location #4 shown as (Figure 10.3d) where the delamination dominates the local energy accumulation with local resonance. The more energy accumulated the higher contrast the damage image is. The entire scanning and post-processing time take several hours depending to the scanning rate and processing power. More details referred to [105], [114], [115].

Figure 10.3 Baseline measured by pulse laser/LDV

Figure 10.4 Wavenumber index (WI) and zero-lag cross-correlation (ZLCC) comparison
10.3 Thermal Excitation – High-power Diode Laser – LSP/ADPNet

Following the same procedures as the inspection of honeycomb composite sandwich panel for LSP. An advanced damage processing network (ADPNet) was proposed to observe the variation of speckle clouds in time sequence and interpret the images in high dimensions. ADPNet comprised of the Riesz bp transform, non-linear filter bank and unsupervised image segmentation to quantify/characterize barely visible impact damages (BVID) on a C-17 Globemaster III composite aileron. It is also demonstrated to be more accurate and robust than Digital Image Correlation (DIC) for minute deformation (sub-nano to nano meter) measurement and large area (meter-by-meter) inspection under industrial environment.

10.3.1 Experimental Setup

Thermal loading was uniformly applied by a 150W heat lamp (110V ceramic infrared heat emitter) and monitored by a thermal camera (FLIR A310, radiometric thermal imaging IR camera) with resolution $320 \times 320$ pixel to ensure the composite aileron did not overheat. The heat lamp was placed with stand-off distance 10 inches to gradually heat the aileron to 60 degrees and cool down in room temperature. Speckle clouds were recorded every second during the cool down process with duration 30 seconds, say total 30 images per loading cycle.

Figure 10.5  (a) High-power diode laser controller with polarizer, cylindrical lens and diffuser set, (b) LSP setup with high-power diode laser for large-area inspection

A high-power diode laser (Thorlabs L638P700M, 638nm, 700mW) was selected to cover the entire aileron span. The L638P700M requires a diode mount (Thorlabs LTC56B) with lens adapter, a driver (Thorlabs ITC5102 or LDC240C) for current control and a temperature controller (Thorlabs ITC5102 or TED200C) for active cooling due to its high power throughout. The high-power diode laser with tunable cylindrical lens system is shown as (Figure 10.5 left) and the entire system setup is shown as (Figure 10.5 right).
Figure 10.6    Cylindrical lens for one side collimation

Same as before to deploy speckles by high power diode laser shown as (Figure 10.5) and the additional lens set collimates the laser to desired AOI/FOV. Because of the properties of one side divergence, a cylindrical lens was selected shown as (Figure 10.6).

A high-resolution CMOS camera (Basler ace, acA3800-10gm, 3840×2784, 10 fps, monochrome) was selected with a 12-bit (0 to 4095) pixel depth so that more minute intensity variations can be detected compared to a traditional CMOS camera with only an 8-bit (0 to 255) pixel depth. Images are acquired by an Ethernet frame grabber (PXIe-8234) with vision acquisition software in LabVIEW.

A customized front panel GUI was made to monitor speckle clouds recorded on an integrated controller (PXIe-8133) and high bandwidth chassis (PXIe-1062Q). Real-time image processing was implemented on a FlexRIO module (PXIe-7975) with LabVIEW FPGA.

10.3.2 The Insight of ADPNet – Advanced Damage Processing Network

Under the machine learning framework, the majority of feature related parameters are learned iteratively. Since there are several hyperparameters, it’s possible for a learning algorithm to optimize the selection of them. The optimization procedure yields an optimal solution to minimize the predefined loss function such as PSNR to evaluate the contrast, cross-entropy to observe the additive information, log-likelihood and KL divergence to estimate the similarities, etc.

The proposed ADPNet includes the following kernels: (1) several pre-processors such as RANSAC (matching geo-metric features to remove translation and rotation by RANdom SAmple Consensus), Riesz bp transform (converting image into hypercomplex domain); (2) an image correlation function with SSIM (comparing image similarities); and (3) several post-processors such as PCA (extracting and projecting image into maximum variance direction by Principal Component Analysis), image fusion (decorrelating geometry features from damage imaging), non-linear filter (denoising and removing speckles) and unsupervised image segmentation (clustering image features and highlighting damage profiles)
10.3.3 Results and Comparisons – LSP/ADPNet

The processing sequence for LSP/ADPNet is shown as below from (Figure 10.7) to (Figure 10.10). Each step can be done recursively with a given loss function so that the maximum rewards can be achieved. The recursive form Riesz bp transform is denoted as a higher order extended Riesz transform to expand the dimensionality for highly non-linear features.

Speckle clouds are recorded in time sequence shown in (Figure 10.7) during the cool down since intensity is the only quantity that can be captured by camera. Any deformation will cause a variation of intensity but with different variance. The higher variance the higher potential a certain region has hidden damages.

Before correlating speckle clouds in time sequence, instantaneous phase (also called local phase, LP) or other quantities had to be extracted from the intensity map. Once the phase maps were correlated shown in (Figure 10.8), fringes were revealed for quantitative analysis if necessary but mostly only global fringes can be observed when the deformation was weakly induced by the damage.

The correlation map is not only a speckle variation map but also includes geometric features. Since we have a set of correlation images, they can be interpreted with principal component analysis (PCA) in very high dimensions to separate the geometric features (low variance) and damage images (high variance) shown in (Figure 10.9 left and middle). Once we have the two of them, the geometric features can be fused back to the damage image to enrich the information (Figure 10.9 right).

Non-linear filters can further filter out speckles without losing the contrast and feature sharpness. Some parametric studies of median filter (contrast enhancement), non-local mean filter (balanced judgment) and Laplacian filter (edge detection) are shown in (Figure 10.10). Unsupervised image segmentation as a higher-level cluster (PCA as a lower-level cluster) can be applied on top of the filter to explore hidden features. [117]

The PCA procedures are shown as

1) Align pixels into a long vector $X_i \in \mathbb{R}^d$ and form the design matrix

2) Treat every pixel as a random variable with mean and covariance $X_i \sim (\mu_{d \times 1}, \Sigma_{d \times d})$

3) Decompose covariance into $\Sigma_{d \times d} = P_{d \times d} \Sigma_{d \times d} P_{d \times d}^T \rightarrow X_i \sim (\mu_{d \times 1}, \Sigma'_{d \times d} = P_{d \times d} \Sigma_{d \times d} P_{d \times d}^T)$

4) Linear transformation of the random variable $P^T X_i \sim (\mu_{d \times 1}, P^T \Sigma_{d \times d} P = P^T P D_{d \times d} P^T P = D_{d \times d})$

5) Extract the 1st few principal components as eigenvectors
Figure 10.7  Speckle clouds recorded in time sequence during cool down

Figure 10.8  Instantaneous phase extracted via Riesz bp transform to create a phase map for each image so that they can be correlated to each other in time sequence

Figure 10.9  Image projection on the largest and the smallest principal directions by PCA to highlight the large variation regions and keep geometric features for imaging fusion

Figure 10.10  Post-processed image with different types of non-linear filters for segmentation and feature separation
10.3.4 Summary

Recent works using a pulsed laser/LDV scan system and C-scan provided qualitative and quantitative baselines for comparison with the proposed images from LSP for the composite BVID inspection shown as below (Figure 10.11). Pulse laser/LDV is also a remote inspection technique but it’s time consuming and structure state has to be kept during the measurement. C-scan which in term as a contact inspection technique provides very high accuracy through the thickness but the inspection area is quite small. With the aim of laser speckle cloud as a transducer to convert minute deformations into intensities and recursively update the results with acquiring images, false positive rate is decreasing when more images are taken. DIC image is also included with multiple false positive detections, without pivot and stiffener features.

The Riesz bp transform aims at converting image intensity into the hypercomplex domain and extracting its instantaneous phase for further processing. When the structure was assigned with transient thermal excitation, the deformation has a modulated waveform, and the intrinsic phase is embedded within its spatial carriers. Any inhomogeneous deformation reveals discontinuity and subsequently changes the phase. Then, correlating the phase maps in time sequence will highlight the location of damages and any feature that has large contribution to phase change. PCA as a lower-level cluster further separates damage features (large phase contribution, large variance) and geometry features (small phase contribution, small variance) to lower down the false positive detection rate (lower sensitivity but higher specificity). Image fusion reversely fuses the geometry features back to the damage image map as the final image for demonstration.
III. Waves

After the exploration of vision-based large-area damage detection techniques, let’s get started a theoretical page for waves as an information carrier either under vibration or propagation. The physics behind will be explained, discussed and visualized in the following chapters.

There exist many different wave phenomena that can be observed in the nature. From a whispering sound to earthquake waves, the mechanical waves involved can be studied from a unified framework. Thus, scalar waves are studied in fields like acoustics, and moreover, if shear elasticity is considered, more exotic phenomena can be observed and studied in electrodynamics, geophysics, or ultrasonics. [118], [119]

The polarization of the waves and the boundary conditions (infinite, semi-infinite and finite) lead to different types of waves but all of them are comprised by two fundamental waves: P-wave and S-wave.
11 Infinite Boundary Conditions

Bulk wave is the most fundamental wave that constitute of other waves. According to its polarization and particle motion direction, it can be categorized into pressure (longitudinal) and shear (transverse) waves existing in traction free boundary conditions. The notations conventions are showing as

- P-wave (also called longitudinal wave) with propagation speed $c_L$
- S-wave (also called transverse wave, including SV and SH) with propagation speed $c_T$

11.1 Pressure Waves – Longitudinal – P

Simulations of longitudinal waves can be exerted by finite differences method (FDM) or finite element method (FEM) explicit dynamics in time domain. Thus, accurate predictions of scalar acoustic waves can be performed showing particle displacements in the direction of the propagation direction. This wave can be observed in both solid elastic and fluid acoustic media. The propagation speed (speed of sound) of the acoustic waves in fluids can be expressed as:

$$c_L = \sqrt{\frac{K}{\rho}}$$  \hspace{1cm} \text{Eq. (11.1)}

where $K$ is the fluid compressibility and $\rho$ is medium density. In elastic solids, the P-wave speed is

$$c_L = \sqrt{\frac{K + \frac{\lambda + 2\mu}{2\rho}}{\rho}} = \sqrt{\frac{\lambda + 2\mu}{\rho}} = \sqrt{\frac{E(1-\nu)}{(1+\nu)(1-2\nu)\rho}}$$  \hspace{1cm} \text{Eq. (11.2)}

where $\lambda$ is Lame’s 1st parameter and $\mu$ is 2nd parameter also called shear modulus. The particle motion is parallel to the propagation direction.

11.2 Shear Waves – Transverse – SV and SH

On the other hand, particle motion is perpendicular to wave propagation direction. Shear waves can be vertically (SV) or horizontally (SH) polarized with propagation speed shown as below

$$c_T = \frac{\mu}{\sqrt{\rho}}$$  \hspace{1cm} \text{Eq. (11.3)}

The particle motions of SV and SH waves are perpendicular to the propagation direction.
11.3 Equation of Motion and Constitutive Relations

In the absence of body forces the components of the displacement vector in a homogeneous, isotropic, linearly elastic medium are governed by the following equation of motion that can be expressed by

\[ \mu \nabla^2 \mathbf{u} + (\lambda + \mu) \nabla \nabla \cdot \mathbf{u} = \rho \mathbf{u}^* \]  \hspace{1cm} \text{Eq. (11.1)}

where \( \mathbf{u} = (u_1, u_2, u_3) \) is the displacement in the medium at point \( \mathbf{x} = (x_1, x_2, x_3) \) and \( t \) is time, \( \rho \) is the density, \( \lambda \) and \( \mu \) are the Lame’s constant and shear modulus, respectively. Above can be derived from equilibrium equation in elasticity with \textit{traction free boundary conditions}

\[ \tau_{ij} + f_i = \rho u_i^* \]  \hspace{1cm} \text{Eq. (11.2)}

In the Helmholtz decomposition, the displacement vector is expressed by the sum of a scalar potential \( \phi \) and a vector potential \( \psi \)

\[ \mathbf{u} = \nabla \phi + \nabla \times \psi \]  \hspace{1cm} \text{Eq. (11.3)}

Note that Eq. (11.3) three components of the displacement vector to four other functions: the scalar potential and the three components of the vector potential. This indicates that \( \phi \) and the components of \( \psi \) should be under an additional constraint condition.

The condition \( \nabla \cdot \psi = 0 \) can provide the sufficient additional condition to uniquely determine the three components of \( \mathbf{u} \) from the four components of \( \phi \) and \( \psi \), but it is not a necessary condition. Some other relation between \( \phi \) and \( \psi \) must be specified if \( \nabla \cdot \psi = 0 \) is not used.

Substituting Eq. (11.3) into Eq. (11.1) yields

\[ \mu \nabla^2 \left( \nabla \phi + \nabla \times \psi \right) + (\lambda + \mu) \nabla \nabla \cdot \left( \nabla \phi + \nabla \times \psi \right) = \rho \left( \nabla \phi^* + \nabla \times \psi^* \right) \]  \hspace{1cm} \text{Eq. (11.4)}

Since \( \nabla \cdot \nabla \phi = \nabla^2 \phi \) and \( \nabla \cdot \nabla \times \psi = 0 \), the above equation upon rearranging terms leads to

\[ \nabla \left[ (\lambda + 2\mu) \nabla^2 \phi - \rho \phi^* \right] + \nabla \times \left[ \mu \nabla^2 \psi - \rho \psi^* \right] = 0 \]  \hspace{1cm} \text{Eq. (11.5)}
Clearly the equation of motion is satisfied if the potentials displacement potentials $\phi$ and $\psi$ satisfy the uncoupled wave equations shown as below

$$\nabla^2 \phi - \frac{1}{c_L^2} \frac{\partial^2 \phi}{\partial t^2} = 0 \quad \text{Eq. (11.6)}$$

$$\nabla^2 \psi - \frac{1}{c_T^2} \frac{\partial^2 \psi}{\partial t^2} = 0 \quad \text{Eq. (11.7)}$$

where

- $c_L = \sqrt{(\lambda + 2\mu)/\rho} = \sqrt{[E(1-v)]/[(1+v)(1-2v)]\rho}$
- $c_T = \sqrt{\mu/\rho} = \sqrt{E/[2(1+v)\rho]}$
- $\alpha = c_L / c_T = \sqrt{(\lambda + 2\mu)/\mu} = \sqrt{2(1-v)/(1-2v)}$

The scalar potential $\phi$ represents longitudinal waves traveling with a wave speed $c_L$ while the vector potential $\psi$ denotes transverse waves traveling with a wave speed $c_T$. Since Eq. (11.6) and Eq. (11.7) are simpler in form than the displacements equations of motion, elastic wave propagation problems in isotropic solids are usually approached by first expressing the displacement field in terms of these two potentials.

Both longitudinal and transverse waves are referred to as bulk waves since they propagate in the volume of bulk solids. With the identity $\nabla \times (\nabla \phi) = 0$ and $\nabla \cdot \nabla \times \psi = 0$ in which

- longitudinal waves are also called P-waves, pressure waves, compressional waves, dilatational waves, or irrotational waves, and
- transverse waves are often called shear waves, distortional waves, equivoluminal waves, or rotational waves.

Since both bulk wave speeds in the equations are independent of frequency, harmonic wave solutions of the type given below represent non-dispersive waves. This is strictly true, however, if the solid is perfectly elastic so that no frequency dependent energy attenuation or dissipation mechanisms are present.
The stress wave field at large distances (compared to the dominant wavelength or smallest characteristic dimension in the medium) from an excitation source can, for many purposes, be quite accurately represented by a plane wave in a Cartesian coordinate system:

\[
\mathbf{u} = \mathbf{a} \cdot f (\mathbf{k} \cdot \mathbf{x} - \omega t) = \mathbf{a} \cdot f (k \mathbf{\hat{k}} \cdot \mathbf{x} - \omega t) = \mathbf{a} \cdot f \left( \frac{\omega}{c_p} \mathbf{\hat{k}} \cdot \mathbf{x} - \omega t \right)
\]

\[
= \mathbf{a} \cdot f (\mathbf{\hat{k}} \cdot \mathbf{x} / c_p - t)
\]

where

- \( \mathbf{u} = (u_1, u_2, u_3) \) is displacement vector
- \( \mathbf{x} = (x_1, x_2, x_3) \) is location vector
- \( \mathbf{a} \) is defined by the vector of the direction of motion (displacement)
- \( \mathbf{\hat{k}} \) is the unit wavevector toward the direction of a wavevector \( \mathbf{k} = k \mathbf{\hat{k}} \)
- \( k = |\mathbf{k}| \) is wavenumber
- \( c_p \) is phase velocity

It is worth noting that \( \mathbf{\hat{k}} \cdot \mathbf{x} = constant \) describes a plane normal to the unit propagation wavevector \( \mathbf{\hat{k}} \) and the constant represents the distance of the plane from the origin (along the normal). If the position of the plane is associated with \( t = 0 \) then for \( t > 0 \) the plane moves with speed \( c_p \) in the direction of \( \mathbf{\hat{k}} \) hence the equation of the plane at time \( t \) described by \( \mathbf{\hat{k}} \cdot \mathbf{x} - c_p t = constant \) which is called a wavefront.

Since at any instant of time the wave crests lie in parallel planes, the motion represented by Eq. (11.8) is called a train of plane waves. Substituting Eq. (11.8) into Eq. (11.1) yields

\[
\left[ \mu \mathbf{a} + (\lambda + \mu)(\mathbf{\hat{k}} \cdot \mathbf{a})\mathbf{\hat{k}} - \rho c_p^2 \mathbf{a} \right] f'' (\mathbf{\hat{k}} \cdot \mathbf{x} / c_p - t) = 0 \tag{11.9}
\]

or

\[
(\mu - \rho c_p^2)\mathbf{a} + (\lambda + \mu)(\mathbf{\hat{k}} \cdot \mathbf{a})\mathbf{\hat{k}} = 0 \tag{11.10}
\]
Since \( \mathbf{k} \) and \( \mathbf{a} \) are two different vectors in Eq. (11.10) which can be satisfied in two ways only:

\[
\mathbf{a} \parallel \mathbf{k} \quad \text{or} \quad \mathbf{a} \cdot \mathbf{k} = 0 \quad \text{Eq. (11.11)}
\]

a) For \( \mathbf{a} \parallel \mathbf{k} \)

In this case, the time-varying displacement is parallel to the direction of propagation and the wave is therefore called \textit{bulk longitudinal wave}.

\[
c_p = c_L = \sqrt{\frac{\lambda + 2\mu}{\rho}} \quad \text{Eq. (11.12)}
\]

b) For \( \mathbf{a} \cdot \mathbf{k} = 0 \)

In this case, the time-varying displacement is normal to the direction of propagation and the wave is therefore called \textit{bulk transverse (shear) wave}. The displacement can have any direction in a plane normal to the direction of propagation.

\[
c_p = c_T = \sqrt{\frac{\mu}{\rho}} \quad \text{Eq. (11.13)}
\]

When the \((x, y)\) plane is normally chosen to contain the vector \( \mathbf{k} \), motions can be either in the \((x, y)\) plane or normal to the \((x, y)\) plane (or along the \(z\) direction). These transverse motions propagating at the same speed are called shear vertical (SV) and shear horizontal (SH) polarized waves, respectively. Note that to describe each type of waves, two constants (such as amplitude and phase) are needed to specify each wave type, in addition to the direction of propagation.

<table>
<thead>
<tr>
<th>Materials</th>
<th>( E ) (GPa)</th>
<th>( \nu )</th>
<th>( \rho ) (kg×10^3/m^3)</th>
<th>( c_L ) (km/sec)</th>
<th>( c_T ) (km/sec)</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td>air</td>
<td>0.0012</td>
<td>0.34</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>water</td>
<td>1.00</td>
<td>1.48</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>aluminum</td>
<td>73</td>
<td>0.33</td>
<td>2.77</td>
<td>6.25</td>
<td>3.15</td>
<td>1.99</td>
</tr>
<tr>
<td>steel</td>
<td>200</td>
<td>0.20</td>
<td>7.83</td>
<td>5.33</td>
<td>3.26</td>
<td>1.64</td>
</tr>
<tr>
<td>copper</td>
<td>115</td>
<td>0.34</td>
<td>8.90</td>
<td>4.46</td>
<td>2.20</td>
<td>2.03</td>
</tr>
<tr>
<td>glass</td>
<td>65</td>
<td>0.22</td>
<td>2.50</td>
<td>5.45</td>
<td>3.26</td>
<td>1.67</td>
</tr>
</tbody>
</table>
where \( \alpha = c_L / c_T = \sqrt{2(1-\nu)/(1-2\nu)} \) and \( \xi = c_T/c_L = \sqrt{(1-2\nu)/2(1-\nu)} \).

Harmonic waves are steady-state waves; waves which are not steady-state are said to be transient waves (pulses). However, the transient waves in linear elastic materials can be obtained by superimposing harmonic waves in Fourier integrals. A plane harmonic wave propagating with phase velocity \( c_p \) in a direction of wave vector \( \hat{k} \) is often convenient to represent the wave in a complex form

\[
\mathbf{u} = \mathbf{a} \exp \theta(\mathbf{x},t) = \mathbf{a} \exp i(\mathbf{k} \cdot \mathbf{x} - \omega t) = \mathbf{a} \exp i\omega \left( \frac{\hat{k} \cdot \mathbf{x}}{c_p} - t \right)
\]

Eq. (11.14)

where \( \mathbf{u} \) is a complex vector consisting of amplitude \( \mathbf{a} \) and phase angle \( \theta(\mathbf{x},t) \) consisting of phase velocity \( c_p = \omega / k \), angular (circular) frequency \( \omega \) and wavevector \( \mathbf{k} \).

In all linear operations on the complex wave form Eq. (11.14), the real part of the derived wave solutions is equal to the solutions of the same operations applied to the real part of the original wave form. The actual solution of Eq. (11.14) is

\[
\text{Re}[\mathbf{u}] = \text{Re}[\mathbf{a} \exp i(\mathbf{k} \cdot \mathbf{x} - \omega t)] = |\mathbf{a}| \cos(\mathbf{k} \cdot \mathbf{x} - \omega t + \eta)
\]

Eq. (11.15)

where \( \eta = \arg \mathbf{a} \)

The quantity of \( \theta(\mathbf{x},t) \) gives the relationship between \( \mathbf{x} \) and \( t \), and is generally called the phase of the waves; it determines the position on the cycle between a crest, where \( \mathbf{u} \) is maximum, and a trough, where \( \mathbf{u} \) is a minimum.

\[
\theta(\mathbf{x},t) = \mathbf{k} \cdot \mathbf{x} - \omega t
\]

Eq. (11.16)

In this phase wave equation, phase surfaces \( \theta = \text{constant} \) are parallel planes. Hence, Eq. (11.14) represents a plane wave whose planes of constant phases are normal to the \( \mathbf{k} \).

The gradient of \( \theta \) in space is the wavenumber \( k = |\mathbf{k}| \), whose direction is normal to the planes and whose magnitude is the average number of crests per \( 2\pi \) units of distance in that direction. Similarly, the gradient of \( \theta \) in time is the frequency \( \omega \), the average number of crests per \( 2\pi \) units of time.
\[
\frac{\partial \theta}{\partial x} = \nabla \theta(x, t) = \nabla (k \cdot x - \omega t) = \frac{\partial}{\partial x_j} \hat{e}_j k_i x_i = k_j x_i \hat{e}_i = k_i \hat{e}_i = k \quad \text{Eq. (11.17)}
\]

\[
\frac{\partial \theta}{\partial t} = \omega \quad \text{Eq. (11.18)}
\]

where \( \omega \) and \( k \) are related to the period \( T \) and wavelength \( \lambda \). Therefore, the following analog between time and spatial variables are shown as below

\[
\begin{array}{c|c|c}
\text{Period} & T & \text{Wavelength} \\
\text{Frequency} & \omega & \lambda \\
\text{Frequency} & \omega = \frac{2\pi}{T} & \text{Wavenumber} & k = \frac{2\pi}{\lambda} \quad \text{Eq. (11.19)}
\end{array}
\]

The wave motion is recognized from Eq. (11.14). Any particular phase surface is moving with normal velocity \( \omega / k \) in the direction of \( k \). To emphasize the direction of the phase velocity, a phase velocity vector can be defined as

\[
c_p = \frac{\omega}{k} \hat{k} \quad \text{Eq. (11.20)}
\]

where \( \hat{k} \) is directional unit vector of wavevector \( k = k \hat{k} \)

In terms of potentials, the harmonic wave solutions can be represented by

\[
\phi = A \exp i (k \cdot x - \omega t) \quad \text{Eq. (11.21)}
\]

\[
\psi = B \exp i (k \cdot x - \omega t) \quad \text{Eq. (11.22)}
\]

where scalar \( A \) and vector \( B \) are arbitrarily complex constants, \( k \) and \( \kappa \) are wavevectors for scalar and vector potentials, \( k = |k| \) and \( \kappa = |\kappa| \) are wavenumbers, \( c_L = \omega / k \) and \( c_T = \omega / \kappa \) are bulk wave velocities.
12 Semi-infinite Boundary Conditions

Surface wave is one side guided at the interface and at the side without boundary conditions. Surface waves exist in many ways such as Rayleigh wave, Love wave, gravity, water wave, etc. Note, not all the surface waves are non-dispersive.

12.1 Rayleigh Waves – R

Combinations of transversal and longitudinal waves under certain boundary condition can create a variety of wave dynamics. Thus, on a fluid/solid interphase Rayleigh superficial waves appear with a propagation speed of about 0.95 times $c_f$. A well-known mode of elastic energy propagation is the Rayleigh wave, which can exist on the free surface of a solid of infinite depth shown as (Figure 12.1). Note the particle trace describes a counterclockwise ellipse in the sagittal plane of the propagation direction.

The particle motion of the surface wave is elliptical and shown as below

![Rayleigh surface wave particle motion](image)

Figure 12.1 Rayleigh surface wave particle motion

For simple isotropic materials the mechanical displacements in this wave are elliptical, lie in a plane containing the propagation direction and the normal to the surface, and decay in an exponential manner to negligible values within a few wavelengths of the surface. The velocity of propagation is somewhat less than the bulk shear velocity associated with the material and the waves are non-dispersive as long as the material itself can be regarded as an elastic continuum. The surface waves are very sensitive to surface defects because the ultrasonic energy is concentrated in a small region at the surface of a solid. The penetration depth of a surface wave is of the order of a wavelength.

At a free surface of an elastic medium, a type of guided wave, which propagates over the free surface of a solid and the disturbance is confined primarily to the vicinity of the surface. These surface waves are called Rayleigh waves. The characteristic of surface waves is that the amplitude of the displacement decays
exponentially with distance from the free surface. Consider a plane strain ($u_z = 0$) two-dimensional case at a stress-free plane surface ($x_3 = 0$) of a semi-infinite elastic medium occupying a region $x_3 \geq 0$ (positive downwards), the displacements fields are assumed decaying in $x_3$ direction and propagates in $x_1$ direction shown as below

$$u_1 = A \exp(-bx_3) \exp(i(kx_1 - \omega t)) \quad \text{Eq. (12.1)}$$

$$u_3 = B \exp(-bx_3) \exp(i(kx_1 - \omega t)) \quad \text{Eq. (12.2)}$$

where $A$ and $B$ are complex constants, and $b$ is real positive so that the amplitude of the waves decreases exponentially with increasing $x_3$ and tends to zero as $x_3 \to \infty$. Rewriting the equation of motion, Eq. (11.1) using definition of $c_L$ and $c_T$ gives

$$c_T^2 \nabla^2 u + (c_L^2 - c_T^2) \nabla \nabla \cdot u = u'' \quad \text{Eq. (12.3)}$$

where $\nabla = \frac{\partial}{\partial x_i}$ and $\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y^2}$

Substituting Eq. (12.1) and Eq. (12.2) into the above equation, canceling the common terms, and rearranging terms yield

u_1 \to \left[ c_L^2 b^2 + (c^2 - c_T^2)k^2 \right] A + i\left( c_L^2 - c_T^2 \right) b k B = 0 \quad \text{Eq. (12.4)}

u_3 \to -i\left( c_L^2 - c_T^2 \right) b k A + \left[ c_L^2 b^2 + (c^2 - c_T^2)k^2 \right] B = 0 \quad \text{Eq. (12.5)}

The nontrivial solution of the above homogeneous equations gives the vanishing the determinant of the coefficients, which may be expressed in the form

$$c_L^2 c_T^2 b^4 + \left( c^2 c_L^2 - 2c_L^2 c^2 + c^2 c_T^2 \right) b^2 k^2 + \left( c^2 - c_L^2 \right) \left( c^2 - c_T^2 \right) k^4 = 0 \quad \text{Eq. (12.6)}$$

After arrangement

$$\left[ c_L^2 b^2 - (c_L^2 - c^2)k^2 \right] \left[ c_T^2 b^2 - (c_T^2 - c^2)k^2 \right] = 0 \quad \text{Eq. (12.7)}$$

This gives the following two roots for $b$
\[ b_1 = k \left(1 - \frac{c^2}{c_L^2}\right)^{1/2} \quad b_2 = k \left(1 - \frac{c^2}{c_T^2}\right)^{1/2} \quad \text{Eq. (12.8)} \]

in which implies that \( b \) is real requires \( c < c_T < c_L \).

The depth of penetration of Rayleigh waves into the medium \((x_3 \geq 0)\) depends on the decay coefficients \(b_1\) and \(b_2\) which are controlled by the frequency. Substituting Eq. (12.8) into Eq. (12.4) and Eq. (12.5) gives the ratio of \(B / A\)

\[
\frac{B_1}{A_1} = \frac{-b_1}{ik} \quad \text{and} \quad \frac{B_2}{A_2} = \frac{ik}{b_2} \quad \text{Eq. (12.9)}
\]

Therefore, a general solution that satisfies the equation of motion may be written as

\[
\begin{align*}
\text{real} & \quad \rightarrow \quad u_1 = A_1 e^{-b_{33} x_3} \exp\left(i k x_1 - \omega t\right) + A_2 e^{-b_{33} x_3} \exp\left(i k x_1 - \omega t\right) \quad \text{Eq. (12.10)} \\
\text{imaginary} & \quad \rightarrow \quad u_3 = B_1 e^{-b_{33} x_3} \exp\left(i k x_1 - \omega t\right) + B_2 e^{-b_{33} x_3} \exp\left(i k x_1 - \omega t\right) \quad \text{Eq. (12.11)}
\end{align*}
\]

where

- displacement vector \(u\) is complex
- \(A_1\) and \(A_2\) are arbitrary constants
- phase velocity \(c\) is chosen to satisfy the free boundary conditions shown as below

\[
\sigma_{33} = \tau_{31} = 0 \quad \text{at} \quad x_3 = 0 \quad \text{Eq. (12.12)}
\]

By substituting Eq. (12.10) and Eq. (12.11) into the expression of \(\sigma_{33}\) and \(\tau_{31} = \tau_{13} = 0\) at \(x_3 = 0\), one obtains after some manipulation

\[
\begin{align*}
u_1 & \quad \rightarrow \quad 2b_1 A_1 + \left(2 - \frac{c^2}{c_T^2}\right) k^2 \frac{A_2}{b_2} = 0 \quad \text{Eq. (12.13)} \\
u_3 & \quad \rightarrow \quad \left(2 - \frac{c^2}{c_T^2}\right) A_1 + 2A_2 = 0 \quad \text{Eq. (12.14)}
\end{align*}
\]
For a nontrivial solution, the determinant of the coefficients $A_1$ and $A_2$ must vanish and yield a well-known equation for the phase velocity of Rayleigh waves

\[
\left(2 - \frac{c^2}{c_T^2}\right)^2 - 4\left(1 - \frac{c^2}{c_L^2}\right)^{1/2}\left(1 - \frac{c^2}{c_T^2}\right)^{1/2} = 0
\]

Eq. (12.15)

The Rayleigh wave speed $c = c_R < c_T$ is a real root of the equation. Since the wave number does not enter the equation, the surface waves at a free surface of an elastic half-space are thus nondispersive. A simple equation that gives a good fit for this root in terms of Poisson’s ratio $\nu$ and the shear wave speed $c_T$ is

\[
c_R = \frac{0.862 + 1.14\nu}{1 + \nu} c_T
\]

Eq. (12.16)

when $\nu = 0.33$ and $c_R = 0.932 c_T$.

### 12.2 Love Waves – LO

Under certain conditions, such a layered media, other wave phenomena can be observed. In the case of a sound speed inversion, as communally occurs in the mantle of the surface of the earth, horizontally polarized surface waves can be trapped at a free elastic boundary. As the vertically polarized surface Rayleigh waves, the divergence of the wave is $1/\sqrt{\tau}$, so love waves can propagate long distances.

### 12.3 Gravity Waves – GA

Instead, gravity waves in the surface of the deep see water describes a clockwise ellipse in the sagittal plane of the propagation direction. There are several types of water wave under different assumptions and the dispersion relations are shown as

\[
\omega^2 = gk \tanh (kh) \quad \text{or} \quad \lambda = \frac{g}{2\pi} T^2 \tanh \left(2\pi \frac{h}{\lambda}\right)
\]

Eq. (12.17)

where $g$ is the acceleration of gravity, $k$ is wavenumber, and $h$ is water depth.
13 Finite Boundary Conditions

Here we add one more boundary condition on the semi-infinite medium and it becomes a two-sided guided wave which is dispersive. This wave was guided by the top and bottom boundaries and widely applied in the field of NDI/NDT and SHM especially for composite with laminate structures.

13.1 Lamb – PSV

Lamb waves exist in an elastic finite thickness plate. Transversal waves are trapped between the two interphases of the plate (upper and lower), and the relation thickness and wavelength induce different modes, called Lamb waves. Two main families of modes are observed, the extensional or symmetric modes $S_n$ and the flexural or anti-symmetrical $A_n$ modes. Their dispersion relations are show as

$$S(k, p, q) = (k^2 - q^2)^2 \tan(qh / 2) + 4k^2 pq \tan(ph / 2)$$  \hspace{1cm} \text{Eq. (13.1)}

$$A(k, p, q) = 4k^2 pq \tan(qh / 2) + (k^2 - q^2)^2 \tan(ph / 2)$$  \hspace{1cm} \text{Eq. (13.2)}

In the case of Lamb waves, the propagation speed strongly depends on the excited mode, the propagation in these layered media is dispersive. In high frequency range, both the phase velocity $c_p$ and group velocity $c_g$ of Lamb waves converge to the velocity of Rayleigh waves $c_R$.

Lamb waves are formed by the interference of P and SV waves so called PSV waves that can be further decompose into symmetric and anti-symmetric modes. Their particle motions are shown as below.

![Figure 13.1](image-url)  
(a)  
(b)

Figure 13.1  Lamb waves (PSV) with (a) symmetric and (b) anti-symmetric modes
IV. Planar and Cylindrical Waveguides in Elastic Solids

Bulk waves are converted into different types of guided waves according to the geometry of waveguides. No matter the wave is in planar or cylindrical waveguides, it’s the combination of P, SV and SH, but having different names when certain assumptions (displacement, strain or stress) are satisfied.

In plate – planar waveguides

- Sym PSV: \( \text{Sym}(K, P, Q) = (K^2 - Q^2)^2 \tan(Q/2) + 4K^2 PQ \tan(P/2) = 0 \)
- Anti PSV: \( \text{Anti}(K, P, Q) = 4K^2 PQ \tan(Q/2) + (K^2 - Q^2)^2 \tan(P/2) = 0 \)
  - Sym SH: \( \text{SymSH}(Q) = \cos(Q/2) = 0 \)
  - Anti SH: \( \text{AntiSH}(Q) = \sin(Q/2) = 0 \)

In rod – cylindrical waveguides

- Longitudinal: \( \text{Long}(K, P, Q) = 2P (Q^2 + K^2) J_1(P)J_1(Q) - (Q^2 - K^2)^2 J_0(P)J_1(Q) - 4K^2 PQ J_1(P)J_0(Q) = 0 \)
- Flexural: \( \text{Flex}(K, P, Q) = J_1(P)J_1(Q)(f_1L_0^2 + f_2L_PL_Q + f_3L_Q + f_4L_P + f_5) = 0 \)
- Torsional: \( \text{Tors}(Q) = QJ_0(Q) - 2J_1(Q) = 0 \)

In plate, the wave exists as Lamb PSV and SH waves with symmetrical or anti-symmetric modes, so as in cylindrical rods, existing as longitudinal, flexural, torsional, torsional modes. Their properties are discussed in the following chapters.
14 Planar Waveguides – PSV Waves

The first two elastic waves, symmetric PSV and anti-symmetric PSV are discussed in this chapter.

For a plate bounded by the surfaces \( z = \pm h / 2 \) and is of infinite extent in the \( x \) and \( y \) directions, the harmonic wave motion can be divided into two classes of wave motions: plane strain and anti-plane shear (or shear horizontal) motions.

### 14.1 Formation of Lamb Waves

A shear horizontal (SH) wave can propagate alone in the \( x \) direction and its polarization is unchanged on reflection or refraction. This is not the case for a longitudinal pressure (P) or shear vertical (SV) wave – these waves are coupled at a surface. Thus, reflection at a free surface is associated with partial conversion, so that an incident P (SV) wave gives a reflected P (SV) wave and a converted SV (P) wave shown as (Figure 14.2). The P and SV waves traveling along \( x \) are successively reflected at each surface of the plate.

For harmonic wave motion under plane strain conditions in the \( xz \) plane of an elastic plate, the guided wave field can be represented by a standing wave in the \( z \) direction and a propagating wave in the \( x \) direction and we have

\[
\begin{align*}
\text{Sym PSV} & : \quad \text{Sym}(K, P, Q) = (K^2 - Q^2)^2 \tan(Q/2) + 4K^2 PQ \tan(P/2) = 0 \\
\text{Anti PSV} & : \quad \text{Anti}(K, P, Q) = 4K^2 PQ \tan(Q/2) + (K^2 - Q^2)^2 \tan(P/2) = 0 \\
\text{Sym SH} & : \quad \text{SymSH}(Q) = \cos(Q/2) = 0 \\
\text{Anti SH} & : \quad \text{AntiSH}(Q) = \sin(Q/2) = 0
\end{align*}
\]
\[ u_y = 0 \quad \text{and} \quad \frac{\partial}{\partial y} (\text{ )} = 0 \quad \text{Eq. (14.2)} \]

We recall Eq. (11.3) with index representation and the displacement components by scalar and vector potentials

\[ u_i = \phi, + \psi_{k,j},e_{ij} \quad \text{Eq. (14.3)} \]

In the \((x, y, z)\) Cartesian coordinate system we expand above equation and have

\[
\begin{align*}
    u_x &= u = \frac{\partial \phi}{\partial x} + \frac{\partial \psi_z}{\partial y} - \frac{\partial \psi_y}{\partial z} \quad \text{Eq. (14.4)} \\
    u_y &= v = \frac{\partial \phi}{\partial y} - \frac{\partial \psi_z}{\partial x} + \frac{\partial \psi_x}{\partial z} \quad \text{Eq. (14.5)} \\
    u_z &= w = \frac{\partial \phi}{\partial z} + \frac{\partial \psi_y}{\partial x} - \frac{\partial \psi_x}{\partial y} \quad \text{Eq. (14.6)}
\end{align*}
\]

The scalar potential \(\phi\) and the components \(\psi_x, \psi_y\) and \(\psi_z\) of the vector potential \(\psi\) satisfy above displacement equations. Applying plane strain conditions in the \(y\) direction, displacement Eq. (14.3) to Eq. (14.5) can be written as

\[
\begin{align*}
    u &= \phi_x - \psi_{y,z} = \frac{\partial \phi}{\partial x} - \frac{\partial \psi_y}{\partial z} \quad \text{Eq. (14.7)} \\
    v &= -\psi_{z,x} + \psi_{z,x} = -\frac{\partial \psi_z}{\partial x} + \frac{\partial \psi_x}{\partial z} = 0 \quad \text{Eq. (14.8)} \\
    w &= \phi_z + \psi_{x,y} = \frac{\partial \phi}{\partial z} + \frac{\partial \psi_y}{\partial x} \quad \text{Eq. (14.9)}
\end{align*}
\]

where the subscript of the function \(\psi_y \rightarrow \psi\) has been omitted for simplicity at later description. Recall the definition of strain

\[ \varepsilon_{ij} = \frac{1}{2} (u_{ij} + u_{ji}) \quad \text{Eq. (14.10)} \]

then the strain under plane strain assumptions can be written as
\[
eq \varepsilon_{xx} = \frac{\partial u}{\partial x} = \frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \psi}{\partial x \partial z} \quad \text{Eq. (14.11)}
\]
\[
\varepsilon_{zz} = \frac{\partial w}{\partial z} = \frac{\partial^2 \phi}{\partial z^2} + \frac{\partial^2 \psi}{\partial x \partial z} \quad \text{Eq. (14.12)}
\]

Recall Hook’s law for isotropic elastic material under small strain or linear strain assumption

\[
\tau_{ij} = \lambda \varepsilon_{kk} \delta_{ij} + 2\mu \varepsilon_{ij} \quad \text{Eq. (14.13)}
\]
\[
2\mu \varepsilon_{ij} = \tau_{ij} - \frac{\lambda}{2\mu + 3\lambda} \tau_{kk} \quad \text{Eq. (14.14)}
\]

The in-plane stress components can be expressed by the potentials from Hooke’s law as

\[
\tau_{xx} = \sigma_x = \lambda \left( \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} \right) + 2\mu \frac{\partial u}{\partial x} = \lambda \left( \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} \right) + 2\mu \left( \frac{\partial^2 \phi}{\partial x^2} - \frac{\partial^2 \psi}{\partial x \partial z} \right) \quad \text{Eq. (14.15)}
\]
\[
\tau_{zz} = \sigma_z = \lambda \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) + 2\mu \frac{\partial w}{\partial z} = \lambda \left( \frac{\partial^2 \phi}{\partial z^2} + \frac{\partial^2 \phi}{\partial x^2} \right) + 2\mu \left( \frac{\partial^2 \phi}{\partial z^2} + \frac{\partial^2 \psi}{\partial x \partial z} \right) \quad \text{Eq. (14.16)}
\]
\[
\tau_{xz} = \tau_{zx} = \mu \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) = \mu \left( 2 \frac{\partial^2 \phi}{\partial x \partial z} + \frac{\partial^2 \psi}{\partial x^2} - \frac{\partial^2 \psi}{\partial z^2} \right) \quad \text{Eq. (14.17)}
\]

As discussed before, the potentials \( \phi \) and \( \psi \) satisfy wave equations

\[
\nabla^2 \phi = \frac{1}{c_L^2} \phi'' \rightarrow \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{1}{c_L^2} \frac{\partial^2 \phi}{\partial t^2} \quad \text{Eq. (14.18)}
\]
\[
\nabla^2 \psi = \frac{1}{c_T^2} \psi'' \rightarrow \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{c_T^2} \frac{\partial^2 \psi}{\partial t^2} \quad \text{Eq. (14.19)}
\]

where

- \( c_L = \sqrt{(\lambda + 2\mu) / \rho} = \sqrt{[E(1-\nu)] / [(1+\nu)(1-2\nu)\rho]} \)
- \( c_T = \sqrt{\mu / \rho} = \sqrt{E / [2(1+\nu)\rho]} \)
The Lamb waves have dual characteristics of being standing waves across the thickness – \( z \) direction and thus traveling waves only in the propagation direction – \( x \) direction. The wave solution can be considered and assumed in the following complex form

\[
\phi = \Phi(z) \exp\left(ikx - \omega t\right) \quad \text{Eq. (14.20)}
\]

\[
\psi = \Psi(z) \exp\left(ikx - \omega t\right) \quad \text{Eq. (14.21)}
\]

These functions represent waves traveling coherently in the \( x \) direction with the same angular frequency \( \omega \) and the same wavenumber \( k \). Substituting Eq. (14.19) and Eq. (14.20) into Eq. (14.17) and Eq. (14.18) leads to

\[
\Phi(z) = A_i \sin(pz) + A_z \cos(pz) \quad \text{Eq. (14.22)}
\]

\[
\Psi(z) = B_i \sin(qz) + B_z \cos(qz) \quad \text{Eq. (14.23)}
\]

Solving the wave equations gives

\[
k^2 + p^2 = \frac{\omega^2}{c_L^2} \quad \text{and} \quad k^2 + q^2 = \frac{\omega^2}{c_T^2} \quad \text{Eq. (14.24)}
\]

The parameters \( p \) and \( q \) in potentials are

\[
p^2 = \frac{\omega^2}{c_L^2} - k^2 \quad \text{and} \quad q^2 = \frac{\omega^2}{c_T^2} - k^2 \quad \text{Eq. (14.25)}
\]

where \( p \) and \( q \) stands for the transverse wavenumber of the bulk pressure and shear wave, respectively.

Observing the displacements from Eq. (14.6) and Eq. (14.8) by substituting potentials from Eq. (14.19) to Eq. (14.22) into, motions can be separated into symmetric and antisymmetric modes artificially according to even or odd function properties

a) Symmetric modes, in which the longitudinal component is an even function of \( z \) and the transverse component is an odd function of \( z \). The longitudinal displacement is symmetric, and the transverse displacement is antisymmetric with respect to the mid-plane of the plate are opposite shown as (Figure 14.3a)
b) Antisymmetric modes, in which the longitudinal component is an odd function of $z$ and the transverse component is an even function of $z$. The transverse displacement is symmetrical, and the longitudinal displacement is antisymmetric with respect to the mid-plane of the plate shown as (Figure 14.3b).

![Figure 14.3](image.png)  
(a) Lamb waves in (a) symmetric modes and (b) anti-symmetric.

The potential, displacement, strain and stress for symmetric and anti-symmetric modes are discussed in the following sections. Their mode shaper in each dispersion regions are demonstrated as below.

![Figure 14.4](image.png)  
Figure 14.4  Mode shapes in region III (1st row), II (2nd row) and I (3rd row)
### 14.1.1 Symmetric Modes

1) **Potentials**

\[
\Phi = A_2 \cos(pz) \\
\Psi = B_1 \sin(qz)
\]  
Eq. (14.26)

2) **Displacements**

\[
u = 0
\]
\[
w = -pA_2 \sin(pz) + ikB_1 \sin(qz)
\]
Eq. (14.27)

3) **Extensional strains**

\[
\varepsilon_{xx} = -k^2 A_2 \cos(pz) - ikqB_1 \cos(qz)
\]
\[
\varepsilon_{xy} = 0
\]
\[
\varepsilon_{zz} = -p^2 A_2 \cos(pz) + ikB_1 \cos(qz)
\]
Eq. (14.28)

4) **Shear strains**

\[
\varepsilon_{xy} = \frac{1}{2} \gamma_{xy} = 0
\]
\[
\varepsilon_{yz} = \frac{1}{2} \gamma_{yz} = 0
\]
\[
\varepsilon_{zx} = \frac{1}{2} \gamma_{zx} = \frac{1}{2} \left[ -ikpA_2 \sin(pz) + q^2 B_1 \sin(qz) \right]
\]
Eq. (14.29)

5) **Stresses**

\[
\tau_{xx} = -\lambda (k^2 + p^2) A_2 \cos(pz) - 2\mu \left[ k^2 A_2 \cos(pz) + ikqB_1 \cos(qz) \right]
\]
\[
= -(\lambda k^2 + \lambda p^2 + 2\mu) k^2 A_2 \cos(pz) - 2\mu ikqB_1 \cos(qz)
\]
\[
\tau_{zz} = -\lambda (k^2 + p^2) A_2 \cos(pz) - 2\mu \left[ \mu^2 A_2 \cos(pz) - ikqB_1 \cos(qz) \right]
\]
\[
= -(\lambda k^2 + \lambda p^2 + 2\mu) \mu^2 A_2 \cos(pz) + 2\mu ikqB_1 \cos(qz)
\]
Eq. (14.30)
\[
\tau_{zx} = \mu \left[ -2ikpA_2 \sin(pz) - (k^2 - q^2) B_1 \sin(qz) \right]
\]

where the term \( \exp i(kx - \omega t) \) has been omitted in the expression of displacements, strains and stresses.
14.1.2 Anti-symmetric Modes

1) Potentials

\[ \Phi = A_1 \sin(pz) \]
\[ \Psi = B_2 \cos(qz) \]  
Eq. (14.31)

2) Displacements

\[ u = ikA_1 \sin(pz) + qB_2 \sin(qz) \]
\[ v = 0 \]  
Eq. (14.32)
\[ w = pA_1 \cos(pz) + ikB_2 \cos(qz) \]

3) Extensional strains

\[ \varepsilon_{xx} = -k^2 A_1 \sin(pz) + ikB_2 \sin(qz) \]
\[ \varepsilon_{xy} = 0 \]  
Eq. (14.33)
\[ \varepsilon_{zz} = -p^2 A_1 \sin(pz) - ikB_2 \sin(qz) \]

4) Shear strains

\[ \varepsilon_{xy} = \frac{1}{2} \gamma_{xy} = 0 \]
\[ \varepsilon_{xz} = \frac{1}{2} \gamma_{xz} = 0 \]  
Eq. (14.34)
\[ \varepsilon_{yz} = \frac{1}{2} \gamma_{yz} = \frac{1}{2} \left[ -ikpA_1 \cos(pz) + q^2 B_2 \cos(qz) \right] \]

5) Stresses

\[ \tau_{xx} = -\lambda(k^2 + p^2)A_1 \sin(pz) - 2\mu \left[ k^2 A_1 \sin(pz) - ikqB_2 \sin(qz) \right] \]
\[ = -(\lambda k^2 + \lambda p^2 + 2\mu k^2)A_1 \sin(pz) + 2\mu ikqB_2 \sin(qz) \]
\[ \tau_{xz} = -\lambda(k^2 + p^2)A_1 \sin(pz) - 2\mu \left[ p^2 A_1 \sin(pz) + ikqB_2 \sin(qz) \right] \]  
Eq. (14.35)
\[ = -(\lambda k^2 + \lambda p^2 + 2\mu p^2)A_1 \sin(pz) - 2\mu ikqB_2 \sin(qz) \]
\[ \tau_{yz} = \mu \left[ 2ikpA_1 \cos(pz) - (k^2 - q^2)B_2 \cos(qz) \right] \]

where the term \( \exp \left( ikx - \omega t \right) \) has been omitted in the expression of displacements, strains and stresses.
14.2 Material Constants and Frequency Parameters Relation

Observing Eq. (14.29) and Eq. (14.34), the first term of extensional stresses consist four parameters \((\lambda, \mu, p, q)\) which can be separated into two groups, material constant group and frequency constant group.

In material constant group consists of six constants \((K, E, \lambda, \mu, \nu, M)\) and they are bulk modulus, Young’s modulus, Lame’s 1st parameter, Lame’s 2nd parameter or shear modulus, Poisson’s ratio and P-wave modulus respectively. Any one of the six constants can be expressed in terms of any two of the other constants inferring that there are only two independent variables.

Frequency constant group consists of four constants \((\omega, k, p, q)\) and they are circumferential/temporal frequency, wavenumber/spatial frequency, transverse wavenumber for pressure wave and transverse wavenumber for shear wave respectively. Also, only two out of four are independent variables.

The change of variables is mathematically equivalent but aims to help to understand the physical meanings when representing in different variables. From Eq. (14.29) and Eq. (14.34), the first term of \(\tau_{xx}\) and \(\tau_{zz}\) are \(\lambda k^2 + \lambda p^2 + 2\mu k^2\) and \(\lambda k^2 + \lambda p^2 + 2\mu p^2\) with even constant pairs \((\lambda, \mu)\) and \((p, k)\) which can be rearranged with uneven constant pairs \((\mu)\) and \((p, q, k)\). This can be done by the following procedures.

Bulk wave velocities are

\[
\begin{align*}
c_L^2 &= \frac{\lambda + 2\mu}{\rho} \quad \text{and} \\
c_T^2 &= \frac{\mu}{\rho}
\end{align*}
\]

Eq. (14.36)

Transverse wavenumbers are

\[
\begin{align*}
p^2 &= \frac{\omega^2}{c_L^2} - k^2 = \omega^2 \left(\frac{1}{c_L^2} - \frac{1}{c_p^2}\right) \\
q^2 &= \frac{\omega^2}{c_T^2} - k^2 = \omega^2 \left(\frac{1}{c_T^2} - \frac{1}{c_p^2}\right)
\end{align*}
\]

Eq. (14.37)

Parameters in \(\tau_{xx}\) can be rearranged as
\[
\lambda k^2 + \lambda p^2 + 2 \mu k^2 \\
= \lambda(k^2 + p^2) + 2 \mu k^2 = \mu \left(\frac{c_L^2}{c_T^2} - 2\right)(k^2 + p^2) + 2 \mu k^2 \\
= \mu \frac{\omega^2}{c_T^2} - 2 \mu p^2 = 2 \mu (q^2 - k^2)
\]

Eq. (14.38)

Parameters in \(\tau_{zz}\) can be rearranged as

\[
\lambda k^2 + \lambda p^2 + 2 \mu p^2 \\
= \lambda(k^2 + p^2) + 2 \mu p^2 = \mu \left(\frac{c_L^2}{c_T^2} - 2\right)(k^2 + p^2) + 2 \mu p^2 \\
= \mu \frac{\omega^2}{c_T^2} - 2 \mu k^2 = 2 \mu (q^2 - k^2)
\]

Eq. (14.39)

Summary above parameters relations

\[
\lambda k^2 + \lambda p^2 + 2 \mu k^2 = \mu(k^2 + q^2 - 2 p^2) \\
\lambda k^2 + \lambda p^2 + 2 \mu p^2 = \mu(q^2 - k^2)
\]

Eq. (14.40)

Rearranging Eq. (14.29) in terms of \((\lambda, \mu, k, p)\) into Eq. (14.34) in terms of \((\mu, k, p, q)\)

1) Symmetric modes

\[
\tau_{xx} = \mu \left[\left(2 p^2 - q^2 - k^2\right)A_2 \cos(pz) - 2ikqB_1 \cos(qz)\right] \\
\tau_{zz} = \mu \left[(k^2 - q^2)A_2 \cos(pz) + 2ikqB_1 \cos(qz)\right] \\
\tau_{zx} = \mu \left[-2ikpA_2 \sin(pz) - (k^2 - q^2)B_1 \sin(qz)\right] \\
\]

Eq. (14.41)

2) Anti-symmetric modes

\[
\tau_{xx} = \mu \left[\left(2 p^2 - q^2 - k^2\right)A_2 \sin(pz) + 2ikqB_2 \sin(qz)\right] \\
\tau_{zz} = \mu \left[(k^2 - q^2)A_2 \sin(pz) - 2ikqB_2 \sin(qz)\right] \\
\tau_{zx} = \mu \left[2ikpA_2 \cos(pz) - (k^2 - q^2)B_2 \cos(qz)\right] \\
\]

Eq. (14.42)

This representation has more insights of in-plate interference and the relations of propagation modes (longitudinal and transverse wavenumbers) in slowness curve that will be discussed in the later sections.
14.3 Dispersion Relations

The dispersion relation can be obtained by imposing the traction-free boundary conditions

$$\tau_{zz} = \tau_{xx} = 0 \quad \text{at} \quad z = \pm h / 2$$  \hspace{1cm} \text{Eq. (14.43)}

Substituting Eq. (14.42) into symmetric and antisymmetric modes in Eq. (14.40) and Eq. (14.41) gives two homogeneous equations for coefficients $A$ and $B$. A necessary and sufficient condition for the existence of a solution to these equations is that the determinant of the coefficients is zero. The implicit equation between $\omega$ and $k$ formed by setting the determinant to zero is usually called as dispersion relation.

The dispersion relation for symmetric modes is given by

$$\tan(\frac{qh}{2}) = -\frac{4k^2 pq}{(k^2 - q^2)^2}$$  \hspace{1cm} \text{Eq. (14.44)}

or

$$(k^2 - q^2)^2 \cos(\frac{ph}{2}) \sin(\frac{qh}{2}) + 4k^2 pq \sin(\frac{ph}{2}) \cos(\frac{qh}{2}) = 0$$  \hspace{1cm} \text{Eq. (14.45)}

For antisymmetric modes

$$\tan(\frac{qh}{2}) = -\frac{(k^2 - q^2)^2}{4k^2 pq}$$  \hspace{1cm} \text{Eq. (14.46)}

or

$$(k^2 - q^2)^2 \sin(\frac{ph}{2}) \cos(\frac{qh}{2}) + 4k^2 pq \cos(\frac{ph}{2}) \sin(\frac{qh}{2}) = 0$$  \hspace{1cm} \text{Eq. (14.47)}

To unify the representation, the dispersion relations Eq. (14.43) and Eq. (14.45) can be combined as

$$\tan(\frac{qh}{2} + \gamma) = -\frac{4k^2 pq}{(k^2 - q^2)^2}$$  \hspace{1cm} \text{Eq. (14.48)}

where $\gamma = 0$ or $\pi/2$

Another representation is by adding $4k^2 q^2$ into $(k^2 - q^2)^2$ to make $(k^2 + q^2)^2$ showing that
\[
(k^2 - q^2)^2 + 4k^2 q^2 = 4k^2 q^2 - 4k^2 pq \frac{\tan(p h/2 + \gamma)}{\tan(q h/2 + \gamma)}
\]
Eq. (14.49)

and rearranging as
\[
(k^2 + q^2)^2 = 4k^2 q^2 \left(1 - \frac{p}{q} \frac{\tan(p h/2 + \gamma)}{\tan(q h/2 + \gamma)} \right)
\]
Eq. (14.50)

where \( \gamma = 0 \) or \( \pi/2 \)

Then, Eq. (14.49) the dispersion relation for Lamb waves describing the relation temporal frequency \( \omega \) and spatial frequencies \( (k, p, q) \).

### 14.4 Representations of Dispersion Relations

As discussed in previous section, only two out of the four variables, \((\omega, k, p, q)\), are independent. The original form \((\omega, k, p, q)\) reveals the insights and interaction of wave formation; \((\tilde{\omega}, \tilde{k}, \tilde{p}, \tilde{q})\) is simply the non-dimensional form of \((\omega, k, p, q)\); the combined variables \((\Omega, K)\) is also non-dimensional and only focuses on in-plane propagation with longitudinal wavenumber \(K\); and \((\tilde{d}, \zeta)\) treats the normalized slowness \(\zeta = c_T / c_p\) (the inverse of phase velocity) as a function of half normalized frequency \(\tilde{d} = \Omega / 2\).

Dispersion relations could be represented into the following forms:

1) Original form represented by \((\omega, k, p, q)\)

2) Non-dimensional/normalized form represented by \((\tilde{\omega}, \tilde{k}, \tilde{p}, \tilde{q})\)

3) Non-dimensional/normalized form represented by \((\Omega, K)\)

4) Non-dimensional/normalized form represented by \((\tilde{d}, \zeta)\)

1) Original form represented by \((\omega, k, p, q)\)

Substituting \(k^2 + q^2 = \omega^2 / c_T^2\) into Eq. (14.44) and Eq. (14.46) to obtain an unified dispersion relation in terms of \((\omega, k, p, q)\) shown as
\[
\frac{\omega^4}{c_r^4} = 4k^2q^2 \left( 1 - \frac{p}{q} \frac{\tan(ph/2 + \gamma)}{\tan(qh/2 + \gamma)} \right)
\]
Eq. (14.51)

where \( \gamma = 0 \) and \( \pi / 2 \) represent symmetric and antisymmetric modes, respectively. Eq. (14.50) is commonly known as Rayleigh-Lamb dispersion relation and \( h \) is the overall thickness of the plate.

Dimensional variables are defined by

\[
\omega = 2\pi f, \quad k = 2\pi / \lambda = \omega / c_p
\]
Eq. (14.52)

\[
p^2 = \omega^2 / c_L^2 - k^2, \quad q^2 = \omega^2 / c_T^2 - k^2
\]
Eq. (14.53)

For symmetric modes

\[
S(k, p, q) = (k^2 - q^2)^2 \tan(qh / 2) + 4k^2 pq \tan(ph / 2)
\]
Eq. (14.54)

For antisymmetric modes

\[
A(k, p, q) = 4k^2 pq \tan(qh / 2) + (k^2 - q^2)^2 \tan(ph / 2)
\]
Eq. (14.55)

2) Non-dimensional form represented by \((\bar{\omega}, \bar{k}, \bar{p}, \bar{q})\)

The non-dimensional form of the dispersion relation in terms of \((\bar{\omega}, \bar{k}, \bar{p}, \bar{q})\) gives

\[
\bar{\omega}^4 = 4\bar{k}^2\bar{q}^2 \left( 1 - \frac{\bar{p}}{\bar{q}} \frac{\tan(\bar{p} / 2 + \gamma)}{\tan(\bar{q} / 2 + \gamma)} \right)
\]
Eq. (14.56)

where the non-dimensional variables are defined by

\[
\bar{\omega} = \omega h / c_r, \quad \bar{k} = kh
\]
Eq. (14.57)

\[
\bar{p}^2 = (ph)^2 = \xi^2 \bar{\omega}^2 - \bar{k}^2, \quad \bar{q}^2 = (qh)^2 = \bar{\omega}^2 - \bar{k}^2, \quad \xi = c_r / c_L
\]
Eq. (14.58)

For symmetric modes

\[
S(\bar{k}, \bar{p}, \bar{q}) = (\bar{k}^2 - \bar{q}^2)^2 \tan(\bar{q} / 2) + 4\bar{k}^2 \bar{p}\bar{q} \tan(\bar{p} / 2)
\]
Eq. (14.59)

120
For antisymmetric modes

\[ A(\bar{k}, \bar{p}, \bar{q}) = 4\bar{k}^2 \bar{p} \bar{q} \tan(\bar{q}/2) + (\bar{k}^2 - \bar{q}^2)^2 \tan(\bar{p}/2) \]  

Eq. (14.60)

3) Non-dimensional form represented by \((\Omega, K)\)

As the Lamb wave dispersion relation has been given in Eq. (14.50), the dispersion curves which give the relationship between frequency and wavenumber \(k\) can be obtained. There are four parameters \((\omega, k, p, q)\) as discussed before but only two of them are independent parameters which infer dispersion relation could be further simplified into two arbitrary selected parameters, for example \((\omega, k)\).

To plot the dispersion curves, it is convenient to introduce the non-dimensional variables. The non-dimensional form of the dispersion relation in terms of \((\Omega, K)\) gives

\[
\Omega^4 = 4K^2 \left( \Omega^2 - K^2 \right) \left[ 1 - \sqrt{\frac{\xi^2 \Omega^2 - K^2}{\Omega^2 - K^2}} \tan \left( \sqrt{\frac{\xi^2 - \left( \frac{K}{\Omega} \right)^2}{2}} \right) \right] \tan \left( \sqrt{1 - \left( \frac{K}{\Omega} \right)^2} \Omega/2 \right) \]  

Eq. (14.61)

where, non-dimensional variables are defined by

\[
\Omega = \frac{\omega h}{c_t}, \quad K = kh, \quad \xi = \frac{c_t}{c_l} = \sqrt{\frac{1-2v}{2(1-v)}} \]  

Eq. (14.62)

and transverse wavenumber in terms of \((\Omega, K)\) are

\[
p^2 = \frac{1}{h^2} \left( \xi^2 \Omega^2 - K^2 \right) \rightarrow \frac{p h}{2} = \sqrt{\frac{\xi^2 - \left( \frac{K}{\Omega} \right)^2}{2}} \Omega \]  

Eq. (14.63)

\[
q^2 = \frac{1}{h^2} \left( \Omega^2 - K^2 \right) \rightarrow \frac{q h}{2} = \sqrt{1 - \left( \frac{K}{\Omega} \right)^2} \Omega/2 \]  

Eq. (14.64)

For symmetric modes

\[
S(\Omega, K) = \left[ 1 - 2 \left( \frac{K}{\Omega} \right)^2 \right] \tan(\sqrt{1 - \left( \frac{K}{\Omega} \right)^2} \Omega/2) + 4 \left( \frac{K}{\Omega} \right)^2 \sqrt{1 - \left( \frac{K}{\Omega} \right)^2} \sqrt{\xi^2 - \left( \frac{K}{\Omega} \right)^2} \tan(\sqrt{\xi^2 - \left( \frac{K}{\Omega} \right)^2} \Omega/2) \]  

Eq. (14.65)
For antisymmetric modes

\[ A(\Omega, K) = 4(K / \Omega)^2 \sqrt{1-(K / \Omega)^2} \sqrt{\xi^2 - (K / \Omega)^2 \tan(\sqrt{1-(K / \Omega)^2} \Omega/2) + \left[ 1 - 2(K / \Omega)^2 \right] \tan(\sqrt{\xi^2 - (K / \Omega)^2} \Omega/2)} \]  

Eq. (14.66)

4) Non-dimensional form represented by \((\bar{d}, \zeta)\)

The non-dimensional form of the dispersion relation in terms of \((\bar{d}, \zeta)\) gives a unity relation shown as

\[ 1 = 4\zeta^2 \left(1 - \zeta^2\right) \left[1 - \frac{\xi^2 - \zeta^2}{1 - \zeta^2} \tan \left(\sqrt{\xi^2 - \zeta^2 \bar{d}} + \gamma\right)\right] \tan \left(\sqrt{1 - \zeta^2 \bar{d}} + \gamma\right) \]  

Eq. (14.67)

where non-dimensional variables are defined by

\[ \bar{d} = \frac{\Omega}{2}, \quad \zeta = \frac{K}{\Omega}, \quad \frac{\xi}{c_T} = \frac{kh}{2d}, \quad \frac{\xi}{c_L} = \frac{(1-2\nu)}{2(1-\nu)} \]  

Eq. (14.68)

Transverse wavenumber in terms of \((\bar{d}, \zeta)\) are

\[ p^2 = \left(\frac{2\bar{d}}{h}\right)^2 \left(\xi^2 - \zeta^2\right) \rightarrow \frac{ph}{2} = \sqrt{\xi^2 - \zeta^2 \bar{d}} \]  

Eq. (14.69)

\[ q^2 = \left(\frac{2\bar{d}}{h}\right)^2 (1-\zeta^2) \rightarrow \frac{qh}{2} = \sqrt{1 - \zeta^2 \bar{d}} \]  

Eq. (14.70)

For symmetric modes

\[ S(\bar{d}, \zeta) = (2\xi^2 - 1)^2 \tan \left(\sqrt{1 - \xi^2 \bar{d}} + 4\xi^2 \sqrt{1 - \xi^2} \sqrt{\xi^2 - \zeta^2} \tan \left(\sqrt{\xi^2 - \zeta^2 \bar{d}}\right)\right) \]  

Eq. (14.71)

For antisymmetric modes

\[ A(\bar{d}, \zeta) = 4\xi^2 \sqrt{1 - \xi^2} \sqrt{\xi^2 - \zeta^2} \tan \left(\sqrt{1 - \xi^2 \bar{d}}\right) + (2\xi^2 - 1)^2 \tan \left(\sqrt{\xi^2 - \zeta^2 \bar{d}}\right) \]  

Eq. (14.72)
14.5 Displacement Fields

Applying one of the traction-free boundary conditions $\tau_{zz} = 0$ or $\tau_{xz} = 0$ at $z = \pm h / 2$ into Eq. (14.40) and Eq. (14.41) then two pairs of coefficient relations could be obtained

For $\tau_{zz} = 0$

$$A_2 = -\frac{2ikq}{k^2 - q^2} \cos\left(\frac{qh}{2}\right) B_1 \quad \text{and} \quad A_i = \frac{2ikq}{k^2 - q^2} \sin\left(\frac{qh}{2}\right) B_2$$  \hspace{1cm} \text{Eq. (14.73)}

For $\tau_{xz} = 0$

$$A_2 = -\frac{k^2 - q^2}{2ikp} \sin\left(\frac{qh}{2}\right) B_1 \quad \text{and} \quad A_i = \frac{k^2 - q^2}{2ikp} \cos\left(\frac{qh}{2}\right) B_2$$  \hspace{1cm} \text{Eq. (14.74)}

Selected Eq. (14.72) substitute into displacement Eq. (14.25) and Eq. (14.30)

For symmetric modes

$$u = qB_1 \left( \frac{2k^2}{k^2 - q^2} \cos\left(\frac{qh}{2}\right) \cos\left(pz\right) - \cos\left(qz\right) \right)$$  \hspace{1cm} \text{Eq. (14.75)}

$$w = ikB_1 \left( \frac{2pq}{k^2 - q^2} \cos\left(\frac{qh}{2}\right) \sin\left(pz\right) + \sin\left(qz\right) \right)$$  \hspace{1cm} \text{Eq. (14.76)}

For antisymmetric modes

$$u = qB_2 \left( -\frac{2k^2}{k^2 - q^2} \sin\left(\frac{qh}{2}\right) \sin\left(pz\right) + \sin\left(qz\right) \right)$$  \hspace{1cm} \text{Eq. (14.77)}

$$w = ikB_2 \left( \frac{2pq}{k^2 - q^2} \sin\left(\frac{qh}{2}\right) \cos\left(pz\right) + \cos\left(qz\right) \right)$$  \hspace{1cm} \text{Eq. (14.78)}

The general solutions with unified representations with constants $A, B, C, D, E, F, G$

$$u = AU(z) \exp\left(ikx - \omega t\right)$$  \hspace{1cm} \text{Eq. (14.79)}

$$w = BW(z) \exp\left(ikx - \omega t\right)$$  \hspace{1cm} \text{Eq. (14.80)}

where

123
\[ U(z) = q \left( \frac{2k^2}{k^2 - q^2} \cos \left( \frac{qh}{2} + \gamma \right) \cos \left( pz + \gamma \right) - \cos \left( qz + \gamma \right) \right) \]  
Eq. (14.81)

\[ W(z) = i k \left( \frac{2pq}{k^2 - q^2} \cos \left( \frac{qh}{2} + \gamma \right) \sin \left( pz + \gamma \right) + \sin \left( qz + \gamma \right) \right) \]  
Eq. (14.82)

The strain components can be obtained as

\[ \varepsilon_{xz} = CE_{xz}(z) \exp(i kx - \omega t) \]  
Eq. (14.83)

\[ \varepsilon_{zx} = \frac{1}{2} \gamma_{xz} = DE_{xz}(z) \exp(i kx - \omega t) \]  
Eq. (14.84)

where

\[ E_{xz}(z) = ikq \left( \frac{2p^2}{k^2 - q^2} \cos \left( \frac{qh}{2} + \gamma \right) \cos \left( pz + \gamma \right) + \cos \left( qz + \gamma \right) \right) \]  
Eq. (14.85)

\[ E_{zx}(z) = q^2 \left( -\frac{2k^2}{k^2 - q^2} \frac{p}{q} \cos \left( \frac{qh}{2} + \gamma \right) \sin \left( pz + \gamma \right) + \sin \left( qz + \gamma \right) \right) \]  
Eq. (14.86)

The stress components can be obtained as

\[ \tau_{xz} = Et_{xz}(z) \exp(i kx - \omega t) \]  
Eq. (14.87)

\[ \tau_{zx} = Ft_{xz}(z) \exp(i kx - \omega t) \]  
Eq. (14.88)

\[ \tau_{zz} = Gt_{xz}(z) \exp(i kx - \omega t) \]  
Eq. (14.89)

where

\[ t_{xz}(z) = -2 \mu kq \left( \frac{2p^2 - q^2 - k^2}{k^2 - q^2} \cos \left( \frac{qh}{2} + \gamma \right) \cos \left( pz + \gamma \right) + \cos \left( qz + \gamma \right) \right) \]  
Eq. (14.90)

\[ t_{zx}(z) = 2 \mu kq \left( -\frac{\cos \left( \frac{qh}{2} + \gamma \right)}{\cos \left( \frac{ph}{2} + \gamma \right)} \cos \left( pz + \gamma \right) + \cos \left( qz + \gamma \right) \right) \]  
Eq. (14.91)

\[ t_{xz}(z) = -\mu \left( \frac{4k^2 pq}{k^2 - q^2} \cos \left( \frac{qh}{2} + \gamma \right) \sin \left( pz + \gamma \right) + \left( k^2 - q^2 \right) \sin \left( qz + \gamma \right) \right) \]  
Eq. (14.92)
14.6 Understanding of Spatial Frequencies – Wavenumbers

Two types of wavenumbers: longitudinal $k$ and transverse $(p, q)$ wavenumber will be discussed in this section. Wavenumbers provide the physical insights and are closely related to the type of mode shapes and sectional displacement fields. The wave formation can also be explained. For example, the mode shapes are characterized by the composition of wavenumbers $(k, p, q)$ for a 5mm Aluminum plate at 200 kHz under excitation shown as below.

![Image](image_url)

**Figure 14.5** The fundamental modes, $A_o$ and $S_o$, and their particle motions

The formation of Lamb waves consists longitudinal wave $k_p$ and transverse shear wave $k_{SY}$ with longitudinal/propagation wavevector $k$ and transverse/oscillation wavenumber $p$ and $q$. For a given excitation frequency $\omega$, Lamb wave will have several symmetric and anti-symmetric modes according to its excitation frequency, and each mode has its own wavevector, $k_{sym}$ and $k_{anti}$, with dispersion relation of wavevector norm $||k||$ and its excitation frequency $\omega$ shown as below.

\[
(\omega, ||k||) \xrightarrow{form\ from\ Bults} k_p = (k, p) \quad k_{SY'} = (k, q) \quad k = (k_x, k_y) \quad k_{anti} \quad \text{Eq. (14.93)}
\]

where $||k_p|| = \sqrt{k_x^2 + p^2}$ and $||k_{SY'}|| = \sqrt{k_x^2 + k_y^2 + q^2}$

When the wavefront is aligned with coordinate axis on a plane strain cross section, the wavevector $k$ will degenerate into a wavenumber $k$ so that the dispersion relation can be discussed in a more general way in complex domain.
\[ (\omega, k) \xrightarrow{\text{form from bulks}} k_p = (k, p) \xrightarrow{\text{interfere into Lamb}} k \xrightarrow{\text{separate into}} k_{\text{sym}} \quad \text{Eq. (14.94)} \]

where \( ||k_p|| = \sqrt{k^2 + p^2} = \frac{\omega}{c_L} \) and \( ||k_{\text{sym}}|| = \sqrt{k^2 + q^2} = \frac{\omega}{c_T} \)

### 14.6.1 Transverse Wavenumbers and Slowness Curves

From the original definition of \( p \) and \( q \) given by Eq. (14.24) to solve the wave equations Eq. (14.17) and Eq. (14.18), it may be seen that the \((\omega, k)\) plane can be divided into three regions shown as (Figure 14.6) according to the arguments \( p \) and \( q \) being real and/or purely imaginary, or depending on whether the phase velocity \( c_p \) exceeds the longitudinal wave velocity \( c_L \) or the transverse wave velocity \( c_T \).

\[ \frac{p^2}{\omega^2} = \left( \frac{1}{c_L^2} - \frac{k^2}{\omega^2} \right) = \left( \frac{1}{c_L^2} - \frac{1}{c_p^2} \right) \quad \text{and} \quad \frac{q^2}{\omega^2} = \left( \frac{1}{c_T^2} - \frac{k^2}{\omega^2} \right) = \left( \frac{1}{c_T^2} - \frac{1}{c_p^2} \right) \quad \text{Eq. (14.95)} \]

The discussion is with physical meaning when \( k \) is real to describe a propagation wave toward positive direction which implies \( c_p \) is real and positive. When \( k \) is pure imaginary or complex, the imaginary term makes the wave to be evanescent and decay in space so that \( c_p \) would no longer exist.

![Figure 14.6](image)

Three regions for the propagating Lamb waves in the real (a) dimensional \((\omega, k)\) and (b) non-dimensional \((\Omega, K)\) planes: (1) \( c_p > c_L \), \( p \) and \( q \) are real; (2) \( c_L > c_p > c_T \), \( p \) is pure imaginary and \( q \) is real; and (3) \( c_T > c_p \), \( p \) and \( q \) are pure imaginary.

Hence the three regions shown in (Figure 14.6) are categorized by
a) \( c_p > c_L > c_T \) or \( \frac{k}{\omega} < \frac{1}{c_L} < \frac{1}{c_T} \), then \( p \) and \( q \) are real

\[
\frac{1}{c_p^2} + \frac{p^2}{\omega^2} = \frac{1}{c_L^2} \quad \text{and} \quad \frac{1}{c_p^2} + \frac{q^2}{\omega^2} = \frac{1}{c_T^2} \quad \text{Eq. (14.96)}
\]

b) \( c_L > c_p > c_T \) or \( \frac{1}{c_L} < \frac{k}{\omega} < \frac{1}{c_T} \), then \( q \) is real and \( p = i \, p_l \) is pure imaginary

\[
\frac{1}{c_p^2} - \frac{p_l^2}{\omega^2} = \frac{1}{c_L^2} \quad \text{and} \quad \frac{1}{c_p^2} + \frac{q^2}{\omega^2} = \frac{1}{c_T^2} \quad \text{Eq. (14.97)}
\]

c) \( c_L > c_T > c_P \) or \( \frac{1}{c_L} < \frac{k}{\omega} < \frac{1}{c_T} \), then \( p = i \, p_l \) and \( q = i \, q_l \) are pure imaginary.

\[
\frac{1}{c_p^2} - \frac{p_l^2}{\omega^2} = \frac{1}{c_L^2} \quad \text{and} \quad \frac{1}{c_p^2} - \frac{q_l^2}{\omega^2} = \frac{1}{c_T^2} \quad \text{Eq. (14.98)}
\]

Instead of only dividing the \((\omega, k)\) plane into three regions according to value of \( c_p \), dispersion relation provides a larger point of view under the traction free conditions on the top and bottom surfaces of the plate.

Figure 14.7   Lamb wave dispersion relations with plane divisions
Mapping plane divisions from \((\omega, k)\) to \((\Omega, K)\) with dispersion relations could lead to (Figure 14.7) which provides the information of real/real, imaginary/real or imaginary/imaginary combination for transverse wavenumber \(p\) and \(q\). With each selected \(c_p\), transverse wavenumber \(p\) and \(q\) will be decided to fall into one of the three regions \(I\), \(II\) and \(III\).

\[
\begin{align*}
\text{(a)} & \quad k_z/\omega &  I & II & III \\
\text{(b)} & \quad k_z+k_z=0 &  I & II & III \\
\text{(c)} & \quad k_z-k_z=0 &  I & II & III
\end{align*}
\]

where \(c_p > c_z > c_T\)

\[
\begin{align*}
\mathbf{k}_p/\omega & = \left( \frac{1}{c_p}, \frac{p}{\omega} \right) \\
\mathbf{k}_z/\omega & = \left( \frac{1}{c_z}, \frac{q}{\omega} \right)
\end{align*}
\]

where \(c_z > c_p > c_T\)

\[
\begin{align*}
\mathbf{k}_p/\omega & = \left( \frac{1}{c_p}, \frac{p}{\omega} \right) \\
\mathbf{k}_z/\omega & = \left( \frac{1}{c_z}, \frac{q}{\omega} \right)
\end{align*}
\]

\[
\begin{align*}
\text{(a)} & \quad k_z/\omega &  I & II & III \\
\text{(b)} & \quad k_z+k_z=0 &  I & II & III \\
\text{(c)} & \quad k_z-k_z=0 &  I & II & III
\end{align*}
\]

where \(c_z > c_T \)

\[
\begin{align*}
\mathbf{k}_p/\omega & = \left( \frac{1}{c_p}, \frac{p}{\omega} \right) \\
\mathbf{k}_z/\omega & = \left( \frac{1}{c_z}, \frac{q}{\omega} \right)
\end{align*}
\]

**Figure 14.8** Slowness curve Lamb waves in an isotropic plate of thickness \(h\)

From the definition of \(p\) and \(q\) given by Eq. (14.94), the slowness curve for Lamb waves in an isotropic plate with thickness \(h\) can be obtained and are shown in (Figure 14.8) where (a) Both the wave vectors \(\mathbf{k}_p\),
and $k_{SV}$ locate on the circular slowness curve, i.e., $p$ and $q$ are both real, (b) wave vector $k_p$ locates on hyperbolic slowness curve and $k_{SV}$ locates on circular slowness curve, i.e., $p$ is imaginary and $q$ is real, (c) both wave vectors $k_p$ and $k_{SV}$ locate on hyperbolic slowness curves, i.e., $p$ and $q$ are both imaginary.

In the slowness curves, the wavenumber $k_x = k$ is always real for propagating modes, and thus the transverse wavenumber $k_z$, which is either $p$ (P-wave) or $q$ (SV-wave) then the wave vector for P and SV waves can locate on the slowness curve as $k = (k_x, k_z)$ so that $k_p = (k, p)$ and $k_{SV} = (k, q)$.

### 14.6.2 Longitudinal Wavenumber

From above relations, the composition of wavenumber $k$ (real, pure imaginary or complex) will determine whether a wave is propagating or evanescent. From the complex dispersion curve, the wavenumber $k$ can be real, pure imaginary or complex values.

1) **Case 1: $k$ is real**

From Eq. (14.78) and Eq. (14.79), the generalized displacements can be rewritten as

$$u = A U \exp \left( k x - \omega t \right)$$

where $u = (u, w)$, $U = (U(z), W(z))$ and $A$ is amplitude coefficient

When $k$ is pure real, the displacements are sinusoidal function for both space and time and thus Lamb waves with pure real wavenumbers are *propagating Lamb waves*. From (Figure 14.9), at any given frequency, there is a finite number of propagating Lamb modes.

2) **Case 2: $k$ is pure imaginary where $k = i k_I$ for real $k_I$ and $k_I > 0$**

$$u = A U \exp \left( -k_I x - \omega t \right)$$

It is seen from Eq. (14.99), the displacements exhibit exponentially decay from positive $x$ direction and the values of the wavenumbers determine the decay rate. The Lamb wave modes with pure imaginary wavenumbers are *non-propagating (evanescent) Lamb waves*. From (Figure 14.9), at any given frequency, there is a finite number evanescent Lamb wave modes with pure imaginary wavenumbers.
3) Case 3: $k$ is complex where $k = k_R + ik_I$ for real $k_R$, $k_I$ and $k_I > 0$

$$u = AU \exp[-k_Ix + i(k_Rx - \omega t)]$$  \hspace{1cm} \text{Eq. (14.101)}

Eq. (14.100) can be characterized as waves propagating with a sinusoidal variation described by the real part of the wavenumbers, modulated by an exponential decaying function controlled by the imaginary part of the wavenumbers. Therefore, the Lamb wave modes with complex wavenumbers are non-propagating (evanescent) Lamb waves as well.

14.6.3 Longitudinal Wavenumber – Real

Clearly the non-dimensional dispersion curves in linear elastic isotropic medium are only functions of Poisson’s ratio. Since Eq. (14.60) is a transcendental equation and no analytical solution is available, a Muller numerical algorithm (Stoer, 1993) is used to solve the dispersion relation.

Note, in above equations, trivial solutions exist and have to be eliminated which are not modal solutions in dispersion curve. Taking dispersion relation in form 3) for example, both $\Omega / K = 1$ and $\Omega / K = \alpha$ will make $S(\Omega, K)$ and $A(\Omega, K)$ equal to zero. Somehow, trivial solutions are asymptotes for high order modes and are good references to examine the characteristic of mode shapes.

where

- $\Omega / K = c_R / c_T < 1$ is the asymptote for $A_0$ and $S_0$
- $\Omega / K = c_T / c_T = 1$ is the asymptote for all the other modes
- $\Omega / K = \alpha = c_L / c_T > 1$ is the transition asymptote for all the other mode

Non-dimensional dispersion curves for an isotropic plate obtained from Eq. (14.63) and Eq. (14.64) is shown in (Figure 14.9). Phase and group velocity are shown in (Figure 14.10) and (Figure 14.11).

The dispersion relation results in an infinite number of wave modes defined by $S_0, S_1, S_2, \cdots, A_0, A_1, A_2, \cdots$ for symmetric and antisymmetric modes, respectively. The order of these modes is defined later and is affected by coefficient $\alpha$.

When $\nu = 0.3$ the order are $S_0, S_1, S_2, S_4, S_3, S_6, S_8 \cdots$ and $A_0, A_1, A_3, A_2, A_5, A_7, A_4, A_9 \cdots$
As shown in (Figure 14.9), the dispersion curves for modes of the same family do not cross each other, though a curve from a symmetric (anti-symmetric) mode may cross a curve from an anti-symmetric (symmetric) mode.

![Non-dimensional Dispersion Relations (K, Ω)](image)

**Figure 14.9**   Lamb wave propagating mode dispersion curves

The order of the curves in the same family depends on the ratio \( \alpha = \frac{c_L}{c_T} \). For example, the curve for mode \( S_2 \) is above or below that for \( S_1 \) according to whether \( c_T \) is less or more than \( c_L/2 \). For the \( S_1 \) mode, and for small values of \( K = kh \), \( \Omega \) decreases with \( K \); the group velocity \( C_g = d\Omega / dK \) (so as \( c_g = d\omega / dk \) ) is therefore negative, that is, in the opposite direction to the phase velocity. Thus the wave energy is transported in the direction opposite to that of the wave propagation.

(Figure 14.10) and (Figure 14.11) show the phase velocity \( C_p = \Omega / K \) and the group velocity \( C_g = d\Omega / dK \). The phase velocity of \( A_0 \) mode monotonically increases with the frequency. Note, as the
frequency increasing, phase velocity for $A_0$ and $S_0$ will saturate to $c_p / c_T$ and all the other modes will saturate to $c_T / c_T = 1$.

Figure 14.10  Lamb wave propagating mode dispersion curves – phase velocity

Similarly, from the definition of group velocity $c_g = d\omega / dk$, it is easy to obtain that nondimensional group velocity $C_g = d\Omega / dK$ which is also equal to $C_g = c_g / c_T$. Knowing the relation of $\Omega \sim K$, one can easily calculate the group velocity dispersion by using differentiating operation $C_g = d\Omega / dK$. (Figure 14.11) displays the group velocity dispersion $C_g \sim \Omega$ calculated from $\Omega \sim K$. Again, the root-finding for solving the transcendental equations need not be performed in this step.

Theoretically, the group velocity dispersion can also be computed directly from transcendental equations without need of taking the derivative of dispersion relations $\Omega \sim K$. This will be discussed in the later sections.
14.6.4 Longitudinal Wavenumber – Real – Wavelength Point of View

The purpose to plot the dispersion curves in wavelength due to the intuitive insights to damage inspection capabilities. Substituting the wavenumber to wavelength with the relation $k = \lambda / 2\pi$. Since we have the relation of $(\omega, k)$ from dispersion relation, every plot with respect to $\omega$ can be plot with respect to $k$, and then to $\lambda$. This provides another point of view to relate every quantity to wavelength. The alternative plots are shown as

- dispersion relation $(\omega, k)$ was plot as $(\lambda, \omega)$ in (Figure 14.12) and (Figure 14.15)
- phase velocity $(\omega, c_p)$ was plot as $(\lambda, c_p)$ in (Figure 14.13) and (Figure 14.16)
- group velocity $(\omega, c_g)$ was plot as $(\lambda, c_g)$ in (Figure 14.14) and (Figure 14.17).

Figure 14.11 Lamb wave propagating mode dispersion curves – group velocity
Figure 14.12  Wavenumber and wavelength dispersion curves – SYM

Figure 14.13  Wavenumber and wavelength dispersion curves – SYM – phase velocity

Figure 14.14  Wavenumber and wavelength dispersion curves – SYM – group velocity
Figure 14.15  Wavenumber and wavelength dispersion curves – ASYM

Figure 14.16  Wavenumber and wavelength dispersion curves – ASYM – phase velocity

Figure 14.17  Wavenumber and wavelength dispersion curves – ASYM – group velocity
14.6.5  **Longitudinal Wavenumber – Pure Imaginary**

When wavenumber is pure imaginary, the dispersion curve tends to be the characteristics for evanescent waves which will exponentially decay in space. At left-half plane (pure imaginary), asymptotes are circular and elliptic functions; at right-half plane (real), asymptotes are hyperbolic functions.

The cut-off frequencies at $K = 0$ are transition points of propagating and evanescent waves. Most of the time, they are also zero group velocity (ZGV) points but there’re exceptions for certain modes and can be found either analytically or numerically.

![Asymptotes on pure imaginary (evanescent) & real (propagating) planes](image)

- $P^2 = (ph)^2 = \xi^2 \Omega^2 - K^2$
- $Q^2 = (qh)^2 = \Omega^2 - K^2$

*Imaginary plane*

*Circular & elliptical function*

*Real plane*

*Hyperbolic function*

\[ W_c = n\pi = Q \]

\[ W_c = \alpha n\pi = \alpha P \]

\[ Q \text{ is SH wave} \]

Figure 14.18  Asymptotes on pure imaginary (evanescent) & real (propagating) planes
Asymptomes are extracted by convergence observation of dispersion relations. Non-dimensional transverse wavenumber \( P^2 = (ph)^2 \) and \( Q^2 = (qh)^2 \) are defined as below in Figure 14.18. There’re two branches of cut-off frequencies where one is determined by \( \alpha P \) and the other is determined by \( Q \). Dispersion curves is shown in Figure 14.19 and in Figure 14.20 with asymptotes.

Figure 14.19  Dispersion curves on pure imaginary (evanescent) & real (propagating) planes
Asymptotes are the boundaries for propagating and evanescent waves. For propagating waves (excepting $A_o$), they converge to $C_p = C_g = \alpha$ at medium frequency and then converge to $C_p = C_g = 1$ at high frequency range. For evanescent waves, each curved-edge parallelogram defines the cross-section points (two non-adjacent points in each region) and evolution of the symmetric (blue) and anti-symmetric (red) evanescent waves.

Figure 14.20  Dispersion curves on pure imaginary (evanescent) & real (propagating) planes with asymptotes
14.6.6  Longitudinal Wavenumber – Complex

Complex roots exist in 3D space and can be mapped into real and imaginary planes for observation shown as (Figure 14.21). The real part of complex roots is less than $2\pi$ where imaginary part fulfills the entire domain. This infers a weakly oscillating (long wavelength) and strongly decaying (large decay coefficient) modes in evanescent waves.

(a) (b)

Figure 14.21  Complex roots projected on (a) real and (b) imaginary planes

Combining real and pure imaginary roots with complex roots constructs a full picture of dispersion curves in 3D space. Complex roots play a role as a mode conversion bridge to connect dispersion curves together.

Dispersion curves can be categorized into three types shown as (Figure 14.22) and (Figure 14.23).

1) Real roots on real plane: most of the curves are monotonic increasing except some with ZGV points.
2) Pure imaginary roots on imaginary plane: half-loop (connecting two real curves) and serpentine (extending real curve).
3) Complex roots in 3D space: like a channel connecting to a ZGV point (could be on real plane, ex. $S_1$, $A_3$ or on imaginary plane, ex. $S'_{4,3}$) or bridging two ZGV points (ex. $A_4$ to $A_5$).
Figure 14.22  Real, pure imaginary and complex roots in different 3D perspectives
Figure 14.23  Real, pure imaginary and complex roots in 3D perspective with symmetry plane
14.7  Mode Numbering & Conversion

Mode naming conventions and wave formation/conversion by tracing its frequency.

1)  Real roots

The name of mode is defined by the number of its cut-off frequency and their order is not following the number order because of coefficient $\alpha$ is in their odd order modes.

Table 14.1  Mode numbering and cut-off frequencies

<table>
<thead>
<tr>
<th>Mode</th>
<th>$S_{2n}$</th>
<th>$S_{2n+1}$</th>
<th>$A_{2n}$</th>
<th>$A_{2n+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega c h / \pi c_T$</td>
<td>$2n$</td>
<td>$(2n+1)\alpha$</td>
<td>$2n\alpha$</td>
<td>$2n+1$</td>
</tr>
</tbody>
</table>

For example, referring to (Figure 14.23) when $\nu = 0.3$ the mode orders are

- $S_0, S_1, S_2, S_3, S_6, S_8, S_5, S_{10}, S_{12}, S_7, S_{14}$... for symmetric modes
- $A_0, A_1, A_2, A_5, A_7, A_4, A_9, A_{11}, A_6, A_{13}, A_8, A_{15}$... for anti-symmetric modes

2)  Pure imaginary roots

Due to there are two types of curves, half-loop and serpentine. Half-loop connects two real modes and is named as the composition of their name, ex. $\text{imag } S_{4,3}$ connects $S_4$ and $S_3$; $\text{imag } S_{8,5}$ connects $S_8$ and $S_5$. Serpentine extends the real mode and inherits the name from real modes, ex. $\text{imag } S_6$, $\text{imag } S_{10}$, $\text{imag } A_5$, $\text{imag } A_9$ are extensions of their real modes.

3)  Complex roots

Before naming complex mode let’s get back to the formation of waves. We already know there are two types of waves: propagating and evanescent waves, and evanescent wave can be further categorized into exponential decay (pure imaginary roots) and oscillation with exponential decay (complex roots) waves. Then which types of waves should be the one when a wave is originally formed.

14.7.1  Mode Conversion

When tracing the frequency from zero, every mode of waves initiates from a complex mode, converting to pure imaginary mode and then to real mode. The up-conversion (Figure 14.24) and down-conversion (Figure 14.25) are demonstrated with $\nu = 0.31$ shown as below.
The naming rules follow the final stage of real modes by up-converting a complex mode shown as (Figure 14.24) For example, a complex mode which at final stage will convert to real $S_1$ and real $S_2$ then is named as complex $S_{1,2}$; a branch a complex mode which will all the way convert to imag $S_{8,5}$ and imag $S_{12,7}$ is names as complex $S_{8,5,12,7}$. Since the pure imaginary mode either connects to one real mode or bridges two real modes, it’s fairly easy to name the mode, for example, real $S_6$ and pure imaginary imag $S_6$ has the same name; imag $S_{8,5}$ connects real $S_8$ and real $S_5$, and the lower mode ($S_8 < S_5$) comes first.
The principle can be applied to the mode down-conversion with an identical result as mode up-conversion. Energy is branching out into several modes during up-conversion (Figure 14.24) and down-conversion (Figure 14.25) works like collecting the real modes from high frequency all the way down to zero.

### 14.7.2 The Concept of First Derivative

Zero group velocity (ZGV) point is defined as \( \frac{d\Omega}{dK} = 0 \) which is also the mode transition point for the dispersion curves in entire complex domain. Since each mode starts from complex, converting into pure imaginary, then converting into real, so there are several transition points for mode conversion. The transition points according to their properties can be categorized into

1) **At pure imaginary bridging mode – single ZGV point**

   There is a transition point at every complex to pure imaginary, for example the complex \( S_{1,2} \) has a transition point (Figure 14.25 red to blue) to imag \( S_{1,2} \); complex \( S_{4,3} \) has another one when connecting to imag \( S_{4,3} \); on top of the imag \( S_{8,5} \) there is another one where the bottom of imag \( S_{8,5} \) belongs to the next category – the diverging mode.

2) **At pure imaginary diverging mode – multiple ZGV points**

   There are multiple transition points along the diverging mode imag \( S_{6} \) at the transition to complex \( S_{4,3} \), complex \( S_{6,8,5} \), complex \( S_{8,5,12,7} \), and so on, and to real \( S_{6} \).

   *When observing the modes in real plane, you feel there is no connection between them, but they are actually connected in complex domain and belong to different branches. This explains the energy conversion in physics should be continuous instead of discrete.*

3) **At real mode – none, single or multiple ZGV points**

   Surprisingly, the number of ZGV points has an analytical solution [120] and is a function of Poisson’s ratio. Some examples are shown as below. There are three ZGV points for \( \nu = 0.265 \) (Figure 14.26); there are two for \( \nu = 0.3 \) (Figure 14.27) and there are three for \( \nu = 0.34 \) (Figure 14.28).
Figure 14.26  ZGV points for $\nu = 0.265$

Figure 14.27  ZGV points for $\nu = 0.3$

Figure 14.28  ZGV points for $\nu = 0.34$
14.7.3 Analytic Solution for Group Velocity

Recall dispersion relations from (Section 0) in the following forms:

1) Original form represented by \((\omega, k, p, q)\)
2) Non-dimensional form represented by \((\bar{\omega}, \bar{k}, \bar{p}, \bar{q})\)
3) Non-dimensional form represented by \((\Omega, K)\)
4) Non-dimensional form represented by \((\bar{d}, \zeta)\)

Unified dispersion relation is given from Eq. (14.46) and can be rearranged as

\[
F(\omega, k) = (k^2 - q^2)^2 \tan(qh / 2 + \gamma) + 4k^2 pq \tan(ph / 2 + \gamma) \quad \text{Eq. (14.102)}
\]

With this definition, any dispersion relation linking frequency and wavenumber is suitable to calculate group velocity. From implicit differentiating, considering the dispersion function as \(dF(\omega, k) = 0\) therefore chain rules told us

\[
dF(\omega, k) = \frac{\partial F(\omega, k)}{\partial \omega} d\omega + \frac{\partial F(\omega, k)}{\partial k} dk \quad \text{Eq. (14.103)}
\]

So that group velocity is given by

\[
c_g = \frac{d\omega}{dk} = -\frac{\partial F(\omega, k)/\partial k}{\partial F(\omega, k)/\partial \omega} \quad \text{Eq. (14.104)}
\]

Then, expanding above equation gives

\[
c_g = \frac{k}{\omega} c_p^2 c_r^2 \left( \frac{A}{B} \right) = \frac{1}{c_p} c_r^2 c_r^2 \left( \frac{A}{B} \right) \quad \text{Eq. (14.105)}
\]
where

\[ A = 8pq\omega^2(2 - \frac{c_p^2}{c_T^2})\tan(qh / 2 + \gamma) + 4\left[2p^2q^2c_p^2 - (p^2 + q^2)\omega^2\right]\tan(ph / 2 + \gamma) \]
\[ -\frac{hp}{2}\left[\frac{\omega^4(2 - c_p^2 / c_T^2)^2}{c_p^2\cos^2(qh / 2 + \gamma)} + \frac{4q^2\omega^2}{\cos^2(ph / 2 + \gamma)}\right] \]  
Eq. (14.106)

\[ B = 4pq\omega^2c_L^2(2 - \frac{c_p^2}{c_T^2})\tan(qh / 2 + \gamma) - 4\omega^2(c_T^2q^2 + c_p^2p^2)\tan(ph / 2 + \gamma) \]
\[ -\frac{hp}{2}\left[\frac{c_L^2\omega^4(2 - c_p^2 / c_T^2)^2}{c_p^2\cos^2(qh / 2 + \gamma)} + \frac{4q^2\omega^2}{\cos^2(ph / 2 + \gamma)}\right] \]
Eq. (14.107)

Expanding in form (2) with non-dimensional parameters \((\bar{\omega}, \bar{k}, \bar{p}, \bar{q})\) gives

\[ C_g = \frac{d\bar{\omega}}{dk} = \frac{d\omega}{dk} \frac{1}{c_T} = \frac{c_g}{\bar{\omega}} \left(\frac{A}{B}\right) \]
Eq. (14.108)

where

\[ A = 8\bar{p}\bar{q}(2\bar{k}^2 - \bar{\omega}^2)\tan(\bar{q} / 2 + \gamma) + 4\left[2\bar{p}^2\bar{q}^2 - (\bar{p}^2 + \bar{q}^2)\bar{k}^2\right]\tan(\bar{p} / 2 + \gamma) \]
\[ -\frac{\bar{p}}{2}\left[\frac{(2\bar{k}^2 - \bar{\omega}^2)^2}{\cos^2(\bar{q} / 2 + \gamma)} + \frac{4\bar{q}^2\bar{k}^2}{\cos^2(\bar{p} / 2 + \gamma)}\right] \]
Eq. (14.109)

\[ B = 4\bar{p}\bar{q}(2\bar{k}^2 - \bar{\omega}^2)\tan(\bar{q} / 2 + \gamma) - 4\bar{k}^2(\bar{p}^2 + \xi^2\bar{q}^2)\tan(\bar{p} / 2 + \gamma) \]
\[ -\frac{\bar{p}}{2}\left[\frac{(2\bar{k}^2 - \bar{\omega}^2)^2}{\cos^2(\bar{q} / 2 + \gamma)} + \frac{4\xi^2\bar{q}^2\bar{k}^2}{\cos^2(\bar{p} / 2 + \gamma)}\right] \]
Eq. (14.110)

and non-dimensional variables are defined by

\[ \bar{\omega} = \omega h / c_T, \bar{k} = kh \]
Eq. (14.111)

\[ \bar{p}^2 = (ph)^2 = \xi^2\bar{\omega}^2 - \bar{k}^2, \bar{q}^2 = (qh)^2 = \bar{\omega}^2 - \bar{k}^2, \xi = c_T / c_L \]
Eq. (14.112)
Expanding in combining form (3) and (4) with non-dimensional parameters \((\Omega, \zeta)\) gives

\[
C_g = \frac{d\Omega}{dK} = \frac{d\omega}{dk} \frac{1}{c_T} = K \left( \frac{I + II + III}{IV + V + VI} \right)
\]

Eq. (14.113)

where

\[
I : 8\sqrt{\xi^2 - \zeta^2} \sqrt{1 - \zeta^2} (1 - 2\zeta^2) \tan(\sqrt{1 - \zeta^2} \Omega/2 + \gamma)
\]

Eq. (14.114)

\[
II : 4\left[ \xi^2 (\xi^2 - 2\zeta^2 + 1) - 2(\xi^2 - \zeta^2)(1 - \zeta^2) \right] \tan(\sqrt{\xi^2 - \zeta^2} \Omega/2 + \gamma)
\]

Eq. (14.115)

\[
III : \frac{\Omega}{2} \sqrt{\xi^2 - \zeta^2} \left( \frac{(1 - 2\zeta^2)^2}{\cos^2(\sqrt{1 - \zeta^2} \Omega/2 + \gamma)} + \frac{4\zeta^2(1 - \zeta^2)}{\cos^2(\sqrt{\xi^2 - \zeta^2} \Omega/2 + \gamma)} \right)
\]

Eq. (14.116)

\[
IV : 4\sqrt{\xi^2 - \zeta^2} \sqrt{1 - \zeta^2} (1 - 2\zeta^2) \tan(\sqrt{1 - \zeta^2} \Omega/2 + \gamma)
\]

Eq. (14.117)

\[
V : 4\xi^2 (2\xi^2 - \zeta^2 - \zeta^2) \tan(\sqrt{\xi^2 - \zeta^2} \Omega/2 + \gamma)
\]

Eq. (14.118)

\[
VI : \frac{\Omega}{2} \sqrt{\xi^2 - \zeta^2} \left( \frac{(1 - 2\zeta^2)^2}{\cos^2(\sqrt{1 - \zeta^2} \Omega/2 + \gamma)} + \frac{4\xi^2 \xi^2(1 - \zeta^2)}{\cos^2(\sqrt{\xi^2 - \zeta^2} \Omega/2 + \gamma)} \right)
\]

Eq. (14.119)

and non-dimensional variables are defined by

\[
\Omega = \omega h/c_T, \quad K = kh, \quad \zeta = K/\Omega = \sqrt{\rho_p} = c_T/c_p, \quad \xi = c_T/c_L
\]

Eq. (14.120)
Further simplifying in terms of \((\bar{d}, \zeta)\) as form (4) gives

\[
C_g = \frac{d \Omega}{dK} = \frac{d \omega}{dk} \frac{1}{c_r} = \frac{K}{\Omega} \left( \frac{I + II + III}{IV + V + VI} \right) = \zeta \left( \frac{I + II + III}{IV + V + VI} \right)
\]

Eq. (14.121)

where

\[
I : 8 \sqrt{\xi^2 - \zeta^2} \sqrt{1 - \zeta^2} (1 - 2\zeta^2) \tan(\bar{d} \sqrt{1 - \zeta^2} + \gamma)
\]

Eq. (14.122)

\[
II : 4 \left[ \xi^2 (\xi^2 - 2\zeta^2 + 1) - 2(\xi^2 - \zeta^2)(1 - \zeta^2) \right] \tan(\bar{d} \sqrt{\xi^2 - \zeta^2} + \gamma)
\]

Eq. (14.123)

\[
III : \bar{d} \sqrt{\xi^2 - \zeta^2} \left( \frac{(1 - 2\zeta^2)^2}{\cos^2(\bar{d} \sqrt{1 - \zeta^2} + \gamma)} + \frac{4\xi^2 (1 - \zeta^2)}{\cos^2(\bar{d} \sqrt{\xi^2 - \zeta^2} + \gamma)} \right)
\]

Eq. (14.124)

\[
IV : 4 \sqrt{\zeta^2 - \xi^2} \sqrt{1 - \zeta^2} (1 - 2\zeta^2) \tan(\bar{d} \sqrt{1 - \zeta^2} + \gamma)
\]

Eq. (14.125)

\[
V : 4 \xi^2 (2\xi^2 - \xi^2 - \zeta^2) \tan(\bar{d} \sqrt{\xi^2 - \zeta^2} + \gamma)
\]

Eq. (14.126)

\[
VI : \bar{d} \sqrt{\zeta^2 - \xi^2} \left( \frac{(1 - 2\zeta^2)^2}{\cos^2(\bar{d} \sqrt{1 - \zeta^2} + \gamma)} + \frac{4\xi^2 \zeta^2 (1 - \zeta^2)}{\cos^2(\bar{d} \sqrt{\xi^2 - \zeta^2} + \gamma)} \right)
\]

Eq. (14.127)

and non-dimensional variables are defined by

\[
\bar{d} = \Omega/2 = \omega h/2c_r, \quad K = kh, \quad \zeta = K/\Omega = 1/C_p = c_r/c_p, \quad \xi = c_r/c_L
\]

Eq. (14.128)
14.7.4 Dispersion Bands

Dispersion curves become dispersion band when there are uncertain variations such as material properties, plate thickness, non-monotone band limited excitation, etc. Below just a demo of dispersion bands for a material with Poisson’s ratio variability between 0.325 and 0.335. The bandwidth of the dispersion band varies with frequency, symmetrical and anti-symmetrical modes. Dispersion bands (Figure 14.29), phase velocity (Figure 14.30) and group velocity (Figure 14.31) are shown as below.

---

**Figure 14.29** For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands

**Figure 14.30** For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands – phase velocity

**Figure 14.31** For $\nu = 0.325$ to $\nu = 0.335$ dispersion bands – group velocity

---
14.7.5 Mode Tracing and Filter Banks

The applications of Riesz bp transform is to generate damage sensitive hypercomplex parameters and filters as a filter bank to separate the waves in different modes. Recall the dispersion curves, for example, a plate under 1.25 MHz excitation (Figure 14.32) has five different modes and we can generate a five frequency bands to isolate the waves in each frequency band from each other so that they can be examined separately.

![Disperion Curves](image)

Figure 14.32 Example of how to assign five filters as a filter bank at 1.25 MHz excitation

**Demo – looking for local resonance from hypercomplex parameters**

A piezo shaker is attached on an aluminum plate with continuous sweeping excitation and wavefield (Figure 14.33 right) was measured by shearography (credited by isi-sys). The waves travel across the object surface and excite the five local subsurface defects on the top. The local defects vibrate differently, depending on their local mechanical properties (stiffness). For comparison to verify the measured wavefield, transient finite element model (Figure 14.33 left) was built to visualize the strain field.

In order to highlight the damage feature, Riesz bp transform was applied to extract the local phase (LP) (Figure 14.34) and local energy (LE) (Figure 14.35) over the recorded image (Figure 14.33 right) which is already the local phase (LP). The process of LP over LP is for feature matching where LE over LP is for enveloping. The central frequency of the Riesz bp was selected from 5 mm over the next nine one-third octave bands (total ten bands) up to 40 mm. Damage location can be clearly seen at about 10 to 15 mm.
Figure 14.33  SG simulated strain field and measured local phase (LP)

Figure 14.34  Riesz bp local phase (LP) of measured local phase (LP)

Figure 14.35  Riesz bp local energy (LE) of measured local phase (LP)
Demo – wavefields separation and enveloping by Riesz bp

This demo aims to test the capabilities of Riesz bp filter to separate the modes when there are multiple reflections. The testing conditions (Figure 14.36) are shown as below. The processing procedures are shown as (Figure 14.37) to process the images in spatial or frequency domain. For a given excitation frequency, the central frequency of the filter is assigned according to the wavenumber for each mode, in this case, the excitation is below the cut-off so there are only two wavenumbers.

![Figure 14.36 Excitation conditions for plate with three dummy damages](image)

- 200kHz toneburst in plane excitation
  - 500 x 500 x 5mm Al plate
  - Three small damages (D = 5mm)
  - 25us excitation + 1000us propagation
  - A0 and S0
  - Multiple reflection at BCs

![200kHz toneburst in plane excitation](image)

- For a given excitation frequency
- Each frequency is corresponding to two wavenumbers
- So we need to pre-select two spatial frequencies/wavenumbers for Gabor BP filter
- Mode selection – separation problem
- Riesz – 2D envelop

![Riesz bp can be applied in either spatial or frequency domain to separate the wavefields](image)

Wavefields in different stages: initiating stage (Figure 14.38), first reflection stage (Figure 14.39) and multiple reflection stage (Figure 14.40) are assigned with Riesz bp to extract hypercomplex parameters and separate the modes (symmetric and anti-symmetric). The directional quadrature $R_x$ and $R_y$ enhance the wavefront direction with carriers where local energy (LE) takes the 2D envelope on the carriers so that the energy propagation can be visualized.
Figure 14.38  Wavefield separation into ASYM and SYM modes – initiating stage

Figure 14.39  Wavefield separation into ASYM and SYM modes – first reflection stage

Figure 14.40  Wavefield separation into ASYM and SYM modes – multiple reflection stage
14.8 Dispersion Curves Simulator

Dispersion curves simulator (Figure 14.41) was developed in LabVIEW with hybrid state machine and dynamic programming in parallel. MATLAB script was embedded to make backend code more compact. Key features including ZGV detection, particle motion animation and mode tracing, etc.

Figure 14.41 Dispersion curves simulator – GUI front panel
Block diagram (Figure 14.42) is designed for data flow tracing. Web publishing function is also included by a third party add-on, LabSocket, to extend the accessibilities to any web browser and any platform.

Figure 14.42  Dispersion curves simulator – block diagram
15 Planar Waveguide – SH Waves

The last two elastic waves, symmetric SH and anti-symmetric SH are discussed in this chapter.

In plate – planar waveguides

- Sym PSV $\text{Sym}(K, P, Q) = (K^2 - Q^2)^2 \tan(Q/2) + 4K^2PQ \tan(P/2) = 0$
- Anti PSV $\text{Anti}(K, P, Q) = 4K^2PQ \tan(Q/2) + (K^2 - Q^2)^2 \tan(P/2) = 0$
- Sym SH $\text{SymSH}(Q) = \cos(Q/2) = 0$
- Anti SH $\text{AntiSH}(Q) = \sin(Q/2) = 0$

Figure 15.1 Elastic waves in planar waveguides

Another type of wave motion in the plate involves in-plane shear motion. In a plate defined surface normal as $z$ direction, the wave polarized in the $(x, z)$ plane with displacement $(u, w)$ decoupled to the wave polarized in $y$ direction with displacement $v$. The particle motion of shear horizontal (SH) waves are normal to the propagation $x$ direction. The displacement in $y$ direction is assumed in the form

$$v = f(z) \exp(i(\omega t - kx)) \quad \text{Eq. (15.2)}$$

This solution must satisfy the wave equation

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial z^2} = \frac{1}{c_y^2} \frac{\partial^2 v}{\partial t^2} \quad \text{Eq. (15.3)}$$

and the boundary condition at $z = \pm h / 2$

$$\tau_{yz} = \mu \frac{\partial v}{\partial z} = 0 \quad \text{Eq. (15.4)}$$

Substituting Eq. (15.2) into Eq. (15.3) and solving for $f(z)$ give

$$f(z) = B_1 \sin qz + B_2 \cos qz \quad \text{Eq. (15.5)}$$

Applying boundary conditions Eq. (15.4) yields
\[ B_1 \cos\left(\frac{q h}{2}\right) \pm B_2 \sin\left(\frac{q h}{2}\right) = 0 \]  
Eq. (15.6)

Eq. (15.6) can be satisfied in two ways:

\[ B_1 = 0 \quad \text{and} \quad \sin\left(\frac{q h}{2}\right) = 0 \]  
Eq. (15.7)

\[ B_2 = 0 \quad \text{and} \quad \cos\left(\frac{q h}{2}\right) = 0 \]  
Eq. (15.8)

If \( B_1 = 0 \), the expression for \( f(z) \) shows that the displacement is symmetric with respect to the mid-plane of the plate. The displacement is antisymmetric if \( B_2 = 0 \). In both cases, the frequencies follow

\[ q h = n\pi \]  
Eq. (15.9)

where \( n = 0, 2, 4, \ldots \) for symmetric modes, and \( n = 1, 3, 5, \ldots \) for antisymmetric modes. Using the expression of \( q \), the dispersion relation can be written in an analytical form as

\[ q^2 = \left(\frac{\omega}{c_T}\right)^2 - k^2 \quad \Rightarrow \quad (n\pi)^2 = \left(\frac{\omega h}{c_T}\right)^2 - (kh)^2 \]  
Eq. (15.10)

The hyperbolic dispersion relation can be normalized and is given by

\[ Q^2 = \Omega^2 - K^2 \]  
Eq. (15.11)

Note that except the fundamental mode \( n = 0 \), all other higher modes are dispersive.

The displacement (Rose and Cho, Materials Evaluation, pp. 1234-1238, 2001) can be written as

\[ v^n(x, z, t) = \begin{cases} 
B_n \cos\left(\frac{n\pi z}{h}\right) & \text{for symmetric modes} \\
A_n \sin\left(\frac{n\pi z}{h}\right) & \text{for anti-symmetric modes}
\end{cases} \]  
Eq. (15.12)

And the stresses are
where the term $\exp\left(\omega t - k_n x\right)$ has been omitted in the expression of displacements and stresses shown above. $A_n$ and $B_n$ are unknown coefficients.

Note that the amplitude of SH modes is independent of frequency and wavenumber. Hence the wave structure of the SH mode does not vary along the entire dispersion curve. In contrast to Lamb wave, the wavefield is a function of $(\omega, k, p, q)$. 
Figure 15.3  SH wave dispersion curves for all propagating modes – phase velocity

Figure 15.4  SH wave dispersion curves for all propagating modes – group velocity
Four types of cylindrical waves are discussed in this chapter.

The waves considered so far propagate in plane waveguides, either near a plane surface (Rayleigh or Love waves), or between parallel plane surfaces (Lamb waves). The analysis of these waveguides is relatively simple because they have an infinite cross-section. However, there are also waveguides with finite cross-sections, for waves traveling over large distances. In the following the waveguide in circular cylinders is considered.

![Cylindrical coordinates](image)

Figure 16.2  Cylindrical coordinates \((r, \theta, z)\) and the radius of the cylinder is \(a\)

Similar to the procedure used in the case of the plate, it is natural to adopt cylindrical coordinates \((r, \theta, z)\) as in (Figure 16.2) and consider wave propagation along the longitudinal axis \(z\).

\[
\nabla^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2} \tag{Eq. (16.1)}
\]

It is assumed that the scalar potential can be written in a separable of variable form

\[
\phi = f(r)g(\theta) \exp(i(kz - \omega t)) \tag{Eq. (16.2)}
\]
Substituting Eq. (16.2) into Eq. (11.6) gives

\[ \frac{d^2 g}{d\theta^2} + n^2 g(\theta) = 0 \quad \text{Eq. (16.3)} \]

\[ \frac{d^2 f}{dr^2} + \frac{1}{r} \frac{df}{dr} + \left( \frac{\omega^2}{c_L^2} - k^2 \right) f - \frac{n^2}{r^2} f(r) = 0 \quad \text{Eq. (16.4)} \]

where \( n \) is integer

The development of the vectorial Laplacian \( \nabla^2 \psi \) is described in the appendix. For the harmonic motion three equations are given by

\[ \nabla^2 \psi_r = \frac{\psi_r}{r^2} - 2 \frac{\partial \psi_\theta}{\partial \theta} + \frac{\omega^2}{c_T^2} \psi_r = 0 \quad \text{Eq. (16.5)} \]

\[ \nabla^2 \psi_\theta = \frac{\psi_\theta}{r^2} + 2 \frac{\partial \psi_r}{\partial \theta} + \frac{\omega^2}{c_T^2} \psi_\theta = 0 \quad \text{Eq. (16.6)} \]

\[ \nabla^2 \psi_z + \frac{\omega^2}{c_T^2} \psi_z = 0 \quad \text{Eq. (16.7)} \]

The components \( \psi_r \) and \( \psi_\theta \) are coupled in Eq. (16.5) and Eq. (16.6). In Eq. (16.7) component \( \psi_z \) is decoupled, has a similar form as Eq. (16.3) for the scalar potential \( \phi \).

The solutions of Eq. (16.3) for \( g(\theta) \) are trigonometric functions. Solutions of Eq. (16.4) for \( f(r) \) are typical Bessel functions. Since the field variables must be finite at \( r = 0 \), the Bessel functions must be of the first kind showing that

\[ J_n(pr) \quad \text{with} \quad p^2 = \frac{\omega^2}{c_L^2} - k^2 \quad \text{Eq. (16.8)} \]

thus, the solution for \( \phi \) can be written as

\[ \phi = AJ_n(pr) \cos(n\theta) \exp i(kz - \omega t) \quad \text{Eq. (16.9)} \]

and \( \psi_z \) has the similar solution
\[ \psi_z = B J_n(qr) \cos(n\theta) \exp(i(kz - \omega t)) \quad \text{with} \quad q^2 = \frac{\omega^2}{c^2} - k^2 \quad \text{Eq. (16.10)} \]

The expressions for \( \psi_z \) and \( \psi_\theta \) are \textit{a priori} of the form

\[ \psi_r = \psi_r(r) \sin(n\theta) \exp(i(kz - \omega t)) \quad \text{Eq. (16.11)} \]
\[ \psi_\theta = \psi_\theta(r) \cos(n\theta) \exp(i(kz - \omega t)) \quad \text{Eq. (16.12)} \]

where the presence of \( \sin(n\theta) \) in the first expression implies \( \cos(n\theta) \) in the second, because the coupling terms in Eq. (16.6) and Eq. (16.7) have differentials with respect to \( \theta \) with different signs. Substitution gives

\[ \frac{d^2 \psi_r}{dr^2} + \frac{1}{r} \frac{d \psi_r}{dr} + \frac{1}{r^2} (-n^2 \psi_r + 2m \psi_\theta - \psi_r) + q^2 \psi_r = 0 \quad \text{Eq. (16.13)} \]
\[ \frac{d^2 \psi_\theta}{dr^2} + \frac{1}{r} \frac{d \psi_\theta}{dr} + \frac{1}{r^2} (-n^2 \psi_\theta + 2m \psi_r - \psi_\theta) + q^2 \psi_\theta = 0 \quad \text{Eq. (16.14)} \]

where the second equation is the same as the first except for interchanging the indices \( r \) and \( \theta \). Subtracting and adding these equations give two new equations whose solutions are, respectively,

\[ \begin{cases} \psi_r - \psi_\theta = 2C J_{n+1}(qr) \\ \psi_r + \psi_\theta = 2C_1 J_{n-1}(qr) \end{cases} \quad \text{Eq. (16.15)} \]

from which we have

\[ \begin{cases} \psi_r = C_1 J_{n-1}(qr) + C J_{n+1}(qr) \\ \psi_\theta = C_1 J_{n-1}(qr) - C J_{n+1}(qr). \end{cases} \quad \text{Eq. (16.16)} \]

Four unknown constants, \( A, B, C \), and \( C_1 \) appear in the potentials \( \phi, \psi_z, \psi_r \) and \( \psi_\theta \). However, the boundary conditions, expressing the fact that the surface \( r = a \) is free, only give the three relations \( \sigma_{rr} = 0 \), \( \tau_{rz} = 0 \) and \( \tau_{r\theta} = 0 \). Fortunately, we are free to impose another condition between the components \( \psi_r, \psi_\theta \) and \( \psi_z \), and the choice \( \psi_r = -\psi_\theta \) is convenient because it gives \( C_1 = 0 \). The expressions for \( \phi \) and the components of \( \psi \) then become
\[ \phi = AJ_n(pr) \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \exp i(kz - \omega t) \]  
Eq. (16.17)

\[ \psi_z = BJ_n(qr) \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \exp i(kz - \omega t) \]  
Eq. (16.18)

\[ \psi_r = CJ_{n+1}(qr) \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \exp i(kz - \omega t) \]  
Eq. (16.19)

\[ \psi_\theta = -CJ_{n+1}(qr) \begin{bmatrix} \cos n\theta \\ \sin n\theta \end{bmatrix} \exp i(kz - \omega t) \]  
Eq. (16.20)

### 16.1 Boundary Conditions

The displacements in terms of potentials are given by

\[ u_r = \frac{\partial \phi}{\partial r} + \frac{1}{r} \frac{\partial \psi_z}{\partial \theta} - \frac{\partial \psi_\theta}{\partial z} \]  
Eq. (16.21)

\[ u_\theta = \frac{1}{r} \frac{\partial \phi}{\partial \theta} + \frac{\partial \psi_r}{\partial z} - \frac{\partial \psi_z}{\partial r} \]  
Eq. (16.22)

\[ u_z = \frac{\partial \phi}{\partial z} + \frac{1}{r} \frac{\partial (r\psi_\theta)}{\partial r} - \frac{1}{r} \frac{\partial \psi_r}{\partial \theta} \]  
Eq. (16.23)

The local dilatation \( \varepsilon \) is

\[ \varepsilon = \nabla \cdot \mathbf{u} = \frac{1}{r} \frac{\partial (ru_r)}{\partial r} + \frac{1}{r} \frac{\partial u_\theta}{\partial \theta} + \frac{\partial u_z}{\partial z} \]  
Eq. (16.24)

and the stresses, involving \( \varepsilon \) and the Lamé constants \( \lambda \) and \( \mu \), are

\[ \sigma_{rr} = \lambda \varepsilon + 2\mu \frac{\partial u_r}{\partial r} \]  
Eq. (16.25)

\[ \tau_{r\theta} = \mu \left[ \frac{1}{r} \left( \frac{\partial u_r}{\partial \theta} - u_\theta \right) + \frac{\partial u_\theta}{\partial r} \right] \]  
Eq. (16.26)

\[ \tau_{rz} = \mu \left( \frac{\partial u_r}{\partial z} + \frac{\partial u_z}{\partial r} \right) \]  
Eq. (16.27)

Setting these three stresses to zero at the free surface \( r = a \) yields three linear homogeneous equations in the three unknowns \( (A,B,C) \) giving non-zero solutions if the determinant of the coefficients is zero.
Expressing this explicitly gives the dispersion relation relating to \((\omega, k)\) and the circumferential order \(n\).

For any \(n\) and real-valued \(k\) there is an infinite number of roots, which are the frequencies of modes propagating along \(z\).

### 16.2 General Dispersion Relations

The terminology used in the section is that for each value of \(n \geq 0\) determine a family of modes. In the case of \(n = 0\), the general dispersion relation reduces to two. The first provides the dispersion of longitudinal waves, axially symmetric, with the displacements \(u_r\) and \(u_z\) independent of \(\theta\). The second provides torsional waves with circumferential displacement \(u_\theta\) independent of \(\theta\).

When \(n = 1\), for example, the ordinary family of flexural modes results with displacements \((u_r, u_\theta, u_z)\) dependent on \((r, \theta, z)\). This family of modes has features similar to the family of flexural modes in a plate.

For a value of \(n \geq 2\), the resulting family of modes will be called the family of flexural modes of circumferential order \(n\). These families of modes have no counterparts in the plate.

### 16.3 Cylindrical Wave – Longitudinal Modes

In this case, with \(n = 0\), the particle motion is described by two displacement components \(u_r\) and \(u_z\), which are independent of \(\theta\). The expressions for the potentials in Eq. (16.17) to Eq. (16.20) become

\[
\phi = AJ_0 (pr) \exp i(kz - \omega t) \quad \text{Eq. (16.28)}
\]

\[
\psi_\theta = - C J_1 (qr) \exp i(kz - \omega t) \quad \text{Eq. (16.29)}
\]

\[
\psi_r = \psi_z = 0 \quad \text{Eq. (16.30)}
\]

Using the relations

\[
\frac{d}{dx} [J_0(x)] = - J_1(x) \quad \text{and} \quad \frac{d}{dx} [J_1(x)] = J_0(x) - \frac{J_1(x)}{x} \quad \text{Eq. (16.31)}
\]

the radial and axial components of displacement in Eq. (16.21) to Eq. (16.23) become

\[
u_r = - [pAJ_1 (pr) + ikCJ_1 (qr)] \exp i(kz - \omega t) \quad \text{Eq. (16.32)}
\]
\[ u_o = 0 \quad \text{Eq. (16.33)} \]
\[ u_z = -\left[i k A J_o(pr) + q C J_o(qr)\right] \exp(i kz - \omega t) \quad \text{Eq. (16.34)} \]

These expressions, and the dilatation \( \varepsilon = -(p^2 + k^2) A J_o(pr) \), are substituted into Eq. (16.25) and Eq. (16.27) to obtain the stresses \( \sigma_{rr} \) and \( \tau_{rz} \). Setting these stresses to zero at the surface \( r = a \) yields

\[
\left\{ \left[-\lambda(p^2 + k^2) - 2\mu p^2\right] J_o(pa) + 2\mu p J_1(pa)/a \right\} A \\
+ 2\mu ik \left[-qJ_o(qa) + J_1(qa)/a \right] C = 0
\]

Eq. (16.35)

\[
-2\mu ikpJ_1(pa)A - \mu(q^2 - k^2)J_1(qa)C = 0
\]

Eq. (16.36)

Using the equality \( \lambda(p^2 + k^2) + 2\mu p^2 = \mu(q^2 - k^2) \), these equations become

\[
\left[-(q^2 - k^2)J_o(pa) + 2pJ_1(pa)/a \right] A + 2ik \left[-qJ_o(qa) + J_1(qa)/a \right] C = 0 \quad \text{Eq. (16.37)}
\]

\[
2ikpJ_1(pa)A + (q^2 - k^2)J_1(qa)C = 0 \quad \text{Eq. (16.38)}
\]

and these are satisfied only if the determinant of coefficients is zero, giving

\[
\frac{2p}{a} (q^2 + k^2)J_1(pa)J_1(qa) - (q^2 - k^2)J_0(pa)J_1(qa) - 4k^2 pq J_1(pa)J_0(qa) = 0 \quad \text{Eq. (16.39)}
\]

Using the non-dimensional variables \((K, P, Q)\), Eq. (16.39) is written as

\[
2P \left(Q^2 + K^2\right) J_1(P)J_1(Q) - (Q^2 - K^2) J_0(P)J_1(Q) - 4K^2 PQ J_1(P)J_0(Q) = 0 \quad \text{Eq. (16.40)}
\]

where \( P^2 = \xi^2 \Omega^2 - K^2 = pa \), \( Q^2 = \Omega^2 - K^2 = qa \), \( \Omega = \omega a/c_r \), \( K = ka \)

This dispersion relation is often called the Pochhammer-Chree equation.

With \((p, q)\) given by Eq. (16.8) and Eq. (16.10), it contains the angular frequency \( \omega \), the wavenumber \( k \), the velocities of bulk longitudinal and transverse waves \( c_L \) and \( c_T \), and the radius \( a \).
Depending on whether the phase velocity $c_p = \omega / k$ is greater than $c_L$ or $c_T$, the coefficients $(p,q)$ are real or pure imaginary. In the latter case, the equations may be rewritten using the identity $J_n(ix) = i^n I_n(x)$, where $I_n(x)$ denotes the modified Bessel functions of the first kind.

The dispersion curves shown as (Figure 16.3) with the relations of $\omega(k)$ for the first few longitudinal modes in a cylinder with specified Poisson’s ratio and the phase velocity $c_p = \omega / k$ and group velocity $c_g = d\omega / dk$, normalized to the bulk transverse wave velocity $c_T$, are shown in (Figure 16.4) and (Figure 16.5) as functions of the normalized angular frequency $\Omega = \omega a / c_T$ for the first few longitudinal modes.

The capital represents the non-dimensional quantities $\Omega = \omega a / c_T$ and $K = ka$, respectively. All the modes are dispersive. The first mode is the only one with no cut-off frequency.

Figure 16.3  Cylindrical wave longitudinal mode dispersion curves in rod
Figure 16.4  Cylindrical longitudinal wave dispersion curves in rod – phase velocity

Figure 16.5  Cylindrical longitudinal wave dispersion curves in rod – group velocity
16.4 Cylindrical Wave – Flexural Modes

Here the particle motion is described by the three displacement components \((u_r, u_\theta, u_z)\), which vary with \(\theta\) as \(\sin(n\theta)\) and \(\cos(n\theta)\). To illustrate the vibration of a mode propagating along the cylinder, it is sufficient to consider the case \(n = 1\). The mode is described by the potentials of Eq. (16.17) to Eq. (16.20), omitting the exponential propagation term for convenience, so that

\[
\begin{align*}
\phi &= AJ_1(pr) \cos \theta & \text{Eq. (16.41)} \\
\psi_r &= CJ_2(qr) \sin \theta & \text{Eq. (16.42)} \\
\psi_\theta &= -CJ_2(qr) \cos \theta & \text{Eq. (16.43)} \\
\psi_z &= BJ_1(qr) \sin \theta & \text{Eq. (16.44)}
\end{align*}
\]

Using Eq. (16.21) to Eq. (16.23), these give

\[
\begin{align*}
u_r &= U_r(r) \cos \theta & \text{Eq. (16.45)} \\
u_\theta &= U_\theta(r) \sin \theta & \text{Eq. (16.46)} \\
u_z &= U_z(r) \cos \theta & \text{Eq. (16.47)}
\end{align*}
\]

where \(\sigma_r\), \(\tau_\theta\) and \(\tau_z\) are dependent only on \(r\). Inspection of the components \((u_r, u_\theta, u_z)\), as functions of the angle \(\theta\), explains why these waves are called flexural. Thus, for \(\theta = 0\) we have \(u_\theta = 0\), so points in the \((x,z)\) plane are displaced in this plane. For \(\theta = \pm \pi / 2\) only \(u_\theta\) is non-zero, so the points in the \((y,z)\) plane are displaced parallel to the \((x,z)\) plane. As given by Pao and Mindlin (1960), the frequency equation for the lowest order family of flexural modes in a non-dimensional form is given by

\[
J_1(P)J_1^2(Q) \left( f_1L_Q^2 + f_2L_pL_Q + f_3L_Q + f_4L_p + f_5 \right) = 0
\]

where

- \(f_1 = 2(Q^2 - K^2)^2\)
- \(f_2 = 2Q^2(5K^2 + Q^2)\)
- \(f_3 = Q^6 - 10Q^4 - 2Q^4K^2 + 2Q^2K^2 + Q^2K^4 - 4K^4\)
\[ f_4 = 2Q^3 \left( 2Q^3K^2 - Q^2 - 9K^2 \right) \]
\[ f_5 = Q^2 \left( -Q^4 + 8Q^2 - 2Q^2K^2 + 8K^2 - K^4 \right) \]
\[ L_x = xJ_0(x) / J_1(x) \quad \text{and} \quad x = P \text{ or } Q \]

As for other families of modes, the dispersion relation is obtained by substituting the expressions for the displacement components into the formulae for the stresses \( \sigma_{rr} \), \( \tau_{r\theta} \), and \( \tau_{rz} \), and then equating these stresses to zero at \( r = a \). This gives three equations in \( A \), \( B \), and \( C \), which are satisfied only if the determinant of the coefficients is zero. The solutions provide the various modes of the family.

The dispersion curves shown as (Figure 16.6) to (Figure 16.8) of flexural waves. Although the dispersion curves are more complicated than those of plates, involving more branches, the geometry of dispersion curves are similar to those in plates in lower branches of modes. At high frequencies, both phase and group velocities approach Rayleigh wave velocity. The behavior of higher modes is similar to the case of plate with all modes having the shear velocity as the limiting value of the phase and group velocities. However, one consequence of the cylindrical geometry is that the displacement field at low frequencies at \( k = 0 \) is no longer trigonometric functions.
Figure 16.7  Cylindrical wave flexural mode dispersion curves in rod – phase velocity

Figure 16.8  Cylindrical wave flexural mode dispersion curves in rod – group velocity
16.5 Cylindrical Wave – Torsional Modes

For torsional waves, \( u_r = 0 \) and \( u_\theta \) is independent of \( \theta \). It is derived from a single potential component \( \psi_\theta \). From Eq. (16.18) the potential may be written as

\[
\psi_\theta = -BJ_0(qr) \exp \left( i(kz - \omega t) / q^2 \right)
\]  

Eq. (16.49)

Giving

\[
u_\theta = -\frac{\partial \psi_\theta}{\partial r} = BJ_1(qr) \exp \left( i(kz - \omega t) / q \right)
\]

Eq. (16.50)

The stress \( \tau_{r\theta} \) must be zero at the free surface \( r = a \), so

\[
\tau_{r\theta} = \mu \left( \frac{\partial u_\theta}{\partial r} - \frac{u_\theta}{r} \right)_{r=a} = 0
\]

Eq. (16.51)

and with Eq. (16.31) the derivative of Bessel function gives

\[
qa J_0(qa) - 2J_1(qa) = 0
\]

Eq. (16.52)

\[
Q J_0(Q) - 2J_1(Q) = 0
\]

Eq. (16.53)

The equation gives an infinite set of roots. The first three values are \( q_1a \approx 5.136 \), \( q_2a = 8.417 \), \( q_3a = 11.62 \). One solution of the equation is \( q = 0 \). By taking the limit \( q \to 0 \) of Eq. (16.50) yields

\[
u_\theta = \frac{1}{2} Br \exp \left( i(kz - \omega t) \right)
\]

Eq. (16.54)

This is the lowest order torsional mode. Each circular section of the cylinder rotates about its center, since the displacement is proportional to the radial coordinate. From Eq. (16.55), the value \( q = 0 \) also requires the velocity to be a constant, equal to \( c_\tau \). Hence the fundamental torsional mode is not dispersive. The higher order torsional modes are dispersive, with velocities followed from the definition of as

\[
\left( \frac{\omega a}{c_\tau} \right)^2 = (ka)^2 + (qa)^2
\]

Eq. (16.55)
or in non-dimensional representation

$$\Omega^2 = K^2 + Q_n^2$$  \hspace{1cm} \text{Eq. (16.56)}$$

where $q, a$ are the solutions of Eq. (16.52) and Eq. (16.53). This equation has the same form as Eq. (15.11) for the SH modes. The dispersion relation of the torsional modes is quite similar to that of the anti-symmetric SH modes shown in (Figure 15.2) with the primary difference being that the frequencies of zero propagation are no longer integral multiplies because of non-periodic Bessel zero point spacing.

Figure 16.9  Cylindrical wave torsional mode dispersion curves in rod
Figure 16.10  Cylindrical wave torsional mode dispersion curves in rod – phase velocity

Figure 16.11  Cylindrical wave torsional mode dispersion curves in rod – group velocity
17 Analogy of Elastic Waves in Plane and in Rod

Since a rod can be treated as a wrapped plate, you’ll be impressed how similar their dispersion curves are. Side-by-side comparisons are made in this chapter.

In plate – planar waveguides

- **Sym PSV**  
  \[ Sym(K, P, Q) = (K^2 - Q^2)^2 \tan(Q/2) + 4K^2PQ \tan(P/2) = 0 \]

- **Anti PSV**  
  \[ Anti(K, P, Q) = 4K^2PQ \tan(Q/2) + (K^2 - Q^2)^2 \tan(P/2) = 0 \]

- **Sym SH**  
  \[ SymSH(Q) = \cos(Q/2) = 0 \]

- **Anti SH**  
  \[ AntiSH(Q) = \sin(Q/2) = 0 \]

In rod – cylindrical waveguides

- **Longitudinal**  
  \[ Long(K, P, Q) = 2P(Q^2 + K^2)J_1(P)J_1(Q) - (Q^2 - K^2)^2J_0(P)J_1(Q) - 4K^2PQJ_1(P)J_0(Q) = 0 \]

- **Flexural**  
  \[ Flex(K, P, Q) = J_1(P)J_1(Q)(f_1L_0^2 + f_2L_1L_0 + f_1L_2 + f_4L_0 + f_3) = 0 \]

- **Torsional**  
  \[ Tors(Q) = QJ_0(Q) - 2J_1(Q) = 0 \]

Figure 17.1 Analogy of elastic waves in plane and in rod

Notations

Waves in plate are denoted as PSV or SH, and in rod are denoted as cylindrical wave, CY. In plate, modes can be symmetric, SYM, or anti-symmetric, ASYM. In rod, modes can be longitudinal, flexural, torsion, torsion shown as below.
17.1 Comparison of PSV SYM Modes & CY Longitudinal Modes

Figure 17.2 Dispersion curves analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod

Figure 17.3 Phase velocity analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod

Figure 17.4 Group velocity analogy of (left) PSV SYM modes in plate and (right) CY longitudinal modes in rod
17.2 Comparison to PSV ASYM Modes & CY Flexural Modes

Figure 17.5 Displacement curves analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod

Figure 17.6 Phase velocity analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod

Figure 17.7 Group velocity analogy of (left) PSV ASYM modes in plate and (right) CY flexural modes in rod
17.3 Comparison to SH SYM Modes & CY Torsional Modes

Figure 17.8 Dispersion curves analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod

Figure 17.9 Phase velocity analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod

Figure 17.10 Group velocity analogy of (left) SH SYM modes in plate and (right) CY torsional modes in rod
V. Planar and Cylindrical Waveguides in Optical Dielectrics

Highly directed transport of light in free space is limited by attenuation and beam broadening due to diffraction and source spatial coherence. Waveguides are developed in advancing of the precise fabrication shown as below. This chapter mainly discussed the formation of EM waves and mode theory.

Lens waveguides: (Gobau 1960)

Light beams can be formed and propagated by lens and mirror systems counteracting transversal diffraction but, light beam in free space is broadened by diffraction (beam widening) and need to be periodically refocused by lenses. Diffraction effects in light beams increase with decreasing beam diameter.

Metallic waveguides:

Possible conceptually, but free carrier losses in metals at optical frequencies are too high for long distances. Waveguide dimensions on the $\mu m$ scale is still a technological challenge.

Dielectric waveguides: (1966 – today)

Very low absorption and scattering losses achieved in ultra-pure glasses as dielectrics, and fabrication of $km$ long waveguides with dimension $\sim$ the optical wavelength $\lambda \approx 1 \mu m$ is feasible.
18 Basic Optics

This chapter gives an elementary introduction to polarization, interference, and diffraction characteristics to understand the basic concepts of fiber optics.

18.1 Wave Assumptions

Assume wave propagates in $z$ direction and below are assumptions often made

- TE wave: electrical field $E$ has no $z$ component $\rightarrow E_z = 0$
- TM wave: magnetic field $H$ has no $z$ component $\rightarrow H_z = 0$
- TEM wave: both fields have no $z$ component $\rightarrow E_z = H_z = 0$
- Hybrid: EH (E dominated) or HE (H dominated) wave $\rightarrow E_z \neq 0, H_z \neq 0$

General expression of EM.

- Plane wave – a wave whose wavefront is a plane (as opposed to a cylindrical wave or a spherical wave with curved wavefront)
- Uniform plane wave – a plane wave with field vectors $(E, H)$ that are constant toward wavefront direction/wave propagating direction/wavevector direction
- TEM (transverse electromagnetic) wave – a wave with field vectors $(E, H)$ which are transverse to the wave propagating direction

A TEM wave traveling in $z$ direction has following properties

$$E_x \perp z, \ H_y \perp z \ \text{and} \ E_x \perp H_y$$  \hspace{1cm} \text{Eq. (18.1)}

Arbitrary wave traveling in arbitrary direction can be described as

$$E = A_i \hat{e}_i \exp \left( k \cdot r - \omega t + \phi_i \right)$$  \hspace{1cm} \text{Eq. (18.2)}

toward to

$$k = k_i \hat{e}_i = k_x \hat{e}_x + k_y \hat{e}_y + k_z \hat{e}_z$$  \hspace{1cm} \text{Eq. (18.3)}
where $\varphi_i$ is phase state for each component

$$\|k\| = k = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{2\pi}{\lambda}$$  
Eq. (18.4)

Then a plane wave travels in $z$ direction can be described as

$$E = E_x \hat{e}_x + E_y \hat{e}_y = A_x \hat{e}_x \exp\left(ik_z (z - \omega t)\right) + A_y \hat{e}_y \exp\left(ik_z (z - \omega t + \delta)\right)$$  
Eq. (18.5)

where $\delta = \varphi_y - \varphi_x$

### 18.2 The Definition of Polarization States

Light in the form of a plane wave in space is said to be linearly polarized. Light is a transverse electromagnetic wave, but natural light is generally unpolarized, all planes of propagation being equally probable. If light is composed of two plane waves of equal amplitude by differing in phase by 90°, then the light is said to be circularly polarized. If two plane waves of differing amplitude are related in phase by 90°, or if the relative phase is other than 90° then the light is said to be elliptically polarized.

For discussion convenience, we made a TE wave assumption to make $E_z = 0$ and three types of polarization states can be discussed as below

#### 18.2.1 Linear Polarization

A linearly polarized plane wave is the simplest electromagnetic wave, and if we assume the plane wave to be propagating in the $+z$ direction, the corresponding electric field can be written in the form when $\delta = 0 + 2m\pi$ and $m$ is integer
\[ E = (A_x \hat{\mathbf{e}}_x + A_y \hat{\mathbf{e}}_y) \exp \left( i(kz - \omega t) \right) \]  
Eq. (18.6)

where the electric field vector is assumed to oscillate in the \( x \) direction. In above equation, \( \omega = 2\pi f \) is the angular frequency and \( k \) is the propagation constant.

Then wave is oscillating in \((A_x \hat{\mathbf{e}}_x + A_y \hat{\mathbf{e}}_y)\) direction (fixed direction) \(\rightarrow\) linear polarized

For propagation in free space

\[ k = \frac{\omega}{c} \]  
Eq. (18.7)

where \( c \) represents the velocity of light in free space; when propagating in a medium characterized by refractive index \( n \), we have

\[ k_i = \frac{\omega}{\nu_i} = n_i \frac{\omega}{c} = n_i k \]  
Eq. (18.8)

where

\[ \nu_i = \frac{c}{n_i} \]  
Eq. (18.9)

represents the velocity of the electromagnetic wave in that medium. Equation describes an \( x \)-polarized wave. It is also known as a linearly polarized wave because the electric vector is oscillating along a specific axis shown as (Figure 18.2)

![Figure 18.2 A linear polarized EM wave propagating along the \( z \) direction (wiki)](image-url)
A linearly polarized wave is also known as a plane polarized wave because the electric field is always confined to a particular plane. A plane electromagnetic wave is said to be linearly polarized. The transverse electric field wave is accompanied by a magnetic field wave as illustrated.

18.2.2 Circular Polarization

When \( \delta = \pi / 2 + 2m\pi \) and \( m \) is still integer

\[
E = (A_x \hat{e}_x + iA_y \hat{e}_y) \exp i(kz - \omega t)
\]

Eq. (18.10)

Wave is pointing to direction \((A_x \hat{e}_x + iA_y \hat{e}_y)\) and \((A_x + iA_y)\) is complex represented as a quadrature pair with 90° phase difference \(\rightarrow\) rotating. When \(A_x = A_y = A\) the rotating path is a circle \(\rightarrow\) circular polarized. Circularly polarized light consists of two perpendicular electromagnetic plane waves of equal amplitude and 90° difference in phase. The light illustrated is right-circularly polarized.

![Circular Polarized EM Wave](wiki)

If light is composed of two plane waves of equal amplitude but differing in phase by 90°, then the light is said to be circularly polarized. *If you could see the tip of the electric field vector, it would appear to be moving in a circle as it approached you.* If while looking at the source, the electric vector of the light coming toward you appears to be rotating counterclockwise, the light is said to be right-circularly polarized. If clockwise, then left-circularly polarized light.

The electric field vector makes one complete revolution as the light advances one wavelength toward you. Another way of saying it is that if the thumb of your right hand were pointing in the direction of propagation
of the light, the electric vector would be rotating in the direction of your fingers. Circularly polarized light may be produced by passing linearly polarized light through a quarter-wave plate at an angle of 45° to the optic axis of the plate.

18.2.3 Elliptical Polarization

When $\delta = \pi / 2 + 2m\pi$ and $m$ is still integer. When $A_x \neq A_y$ the rotating path is not a circle $\rightarrow$ elliptical polarized. Elliptically polarized light consists of two perpendicular waves of unequal amplitude which differ in phase by 90°. The illustration shows right-elliptically polarized light.

If the thumb of your right hand were pointing in the direction of propagation of the light, the electric vector would be rotating in the direction of your fingers.

18.3 Production of Plane Polarized Waves

The ordinary light wave, like the one coming from a sodium lamp or from the sun, is unpolarized - that is, the electric vector (on a plane transverse to the direction of propagation) keeps changing its direction in a random manner as shown in (Figure 18.5).

If we allow the unpolarized wave to fall on a polarizer (such as a polaroid), then the wave emerging from the polaroid will be linearly polarized. In (Figure 18.5) the lines shown on the polaroid represent what is usually referred to as the "pass axis" of the polaroid - that is, the electric field perpendicular to the pass axis is gets absorbed. Polaroid sheets are extensively used for producing linearly polarized light waves. As an
interesting corollary, we note that if a second polaroid (whose pass axis is at right angles to the pass axis of the first polaroid) is put immediately after the first polaroid, then no light will come through.

![Figure 18.5](image)

Figure 18.5 Production of plane polarized light from an unpolarized light source using a polarizer.

The vertical lines in the polarizer represent the pass axis as a polarizing filter

### 18.4 Light Propagation through a Quarter Wave Plate

A quarter-wave plate (QWP) is a device that is used in many experiments involving fiber and integrated optics and is basically used to change the SOP of a propagating optical wave. For example, using this device we may transform a circularly polarized wave to a linearly polarized wave and vice versa shown as (Figure 18.6). A QWP is made of an anisotropic medium like calcite or quartz.

We will not go into the details of an anisotropic medium; it suffices here to say that inside a crystal like that of calcite, quartz, $LiNbO_3$, there is a preferred direction known as the optic axis. In a QWP the crystal is cut in such a way that the optic axis is at right angles to the thickness of the plate. If linearly polarized light is incident on a quarter-wave plate at 45° to the optic axis, then the light is divided into two equal electric field components. One of these is retarded by a quarter wavelength by the plate. This produces circularly polarized light. Incident circularly polarized light will be changed to linearly polarized light.

![Figure 18.6](image)

Figure 18.6 The transition from an unpolarized wave to a linearly polarized wave and then to a circularly polarized wave

185
19 EM Waves at Dielectric Interface

Fresnel's equations describe the reflection and transmission of electromagnetic waves at an interface. That is, they give the reflection and transmission coefficients for waves parallel and perpendicular to the plane of incidence. For a dielectric medium where Snell's Law can be used to relate the incident and transmitted angles, Fresnel's Equations can be stated in terms of the angles of incidence and transmission.

For light propagating in dielectric material, its properties can be described with following characteristics. Relates to vector nature of light – wavevector or Poyving vector shown as

\[ \|S\| = \|E \times H\| = \frac{1}{\mu_o} \|E \times B\| = c\epsilon_o \|E\|^2 \]

Eq. (19.1)

19.2 TE and TM Waves

Polarization as TE and TM waves.

Transverse-electric (TE) or s-polarized: E is perpendicular to plane-of-incidence

Coefficients can be calculated as below and illustrated as (Figure 19.1a)

1) For TE reflection

\[ r_\perp = \left( \frac{E_{oc}}{E_{oi}} \right)_\perp = \left( \frac{n_i \cos \theta_i - n_i \cos \theta_t}{n_i \cos \theta_i + n_i \cos \theta_t} \right) = \frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \]

Eq. (19.2)

2) For TE transmission

\[ t_\perp = \left( \frac{E_{ot}}{E_{oi}} \right)_\perp = \frac{2n_i \cos \theta_t}{n_i \cos \theta_i + n_i \cos \theta_t} = \left( \frac{2 \sin \theta_i \cos \theta_t}{\sin(\theta_i + \theta_t)} \right) \]

Eq. (19.3)

3) For TE reflectance

\[ R_\perp = \left( \frac{I_r}{I_i} \right)_\perp = \left( \frac{A \cos \theta_i E_{ot}^2}{A \cos \theta_i E_{oi}^2} \right)_\perp = r_\perp^2 \]

Eq. (19.4)

4) For TE transmittance

\[ T_\perp = \left( \frac{n_i \cos \theta_i E_{ot}^2}{n_i \cos \theta_i E_{oi}^2} \right)_\perp = \left( \frac{n_i \cos \theta_t}{n_i \cos \theta_i} \right)_\perp \]

Eq. (19.5)
Transverse-magnetic (TM) or p-polarized: B is perpendicular to plane-of-incidence

Coefficients can be calculated as below and illustrated as (Figure 19.1b)

5) For TM reflection

$$ r_\parallel = \frac{E_{or}}{E_{oi}}\parallel = \frac{n_i \cos \theta_i - n_i \cos \theta_i}{n_i \cos \theta_i + n_i \cos \theta_i} = \frac{\tan(\theta_i - \theta_i)}{\tan(\theta_i + \theta_i)} $$

Eq. (19.6)

6) For TM transmission

$$ t_\parallel = \left(\frac{E_{or}}{E_{oi}}\parallel\right) = \frac{2n_i \cos \theta_i}{n_i \cos \theta_i + n_i \cos \theta_i} = \frac{2 \sin \theta_i \cos \theta_i}{\sin(\theta_i + \theta_i) \cos(\theta_i - \theta_i)} $$

Eq. (19.7)

7) For TM reflectance

$$ R_\parallel = n_\parallel^{-2} $$

Eq. (19.8)

8) For TM transmittance

$$ T_\parallel = \left(\frac{n_i \cos \theta_i}{n_i \cos \theta_i}\right)t_\parallel^{-2} $$

Eq. (19.9)

Note that these coefficients are fractional amplitudes and must be squared to get fractional intensities for reflection and transmission. The signs of the coefficients depend on the original choices of field directions.

Due to the difference of medium refraction coefficients, there are two conditions to discuss, external (from small index medium to large index medium) and internal (from large index medium to small index medium) reflection.
## 19.3 Reflection and Deflection at a Plane Interface

Plane-of-incidence includes incident, reflected, transmitted. Consider a plane interface formed between two media of refractive indices $n_1$ and $n_2$ (Figure 19.2). When an electromagnetic wave is incident on such an interface, it will, in general, give rise to a reflected wave and a transmitted wave.

The amplitudes of the reflected and transmitted waves can be obtained by using the EM boundary conditions at the interface. The amplitude reflection and transmission coefficients defined as the ratio of the electric field of the reflected and transmitted waves to the electric field of the incident wave are given by [121].

![Figure 19.2](image)

**Figure 19.2** Reflection and transmission of a plane wave incident at an interface between media of refractive indices $n_i = n_r$ and $n_t$, (a) plane of incident, (b) p-polarization, (c) s-polarization.

### 19.3.1 Snell's Law

Snell's Law relates the indices of refraction $n$ of the two media to the directions of propagation in terms of the angles to the normal. Snell's law can be derived from Fermat's Principle or from the Fresnel Equations.

$$\frac{n_1}{n_2} = \frac{\sin \theta_2}{\sin \theta_1} \quad \text{Eq. (19.10)}$$

If the incident medium has the larger index of refraction, then the angle with the normal is increased by refraction. The larger index medium is commonly called the "internal" medium, since air with $n = 1$ is usually the surrounding or "external" medium.

You can calculate the condition for total internal reflection by setting the refracted angle $= 90°$ and calculating the incident angle. Since you can't refract the light by more than 90°, all of it will reflect for angles of incidence greater than the angle which gives refraction at 90°. In special case when material with complex refraction index which can refract light more than 90°. Here provide further information of detail derivation of Snell’s law and wave propagation in medium.
Let’s get started from the traveling speed in medium with refraction index $n_i$ of an incident wave that speed of light in medium will be slowed down to

$$\nu_i = \frac{c}{n_i} = \frac{\lambda_i f}{n_i} = \lambda_i f$$

Eq. (19.11)

where $c$ is speed of light and wavelength in medium will be compressed to

$$\lambda_i = \frac{\lambda}{n_i}$$

Eq. (19.12)

and wavenumber in medium will be expanded to

$$k_i = \frac{2\pi}{\lambda_i} = n_i \frac{2\pi}{\lambda} = n_i k$$

Eq. (19.13)

Note, frequency is the only quantity does not change when traveling into different mediums. The projected wavelength at the interface is shown as

$$\Lambda_i \sin \theta_i = \lambda_i \quad \text{and} \quad \Lambda_i \sin \theta_i = \lambda_i$$

Eq. (19.14)

According to satisfy wavelength continuity at medium interface

$$\Lambda_i = \Lambda_t$$

Eq. (19.15)

wavelength relations at incident and transmitted mediums can be found
Substituting Eq. (19.12) into Eq. (19.16) to get a famous equation Snell’s law

\[
\frac{\lambda_i}{\sin \theta_i} = \frac{\lambda_t}{\sin \theta_t}
\]

Eq. (19.16)

Substituting Eq. (19.13) into Eq. (19.17) to get wavenumber form

\[
k_i \sin \theta_i = k_t \sin \theta_t
\]

Eq. (19.18)

Taking note as

\[
k_{iz} = k_{tz}
\]

Eq. (19.19)

The notation represents wavenumber \( z \) component in \( i \) (incident) and \( t \) (transmitted) medium.

### 19.3.2 Waves in Medium with Refraction Index

Wavelength will be compressed by effective index as

\[
\lambda_i = \frac{\lambda}{n_i}
\]

Eq. (19.20)

Wavenumber will be expanded as

\[
k_i = n_i k = n_i \frac{2\pi}{\lambda}
\]

Eq. (19.21)

Projected wavelength in medium \( n_i \) will be expanded as

\[
\frac{\lambda_i}{\sin \theta} = \Lambda_i = \frac{\lambda}{n_i \sin \theta}
\]

Eq. (19.22)
19.4 External and Internal Reflection

Two cases will be discussed for external and internal reflection shown as Figure 19.4 with following setup

1) Wave is uniform at $y$ direction as an inclined plane field which implies wave doesn’t propagate at $y$ direction and $\lambda \to \infty$

$$k_y = 0$$

Eq. (19.23)

2) Two mediums with

$$n_1 < n_2 \quad \rightarrow \quad \frac{n_1}{n_2} < 1$$

Eq. (19.24)

3) Obey Snell’s law

$$\sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1 \quad \Rightarrow \quad \theta_2 < \theta_1$$

Eq. (19.25)

4) The index notation for wavenumber in medium $i$ can be further defined as

$$\mathbf{k}_i = k_i \hat{e}_j = k_{i\nu} \hat{e}_x + k_{i\gamma} \hat{e}_y + k_{i\zeta} \hat{e}_z$$

and

$$\|\mathbf{k}_i\| = k_i > 0$$

Eq. (19.26)

where original definition without medium notation is

$$\mathbf{k} = k_i \hat{e}_i = k_x \hat{e}_x + k_y \hat{e}_y + k_z \hat{e}_z$$

Eq. (19.27)

In medium $n_1$

$$\mathbf{k}_1 = k_{1\nu} \hat{e}_x + k_{1\gamma} \hat{e}_y + k_{1\zeta} \hat{e}_z = k_{1x} \hat{e}_x + k_{1y} \hat{e}_y + k_{1z} \hat{e}_z$$

Eq. (19.28)

and

$$\|\mathbf{k}_1\| = k_1 = \sqrt{k_{1x}^2 + k_{1z}^2} > 0$$

Eq. (19.29)

In medium $n_2$

$$\mathbf{k}_2 = k_{2\nu} \hat{e}_x + k_{2\gamma} \hat{e}_y + k_{2\zeta} \hat{e}_z = k_{2x} \hat{e}_x + k_{2y} \hat{e}_y + k_{2z} \hat{e}_z$$

Eq. (19.30)
and

\[ \|k_2\| = k_2 = \sqrt{k_{2x}^2 + k_{2z}^2} > 0 \]  \hspace{1cm} \text{Eq. (19.31)}

**Case 1 – from smaller index to larger index**

Wave is transmitting into medium \( n_2 \) so we observe its properties in medium \( n_2 \).

From Snell’s law

\[ \sin \theta_2 = \frac{n_1}{n_2} \sin \theta_1 < 1 \quad \rightarrow \quad \theta_2 < \theta_1 < 90^\circ \]  \hspace{1cm} \text{Eq. (19.32)}

Recall wavenumber component in \( z \) direction

\[ k_{2z} = k_2 \sin \theta_2 < k_2 \quad \rightarrow \quad k_z > k_{2z} > 0 \]  \hspace{1cm} \text{Eq. (19.33)}

and \( x \) component is

\[ k_{2x}^2 = k_2^2 - k_{2z}^2 > 0 \]  \hspace{1cm} \text{Eq. (19.34)}

So both components are real then the common term for wave solution is

\[ \exp i\left(k_{2x}x + k_{2z}z - \omega t\right) \]  \hspace{1cm} \text{Eq. (19.35)}

which describes wave propagating in \( x \) and \( z \) direction.
Case 2 – from larger index to smaller index

Wave is transmitting into medium $n_1$ so we observe its properties in medium $n_1$.

If $\theta_2 < \theta_c$

$$\sin \theta_1 = \frac{n_2}{n_1} \sin \theta_2 < \frac{n_2}{n_1} \sin \theta_1 = 1 \quad \rightarrow \quad \sin \theta_1 < 1 \quad \rightarrow \quad \theta_1 < 90^\circ$$

Eq. (19.36)

Transmitted wave always exists as a propagating wave.

If $\theta_2 > \theta_c$

$$\sin \theta_1 = \frac{n_2}{n_1} \sin \theta_2 > \frac{n_2}{n_1} \sin \theta_1 = 1 \quad \rightarrow \quad \sin \theta_1 > 1$$

Eq. (19.37)

Recall wavenumber component in $z$ direction

$$k_{1z} = k_1 \sin \theta_1 > k_1 \quad \rightarrow \quad k_{1z} > k_1 > 0$$

Eq. (19.38)

and $x$ component is

$$k_{1x}^2 = k_1^2 - k_{1z}^2 < 0 \quad \rightarrow \quad k_{1x}^2 = - (k_{1z}^2 - k_1^2) < 0$$

Eq. (19.39)

We have

$$k_{1x} = \pm i \sqrt{k_{1z}^2 - k_1^2} = \pm i \alpha$$

Eq. (19.40)

where $\alpha$ is real positive and the common term for wave solution is

$$\exp(-\alpha x) \cdot \exp(i (k_{1z} z - \omega t))$$

Eq. (19.41)

which describes a wave propagating in $z$ and decaying in $x$ as an evanescent wave.
19.4.1 External Reflection

Typical reflection and transmission curves for external reflection. External reflection implies that the reflection is from an interface to a medium of more index of refraction such as from air to water. These curves are the graphical representation of the Fresnel equations.

Note that the reflected amplitude for the light polarized parallel to the incident plane is zero for a specific angle called the Brewster angle. The reflected light is then linearly polarized in a plane perpendicular to the incident plane. This polarization by reflection is exploited in numerous optical devices.

Its characteristics can be summarized as

1) Transmission always in phase
2) Negative amplitude is $\pi$ phase shift
3) Peculiar behavior at Brewster angle $\theta_p$

where

- $r_p$ is p-polarized reflected
- $t_p$ is p-polarized transmitted
- $r_s$ is s-polarized reflected
- $t_s$ is s-polarized transmitted
This angle is referred to as the Brewster angle and is such that at this angle of incidence the reflection coefficient is zero for TM / p-polarized wave. No such angle exists for the TE / s-polarized wave. Thus, if an unpolarized beam is incident at the Brewster angle, the reflected light will be linearly polarized, and only s-polarized wave left to reflect with incident angle which is also Brewster angle.

19.4.2 Brewster Angle

Brewster's angle (also known as the polarization angle) is an angle of incidence at which light with a particular polarization is perfectly transmitted through a transparent dielectric surface, with no reflection. When unpolarized light is incident at this angle, the light that is reflected from the surface is therefore perfectly polarized. This special angle of incidence is named after the Scottish physicist Sir David Brewster (1781–1868).

The Fresnel equations predict that light with the p polarization (electric field polarized in the same plane as the incident ray and the surface normal) will not be reflected if the angle of incidence is

$$\theta_p = \tan^{-1}\left(\frac{n_2}{n_1}\right)$$

where $n_i$ is the refractive index of the initial medium through which the light propagates (the "incident medium"), and $n_r$ is the index of the other medium. This equation is known as Brewster's law, and the angle defined by it is Brewster's angle.
Its characteristics can be summarized as

1) Exists for TM polarization only

\[ r_\parallel = \left( \frac{E_{\text{out}}}{E_{\text{in}}} \right)_{\parallel} = \frac{n_i \cos \theta_i - n_i \cos \theta_i}{n_i \cos \theta_i + n_i \cos \theta_i} = \frac{\tan(\theta_i - \theta_i)}{\tan(\theta_i + \theta_i)} \]  
Eq. (19.43)

2) Reflection for TM is zero when \( \theta_i \) is at Brewster angle

3) Most light sources (sun, light bulb, etc.) are unpolarized

4) Reflection is polarized (TE) due to Brewster effect

Polarized sunglasses use the principle of Brewster's angle to reduce glare from the sun reflecting off horizontal surfaces such as water or road. In a large range of angles around Brewster's angle, the reflection of p-polarized light is lower than s-polarized light.

Figure 19.7  
Incident at Brewster angle and partially polarized refraction light

Thus, if the sun is low in the sky, reflected light is mostly s-polarized. Polarizing sunglasses use a polarizing material such as Polaroid sheets to block horizontally polarized light, preferentially blocking reflections from horizontal surfaces. The effect is strongest with smooth surfaces such as water, but reflections from roads and the ground are also reduced.

Photographers use the same principle to remove reflections from water so that they can photograph objects beneath the surface. In this case, the polarizing filter camera attachment can be rotated to be at the correct angle shown as (Figure 19.8).
Photographs taken of a window with a camera polarizer filter rotated to two different angles. In the picture at left, the polarizer is aligned with the polarization angle of the window reflection. In the picture at right, the polarizer has been rotated 90° eliminating the heavily polarized reflected sunlight.

19.4.3 Internal Reflection

The illustration shows typical reflection curves for internal reflection. Internal reflection implies that the reflection is from an interface to a medium of lesser index of refraction such as from water to air. These curves are the graphical representation of the Fresnel equations.

Note that the reflected amplitude for the light polarized parallel to the incident plane is zero for a specific angle called the Brewster angle. The reflected light is then linearly polarized in a plane perpendicular to the incident plane. This polarization by reflection is exploited in numerous optical devices.
Another characteristic of internal reflection is that there is always an angle of incidence $\theta_c$ above which all light is reflected back into the medium. This phenomenon of total internal reflection has many practical applications in optics.

Its characteristics can be summarized as

1) Reflection amplitude depends on TE/TM
2) Reflection is 100% when $\theta_i > \text{critical angle } \theta_c$
3) Agrees with ray optics!
4) Brewster angle will be reached first and then go into TIR

where

- $r_p$ is p-polarized reflected
- $t_p$ is p-polarized transmitted
- $r_s$ is s-polarized reflected
- $t_s$ is s-polarized transmitted

![Figure 19.10 Coefficients for internal reflection](image)
19.4.4 Total Internal Reflection

In fact, reflection of light may occur whenever light travels from a medium of a given refractive index into a medium with a different refractive index. In the most general case, a certain fraction of the light is reflected from the interface, and the remainder is refracted.

Solving Maxwell's equations for a light ray striking a boundary allows the derivation of the *Fresnel equations, which can be used to predict how much of the light is reflected, and how much is refracted in a given situation. This is analogous to the way impedance mismatch in an electric circuit causes reflection of signals*. Total internal reflection of light from a denser medium occurs if the angle of incidence is greater than the critical angle.

When a wave is incident from a denser medium on an interface separating two media, then, since, \( n_i > n_1 \rightarrow n_i > n_2 \) we have Eq. (19.10) from Snell’s law \( \theta_2 > \theta_1 \).

For an angle of incidence \( \theta_1 = \theta_c \) such that

\[
\theta_1 \sin \theta_c = \theta_2
\]

and shown as below

![Figure 19.11 Refraction of light at the interface between two media.](image-url)

*When light reflects off a material denser (with higher refractive index) than the external medium, it undergoes a polarity inversion (external reflection). In contrast, a less dense, lower refractive index material will reflect light in phase (internal reflection). This is an important principle in the field of thin-film optics.*
20 Mode Theory and Ray Tracing

This chapter discussed how to find the eigenvalues and eigenfunctions for EM waves in free space and constrained in 2D boundaries. Fields transition in dielectric medium and at interface are also discussed.

20.1 1D String Vibration in Free Space

The deflection of counter propagating waves can be described as

\[ A(z, t) = A_1 \cos(kz + \omega t) + A_2 \cos(kz - \omega t) \]  
Eq. (20.1)

Boundary free and fixed end boundary conditions on both sides are shown as below

![Boundary conditions](image)

Figure 20.1 (a) Out-going (blue) and in-coming (red) waves; and (b) the formation of standing wave with two-sided boundary conditions

**At one side**

\[ A(z, t) \bigg|_{z=0} = 0 \]  
Eq. (20.2)

Then we can get

\[ A(z, t) = A_1 \cos \omega t + A_2 \cos \omega t = 0 \]  
Eq. (20.3)

Non-trivial solution for all \( t \)

\[ A_1 + A_2 = 0 \quad \rightarrow \quad A_1 = A_2 = A_o \]  
Eq. (20.4)

**At the other side**

\[ A(z, t) \bigg|_{z=d} = 0 \]  
Eq. (20.5)
Then we get

\[ A(z, t) = A_1 \cos(kd + \omega t) + A_2 \cos(kd - \omega t) \]
\[ = A_0 \left[ \cos(kd + \omega t) + \cos(kd - \omega t) \right] \]
\[ = -2A_0 \sin kd \sin \omega t = 0 \]

Non-trivial solution here is also an eigenfunction or characteristic function

\[ \sin kd \sin \omega t = 0 \rightarrow \sin kd = 0 \]

Eigenvalue \( k_m \) can be found as

\[ k_m = \frac{m\pi}{d} \]

where \( m \) is integer

Eigen modes are discrete solutions for motion equation which is wave equation with boundary conditions. An analogy example is quantum mechanics with discrete energy levels for atoms described in Schrodinger equation. One of the interesting phenomena can be observed that eigenfunction is geometric dependent and it remains the same for different kind of waves such as elastic waves, acoustic waves, EM waves, etc.
20.2 2D Fields at Dielectric Interface

In order to understand how does wave travel in free space, let’s get started with following assumptions

1) Plane wave in free space
2) TE / s-polarized
3) Having the same omitting oscillation term \( \exp(i \omega t) \)

**Fields in Single Boundary**

Get started from single boundary shows as below

Incident wave in medium \( n_1 \) and \( n_1 > n_2 \)

\[
E_i = E_i \hat{e}_x = E_{i0} \exp(i (k_i \cdot r)) \hat{e}_x = E_{i0} \exp(i (k_{i1x} x + k_{i1y} y + k_{i1z} z)) \hat{e}_x
\]

Eq. (20.9)

Reflection wave

\[
E_r = E_r \hat{e}_x = E_{r0} \exp(i (k_r \cdot r)) \hat{e}_x = E_{r0} \exp(i (k_{r1x} x + k_{r1y} y + k_{r1z} z)) \hat{e}_x
\]

Eq. (20.10)

In order to simplify notation \( k_i = k_i^0 \hat{j} = k_{i1} \hat{e}_x + k_{i2} \hat{e}_y + k_{i3} \hat{e}_z \)

\[
\begin{align*}
k_{i1x} &= -k_{r1x} = k_x & \rightarrow \text{perfect reflection} \\
k_{i1y} &= k_{ry} = 0 & \rightarrow \text{no leakage in } y \\
k_{i1z} &= k_{rz} = k_z & \rightarrow \text{propagating in } z
\end{align*}
\]

Eq. (20.11)

Only consider TIR
where reflection coefficient $|r_{TE}| = 1$ for TIR and $\phi_r$ is reflection phase angle

Because incident and reflection waves are s-polarized and oscillating toward the same direction so the total field for TE wave is just adding them together

$$E_{total} = E_i + E_r$$

$$= E_{io} \left[ \exp i k_x x + \exp i (\phi_r - k_x x) \right] \exp i k_z z$$

$$= E_{io} \left[ \exp i \left( k_x x - \frac{\phi_r}{2} \right) + \exp i \left( \frac{\phi_r}{2} - k_x x \right) \right] \exp i \left( k_z z + \frac{\phi_r}{2} \right)$$

where

- $\cos \left( k_x x - \frac{\phi_r}{2} \right)$ describes a standing wave along $x$ direction

- $\exp i \left( k_z z + \frac{\phi_r}{2} \right)$ describes a propagating wave along $z$ direction

Note, one side boundary still can form a standing wave when there’s an incident source
20.3 2D Fields in Two Parallel Boundaries

With two parallel boundaries the coordinate system is $d / 2$ below upper boundary shown as below

![Figure 20.3 Two parallel boundaries](image)

Translate $x$ coordinate $d / 2$ and the field becomes

$$E_{total} = 2E_{io} \cos \left( k_x \left( x - \frac{d}{2} \right) \frac{\phi_i}{2} \right) \exp \left( k_z z + \frac{\phi_r}{2} \right)$$  \hspace{1cm} \text{Eq. (20.14)}$$

Rearrange as

$$E_{total} = 2E_{io} \cos \left( k_x \left( x - \frac{d}{2} \right) \frac{\phi_i}{2} \right) \exp \left( k_z z + \frac{\phi_r}{2} \right)$$

$$= 2E_{io} \cos \left( k_x x - \frac{k_x d + \phi_r}{2} \right) \exp \left( k_z z + \frac{\phi_r}{2} \right)$$  \hspace{1cm} \text{Eq. (20.15)}$$

$$= 2E_{io} \left[ \cos \left( k_x x \right) \cos \left( \frac{k_x d + \phi_r}{2} \right) + \sin \left( k_x x \right) \sin \left( \frac{k_x d + \phi_r}{2} \right) \right] \exp \left( k_z z + \frac{\phi_r}{2} \right)$$

where

- $\cos \left( \frac{k_x d + \phi_r}{2} \right)$ is constant coefficient of cos for standing wave
- $\sin \left( \frac{k_x d + \phi_r}{2} \right)$ is constant coefficient of sin for standing wave

Geometry is symmetric so the solution must be symmetric as well.

For even symmetric field
For odd symmetric field

\[
\cos\left(\frac{k_x d + \phi}{2}\right) = 0 \quad \rightarrow \quad \frac{k_x d + \phi}{2} = \frac{\pi}{2} + m\pi \quad \rightarrow \quad k_x d + \phi = \pi + 2m\pi
\]

Unified the representation to contain all solutions and satisfy one condition at a time separately

\[k_x d + \phi = m\pi\]  

Eq. (20.18)

Note it's not a general solution which is the linear combination of even and odd symmetric field.

The eigenfunction for integer \(m\) is

\[k^m_x = \frac{m\pi - \phi}{d}\]  

Eq. (20.19)
20.4 2D Fields in Two Parallel Boundaries – Ray Tracing Approach

There’s another way to derive the eigenfunction by using the compatibility constrain or called ray tracing approach. Select a close path 1-2-3-4-5-1 and trace the phase variation along this loop.

![Wavefront geometry](image)

Figure 20.4  Wavefront geometry

Points and path are defined as:

- Path 1-2 is the same wavefront
- Path 1-4 is the same wavefront
- Point 1 is after reflection
- Point 3 is reflection point
- Point 5 is before reflection

According to the compatibility of phase

\[
0 + k_x d + \phi_r + k_x d + \phi_r = 2m\pi 
\]

We get

\[
0 + k_x d + \phi_r + k_x d + 0 + \phi_r = 2m\pi \quad \rightarrow \quad k_x d + \phi_r = m\pi \quad \rightarrow \quad k_x^m = \frac{m\pi - \phi_r}{d} \quad \text{Eq. (20.20)}
\]

So eigenfunction \( k_x^m \) for integer \( m \) is identical to the result derived from EM waves.
21 Planar Waveguides

In optical dielectrics, only EM waves are considered and the coupling to elastic waves is ignored.

![Figure 21.1 EM waves in planar waveguides](image)

An overview of how the solution goes and how the assumptions affect the solution. For a wave propagates in \( z \) direction in a planar waveguide, below are assumptions often made

1) TE wave: electrical field \( E \) has no \( z \) component \( \rightarrow E_z = 0 \)

2) TM wave: magnetic field \( H \) has no \( z \) component \( \rightarrow H_z = 0 \)

3) TEM wave: both fields have no \( z \) component \( \rightarrow E_z = H_z = 0 \)

4) Hybrid EH or HE wave \( \rightarrow E_z \neq 0, H_z \neq 0 \)

Before getting into planar waveguide, we firstly solve the uniform plane waves.

21.1 Uniform Plane Waves

We start with the general form of Maxwell’s equations in a macroscopic view

**Step 1 – constitutive equations**

\[
\begin{align*}
\nabla \cdot \mathbf{D} &= \rho_f \\
\nabla \cdot \mathbf{B} &= 0 \\
\n\nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t} \\
\n\nabla \times \mathbf{H} &= \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}
\end{align*}
\]

Eq. (21.2)
Constitutive relation

\[
\begin{align*}
D &= \varepsilon_0 E + P = \varepsilon_0 (1 + \chi_e) E = \varepsilon E \\
B &= \mu_0 (H + M) = \mu_0 (1 + \chi_m) H = \mu H
\end{align*}
\]

Eq. (21.3)

Three types of variables

\[
\begin{align*}
\{E, B\} &\quad \rightarrow \quad EM \, fields \\
\{P, M\} &\quad \rightarrow \quad properties \, of \, medium \\
\{\rho_f, J_f\} &\quad \rightarrow \quad " \, free" \, charge \, per \, unit \, volume / \, current \, density
\end{align*}
\]

Eq. (21.4)

Note, \(\rho_f\) and \(J_f\) are independent of medium

**Step 2 – condition assumptions**

Make assumptions to simplicity Maxwell’s equations

\[
\begin{align*}
\{no \, source\} &\quad \rightarrow \quad \rho_f = 0, \ J_f = 0 \\
\{linear \, medium\} &\quad \rightarrow \quad D = \varepsilon E \\
\{isotropic \, homogeneous\} &\quad \rightarrow \quad B = \mu H
\end{align*}
\]

Eq. (21.5)

where \(\varepsilon\) is electric permittivity, \(\mu\) is magnetic permeability and both are medium constants

With no source assumption Eq. (21.2) can be simplified as below for EM waves transmitted in any mediums such as vacuum, air, water, glass, etc.

\[
\begin{align*}
\nabla \cdot E &= 0 \\
\nabla \cdot H &= 0 \\
\nabla \times E &= -\mu \frac{\partial H}{\partial t} \\
\nabla \times H &= \varepsilon \frac{\partial E}{\partial t}
\end{align*}
\]

Eq. (21.6)

Recall vector calculus identity (vector Laplacian)
\[ \nabla \times (\nabla \times \mathbf{X}) = \nabla (\nabla \cdot \mathbf{X}) - \nabla^2 \mathbf{X} \quad \text{Eq. (21.7)} \]

Substitute electrical field \( \mathbf{E} \) into above and get consecutive terms as below

For left term

\[ \nabla \times (\nabla \times \mathbf{E}) = \nabla \times \left( -\mu \frac{\partial \mathbf{H}}{\partial t} \right) = -\mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad \text{Eq. (21.8)} \]

For right term

\[ \nabla (\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} = -\nabla^2 \mathbf{E} \quad \text{Eq. (21.9)} \]

Thus wave equations can be obtain for electrical field and so as magnetic field where wave equations are hard to solve

\[
\begin{cases}
\nabla^2 \mathbf{E} = \mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \\
\nabla^2 \mathbf{H} = \mu \varepsilon \frac{\partial^2 \mathbf{H}}{\partial t^2}
\end{cases} \quad \text{Eq. (21.10)}
\]

**Step 3 – fields description**

We define the operators here for later usage

\[ \mathbf{E} = E_x \hat{\mathbf{x}} + E_y \hat{\mathbf{y}} + E_z \hat{\mathbf{z}} \quad \text{Eq. (21.11)} \]

where \( \hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}} \) are unit vectors along in Cartesian coordinates then the expansion for each operators are shown as below

Curl for vector field

\[ \nabla \times \mathbf{E} = \left( \frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \hat{\mathbf{x}} + \left( \frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} \right) \hat{\mathbf{y}} + \left( \frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right) \hat{\mathbf{z}} \quad \text{Eq. (21.12)} \]

Divergence for vector field
\[ \mathbf{E} \cdot \nabla = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} \]  
Eq. (21.13)

Gradient for scalar field

\[ \nabla E = \frac{\partial E}{\partial x} \hat{\mathbf{x}} + \frac{\partial E}{\partial y} \hat{\mathbf{y}} + \frac{\partial E}{\partial z} \hat{\mathbf{z}} \]  
Eq. (21.14)

Laplace for scalar field

\[ \Delta E = \nabla^2 E = \nabla \cdot \nabla E = \frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} \]  
Eq. (21.15)

Vector Laplacian

\[ \nabla^2 \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla \times (\nabla \times \mathbf{E}) \]  
Eq. (21.16)

**Step 4 – uniform plane field assumption**

Assume the field is uniform in \( xy \) plane with constant wavefront phase which is plane wave assumption and propagates in \( z \) direction so that the field is function of \( (z,t) \) only

\[ E_i(x, y, z, t) \rightarrow E_i(z, t) \]  
Eq. (21.17)

this also implies \( E_i \) is time-varying then

\[ \mathbf{E}(z, t) = E_x(z, t) \hat{\mathbf{x}} + E_y(z, t) \hat{\mathbf{y}} + E_z(z, t) \hat{\mathbf{z}} \]  
Eq. (21.18)

\[ \mathbf{H}(z, t) = H_x(z, t) \hat{\mathbf{x}} + H_y(z, t) \hat{\mathbf{y}} + H_z(z, t) \hat{\mathbf{z}} \]  
Eq. (21.19)

Recall simplified Maxwell’s Eq. (21.6) for electric fields

\[ \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \]  
Eq. (21.20)

Substitute Eq. (21.18) into Eq. (21.20) we get

\[ -\frac{\partial E_y}{\partial z} \hat{\mathbf{x}} + \frac{\partial E_z}{\partial z} \hat{\mathbf{y}} = -\mu \left( \frac{\partial H_x}{\partial t} \hat{\mathbf{x}} + \frac{\partial H_y}{\partial t} \hat{\mathbf{y}} + \frac{\partial H_z}{\partial t} \hat{\mathbf{z}} \right) \]  
Eq. (21.21)
Then

\[
\begin{align*}
\frac{\partial E_y}{\partial z} &= \mu \frac{\partial H_x}{\partial t} \quad (1) \\
\frac{\partial E_x}{\partial z} &= -\mu \frac{\partial H_y}{\partial t} \quad (2) \\
0 &= \mu \frac{\partial H_z}{\partial t} \quad (3)
\end{align*}
\]

Recall simplified Maxwell’s Eq. (21.6) for magnetic fields

\[\nabla \times \mathbf{H} = \epsilon \frac{\partial \mathbf{E}}{\partial t} \quad \text{Eq. (21.23)}\]

Substitute Eq. (21.23) into Eq. (21.19) we get

\[\frac{\partial H_x}{\partial z} \hat{e}_x + \frac{\partial H_y}{\partial z} \hat{e}_y = \epsilon \left( \frac{\partial E_x}{\partial t} \hat{e}_x + \frac{\partial E_y}{\partial t} \hat{e}_y + \frac{\partial E_z}{\partial t} \hat{e}_z \right) \quad \text{Eq. (21.24)}\]

Then

\[
\begin{align*}
\frac{\partial H_y}{\partial z} &= -\epsilon \frac{\partial E_x}{\partial t} \quad (4) \\
\frac{\partial H_x}{\partial z} &= \epsilon \frac{\partial E_y}{\partial t} \quad (5) \\
0 &= \epsilon \frac{\partial E_z}{\partial t} \quad (6)
\end{align*}
\]

**Step 5 – guess the solutions for Maxwell’s equations**

From (3) and (6)

\[H_z = 0 \text{ and } E_z = 0 \quad \text{Eq. (21.26)}\]

We already know the field component \(E_i\) is independent to each other and so is \(H_j\). Separating above equations (1) to (6) into two groups so that two sets of independent solutions could be obtained.
The 1\textsuperscript{st} set of solutions for $E_x$, $H_y$ from equation (2) and (4)

$$
\begin{align*}
E_x & , 
H_y = 0 \\
E_y & = 0 , 
H_y \\
E_z & = 0 , 
H_z = 0
\end{align*}
$$

Eq. (21.27)

The 2\textsuperscript{nd} set of solutions for $E_y$, $H_x$ from equation (1) and (5)

$$
\begin{align*}
E_x & = 0 , 
H_x \\
E_y & , 
H_y = 0 \\
E_z & = 0 , 
H_z = 0
\end{align*}
$$

Eq. (21.28)

Both sets are solutions for equation (1) to (6) and only differ by a rotation angle of 90\textdegree.

**Step 6 – 1D wave equation**

So stands only $(E_x, H_y)$ and take derivative of (1) and substitute into (5)

$$
\frac{\partial^2 E_x}{\partial z^2} = \frac{\partial}{\partial t} \left( \frac{\partial E_x}{\partial z} \right) = -\mu \frac{\partial}{\partial z} \left( \frac{\partial H_y}{\partial t} \right) = -\mu \frac{\partial}{\partial t} \left( \frac{\partial H_y}{\partial z} \right) = \mu \varepsilon \frac{\partial^2 E_x}{\partial t^2}
$$

Eq. (21.29)

1D wave equation for $x$ component propagating at $z$ direction can be obtained

$$
\frac{\partial^2 E_x}{\partial z^2} = \mu \varepsilon \frac{\partial^2 E_x}{\partial t^2}
$$

Eq. (21.30)

So as $y$ component propagating at $z$ direction

$$
\frac{\partial^2 E_y}{\partial z^2} = \mu \varepsilon \frac{\partial^2 E_y}{\partial t^2}
$$

Eq. (21.31)

**Step 7 – solve 1D wave equation**

The one-dimensional wave equation can be solved by separation of variables using a trial solution

$$
E_x = E_x(z,t) = Z(z)\Pi(t)
$$

Eq. (21.32)
Substituting back into Eq. (21.30) and get

\[ Z^*(z)\Pi(t) = \mu \epsilon Z(z)\Pi^*(t) \]  
Eq. (21.33)

Assume above equation equal to a constant \( \omega \) and let \( 1/\mu \epsilon = c^2 \)

\[ \frac{1}{\mu \epsilon} \frac{Z''(z)}{Z(z)} = \frac{\Pi''(t)}{\Pi(t)} = -\omega^2 \]  
Eq. (21.34)

The general solution for a constant coefficient differential equation will be

\[ \frac{Z^*(z)}{Z(z)} = -\frac{\omega^2}{c^2} \rightarrow Z(z) = A \exp \pm i \left( \frac{\omega}{c} z \right) \]  
Eq. (21.35)

and

\[ \frac{\Pi''(t)}{\Pi(t)} = -\omega^2 \rightarrow \Pi(t) = B \exp \pm i \omega t \]  
Eq. (21.36)

Finally, the solution for electrical field is

\[ E_x = E_x(z,t) = Z(z)\Pi(t) = C \exp \left( \pm \frac{\omega}{c} z \pm \omega t \right) = C \exp \left( \pm k z \pm \omega t \right) \]  
Eq. (21.37)

where \( C \) is constant and \( k \) is wavenumber deriving from

\[ \frac{\omega}{c} = \frac{2\pi f}{\lambda f} = \frac{2\pi}{\lambda} = k \]  
Eq. (21.38)

So as for magnetic field \( H_y \) which is perpendicular to electrical field \( E_x \),

\[ H_y = H_y(z,t) = D \exp \left( \pm \frac{\omega}{c} z \pm \omega t \right) = D \exp \left( \pm k z \pm \omega t \right) \]  
Eq. (21.39)

The coefficient ratio can be found by substituting back to constitutive equation Eq. (21.22) or Eq. (21.25)

\[ \frac{C}{D} = -\mu \omega \frac{\omega}{k} = -\mu c = -\sqrt{\frac{\mu}{\epsilon}} \]  
Eq. (21.40)
21.2 Uniform Plane Waves – Summary

Before solving an arbitrary wave in free space without making any assumption let’s take a review of how to deal with Maxwell’s equations when making a plane wave assumption and understand the logic flow

Review of the uniform plane waves

1) Manipulate Maxwell’s equations → wave equations
\[ \nabla^2 \mathbf{E} = \mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \]
\[ \nabla^2 \mathbf{H} = \mu \varepsilon \frac{\partial^2 \mathbf{H}}{\partial t^2} \]

2) Define vector fields
\[ \mathbf{E} = E_x \hat{\mathbf{e}}_x + E_y \hat{\mathbf{e}}_y + E_z \hat{\mathbf{e}}_z \]
\[ \mathbf{H} = H_x \hat{\mathbf{e}}_x + H_y \hat{\mathbf{e}}_y + H_z \hat{\mathbf{e}}_z \]

3) Uniform plane wave assumption → \(xy\) uniform → \(E_x = E_i(z,t)\) & \(H_z = H_i(z,t)\)

4) TEM wave assumption → \(E_z = H_z = 0\)

5) Only \((E_x, H_y)\) and \((E_y, H_x)\) left → propagating toward \(z\) direction

6) Assume variables can be separated → \(E_x = E_i(z,t) = Z(z) \Pi(t)\)

7) Solve wave equation → \(\frac{1}{\mu \varepsilon} Z''(z) = \frac{\Pi''(t)}{\Pi(t)} = -\omega^2\)

8) Get a solution pair
\[ E_x = C \exp\left(\pm \frac{\omega}{c} z \pm \omega t\right) = C \exp\left(\pm k z \pm \omega t\right) \]
\[ H_y = D \exp\left(\pm \frac{\omega}{c} z \pm \omega t\right) = D \exp\left(\pm k z \pm \omega t\right) \]

9) Take a look the solution. \(C\) and \(D\) are constant and satisfy \(xy\) uniform assumption.

The general form of solutions

1) Assume fields oscillating with time → \[ E_i(x,y,z,t) = E_i(x,y,z) \exp \pm i\omega t \]
\[ H_i(x,y,z,t) = H_i(x,y,z) \exp \pm i\omega t \]

2) For arbitrary field there’s no further assumption
3) Rewrite Maxwell's equations

\[
\begin{align*}
\nabla^2 \mathbf{E} &= \mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \\
\nabla^2 \mathbf{H} &= \mu \epsilon \frac{\partial^2 \mathbf{H}}{\partial t^2}
\end{align*}
\]

\[
\begin{align*}
\nabla^2 \mathbf{E} &= -\mu \epsilon \omega^2 \mathbf{E} = -\frac{\omega^2}{c^2} \mathbf{E} \\
\nabla^2 \mathbf{H} &= -\mu \epsilon \omega^2 \mathbf{H} = -\frac{\omega^2}{c^2} \mathbf{H}
\end{align*}
\]

Wave equation in vector form is known as Helmholtz equation or eigenvalue equation and can be solved component by component separately.

\[
\begin{align*}
\nabla^2 \mathbf{E} + k^2 \mathbf{E} &= 0 \\
\nabla^2 \mathbf{H} + k^2 \mathbf{H} &= 0
\end{align*}
\]

4) No need to expand the vector Laplacian \( \nabla^2 \mathbf{X} = \nabla(\nabla \cdot \mathbf{X}) - \nabla \times (\nabla \times \mathbf{X}) \)

5) Wave equation degenerate from vector field into scalar field

\[
\begin{align*}
\nabla^2 E_i + k^2 E_i &= 0 \\
\nabla^2 H_i + k^2 H_i &= 0
\end{align*}
\]

6) Assume variables can be separated for each component \( E_i(x, y, z) = X(x)Y(y)Z(z) \)

Wave solutions will be different according to the assumption made at this stage.

7) Solve wave equations \( \frac{X''}{X} + \frac{Y''}{Y} + \frac{Z''}{Z} + k^2 = 0 \) for each component \( E_i \) and \( H_i \)

8) Guess the solution for each variable \( \rightarrow \frac{X''}{X} = -k^2 \) so that \( X(x) = \exp ik_x x \)

9) Solution for each component

\[
E_i(x, y, z) = X(x)Y(y)Z(z) = \exp i(k_x x + k_y y + k_z z) = \exp i(k \cdot \mathbf{r})
\]

10) Complete solution for each component \( \rightarrow E_i(x, y, z, t) = \exp i(k \cdot \mathbf{r} \pm \omega t) \)

With constant amplitude in space when ignoring oscillating term in space-time) traveling toward \( k \) direction and each component is like a plane wave because of the assumption of variable separation.

11) Complete representation for field is \( \mathbf{E} = E_i \hat{e}_i \) and \( \mathbf{H} = H_i \hat{e}_i \)

For arbitrary field complete equation should be solved however due to the complexity we solve each component independently and decouple to coordinate (should proof the solidity).
21.3 General Solutions

The first step is the most crucial one to identify types of fields and make assumptions for each component. Let’s start over again from wave equations with no source assumption and fields are harmonic oscillating with \( \exp(-i\omega t) \) showing that

\[
\begin{align*}
\nabla^2 \mathbf{E} &= \mu \varepsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \\
\nabla^2 \mathbf{H} &= \mu \varepsilon \frac{\partial^2 \mathbf{H}}{\partial t^2}
\end{align*}
\]

\[
\rightarrow \begin{align*}
\nabla^2 \mathbf{E} &= \omega^2 \mu \varepsilon \mathbf{E} \\
\nabla^2 \mathbf{H} &= \omega^2 \mu \varepsilon \mathbf{H}
\end{align*}
\]

where \( \mu \varepsilon = 1/c^2 \)

Solve each component separately from the vector field \( \mathbf{E} \) and \( \mathbf{H} \)

\[
\begin{align*}
\nabla^2 \mathbf{E} + k^2 \mathbf{E} &= 0 \\
\nabla^2 \mathbf{H} + k^2 \mathbf{H} &= 0
\end{align*}
\]

\[
\rightarrow \begin{align*}
\nabla^2 E_i + k^2 E_i &= 0 \\
\nabla^2 H_i + k^2 H_i &= 0
\end{align*}
\]

Eq. (21.1)

So far, we only know there are 6 equations for \( i = x, y, z \) and 6 unknowns \( E_x, E_y, E_z, H_x, H_y, H_z \) that are not all independent. Apparently, we cannot make an independent assumption without a good reason.

Here brings a question, is there a way to solve a fraction of field components (ex. 2 out of 6) and derive the relations for the rest of field components (ex. 4 out of 6) so that the field equations would be more plausible to be solved under a certain number of independent fields? The idea here is to decompose the equations into transverse and longitudinal fields so that they are no longer coupled.

Selecting \( E_z, H_z \) as independent variables in simple medium showing that

\[
\begin{align*}
E_x, E_y, H_x, H_y = f(E_z, H_z) \\
\nabla^2 E_z + k^2 E_z &= 0 \\
\nabla^2 H_z + k^2 H_z &= 0
\end{align*}
\]

Eq. (21.3)

with tangential continuity

\[
\begin{align*}
\mathbf{n} \times (\mathbf{E}_2 - \mathbf{E}_1) &= 0 \\
\mathbf{n} \times (\mathbf{H}_2 - \mathbf{H}_1) &= 0
\end{align*}
\]

Eq. (21.4)
where we assume $E_x, E_y, H_x, H_y$ are functions of $(E_z, H_z)$ according to the selection of independent variables and $(E_z, H_z)$ are still functions of $(x, y, z)$. Note, they are not transverse wave so far.

In waveguide, wave propagates along $z$ in all regions and has the same wavenumber in order to satisfy compatibilities at the interface. Since the propagation constant is a complex quantity we can write: $\gamma = \alpha \pm i \beta$ when $\alpha$ the real part is called the attenuation constant and $\beta$ the imaginary part is called the phase constant. Assume all fields propagate in $z$ direction only and $\alpha = 0$

$$\exp{\gamma z} = \exp{(\alpha \pm i \beta)z} = \exp{\pm i \beta z} \quad \text{Eq. (21.5)}$$

So we can describe the wave fields in vector form under the assumption longitudinal and transverse waves can be decoupled

$$\left\{ \begin{array}{l} E = E_i \hat{e}_i = E_i^T(x, y) \exp{i \beta z \cdot \hat{e}_i} \\ H = H_i \hat{e}_i = H_i^T(x, y) \exp{i \beta z \cdot \hat{e}_i} \end{array} \right. \quad \text{Eq. (21.6)}$$

and for each component

$$\left\{ \begin{array}{l} E_i = E_i^T(x, y) \exp{i \beta z} \\ H_i = H_i^T(x, y) \exp{i \beta z} \end{array} \right. \quad \text{Eq. (21.7)}$$

where superscript $T$ denotes transverse for each component.

Substitute Eq. (21.10) into Eq. (21.6)

$$\nabla^2 E_z + k^2 E_z$$
$$= (\nabla_{T}^2 + \nabla_z^2) E_z^T(x, y) \exp{i \beta z} + k^2 E_z^T(x, y) \exp{i \beta z}$$
$$= (\nabla_{T}^2 - \beta^2) E_z^T(x, y) \exp{i \beta z} + k^2 E_z^T(x, y) \exp{i \beta z}$$
$$= \nabla_z^2 E_z^T(x, y) \exp{i \beta z} + (k^2 - \beta^2) E_z^T(x, y) \exp{i \beta z}$$
$$= \nabla_z^2 E_z^T(x, y) \exp{i \beta z} + q^2 E_z^T(x, y) \exp{i \beta z} = 0 \quad \text{Eq. (21.8)}$$

where $\nabla = \nabla_{T}^2 + \nabla_z^2$ and $\nabla_{T}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ and $\nabla_z = \frac{\partial^2}{\partial z^2}$

Then we have transverse wave equation for electrical field $E_z$
\[ \nabla^2 E_z^T(x, y) + q^2 E_z^T(x, y) = 0 \]  
\text{Eq. (21.9)}

So as for magnetic field \( H_z \)

\[ \nabla^2 H_z^T(x, y) + q^2 H_z^T(x, y) = 0 \]  
\text{Eq. (21.10)}

Because all the components are function of \( xy \) exclude the propagation \( z \) so they also can be function of \( E_z, H_z \) and separation with time as Ansatz.

Then the function of \( x \) degenerates into a 1D problem and Eq. (21.6) can be written as

\[
\begin{align*}
E_x^T, E_y^T, H_x^T, H_y^T &= f(E_z^T, H_z^T) \\
\nabla^2 E_x^T + q^2 E_x^T &= 0 \\
\nabla^2 H_x^T + q^2 H_x^T &= 0
\end{align*}
\text{Eq. (21.11)}
\]

where \( q^2 = k^2 - \beta^2 \), \( k \) is wavefront wavenumber, \( \beta \) is longitudinal wavenumber and \( q \) is transverse wavenumber. So far we have a further representation of \( E_z, H_z \) and it’s time to deal with \( E_x, E_y, H_x, H_y \).

Recall Maxwell’s Eq. (33.5) and expand the curl of a vector field by Eq. (33.11) with oscillating \( \exp(-i\omega t) \) and propagating \( \exp(i\beta z) \) then we get a series of equations

\[
\begin{align*}
\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} &= -\mu \frac{\partial H_z}{\partial t} \\
\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} &= -\mu \frac{\partial H_x}{\partial t} + i\beta E_z \frac{\partial E_y}{\partial x} = i\omega \mu H_x^T \\
\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} &= -\mu \frac{\partial H_y}{\partial t} \\
\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z} &= -\mu \frac{\partial H_z}{\partial t} + \frac{\partial E_y}{\partial x} = i\omega \mu H_y^T \\
\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} &= \varepsilon \frac{\partial E_z}{\partial t} \\
\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} &= \varepsilon \frac{\partial E_x}{\partial t} + i\beta H_z \frac{\partial E_x}{\partial y} = -i\omega \varepsilon E_x^T \\
\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} &= \varepsilon \frac{\partial E_y}{\partial t} \\
\frac{\partial H_z}{\partial x} - \frac{\partial H_x}{\partial z} &= \varepsilon \frac{\partial E_y}{\partial t} + i\beta H_x \frac{\partial E_y}{\partial y} = -i\omega \varepsilon E_y^T
\end{align*}
\text{Eq. (21.12)}
\]
Then we can express \( E_x, E_y, H_x, H_y \) as functions of \( E_z, H_z \)

From (1) to (5) solve \( H_x - E_y \)

\[
H_x^T = \frac{i}{q^2} \left( \beta \frac{\partial H_y^T}{\partial x} - \omega\epsilon \frac{\partial E_x^T}{\partial y} \right) \quad \text{Eq. (21.13)}
\]

\[
E_y^T = \frac{i}{q^2} \left( \beta \frac{\partial E_y^T}{\partial y} - \omega\mu \frac{\partial H_x^T}{\partial x} \right) \quad \text{Eq. (21.14)}
\]

From (2) to (4) solve \( E_x - H_y \)

\[
E_x^T = \frac{i}{q^2} \left( \beta \frac{\partial E_x^T}{\partial x} + \omega\mu \frac{\partial H_y^T}{\partial y} \right) \quad \text{Eq. (21.15)}
\]

\[
H_y^T = \frac{i}{q^2} \left( \beta \frac{\partial H_y^T}{\partial y} + \omega\epsilon \frac{\partial E_x^T}{\partial x} \right) \quad \text{Eq. (21.16)}
\]

(3) describes the relation of \( E_x - E_y \)

\[
\frac{\partial E_x^T}{\partial x} - \frac{\partial E_y^T}{\partial y} = i\omega\mu H_z^T \quad \text{Eq. (21.17)}
\]

(6) describes the relation of \( H_x - H_y \)

\[
\frac{\partial H_x^T}{\partial x} - \frac{\partial H_y^T}{\partial y} = -i\omega\epsilon E_z^T \quad \text{Eq. (21.18)}
\]

In summary for \( z \) propagating wave decoupled into longitudinal and transverse components
\[ E^T_z = \frac{i}{q^2} \left( \beta \frac{\partial E^T_z}{\partial x} + \omega \mu \frac{\partial H^T_z}{\partial y} \right) \]  \tag{1} \\
\[ E^T_y = \frac{i}{q^2} \left( \beta \frac{\partial E^T_z}{\partial y} - \omega \mu \frac{\partial H^T_z}{\partial x} \right) \]  \tag{2} \\
\[ \nabla^2 E^T_z + q^2 E^T_z = 0 \]  \tag{3} \\
\[ H^T_x = \frac{i}{q^2} \left( \beta \frac{\partial H^T_z}{\partial x} - \omega \epsilon \frac{\partial E^T_z}{\partial y} \right) \]  \tag{4} \\
\[ H^T_y = \frac{i}{q^2} \left( \beta \frac{\partial H^T_z}{\partial y} + \omega \epsilon \frac{\partial E^T_z}{\partial x} \right) \]  \tag{5} \\
\[ \nabla^2 H^T_z + q^2 H^T_z = 0 \]  \tag{6} 

Eq. (21.19)
21.4 Coupled TEM Waves

Observe $\exp(i(k \cdot r)) = \exp(i(qx + \beta z))$ for two types of waves

1) Propagate along $z$ then $\beta^2 > 0$ and $\beta > 0$ for outgoing wave

2) Evanescent along $x$ then $q^2 < 0$ $\rightarrow$ $q = \pm i\alpha$ for $\alpha > 0$

3) In each medium $n_\i$ must satisfy $k^2_\i = \beta^2 + q^2$

Wavenumbers in a three layers medium are shown as below

![Figure 21.2 Three layers of medium for EM waves](image)

Assume fields are uniform along $y$ in medium $n_1, n_2$ and region $I, II, III$

\[
\frac{\partial}{\partial y} = 0 \quad \text{Eq. (21.20)}
\]

**In region $I$**

\[
\beta > 0 \quad q_\i > 0 \rightarrow \beta^2 + q^2_\i = k^2_\i \rightarrow \beta < k_\i \quad \text{Eq. (21.21)}
\]

**In region $II$**

\[
\beta > 0 \quad q_\ii = i\alpha \rightarrow \beta^2 + q^2_\ii = \beta^2 - \alpha^2 = k^2_2 \rightarrow \beta > k_2 \quad \text{Eq. (21.22)}
\]

**In region $III$**

\[
\beta > 0 \quad q_\iii = -i\alpha \rightarrow \beta^2 + q^2_\iii = \beta^2 - \alpha^2 = k^2_2 \rightarrow \beta > k_2 \quad \text{Eq. (21.23)}
\]
so that we have

\[ k_z < \beta < k_i \rightarrow n_z < \frac{\beta}{k} < n_i \]  

Eq. (21.24)

where \( k \) is wavenumber in vacuum, in medium \( i \) denoted as \( k_i \).

Let’s apply condition Eq. (21.20) into Eq. (21.19) and can get two categories of equations

**TE modes**

\[ \frac{\partial}{\partial y} = 0 \rightarrow \begin{cases} E_x' & E_y' & E_z' \\ H_x' & H_y' & H_z' \end{cases} \]

\[ Eq. (21.25) \]

**TM modes**

\[ \frac{\partial}{\partial y} = 0 \rightarrow \begin{cases} E_x'' & E_y'' & E_z'' \\ H_x'' & H_y'' & H_z'' \end{cases} \]

\[ Eq. (21.26) \]

Decomposing the oscillating and propagating components into

\[ \text{decompose} \rightarrow E_i(x, y, z) = E_i^T(x, y) \exp i \beta z \]  

Eq. (21.27)

\[ \frac{\partial}{\partial y} = 0 \rightarrow E_i^T(x, y) = E_i^T(x) \]  

Eq. (21.28)

This assumption makes the above equations *Helmotz solvable.*
**TE modes → electrical field perpendicular to xz plane / incident plane**

\[
H^T_z = \frac{i}{q^2} \left( \beta \frac{\partial H^T_z}{\partial x} \right)
\]

Eq. (21.29)

\[
E^T_y = -\frac{i}{q^2} \left( \omega \mu \frac{\partial H^T_z}{\partial x} \right)
\]

Eq. (21.30)

\[
\frac{\partial^2}{\partial x^2} H^T_z (x) + q^2 H^T_z (x) = 0
\]

Eq. (21.31)

\[
E_z = 0
\]

Eq. (21.32)

where \( H_x, E_y \) are only dependent on \( H_z \)

**TM modes → electrical field parallel to xz plane / incident plane**

\[
E^T_x = \frac{i}{q^2} \left( \beta \frac{\partial E^T_x}{\partial x} \right)
\]

Eq. (21.33)

\[
H^T_y = \frac{i}{q^2} \left( \omega \epsilon \frac{\partial E^T_x}{\partial x} \right)
\]

Eq. (21.34)

\[
\frac{\partial^2}{\partial x^2} E^T_z (x) + q^2 E^T_z (x) = 0
\]

Eq. (21.35)

\[
H_z = 0
\]

Eq. (21.36)

where \( E_x, H_y \) are only dependent on \( E_z \)

In summary, each mode (TE or TM) forms independent solutions
21.5 Decoupled TE and TM Waves

Two sets of equations for TE and TM modes here we select TE mode to solve $H_x,E_y,H_z$ where TM modes will have the same results as TE modes. TEM is just the combination of TE and TM.

Continue to solve above TEM modes in the following steps.

**Step 1. Apply transverse wavenumber for each region with independent coefficient.**

**In region I**

For $-\frac{d}{2} \leq x \leq \frac{d}{2}$ where $q_I > 0$

Solution for Eq. (21.31) at region I is

$$H^T_{z,I} = A_I \exp iq_Ix + B_I \exp -iq_Ix$$

Eq. (21.38)

Then

$$H^T_{x,I} = -\frac{\beta}{q_I} (A_I \exp iq_Ix - B_I \exp -iq_Ix)$$

Eq. (21.39)

and

$$E^T_{y,I} = \frac{\epsilon_\mu}{q_I} (A_I \exp iq_Ix - B_I \exp -iq_Ix)$$

Eq. (21.40)

**In region II**

For $x > \frac{d}{2}$ where $q_{II} = i\alpha$

Solution for Eq. (21.31) at region II is

$$H^T_{z,II} = A_{II} \exp -\alpha x + B_{II} \exp \alpha x$$

Eq. (21.42)

then
\[ H^T_{x,II} = i \frac{B}{\alpha} \left( A_{II} \exp(\alpha x - B_{II} \exp(\alpha x)) \right) \quad \text{Eq. (21.43)} \]

and

\[ E^T_{y,II} = -i \frac{\omega \mu}{\alpha} \left( A_{II} \exp(-\alpha x - B_{II} \exp(-\alpha x)) \right) \quad \text{Eq. (21.44)} \]

**In region III**

For \( x < -\frac{d}{2} \) where \( q_{III} = -i \alpha \)

Eq. (21.45)

Solution for Eq. (21.31) at region III is

\[ H^T_{z,III} = A_{III} \exp(-\alpha x) + B_{III} \exp(-\alpha x) \quad \text{Eq. (21.46)} \]

then

\[ H^T_{x,III} = i \frac{B}{\alpha} \left( A_{III} \exp(-\alpha x) - B_{III} \exp(-\alpha x) \right) \quad \text{Eq. (21.47)} \]

and

\[ E^T_{y,III} = -i \frac{\omega \mu}{\alpha} \left( A_{III} \exp(-\alpha x) - B_{III} \exp(-\alpha x) \right) \quad \text{Eq. (21.48)} \]

Note, \( \exp(\alpha x) \) for \( x > 0 \) and \( \exp(-\alpha x) \) for \( x < 0 \) are not physically possible because of finite energy so \( B_{II} = A_{III} = 0 \)

Then we still have 5 variables \( A_I, B_I, A_{II}, B_{III}, \beta \)

**Step 2. Apply boundary conditions**

Let’s take a look of the fields \( H_x, E_y, H_z \) which can be categorized into

1) \( H_x \) normal component

2) \( E_y, H_z \) tangent component
In order to satisfy tangential continuity boundary conditions

**At region I-II boundary**

\[ E_{y,I} = E_{y,II} \]
\[ H_{z,I} = H_{z,II} \quad \text{at} \quad x = d/2 \]  

Eq. (21.49)

We have

\[ \frac{\omega \mu}{q_I} (A_I \exp i q_I x - B_I \exp -i q_I x) = \frac{i \omega \mu}{\alpha} \left( A_{II} \exp -\alpha x - B_{II} \exp \alpha x \right) \]  

Eq. (21.50)

\[ A_I \exp i q_I x + B_I \exp -i q_I x = A_{II} \exp -\alpha x + B_{II} \exp \alpha x \]  

Eq. (21.51)

At \( x = d/2 \)

\[ A_I \exp i \left( \frac{q_I d}{2} \right) - B_I \exp -i \left( \frac{q_I d}{2} \right) + A_{II} \frac{i q_I}{\alpha} \exp \left( -\frac{\alpha d}{2} \right) + 0 = 0 \]  

Eq. (21.52)

\[ A_I \exp i \left( \frac{q_I d}{2} \right) + B_I \exp -i \left( \frac{q_I d}{2} \right) - A_{II} \exp \left( -\frac{\alpha d}{2} \right) + 0 = 0 \]  

At region I-III boundary

\[ E_{y,I} = E_{y,III} \]
\[ H_{z,I} = H_{z,III} \quad \text{at} \quad x = -d/2 \]  

Eq. (21.53)

We have

\[ \frac{\omega \mu}{q_I} (A_I \exp i q_I x - B_I \exp -i q_I x) = \frac{i \omega \mu}{\alpha} \left( A_{III} \exp -\alpha x - B_{III} \exp \alpha x \right) \]  

Eq. (21.54)

\[ A_I \exp i q_I x + B_I \exp -i q_I x = A_{III} \exp -\alpha x + B_{III} \exp \alpha x \]  

Eq. (21.55)

At \( x = -d/2 \)

\[ A_I \exp -i \left( \frac{q_I d}{2} \right) - B_I \exp i \left( \frac{q_I d}{2} \right) + 0 - B_{III} \frac{i q_I}{\alpha} \exp \left( -\alpha d \right) = 0 \]  

Eq. (21.56)

\[ A_I \exp -i \left( \frac{q_I d}{2} \right) + B_I \exp i \left( \frac{q_I d}{2} \right) + 0 - B_{III} \exp \left( -\frac{\alpha d}{2} \right) = 0 \]  

226
**Step 3. Rearrange in matrix form**

Combine Eq. (21.52) and Eq. (21.56) and rearrange into matrix form as a set of linear equation

\[
\begin{bmatrix}
\exp i\left(\frac{q_I d}{2}\right) & -\exp i\left(\frac{q_I d}{2}\right) & i\frac{q_I}{\alpha} \exp\left(-\frac{\alpha d}{2}\right) & 0 \\
\exp i\left(\frac{q_I d}{2}\right) & \exp i\left(\frac{q_I d}{2}\right) & -\exp\left(-\frac{\alpha d}{2}\right) & 0 \\
-\exp i\left(\frac{q_I d}{2}\right) & -\exp i\left(\frac{q_I d}{2}\right) & 0 & -\frac{iq_I}{\alpha} \exp\left(-\frac{\alpha d}{2}\right) \\
-\exp i\left(\frac{q_I d}{2}\right) & \exp i\left(\frac{q_I d}{2}\right) & 0 & -\exp\left(-\frac{\alpha d}{2}\right)
\end{bmatrix}
\begin{bmatrix}
A_I \\
B_I \\
A_{II} \\
B_{III}
\end{bmatrix} = 0 \quad \text{Eq. (21.57)}
\]

Must have

\[
[M] \begin{bmatrix}
A_I \\
B_I \\
A_{II} \\
B_{III}
\end{bmatrix} = 0 \quad \Rightarrow \quad \text{Det} (M) = 0 \quad \text{Eq. (21.58)}
\]

After rearrangement

\[
\left(1 + \frac{iq_I}{\alpha}\right)^2 \exp iq_I d = \left(1 - \frac{iq_I}{\alpha}\right)^2 \exp -iq_I d \quad \text{Eq. (21.59)}
\]

Then we have

\[
(q_I - i\alpha)^2 \exp iq_I d = (q_I + i\alpha)^2 \exp -iq_I d \quad \text{Eq. (21.60)}
\]

*Then the characteristic equation can be found*

\[
\exp i(2q_I d) = \left(\frac{q_I + i\alpha}{q_I - i\alpha}\right)^2 \quad \text{Eq. (21.61)}
\]
Step 4. Recall the original wavenumber definitions

Take the notation with coordinate index for the wavenumbers. For longitudinal wavenumber with compatible condition along interface

$$\beta = k_{1z} = k_{2z} = k_{3z}$$  \hspace{1cm} \text{Eq. (21.62)}

For transverse wavenumber in each region

$$q_l = k_{lx}$$  \hspace{1cm} \text{Eq. (21.63)}

$$q_{ll} = k_{2x} = \sqrt{k_2^2 - \beta^2} = \sqrt{k_2^2 - k_{2x}^2} = i\alpha$$  \hspace{1cm} \text{Eq. (21.64)}

$$q_{lll} = k_{xx} = \sqrt{k_2^2 - \beta^2} = \sqrt{k_2^2 - k_{2x}^2} = -i\alpha$$  \hspace{1cm} \text{Eq. (21.65)}

Step 5. Eigenfunctions at interface

Then substitute above into wavenumbers we get

$$q_l = k_{lx} = k_i \cos \theta_i = n_i k_o \cos \theta_i$$  \hspace{1cm} \text{Eq. (21.66)}

and

$$\alpha = \left(\beta^2 - k_2^2\right)^{1/2} = \left(k_{1z}^2 - k_2^2\right)^{1/2} = \left(k_i^2 \sin \theta_i^2 - k_2^2\right)^{1/2} = k_o \left(n_i^2 \sin \theta_i^2 - n_2^2\right)^{1/2}$$  \hspace{1cm} \text{Eq. (21.67)}

The right-hand side of Eq. (21.61) can be written as

$$\frac{q_l + i\alpha}{q_l - i\alpha} = \frac{n_i k_o \cos \theta_i + i k_o \left(n_i^2 \sin \theta_i^2 - n_2^2\right)^{1/2}}{n_i k_o \cos \theta_i - k_o \left(n_i^2 \sin \theta_i^2 - n_2^2\right)^{1/2}} = \frac{n_i \cos \theta_i + n_2 \cos \theta_2}{n_i \cos \theta_i - n_2 \cos \theta_2}$$  \hspace{1cm} \text{Eq. (21.68)}

which is exactly the inverse of TE wave reflection coefficient so that

$$\left(\frac{q_l + i\alpha}{q_l - i\alpha}\right)^2 = \left(\frac{1}{n_1}\right)^2 = \left(\frac{1}{|r_{TE}| \exp i\phi_r}\right)^2 = \exp - i \left(2\phi_r\right)$$  \hspace{1cm} \text{Eq. (21.69)}

where \(|r_{TE}| = 1\) when TIR condition is matched
The characteristic Eq. (21.61) can be written as

\[ \exp i(2q_i d) = \exp -i(2\phi_r) \]  

Eq. (21.70)

So that

\[ 2q_i d = -2\phi_r + 2m\pi \rightarrow 2k_x d = -2\phi_r + 2m\pi \]  

Eq. (21.71)

The eigenfunction for integer \( m \) is

\[ k^{m}_x = \frac{m\pi - \phi_r}{d} \]  

Eq. (21.72)

which exactly the same as Eq. (20.19) with a unified representation for even and odd symmetric field along transverse direction at interface only.

**Step 6. Eigenfunctions in medium**

Let’s expand the Eq. (21.61) to satisfy entire domain not limited to interfaces

\[ \exp i(2q_i d) = \left( \frac{q_i + i\alpha}{q_i - i\alpha} \right)^2 \rightarrow \exp i(q_i d) = \pm \left( \frac{q_i + i\alpha}{q_i - i\alpha} \right) \]  

Eq. (21.73)

Remember we solve the fields at interface I-II and I-III and each provides a sets of equations. Here we would like to fide the relation for coefficient \( A_i \) and \( B_i \) by eliminating \( A_{II} \) at region I-II interface from Eq. (21.52) or eliminating \( B_{III} \) at region I-III interface from Eq. (21.56).

Each way can get identical results and here we select Eq. (21.52) and after rearrangement we have

Even function – symmetric modes

\[ \exp i(q_i d) = + \left( \frac{q_i + i\alpha}{q_i - i\alpha} \right) \rightarrow \frac{A_i}{B_i} = -1 \]  

Eq. (21.74)

Odd function – anti-symmetric modes
\[ \exp i(q_id) = -\frac{q_i + i\alpha}{q_i - i\alpha} \rightarrow \frac{A_i}{B_i} = +1 \]  

Eq. (21.75)

Even or odd mode can be identified by observing field Eq. (21.38) to Eq. (21.40) and the characteristic equation can be separated into two groups as below

**Even mode – symmetric fields**

Expand the Eq. (21.74) for even mode

\[ \exp i(q_id) = +\frac{q_i + i\alpha}{q_i - i\alpha} \rightarrow \begin{cases} q_i \cos q_id + \alpha \sin q_id = q_i \\ q_i \sin q_id - \alpha \cos q_id = \alpha \end{cases} \]  

Eq. (21.76)

Rearrange and get

\[ \cos q_id = \frac{1 - \left(\frac{\alpha}{q_i}\right)^2}{1 + \left(\frac{\alpha}{q_i}\right)^2} \rightarrow \tan \left(\frac{q_id}{2}\right) = \left(\frac{\alpha}{q_i}\right) \]  

Eq. (21.77)

So that eigenfunction for symmetric mode is

\[ q_i \tan \left(\frac{q_id}{2}\right) = \alpha \]  

Eq. (21.78)

**Odd mode – anti-symmetric fields**

Following the same procedure by expanding Eq. (21.75) for odd mode

\[ \exp i(q_id) = -\frac{q_i + i\alpha}{q_i - i\alpha} \rightarrow \begin{cases} q_i \cos q_id + \alpha \sin q_id = -q_i \\ q_i \sin q_id - \alpha \cos q_id = -\alpha \end{cases} \]  

Eq. (21.79)

Rearrange and get

\[ \sin q_id = \frac{-2\left(\frac{q_i}{\alpha}\right)}{1 + \left(\frac{q_i}{\alpha}\right)^2} \rightarrow \tan \left(\frac{q_id}{2}\right) = -\left(\frac{q_i}{\alpha}\right) \]  

Eq. (21.80)
So that eigenfunction for anti-symmetric mode is

$$-q_j \cot \left( \frac{q_j d}{2} \right) = \alpha$$  

Eq. (21.81)

**In summary**

The two equations can be combined as

$$q_j \tan \left( \frac{q_j d}{2} - \frac{m\pi}{2} \right) = \alpha$$  

Eq. (21.82)

where \( m \) is integer, even modes for \( m = 0, 2, 4, 6 \ldots \), and odd modes for \( m = 1, 3, 5 \ldots \)
21.6 Dispersion Curves

The eigenfunction Eq. (21.82) with normalized parameters shows that

\[
X \tan \left( X + \frac{m\pi}{2} \right) = Y \tag{Eq. (21.83)}
\]

where \( X = q_h \) and \( Y = \alpha h \)

Define a new normalized parameter \( V \) number as

\[
X^2 + Y^2 = \left( q_h^2 + \alpha^2 \right) h^2 = \left( k_1^2 - k_2^2 \right) h^2 = k^2 h^2 \left( n_1^2 - n_2^2 \right) = V^2 \tag{Eq. (21.84)}
\]

and showing that

\[
V = kh \sqrt{n_1^2 - n_2^2} = kh \cdot NA = \frac{\omega h}{c_p} \cdot NA \tag{Eq. (21.85)}
\]

Then we’re looking for the relation of \((V, \beta / k)\) shown as (Figure 21.3) where \( V \) is frequency or transverse wavenumber, \( \beta \) is longitudinal wavenumber, and \( k \) is intrinsic wavenumber in vacuum.

Figure 21.3 Dispersion curves of planar waveguides
21.7 Analogy of Elastic and EM Waves in Planar Waveguides

For comparison the mode theory between elastic wave in elastic solids and EM wave in dielectric materials, their frequency parameters are shown as (Figure 21.4). Lamb waves are comprised of $k_p$ (P-wave) and $k_{SV}$ (S-wave) where EM waves are guided by $k_1$ (region I as core), $k_2$ (region II as cladding) and $k_3$ (region III as cladding as well) in each layer shown as (Figure 21.5)

\[ K = kh \]
\[ \Omega = k_{SV}h = \omega h / c_T \]
\[ V = kh\sqrt{n_1^2 - n_2^2} = kh \cdot NA = \frac{\omega h}{c_p} \cdot NA \]

Figure 21.4 Lamb and EM waves analogy – frequency parameters

Figure 21.5 Lamb and EM waves analogy – wavevectors
Lamb convention

EM dispersion curves are rotated 90 degree so that the spatial and temporal frequency are aligned.

\[ K = kh \quad \Omega = k_{SV} h = \frac{\omega h}{c_T} \quad \beta / k \quad V = \frac{kh\sqrt{n_1^2 - n_2^2}}{n_2} = \frac{\omega h}{c_p} \cdot NA \]

Figure 21.6  Lamb and EM waves analogy – dispersion curves – Lamb convention

EM convention

Lamb dispersion curves are rotated 90 degree so that the spatial and temporal frequency are aligned.

\[ K = kh \quad \Omega = k_{SV} h = \frac{\omega h}{c_T} \quad \beta / k \quad V = \frac{kh\sqrt{n_1^2 - n_2^2}}{n_2} = \frac{\omega h}{c_p} \cdot NA \]

Figure 21.7  Lamb and EM waves analogy – dispersion curves – EM convention
22 Cylindrical Waveguides

Recall the dispersion relation in planar waveguides shown as (Figure 22.1) so that cylindrical waveguides can follow the same logic looking for the relation of normalized parameters of \((X, Y)\) and \((V, \beta / k)\).

### Dispersion relation for TE sym & anti-sym modes

\[
X \tan \left( X + \frac{m \pi}{2} \right) = Y
\]

- TM mode only
- Weakly guided is not included
- Only describes transverse wavenumber relation
- For \(m = 0, 2, 4, 6, \ldots\)
- Have to relate to longitudinal wavenumber!!

### Define V number

\[X^2 + Y^2 = \left( q_i^2 + \alpha^2 \right) h^2 = \left( k_i^2 - k_s^2 \right) h^2 = k_s^2 h^2 \left( n_1^2 - n_2^2 \right) = V^2\]

so that

\[V = kh \sqrt{n_1^2 - n_2^2} = kh \cdot NA = \frac{oh}{c_p} \cdot NA\]

We’re looking for \((V, \beta)\) where \(V\) transverse \(\beta\) longitudinal

Figure 22.1 Dispersion relation of EM waves in planar waveguides

Then the dispersion relation for all kinds of modes (TE, TM, HE, EH) for cylindrical waveguides is shown as (Figure 22.2). Note, LP modes in not shown yet.

### Dispersion relation for all modes (TE, TM, HE, EH)

\[
\left\{ k_1^2 \tilde{J}_m(X) + k_2^2 \tilde{K}_m(Y) \right\} \cdot \left\{ \tilde{J}_m(X) + \tilde{K}_m(Y) \right\} - m^2 \beta^2 \left( \frac{1}{X^2} + \frac{1}{Y^2} \right)^2 = 0
\]

where

\(J_m\) Bessel 1\(^{st}\) kind

\(K_m\) Modified Bessel 2\(^{nd}\) kind

\(\tilde{J}_m(X) = \frac{J_m(X)}{X \cdot J_m(X)}\) Normalized Derivative of Bessel 1\(^{st}\) kind

\(\tilde{K}_m(Y) = \frac{K_m(Y)}{Y \cdot K_m(Y)}\) Normalized Derivative of Modified Bessel 2\(^{nd}\) kind

We’re looking for \((V, \beta)\) where \(V\) transverse \(\beta\) longitudinal

Figure 22.2 Dispersion relation of EM waves in cylindrical waveguides
22.1 Dispersion Curves and Energy Distributions

The dispersion curves for all kinds of modes (TE, TM, HE, EH) are shown as (Figure 22.3), LP modes are shown as (Figure 22.4), and overlay with linear polarized (LP) modes are shown as (Figure 22.5).

![Dispersion Relations (V, β/k)](image)

Figure 22.3 Dispersion curves of cylindrical waveguides – TE, TM, HE, EH

![Dispersion Relations (V, β/k)](image)

Figure 22.4 Dispersion curves of cylindrical waveguides – LP
Linear polarized modes can be found when the waves are weakly guided showing that \( n_1 \approx n_2 \) and \( k_1^2 \approx k_2^2 \approx \beta^2 \). The eigenfunction in (Figure 22.2) will be simplified to

\[
\frac{X \cdot J_{m-1}(X)}{J_m(X)} = -\frac{Y \cdot K_{m-1}(Y)}{K_m(Y)}
\]

Eq. (22.2)

Then the linear polarized modes and their mode shapes can be found shown as below

Figure 22.6 LP modes and their mode shapes
VI. Conclusions and Future Works

Conclusions, future works, and participated projects are summarized in the following chapters.

23 Conclusions.................................................................................................................................. 239
24 Future Works .................................................................................................................................. 240
25 Publications .................................................................................................................................. 241
26 Participated Project Summary ......................................................................................................... 242
27 Declaration of Conflicting Interests ............................................................................................... 253
Conclusions

In this article, a full-field optical non-destructive inspection (NDI) system based on laser speckle interferometry (LSI) demonstrated with LSP/SG processed with advanced imaging processing algorithms for real-time large-area inspection was developed. With multiple damages visualization capability under excitation such as thermal stressing and steady-state vibration in ultrasonic range. The system is capable of visualizing multiple damages under external disturbances and is sensitive to strain related quantities.

Laser speckle itself is the result of scattered wavefield self-interference from a rough surface and each speckle can be treated ideally as a sensing point from a randomly distributed speckle cloud, has been mostly coined to speckle pattern. By observing the variation of speckle pattern on the structure surface, displacement related quantities in higher dimensions (e.g., hypercomplex envelope, phase between real-valued signal and its quadrature, phase congruency, etc.) can be deduced by Riesz transform with quadrature filters and correlated in time series to highlight the hidden damages in a highly effective way.

The proposed Riesz transform with log-Gabor bandpass filters can also be employed for retrieving the underlying phase distribution in time. Besides, by taking image in sequence with principal component analysis (PCA), damage features can be increasingly clear with high confidence, not limited to quasi-static or static problem but also applicable to transient dynamic problems. This breakthrough overcomes the low probabilities of detection (POD) for traditional LSP/SG.

With the aim of machine learning and advanced image processing algorithm in hypercomplex domain, more physical insights can be discovered and iteratively being smarter than traditional NDI methods, such as ultrasonic C-scan, pulse laser/LDV or even phase array system.
To conclude, a barely visible impact damages (BVID) on honeycomb composite panel was tested with thermal excitation and speckle variations are captured by the CMOS camera. The image sequence was interpreted with the Riesz bp transform and projected into certain directions by PCA. The images processed with Riesz bp showing a very good agreement with X-ray computed tomography (X-ray CT) and ultrasonic C-scan results.

Just like the analogy of mechanical mass-spring-damper (MKC) and electrical resister-inductor-capacitor (RLC) system. The elastic and EM waves in waveguides (as elastic medium and optical dielectrics) also have this kind of analogy.

The analogy of waves in elastic and dielectric waveguides are discussed. Lots of interesting properties that are shared in common including the characteristic equation, wavelength related dispersion, energy distribution, cut-off frequency, mode conversion, the transition from propagation to evanescent, etc. The interrogation of elastic and EM waves are also discussed to avoid spatial aliasing when the structure is under ultrasonic excitation.

This fundamental study helps us understand the physical insights, limitations and where are the upper and lower bounds. Also, the continuity if medium decides the continuity of the dispersion curves. For example, vibration (infinite numbers of discrete eigenvalues due to its finite bounded boundaries) is the special case of propagating waves (infinite numbers of continuous eigenvalues with infinite bounds) in continuous medium. To conclude, the physics is interesting with surprising analogies.

## 24 Future Works

Since the implementation of Riesz transform with log-Gabor bandpass filter supplants the phase shifting optical setup, yielding a compact, portable, and cost-effective system, having the potential to realize large-area real-time inspection on a portable device such as smart phone or other handheld devices.
25 Publications

Publications of journal and conference papers are listed below.

**Journal**


**Journal – to be submitted**


**Conference – invited keynote speaker**


**Conference**


**Conference – to be submitted**

26  **Participated Project Summary**

Participated projects in chronological order are listed below with posted summary in the following pages.

1)  2014 – A Preliminary Study of Laser Ultrasonic (Pulse Laser/LDV) Composite Inspection System for Damage Quantification

2)  2015 – Non-contact Technologies for Non-destructive Inspection (NDI) on Large Composite Structures | Poster

3)  2016 – Laser Ultrasonic Composite Inspection System (LUCIS) for Rapid Damage Qualitative Analysis and Quantification | Poster

4)  2017 – A Real-time System-wide Safety Assurance (SWS) System for Anomaly Detection, Diagnosis and Prognosis (ADP)

5)  2018 – A Robust Vectorial Random Search Root Finding Algorithm for Lamb Wave Dispersion Curves in Complex Domain

6)  2019 – An Analytical Interpretation of Lamb Wave Mode Up/Down Conversion and the Formation of Complex Roots


8)  2020 – An Advanced Phase-based Image Processing Technique via Riesz bp Transform in Hypercomplex Domain

9)  2021 – Real-time Inspection with Advanced Damage Processing Network (ADPNet) for Laser Speckle Interferometry (LSI)

10) 2021 – Barely Visible Impact Damage (BVID) Inspection on C-17 Globemaster III Composite Aileron using Laser Speckle Interferometry
2014 Project – An automated, contactless laser ultrasonic imaging system for barely visible impact damage (BVID) inspection for advanced composite structures has been developed. The current effort focuses on the application of the technology to inspect a honeycomb-core composite-facesheet panel with a BVID. Lamb waves are generated in the impacted panel by a Q-switched Nd:YAG laser, raster scanned by a set of galvano-mirrors over a two-dimensional grid over the damaged area of the panel surface. The out-of-plane velocities are measured through a laser Doppler vibrometer (LDV) that is stationary at a point on the corner of the grid. The ultrasonic wavefield of the scanned area is reconstructed and analyzed for high resolution characterization of impact damage in the composite honeycomb panel.
2015 Project – Laser Doppler vibrometer (LDV) is a key non-contact tool in measuring the surface vibration of a remote target. The vibrometer enables the measurement of wideband response (i.e., the vibration frequencies from DC up to 24 MHz and vibrational velocities of 0.01 μm/s up to 20 m/s). It is widely used to detect any material defects by measuring surface acoustic wave propagation generated by a pump laser demonstrated with pulse laser for wideband excitation. A defect in the material will disturb the propagating wavefields or accumulate the propagating waves around the damage location. By scanning the structure surface by the probe laser executed by laser Doppler vibrometer across the target surface, the defect locations and region can be identified.
2016 Project – The objective of the proposal is to develop a rapid laser ultrasonic composite inspection system (LUCIS) for damage qualitative analysis and quantification. The proposal will set out a three-year research effort to develop, demonstrate, and mature two innovation technologies in field applications including new hardware and software tools to be encompassed in LUCIS. These two new innovations will overcome current major obstacles for rapid large area detection and characterization of either manufacturing or in-service damage in composite structures in terms of sensitivity, resolution, and accuracy under the optical framework of heterodyne interferometry. For visualization, image conditions such as zero-lag cross-correlation (ZLCC) and the RMS of wavefields were applied to identify damage location and evaluate its effective regions.
2017 Project – The Thrust 5 Safety Program focuses on assuring the integrity of flight dynamics and control parameters to utilize all available information from diverse physical and virtual sensors in order to rapidly detect, isolate, and mitigate erroneous behavior within a sensor or sensor suite in real-time. The utilization of information fused across multiple sensors (physical and virtual) and algorithmic redundancy to estimate lost information from failed sensors. The safe operation of unmanned aircraft is essential to their acceptance and efficient use. Integrated Vehicle Health Management (IVHMs) is an important aspect of UAV operation and will enable improved Safety Assurance on such vehicles and can be achieved by comparing in-situ, real-time sensor flight data to prognostic models of components and subsystems to detect and mitigate faults as they occur.
2018 Project – This project aims to build a software tool with graphic user interface (GUI) for Lamb wave dispersion curves in complex domain. A robust vectorial root searching algorithm was applied to fetch complex roots without missing roots. Mode shapes with particle motions are also visualized to provide additional physical insights for students and engineers to work on guided wave problems. Lamb waves not only propagate in plate-like elastic solids but also leak into another medium as evanescent waves. The properties of propagation waves are described by real-valued eigenvalues (or real roots) where evanescent waves are characterized by pure imaginary roots. By observing the eigenvalues in complex domain, the formation, transition and conversion of Lamb wave modes can be easily understood.
2018 Project – A numerical study has been done for planar and cylindrical waveguides in elastic solids and dielectric materials. Before jumping into the coupling of elastic-electrical-magnetic interrogation problem, the analogy of Lamb waves and electrical-magnetic (EM) waves are made for a better understanding of how waves are guided in analytical geometries. For waves in planar waveguides, transcendental functions are solved with multi-start root finding algorithms to avoid local locking in planar waveguides where in cylindrical waveguides the functions turn out to be a set of Bessel functions which aims to take the advantages of this multi-start root finding algorithms even more. Mode shapes are also visualized with particle motion in elastic solids and energy distribution in optical dielectrics with physical insights.
2019 Project – Lamb waves not only propagate in elastic waveguides but also decay in space-time due to its intrinsic properties as evanescent waves. Mathematically, evanescent waves whose wavenumbers are pure imaginary have displacement fields decaying or increasing exponentially in space-time, but only the decay term physically makes sense due to energy conservation. This study aims to visualize the frequency-wavenumber in hypercomplex domain so that the wave formation and mode conversion can be observed in this 3D domain. Mode conversion is not only converting from one mode to another but also a complex procedure transferring/exchanging energy between different propagating and evanescent wave modes. The conversion process is simply the beauty of wave constitution.
2020 Project – This study aims to derive and visualize the relations of Riesz bandpass (Riesz bp) transform and its impulse response. Riesz bp transform is the 2D version Hilbert transform to evaluate the local phase (LP) of an input image. With the hypercomplex representation, the quadrature vector is generated via Riesz transform so that the monogenic signal which is a higher dimensional analytic signal can be formed. Compared to the Hilbert transform, the Riesz bp transform kernel includes a coupling term in its denominator so that the omnidirectional image feature can be fetched without rotating its kernel. Otherwise, Riesz bp transform generates multiple higher dimensional components as candidates for machine learning, feature extraction, and object detection applications.
2021 Project – An Advanced Damage Processing Network (ADPNet) was proposed with laser speckle interferometry (LSI) employed in a non-contact, full-field, vision-based non-destructive inspection (V-NDI) system for large area multiple hidden damages detection under thermal excitation. The LSI system comprises two sub-systems (1) laser speckle photography/photometry (LSP) and (2) shearography (SG). Both sub-systems rely on observing the variation of laser speckles in time sequence without reference arm and are very in-sensitive to ambient noise. Other advances of LSI/ADPNet system are its robust tolerance of laser coherence, larger illumination area, flexible choice of correlation functions, real-time processing on FPGA and more advanced post-processing techniques such as Bayesian updating/inference that can be readily applied.
**2021 Project** – Laser speckle photography/photometry (LSP) relies on observing the variation of laser speckles in time sequence is very sensitive to deformation for displacement field measurement. The robust tolerance of laser coherence, larger illumination area, and flexible choice of correlation functions also make LSP a compact and portable system. Direct phase estimation (DPE) via Riesz transform in hypercomplex domain aims to be the trend to move the phase estimation from optical setup to purely numerical processing. Without the phase stepping like traditional holography/ESPI, displacement fields can be observed simultaneously without repeating the experiment several times. This extends the applicability of this system not only suitable for steady-state but also for transient measurements in industrial environments.
27 Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
REFERENCES


[89] R. Lyons, “DSP Training DSP Class QUADRATURE SIGNALS: COMPLEX, BUT NOT COMPLICATED.”


[131] “Applications of Riesz Transforms and Monogenic Wavelet Frames in Imaging and Image Processing.”


[140] E. W. Weisstein, “Bessel Function of the First Kind”.


APPENDICES

Several useful materials are organized as appendices including Riesz transform and monogenic signal [85], [93], [122]–[132], color image processing [133], [134], Bessel functions for cylindrical waveguides [135]–[141], microscopic and macroscopic Maxwell equations [142]–[158], etc.

Most the equations are presented with 2D or 3D plots to explain the physical insights.
1D Analytic Signal and its Complex Representation

A real-valued bandpass signal (BPS) \( x(t) = A(t) \cos [2\pi f_s t + \phi(t)] \in \mathbb{R} \)

Note, real signal \( x(t) \) is in passband and its phase angle is defined in baseband

Hilbert transform \( \mathcal{H}(x(t)) \in \mathbb{R} \)

Quadrature \( \hat{x}(t) = \mathcal{H}(x(t)) \)

Analytic signal and its complex representation

\[
\begin{align*}
x_a(t) &= x(t) + j\hat{x}(t) = x(t) + j\mathcal{H}(x(t))
\end{align*}
\]

Amplitude

\[
A(t) = \|x_a(t)\| = \sqrt{x(t)^2 + \hat{x}(t)^2}
\]

Phase

\[
\phi(t) = \angle x_a(t) - 2\pi f_s t = \arg(x_a(t)) - 2\pi f_s t = \tan^{-1}\left(\frac{\hat{x}(t)}{x(t)}\right) - 2\pi f_s t
\]

Equivalent lowpass signal (ELPS) or phasor can be represented into a out-of-phase quadrature pair

\[
x_q(t) = x_a(t)e^{-j2\pi f_s t} = A(t)e^{j\phi(t)} = x_q(t) + jx_q(t)
\]

Amplitude measures the envelope of the signal

\[
A(t) = \|x_q(t)\| = \sqrt{x_q(t)^2 + \hat{x}_q(t)^2}
\]

Phase measures the shape of the signal where peaks at \( \phi(t) = 0 \) and troughs at \( \phi(t) = \pi \)

\[
\phi(t) = \angle x_q(t) = \arg(x_q(t)) = \tan^{-1}\left(\frac{x_q(t)}{x_q(t)}\right)
\]
At analytic signal point of view in Euler form

\[ x_a(t) = x(t) + j \hat{x}(t) = x_i(t) e^{j2\pi f_i t} = [x_i(t) + jx_q(t)] e^{j2\pi f_i t} \]

where \( x_i(t) \) is also called complex envelope

28.1 Quadrature Definition in Time Domain

Quadrature \( \hat{x}(t) \) of the analytic signal is generated by Hilbert transform which is defined by the convolution of \( 1/\pi t \)

\[ \mathcal{H}(x)(t) = \frac{1}{\pi t} \otimes x(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \]

Because \( 1/\pi t \) is not integrable, the convolution does not always converge. Instead, Hilbert transform is defined by using Cauchy principal value and denoted as

\[ \mathcal{H}(x)(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \]

This is precisely the convolution of \( x(t) \) with the tempered distribution \( \text{p.v.} 1/\pi t \) (due to Schwartz (1950); see Pandey (1996, Chapter 3)). Alternatively, by changing variables, the principal value integral can be written explicitly (Zygmund 1968, §XVI.1) as

\[ \mathcal{H}(x)(t) = -\frac{1}{\pi} \lim_{\varepsilon \rightarrow 0} \int_{\varepsilon}^{\infty} \frac{x(t+\tau) - x(t-\tau)}{\tau} d\tau \]

28.2 Quadrature Definition in Frequency Domain

The quadrature is easier to calculate in frequency domain by the properties of Fourier transform

\[ \mathcal{F}(\mathcal{H}(x(t)))(\omega) = \mathcal{F}\left(\frac{1}{\pi t}\right)(\omega) \cdot \mathcal{F}(x(t))(\omega) = -j \text{sgn}(\omega) \cdot \mathcal{F}(x(t))(\omega) \]

With simplified notation we implicitly know they are functions of \( \omega \)
\[ \mathcal{F}(\mathcal{H}(x(t))) = \mathcal{F}\left(\frac{1}{\pi t}\right) \cdot \mathcal{F}(x(t)) = -j \text{sgn}(\omega) \cdot \mathcal{F}(x(t)) \]

To deal with \( \omega = 0 \) we have

\[
\text{sgn}(\omega) = \begin{cases} 
1 & \omega > 0 \\
0 & \omega = 0 \\
-1 & \omega < 0 
\end{cases} \quad \rightarrow \quad \mathcal{F}(\mathcal{H}(x(t))) = \begin{cases} 
-j \mathcal{F}(x(t)) & \omega > 0 \\
0 & \omega = 0 \\
j \mathcal{F}(x(t)) & \omega < 0 
\end{cases}
\]

Further we can modify the sign function as a unity function

\[
\text{sgn}(\omega) = \begin{cases} 
\frac{\omega}{\|\omega\|} & \omega \neq 0 \\
0 & \omega = 0 
\end{cases} \quad \rightarrow \quad \mathcal{F}(\mathcal{H}(x(t))) = \begin{cases} 
-j \frac{\omega}{\|\omega\|} \cdot \mathcal{F}(x(t)) & \omega \neq 0 \\
0 & \omega = 0 
\end{cases}
\]

Inverse back to time domain then we get the quadrature

\[
\mathcal{F}^{-1}\left[\mathcal{F}(\mathcal{H}(x(t)))\right] = \mathcal{H}(x(t)) = \begin{cases} 
\mathcal{F}^{-1}\left[-j \frac{\omega}{\|\omega\|} \cdot \mathcal{F}(x(t))\right] & \omega \neq 0 \\
0 & \omega = 0 
\end{cases}
\]

### 28.3 Relation to Gradient

Recall Hilbert transform

\[
\mathcal{F}(\mathcal{H}(x(t))) = \begin{cases} 
-j \frac{\omega}{\|\omega\|} \cdot \mathcal{F}(x(t)) & \omega \neq 0 \\
0 & \omega = 0 
\end{cases}
\]

Compare to the derivative of Fourier transform

\[
\mathcal{F}'(x(t))(\omega) = j\omega \cdot \mathcal{F}(x(t))(\omega)
\]
29 2D Monogenic Signal and its Hypercomplex Representation

A real-valued image channel \( u(x, y) \in \mathbb{R} \) and \( \mathbf{x} = (x, y)^T \in \mathbb{R}^2 \)

Note, the scalar function \( u(x, y) \) can be treated as a 2D passband signal – one of image channels

Riesz transform \( \mathcal{R}(u)(x, y) \in \mathbb{R}^2 \)

Quadrature \( \hat{u}(x, y) = \mathcal{R}(u)(x, y) \)

Monogenic signal and its hypercomplex representation

\[
\begin{align*}
  u_m(x, y) &= u(x, y) + (i, j)^T \cdot \hat{u}(x, y) = u(x, y) + (i, j)^T \cdot \mathcal{R}(u)(x, y)
\end{align*}
\]

Amplitude – local energy (LE)

\[
A(x, y) = \|u_m(x, y)\| = \sqrt{u(x, y)^2 + \|\mathcal{R}(u)(x, y)\|^2}
\]

Phase – local phase (LP)

\[
\varphi(x, y) = \angle u_m(x, y) = \tan^{-1}\left(\frac{\|\mathcal{R}(u)(x, y)\|}{u(x, y)}\right)
\]

Orientation – local orientation (LO)

\[
\theta(x, y) = \angle \mathcal{R}(u)(x, y) = \tan^{-1}\left(\frac{R_s(u)(x, y)}{R_r(u)(x, y)}\right)
\]

Phase vector is defined to contain both phase and

\[
r(x, y) = \frac{\mathcal{R}(u)(x, y)}{\|\mathcal{R}(u)(x, y)\|} \cdot \angle u_m(x, y) = \frac{\mathcal{R}(u)(x, y)}{\|\mathcal{R}(u)(x, y)\|} \cdot \tan^{-1}\left(\frac{\|\mathcal{R}(u)(x, y)\|}{u(x, y)}\right)
\]
29.1 Definition of Multivariate Quadrature in Time Domain

Quadrature vector

$$\mathcal{R}(u)(x, y) = \frac{x}{2\pi \|x\|^3} \otimes u(x, y)$$

Each quadrature component

$$\begin{pmatrix} R_x(u)(x, y) \\ R_y(u)(x, y) \end{pmatrix} = \frac{(x, y)^T}{2\pi(x^2 + y^2)^{3/2}} \otimes u(x, y)$$

29.2 Definition of Multivariate Quadrature in Frequency Domain

Quadrature vector

$$\mathcal{F}(\mathcal{R}(u)(x, y)) = \begin{pmatrix} i \\ j \end{pmatrix} \circ \frac{k}{\|k\|} \cdot \mathcal{F}(u(x, y))$$

Each quadrature component

$$\begin{pmatrix} R_x(u)(x, y) \\ R_y(u)(x, y) \end{pmatrix} = \begin{pmatrix} i \\ j \end{pmatrix} \circ \frac{(k_x, k_y)^T}{(k_x^2 + k_y^2)^{1/2}} \cdot \mathcal{F}(u(x, y))$$

$\cdot$ is dot product, $\circ$ is element-wise multiplication, $\otimes$ is convolution

29.3 Multivariate Operators

*Quadrature operator* converts a real-valued signal into its quadrature.

$$-\begin{pmatrix} i \\ j \end{pmatrix} \circ \frac{k}{\|k\|} \xrightarrow{\text{impulse response}} \frac{x}{2\pi \|x\|^3} \xrightarrow{\text{frequency response}}$$

*Monogenic operator* converts a real-valued signal into a monogenic signal.

$$\begin{pmatrix} 1 + i^T \frac{k}{\|k\|} \end{pmatrix} \xrightarrow{\text{impulse response}} \left( \delta(x) + \begin{pmatrix} i \\ j \end{pmatrix}^T \right) \xrightarrow{\text{frequency response}} \left( \frac{x}{2\pi \|x\|^3} \right)$$
29.4 2D Space Monogenic Wavefield – Ultrasonic Bulk Wave

At every time instance

\[ u_m(x,y)(t) = u(x,y)(t) + \hat{u}(x,y)(t) = u(x,y)(t) + (i, j)^T \cdot \mathcal{R}(u)(x,y)(t) \]

Simplified as

\[ u_m(x,y) = u(x,y) + \hat{u}(x,y) = u(x,y) + (i, j)^T \cdot \mathcal{R}(u)(x,y) \]

Phase in space can be written as

\[ \varphi(x,y) = \mathbf{k} \cdot \mathbf{x} = k_x x + k_y y \]

Instantaneous wavenumber can be obtained by taking gradient of its phase

\[ (k_x, k_y) = \nabla \varphi(x,y) = \left( \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y} \right) \]

29.5 2D Space-Time Monogenic Wavefield – Lamb in Arbitrary Direction

At every time instance

\[ u_m(x,t) = u(x,t) + \hat{u}(x,t) = u(x,t) + (i, j)^T \cdot \mathcal{R}(u)(x,t) \]

Phase in space-time can be written as

\[ \varphi(x,t) = \mathbf{k} \cdot \mathbf{x} - \omega t = k_x x - \omega t \]

Then instantaneous wavenumber and frequency can be obtained by taking gradient of its phase

\[ (k_x, \omega) = \nabla \varphi(x,t) = \left( \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial t} \right) \]
29.6 Visualization of Quadrature Operator

Quadrature operator converts a real-valued signal into its quadrature.

\[
\begin{align*}
-\left( \begin{array}{c} i \\ j \end{array} \right) \frac{k}{\|k\|} & \quad \text{impulse response} & \frac{x}{2\pi \|x\|^3} \\
\end{align*}
\]

**Frequency response and phase**

<table>
<thead>
<tr>
<th>Quadrature operator</th>
<th>Absolute value</th>
<th>Phase angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-i \frac{k_x}{|k|})</td>
<td>(\text{abs} \left( -i \frac{k_x}{|k|} \right) )</td>
<td>(\text{angle} \left( -i \frac{k_x}{|k|} \right) )</td>
</tr>
<tr>
<td>(-j \frac{k_y}{|k|})</td>
<td>(\text{abs} \left( -j \frac{k_y}{|k|} \right) )</td>
<td>(\text{angle} \left( -j \frac{k_y}{|k|} \right) )</td>
</tr>
</tbody>
</table>

**Gray scale**

[Predesigned scaled images of frequency response and phase for different components]
Blue-white-red

Blue-white-red-3D
**Frequency response and impulse response**

*Quadrature operator* converts a real-valued signal into its quadrature.

\[
\begin{pmatrix}
  i \\
  j
\end{pmatrix} \cdot \frac{k}{\|k\|} \quad \xrightarrow{\text{impulse response}} \quad \frac{x}{2\pi \|x\|^3}
\]

<table>
<thead>
<tr>
<th>Quadrature operator</th>
<th>Complex quadrature operator</th>
<th>Complex impulse response</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \begin{pmatrix} -i \cdot \frac{k_x}{|k|} \ -j \cdot \frac{k_y}{|k|} \end{pmatrix} )</td>
<td>( i \begin{pmatrix} -i \cdot \frac{k_x}{|k|} \ -j \cdot \frac{k_y}{|k|} \end{pmatrix} )</td>
<td>( \mathcal{F}^{-1} \left( \frac{k_x}{|k|} \right) = 0 + i \frac{x}{2\pi |x|^3} )</td>
</tr>
<tr>
<td>( \begin{pmatrix} -j \cdot \frac{k_y}{|k|} \ j \cdot \frac{k_x}{|k|} \end{pmatrix} )</td>
<td>( j \begin{pmatrix} -j \cdot \frac{k_y}{|k|} \ j \cdot \frac{k_x}{|k|} \end{pmatrix} )</td>
<td>( \mathcal{F}^{-1} \left( \frac{k_y}{|k|} \right) = 0 + j \frac{y}{2\pi |x|^3} )</td>
</tr>
</tbody>
</table>

**Gray scale**
Blue-white-red

Blue-white-red-3D
### 29.7 Visualization of Monogenic Operator

*Monogenic operator* converts a real-valued signal into a monogenic signal.

**Frequency response and phase**

<table>
<thead>
<tr>
<th>Monogenic operator</th>
<th>Absolute value</th>
<th>Phase angle</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \left( 1 + \frac{k_x}{|k|} + \frac{k_y}{|k|} \right) )</td>
<td>( \text{abs} \left( 1 + \frac{k_x}{|k|} + \frac{k_y}{|k|} \right) )</td>
<td>( \text{angle} \left( 1 + \frac{k_x}{|k|} + \frac{k_y}{|k|} \right) )</td>
</tr>
</tbody>
</table>

---

**Gray scale**

**Blue-white-red**

**Blue-white-red-3D**
**Frequency response and impulse response**

<table>
<thead>
<tr>
<th>Monogenic operator</th>
<th>Impulse response – real</th>
<th>Impulse response – imaginary</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 + \frac{k_x}{|k|} + \frac{k_y}{|k|})</td>
<td>(\delta(x, y))</td>
<td>(i \frac{x}{2\pi |x|^2} + j \frac{y}{2\pi |x|^3})</td>
</tr>
</tbody>
</table>

---

**Gray scale**

- Monogenic filter – frequency response – real
- Monogenic filter – impulse response – real
- Monogenic filter – impulse response – imaginary

**Blue-white-red**

- Monogenic filter – frequency response – real
- Monogenic filter – impulse response – real
- Monogenic filter – impulse response – imaginary

**Blue-white-red-3D**

- Monogenic filter – frequency response – real
- Monogenic filter – impulse response – real
- Monogenic filter – impulse response – imaginary

---

278
29.8 Monogenic Filter and its Quadrature Pair

We start with a bandpass filter which is even symmetric to the origin and introduce 2D log-Gabor filter.

\[ G_{even}(k) = \exp \left( -\frac{1}{2} \log \left( \frac{k}{k_o} \right)^T \log \left( \Sigma_o \right)^{-1} \log \left( \Sigma_o^T \right)^{-1} \log \left( \frac{k}{k_o} \right) \right) \]

Its quadrature can be obtained by applying quadrature operator filter

\[ G_{odd-x}(k) = -i \frac{k_x}{\|k\|} \cdot G_{even}(k) \quad \text{convolution} \quad \text{multiplication} \quad g_{odd-x}(x) = \frac{x}{2\pi \|x\|^3} \otimes g_{even}(x) \]

\[ G_{odd-y}(k) = -j \frac{k_y}{\|k\|} \cdot G_{even}(k) \quad \text{convolution} \quad \text{multiplication} \quad g_{odd-y}(x) = \frac{y}{2\pi \|x\|^3} \otimes g_{even}(x) \]

Frequency response and impulse response – \( B = 1 \) and \( k_o = 0.2k_s \)

<table>
<thead>
<tr>
<th>Even filter</th>
<th>Complex filter odd-x</th>
<th>Complex filter odd-y</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G_{even}(k) )</td>
<td>( i \cdot G_{odd-x}(k) )</td>
<td>( j \cdot G_{odd-y}(k) )</td>
</tr>
<tr>
<td>( g_{even}(x) )</td>
<td>( i \cdot g_{odd-x}(x) )</td>
<td>( j \cdot g_{odd-y}(x) )</td>
</tr>
</tbody>
</table>

Gray scale

![Gray scale images]
Monogenic filter is a vector with hypercomplex numbers and its impulse response is complex as well.

\[
G_m(k) = G_{even}(k) + i \cdot G_{odd-x}(k) + j \cdot G_{odd-y}(k) = \left(1 + \frac{k_x}{||k||} + \frac{k_y}{||k||}\right) \cdot G_{even}(k)
\]

\[
g_m(x) = g_{even}(x) + i \cdot g_{odd-x}(x) + j \cdot g_{odd-y}(x) = \left(\delta(x,y) + i \frac{x}{2\pi ||x||^3} + j \frac{y}{2\pi ||x||^3}\right) \otimes g_{even}(x)
\]
Frequency response and impulse response – $B = 1$ and $k_o = 0.2k_s$

Complex monogenic filter

$$G_m(k) = G_{even}(k) + i \cdot G_{odd-x}(k) + j \cdot G_{odd-y}(k)$$

Impulse response – real

$$g_{even}(x)$$

Impulse response – imaginary

$$i \cdot g_{odd-x}(x) + j \cdot g_{odd-y}(x)$$
29.9 Complex Filter – Degenerated form of Monogenic Quadrature

Another way to form the filter representation is ignoring the monogenic representation and hypercomplex numbers, and just bringing everything back to complex number only. In this way, components can be examined one by one: 

\[ G_{even}(\mathbf{k}), \quad G_{odd-x}(\mathbf{k}) = -i \frac{k_x}{||\mathbf{k}||} \cdot G_{even}(\mathbf{k}), \quad G_{odd-y}(\mathbf{k}) = -i \frac{k_y}{||\mathbf{k}||} \cdot G_{even}(\mathbf{k}), \]

and we defined a complex odd filter as

\[ G_{co}(\mathbf{k}) = G_{odd-x}(\mathbf{k}) + i \cdot G_{odd-y}(\mathbf{k}) = \left( -i \frac{k_x}{||\mathbf{k}||} + \frac{k_y}{||\mathbf{k}||} \right) \cdot G_{even}(\mathbf{k}) \]

29.10 Summary of Logic Flow

The visualization of a log-Gabor bandpass filter with Riesz transform to generate a monogenic filter
30  3D Monogenic Signal and its Hypercomplex Representation – Quaternion

A real-valued scalar function \( u(\mathbf{x}) \in \mathbb{R} \) and \( \mathbf{x} = (x, y, z)^T \in \mathbb{R}^3 \)

Note, the scalar function \( u(\mathbf{x}) \) can be treated as a multivariate passband signal

Riesz transform \( \mathcal{R}(u)(\mathbf{x}) \in \mathbb{R}^3 \)

Quadrature \( \hat{u}(\mathbf{x}) = \mathcal{R}(u)(\mathbf{x}) \)

---

Monogenic signal and its hypercomplex representation

\[
u_m(\mathbf{x}) = u(\mathbf{x}) + (i, j, k)^T \cdot \hat{u}(\mathbf{x}) = u(\mathbf{x}) + (i, j, k)^T \cdot \mathcal{R}(u)(\mathbf{x})
\]

---

Amplitude – local energy (LE)

\[
A(\mathbf{x}) = \|u_m(\mathbf{x})\| = \sqrt{\|u(\mathbf{x})\|^2 + \|\mathcal{R}(u)(\mathbf{x})\|^2}
\]

Phase – local phase (LP)

\[
\varphi(\mathbf{x}) = \angle u_m(\mathbf{x}) = \tan^{-1}\left( \frac{\|\mathcal{R}(u)(\mathbf{x})\|}{\|u(\mathbf{x})\|} \right)
\]

Orientation \((i, j)\) of \( C^3_2 \) – local orientation (LO)

\[
\theta(i, j) = \angle \mathcal{R}(u)(\mathbf{x}) = \tan^{-1}\left( \frac{R_j(u)(\mathbf{x})}{R_i(u)(\mathbf{x})} \right)
\]

---

Phase vector is defined to contain both phase and

\[
\mathbf{r}(\mathbf{x}) = \frac{\mathcal{R}(u)(\mathbf{x})}{\|\mathcal{R}(u)(\mathbf{x})\|}, \quad \angle u_m(\mathbf{x}) = \frac{\mathcal{R}(u)(\mathbf{x})}{\|\mathcal{R}(u)(\mathbf{x})\|} \cdot \tan^{-1}\left( \frac{\|\mathcal{R}(u)(\mathbf{x})\|}{\|\mathcal{R}(u)(\mathbf{x})\|} \right)
\]
30.1 Multivariate Quadrature Definition in Time Domain

Quadrature vector

\[ \mathcal{R}(u)(x) = \frac{x}{\pi^2 \| x \|^4} \otimes u(x) \]

Each quadrature component

\[ \begin{cases} R_x(u)(x, y, z) \\ R_y(u)(x, y, z) \\ R_z(u)(x, y, z) \end{cases} = \frac{(x, y, z)^T}{\pi^2 (x^2 + y^2 + z^2)^2} \otimes u(x, y, z) \]

30.2 Multivariate Quadrature Definition in Frequency Domain

Quadrature vector

\[ \mathcal{F}(\mathcal{R}(u)(x)) = -(i, j, k)^T \odot \frac{k}{\| k \|} \cdot \mathcal{F}(u(x)) \]

Each quadrature component

\[ \mathcal{F} \begin{cases} R_x(u)(x, y) \\ R_y(u)(x, y) \\ R_z(u)(x, y) \end{cases} = -(i, j, k)^T \odot \frac{(k_x, k_y, k_z)^T}{(k_x^2 + k_y^2 + k_z^2)^{1/2}} \cdot \mathcal{F}(u(x, y, z)) \]

\( \cdot \) is dot product, \( \odot \) is element-wise multiplication, \( \otimes \) is convolution

30.3 Multivariate Operators

Quadrature operator converts a real-valued signal into its quadrature.

\[ -(i, j, k)^T \odot \frac{k}{\| k \|} \leftrightarrow \text{impulse response} \quad \frac{x}{\pi^2 \| x \|^4} \leftrightarrow \text{frequency response} \]

Monogenic operator converts a real-valued signal into a monogenic signal.

\[ \left( 1 + i^T \cdot \frac{k}{\| k \|} \right) \leftrightarrow \text{impulse response} \quad \frac{\delta(x) + (i, j, k) \cdot \frac{x}{\pi^2 \| x \|^4}}{\text{frequency response}} \]
30.4 3D Space Monogenic Wavefield – Ultrasonic Bulk Wave

At every time instance

\[ u_m(x, y, z)(t) = u(x, y, z)(t) + \hat{u}(x, y, z)(t) = u(x, y, z)(t) + (i, j, k) \cdot \mathcal{R}(u)(x, y, z)(t) \]

Simplified as

\[ u_m(x, y, z) = u(x, y, z) + \hat{u}(x, y, z) = u(x, y, z) + (i, j, k) \cdot \mathcal{R}(u)(x, y, z) \]

Phase in space can be written as

\[ \varphi(x, y, z) = k \cdot x = k_x x + k_y y + k_z z \]

Instantaneous wavenumber can be obtained by taking gradient of its phase

\[ (k_x, k_y, k_z) = \nabla \varphi(x, y, z) = \left( \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial z} \right) \]

30.5 3D Space-Time Monogenic Wavefield – Lamb in Arbitrary Direction

At every time instance

\[ u_m(x, y, t) = u(x, y, t) + \hat{u}(x, y, t) = u(x, y, t) + (i, j, k) \cdot \mathcal{R}(u)(x, y, t) \]

Phase in space-time can be written as

\[ \varphi(x, y, t) = k \cdot x - \omega t = k_x x + k_y y - \omega t \]

Then instantaneous wavenumber and frequency can be obtained by taking gradient of its phase

\[ (k_x, k_y, \omega) = \nabla \varphi(x, y, t) = \left( \frac{\partial \varphi}{\partial x}, \frac{\partial \varphi}{\partial y}, \frac{\partial \varphi}{\partial t} \right) \]
31 nD Monogenic Signal and its Hypercomplex Representation – Octonion

A real-valued scalar function $u(x) \in \mathbb{R}$ and $x = (x, y, z, \ldots)^T \in \mathbb{R}^d$

Note, the scalar function $u(x)$ can be treated as a multivariate passband signal

Riesz transform $\mathcal{R}(u)(x) \in \mathbb{R}^d$

Quadrature $\hat{u}(x) = \mathcal{R}(u)(x)$

Monogenic signal and its hypercomplex representation

$$u_m(x) = u(x) + (i, j, k, \ldots)^T \cdot \hat{u}(x) = u(x) + (i, j, k, \ldots)^T \cdot \mathcal{R}(u)(x)$$

Amplitude – local energy (LE)

$$A(x) = \|u_m(x)\| = \sqrt{u(x)^2 + \|\mathcal{R}(u)(x)\|^2}$$

Phase – local phase (LP)

$$\varphi(x) = \angle u_m(x) = \tan^{-1} \left( \frac{\|\mathcal{R}(u)(x)\|}{u(x)} \right)$$

Orientation $(i, j)$ of $C_2^d$ – local orientation (LO)

$$\theta(i, j) = \angle \rho \mathcal{R}(u)(x) = \tan^{-1} \left( \frac{R_j(u)(x)}{R_i(u)(x)} \right)$$

Phase vector is defined to contain both phase and

$$r(x) = \frac{\mathcal{R}(u)(x)}{\|\mathcal{R}(u)(x)\|}, \angle u_m(x) = \frac{\mathcal{R}(u)(x)}{\|\mathcal{R}(u)(x)\|} \cdot \tan^{-1} \left( \frac{\|\mathcal{R}(u)(x)\|}{u(x)} \right)$$

286
31.1 Multivariate Quadrature Definition in Time Domain

Singular integral

\[ R_j(u)(x) = c_d \lim_{\epsilon \to 0} \int_{\mathbb{R}^d \setminus B_\epsilon(x)} \frac{(x_j - \tau_j) u(x)}{||x - \tau||^{d+1}} d\tau \]

for \( d \geq 1 \) where \( \omega_{d-1} \) is volume unit \((d-1)\) ball

\[ c_d = \frac{1}{\pi \omega_{d-1}} = \frac{\Gamma[(d+1)/2]}{\pi^{(d+1)/2}} \]

31.2 Multivariate Quadrature Definition in Frequency Domain

Quadrature vector

\[ \mathcal{F}(\mathcal{R}(u)(x)) = -(i, j, k, \ldots)^T \circ \frac{k}{||k||} . \mathcal{F}(u(x)) \]

Each quadrature component

\[ \mathcal{F}\left( \begin{array}{c} R_x(u)(x,y,z, \ldots) \\ R_y(u)(x,y,z, \ldots) \\ R_z(u)(x,y,z, \ldots) \\ \vdots \end{array} \right) = -(i, j, k, \ldots)^T \circ \frac{(k_x,k_y,k_z, \ldots)^T}{(k_x^2+k_y^2+k_z^2+\ldots)^{1/2}} . \mathcal{F}(u(x,y,z, \ldots)) \]

\( \cdot \) is dot product, \( \circ \) is element-wise multiplication, \( \otimes \) is convolution

31.3 Multivariate Operators

**Quadrature operator** converts a real-valued signal into its quadrature.

\[ -(i, j, k, \ldots)^T \circ \frac{k}{||k||} \xrightarrow{\text{impulse response}} c_d \cdot \frac{x}{||x||^{d+1}} \]

**Monogenic operator** converts a real-valued signal into a monogenic signal.

\[ \left(1 + I^T \cdot \frac{k}{||k||}\right) \xrightarrow{\text{impulse response}} \left(\delta(x) + (i, j, k, \ldots) \cdot c_d \cdot \frac{x}{||x||^{d+1}}\right) \]
31.4 Relationship with Laplacian

- Function $f$ is a scalar
- Fourier transform $\mathcal{F}$ is a scalar
- Riesz transform $\mathcal{R}$ is a vector with index component $R_j$

The Riesz transforms of $f$ give the first partial derivatives of an imprecise solution of the equation

$$(-\Delta)^{1/2} u = f$$

where $\Delta = \nabla \cdot \nabla$ is Laplacian, Thus the Riesz transform of $f$ can be written as

$$\mathcal{R}f = \nabla (-\Delta)^{-1/2} f$$

Riesz operator is denoted as

$$\mathcal{R} = \nabla (-\Delta)^{-1/2} \quad \text{index form} \quad \leftarrow \quad \nabla \cdot \nabla \quad \text{vector form} \quad R_j = \frac{\partial}{\partial x_j} (-\Delta)^{-1/2}$$

In particular, one should also have

$$R_i R_j \Delta u = -\frac{\partial^2 u}{\partial x_i \partial x_j}$$

so that the Riesz transforms give a way of recovering information about the entire Hessian of a function from knowledge of only its Laplacian. This is now made more precise. Suppose that $u$ is a Schwartz function. Then indeed by the explicit form of the Fourier multiplier having

$$R_i R_j (\Delta u) = -\frac{\partial^2 u}{\partial x_i \partial x_j}$$

The identity is not generally true in the sense of distributions. For instance, if $u$ is a tempered distribution such that $\Delta u \in L^2(\mathbb{R}^d)$, then one can only conclude for some polynomial $P_{ij}$ showing that

$$\frac{\partial^2 u(x)}{\partial x_i \partial x_j} = -R_i R_j \Delta u(x) + P_{ij}(x)$$
### 31.5 Differential Operators

The generalized quadrature operators are summarized as below.

<table>
<thead>
<tr>
<th>Operator</th>
<th>$(-\Delta)^{-1/2}$</th>
<th>$R_j = -\frac{\partial}{\partial x_j}(-\Delta)^{-1/2}$</th>
<th>$(-\Delta)^{1/2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency response</td>
<td>$\frac{1}{</td>
<td></td>
<td>k</td>
</tr>
<tr>
<td>Impulse response</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d = 1$</td>
<td>$\log</td>
<td></td>
<td>x_j</td>
</tr>
<tr>
<td>$d = 2$</td>
<td>$\frac{1}{2\pi</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>$\frac{1}{2\pi^2</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>$d = 4$</td>
<td>$\frac{1}{4\pi^2</td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>
31.6 Jacobian and Hessian

The definition of Jacobian and Hessian matrix are shown as below.

<table>
<thead>
<tr>
<th>Jacobian</th>
<th>Hessian</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f : \mathbb{R}^n \rightarrow \mathbb{R}^m ) is a column vector</td>
<td>( f : \mathbb{R}^n \rightarrow \mathbb{R} ) is a scalar</td>
</tr>
<tr>
<td>( x \in \mathbb{R}^n ) and ( f(x) \in \mathbb{R}^m )</td>
<td>( x \in \mathbb{R}^n ) and ( f(x) \in \mathbb{R} )</td>
</tr>
</tbody>
</table>

\[
\mathbf{J}_f = \nabla f = \left( \begin{array}{c|c|c} \frac{\partial f_i}{\partial x_1} & \frac{\partial f_i}{\partial x_2} & \cdots & \frac{\partial f_i}{\partial x_n} \\ \hline \frac{\partial f_j}{\partial x_1} & \frac{\partial f_j}{\partial x_2} & \cdots & \frac{\partial f_j}{\partial x_n} \end{array} \right)_{m \times n} \quad \mathbf{H}_f = \left( \begin{array}{c|c|c} \frac{\partial^2 f}{\partial x_i \partial x_1} & \frac{\partial^2 f}{\partial x_i \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_i \partial x_n} \\ \frac{\partial^2 f}{\partial x_1 \partial x_j} & \frac{\partial^2 f}{\partial x_2 \partial x_j} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_j} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_i} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{array} \right)_{n \times n}
\]

The connection in between showing that *the Hessian matrix of a scalar function \( f \) is the Jacobian matrix of the gradient of the function*

\[
\mathbf{H}(f(x)) = \mathbf{J}(\nabla f(x))
\]
32 Bessel Functions

Bessel functions, first defined by the mathematician Daniel Bernoulli and then generalized by Friedrich Bessel, are the canonical solutions \( y(x) \) of the differential equation Eq. (32.1) known as Bessel's differential equation for an arbitrary complex number \( \alpha \), the order of the Bessel function. Although \( \alpha \) and \(-\alpha\) produce the same differential equation for real \( \alpha \), it is conventional to define different Bessel functions for these two values in such a way that the Bessel functions are mostly smooth functions of \( \alpha \).

\[
x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - \alpha^2)y = 0
\]

Eq. (32.1)

The most important cases are for \( \alpha \) an integer or half-integer. Bessel functions for integer \( \alpha \) are also known as cylinder functions or the cylindrical harmonics because they appear in the solution to Laplace's equation in cylindrical coordinates. Spherical Bessel functions with half-integer \( \alpha \) are obtained when the Helmholtz equation is solved in spherical coordinates.

32.1 Applications of Bessel functions

Bessel's equation arises when finding separable solutions to Laplace's equation and the Helmholtz equation in cylindrical or spherical coordinates. Bessel functions are therefore especially important for many problems of wave propagation and static potentials. In solving problems in cylindrical coordinate systems, one obtains Bessel functions of integer order \( (\alpha = n) \); in spherical problems, one obtains half-integer orders \( (\alpha = n + 1/2) \).

For example: electromagnetic waves in a cylindrical waveguide, pressure amplitudes of inviscid rotational flows, heat conduction in a cylindrical object, modes of vibration of a thin circular (or annular) acoustic membrane (such as a drum or other membranophone), diffusion problems on a lattice, solutions to the radial Schrödinger equation (in spherical and cylindrical coordinates) for a free particle, solving for patterns of acoustical radiation, frequency-dependent friction in circular pipelines, dynamics of floating bodies, angular resolution, etc. Bessel functions also appear in other problems, such as signal processing (e.g., see FM synthesis, Kaiser window, or Bessel filter).
32.2 Definitions

Because this is a second-order differential equation, there must be two linearly independent solutions. Depending upon the circumstances, however, various formulations of these solutions are convenient. Different variations are summarized in the table shown as below and described in the following sections.

<table>
<thead>
<tr>
<th>Type</th>
<th>First kind</th>
<th>Second kind</th>
<th>Details in</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bessel functions</td>
<td>( J_\alpha )</td>
<td>( Y_\alpha )</td>
<td>Section 32.3 and 0</td>
</tr>
<tr>
<td>Hankel functions, third kind</td>
<td>( H^{(1)}<em>\alpha = J</em>\alpha + i Y_\alpha )</td>
<td>( H^{(2)}<em>\alpha = J</em>\alpha - i Y_\alpha )</td>
<td>Section 0</td>
</tr>
<tr>
<td>Modified Bessel functions</td>
<td>( I_\alpha )</td>
<td>( K_\alpha )</td>
<td>Section 32.6</td>
</tr>
<tr>
<td>Spherical Bessel functions</td>
<td>( j_n )</td>
<td>( y_n )</td>
<td>Section 32.7</td>
</tr>
<tr>
<td>Spherical Hankel functions</td>
<td>( h^{(1)}_n = j_n + iy_n )</td>
<td>( h^{(2)}_n = j_n - iy_n )</td>
<td>Section 32.8</td>
</tr>
</tbody>
</table>

Bessel functions of the second kind and the spherical Bessel functions of the second kind are sometimes denoted by \( N_n \) and \( n_n \), respectively, rather than \( Y_n \) and \( y_n \).

32.3 Bessel functions of the 1st kind: \( J_\alpha \)

Bessel functions of the first kind, denoted as \( J_\alpha(x) \), are solutions of Bessel's differential equation that are finite at the origin \( x = 0 \) for integer or positive \( \alpha \), and diverge as \( x \) approaches zero for negative non-integer \( \alpha \). It is possible to define the function by its series expansion around \( x = 0 \), which can be found by applying the Frobenius method to Bessel's equation:

\[
J_\alpha(x) = \sum_{m=0}^{\infty} \frac{(-1)^m}{m! \Gamma(m + \alpha + 1)} \left( \frac{x}{2} \right)^{2m+\alpha}
\]

Eq. (32.2)

where \( \Gamma(z) \) is the gamma function, a shifted generalization of the factorial function to non-integer values. The Bessel function of the first kind is an entire function if \( \alpha \) is an integer, otherwise it is a multivalued function with singularity at zero. The graphs of Bessel functions look roughly like oscillating sine or cosine functions that decay proportionally to \( 1 / x^{1/2} \) (see also their asymptotic forms below), although their roots are not generally periodic, except asymptotically for large \( x \).
The series indicates that $-J_i(x)$ is the derivative of $J_n(x)$, much like $-\sin x$ is the derivative of $\cos x$; more generally, the derivative of $J_n(x)$ can be expressed in terms of $J_{n+1}(x)$ by the identities below. For non-integer $\alpha$, the functions $J_{\alpha}(x)$ and $J_{-\alpha}(x)$ are linearly independent shown as (Figure 32.1a) and are therefore the two solutions of the differential equation. On the other hand, for integer order $\alpha$, the following relationship is valid (note that the Gamma function has simple poles at each of the non-positive integers) showing that:

$$J_{-\alpha}(x) = (-1)^n J_n(x)$$

Eq. (32.3)

This means that the two solutions are no longer linearly independent. In this case, the second linearly independent solution is then found to be the Bessel function of the second kind shown as (Figure 32.1a).

![Figure 32.1](image)

Figure 32.1  Bessel functions of (a) the 1st kind, $J_{\alpha}(x)$ for integer orders $\alpha = 0, 1, 2$ and (b) the 2nd kind, $Y_{\alpha}(x)$ for integer orders $\alpha = 0, 1, 2$
32.4 **Bessel functions of the 2nd kind: \( Y_\alpha \)**

The Bessel functions of the second kind, denoted by \( Y_\alpha (x) \), occasionally denoted instead by \( N_\alpha (x) \), are solutions of the Bessel differential equation that have a singularity at the origin \( x = 0 \) and are multivalued. These are sometimes called Weber functions as they were introduced by H. M. Weber (1873), and also Neumann functions after Carl Neumann.

For non-integer \( \alpha \), \( Y_\alpha (x) \) is related to \( J_\alpha (x) \) by:

\[
Y_\alpha (x) = \frac{J_\alpha (x) \cos(\alpha \pi) - J_{-\alpha} (x)}{\sin(\alpha \pi)}
\]

Eq. (32.4)

In the case of integer order \( n \), the function is defined by taking the limit as a non-integer \( \alpha \) tends to \( n \),

\[
Y_n (x) = \lim_{\alpha \to n} Y_\alpha (x)
\]

Eq. (32.5)

\( Y_\alpha (x) \) is necessary as the second linearly independent solution of the Bessel's equation when \( \alpha \) is an integer. But \( Y_\alpha (x) \) has more meaning than that. It can be considered as a 'natural' partner of \( J_\alpha (x) \). Also see the (Section 0) on Hankel functions for more details. When \( \alpha \) is an integer, moreover, as was similarly the case for the functions of the first kind, the following relationship is valid:

\[
Y_{-n} (x) = (-1)^n Y_n (x)
\]

Eq. (32.6)

Both \( Y_\alpha (x) \) and \( J_\alpha (x) \) are holomorphic functions of \( x \) on the complex plane cut along the negative real axis. When \( \alpha \) is an integer, the Bessel functions \( J \) are entire functions of \( x \). If \( x \) is held fixed at a non-zero value, then the Bessel functions are entire functions of \( \alpha \). The Bessel functions of the second kind when \( \alpha \) is an integer is an example of the second kind of solution in Fuchs's theorem.
Another important formulation of the two linearly independent solutions to Bessel’s equation are the Hankel functions of the first and second kind, \( H_{\alpha}^{(1)} \) and \( H_{\alpha}^{(2)} \), defined by:

\[
\begin{align*}
H_{\alpha}^{(1)}(x) &= J_{\alpha}(x) + iY_{\alpha}(x) \\
H_{\alpha}^{(2)}(x) &= J_{\alpha}(x) - iY_{\alpha}(x)
\end{align*}
\]

where \( i \) is the imaginary unit or complex number. These linear combinations are also known as Bessel functions of the third kind; they are two linearly independent solutions of Bessel’s differential equation. They are named after Hermann Hankel.

The importance of Hankel functions of the first and second kind lies more in theoretical development rather than in application. These forms of linear combination satisfy numerous simple-looking properties, like asymptotic formulae or integral representations. Here, 'simple' means an appearance of the factor of the form \( \exp(i \cdot f(x)) \).

The Bessel function of the second kind then can be thought to naturally appear as the imaginary part of the Hankel functions. The Hankel functions are used to express outward- and inward-propagating cylindrical wave solutions of the cylindrical wave equation, respectively (or vice versa, depending on the sign convention for the frequency).

Using the previous relationships, they can be expressed as:

\[
\begin{align*}
H_{\alpha}^{(1)}(x) &= J_{-\alpha}(x) - e^{-i\pi\alpha} J_{\alpha}(x) \\
H_{\alpha}^{(2)}(x) &= J_{-\alpha}(x) - e^{-i\pi\alpha} J_{\alpha}(x)
\end{align*}
\]

If \( \alpha \) is an integer, the limit of Eq. (32.8) can be calculated and obtain Eq. (32.9). The following relationships are valid whether \( \alpha \) is an integer or not:

\[
\begin{align*}
H_{\alpha}^{(1)}(x) &= e^{i\pi\alpha} H_{\alpha}^{(1)}(x) \\
H_{-\alpha}^{(2)}(x) &= e^{-i\pi\alpha} H_{\alpha}^{(2)}(x)
\end{align*}
\]
In particularly, if \( \alpha = m + 1/2 \) where \( m \) is a nonnegative integer, the above relations Eq. (32.9) imply directly to Eq. (32.10) so that

\[
\begin{align*}
J_{-(m+1/2)}(x) &= (-1)^m J_{m+1/2}(x) \\
Y_{-(m+1/2)}(x) &= (-1)^m J_{m+1/2}(x)
\end{align*}
\]

Eq. (32.10)

These are useful in developing the spherical Bessel functions in (Section 32.7) and (Section 32.8) plotted as below where Eq. (32.9) were defined in complex domain.

Figure 32.2  The plots above show the structure of \( H_0^{(1)}(z) \) in the complex plane.

Figure 32.3  The plots above show the structure of \( H_0^{(2)}(z) \) in the complex plane.
32.6 Modified Bessel functions: \( I_\alpha \) and \( K_\alpha \)

The Bessel functions are valid even for complex arguments \( x \), and an important special case is that of a pure imaginary argument. In this case, the solutions to the Bessel equation are called the modified Bessel functions (or occasionally the hyperbolic Bessel functions) of the first and second kind, and are defined by:

\[
\begin{align*}
I_\alpha(x) &= i^{-\alpha} J_\alpha(ix) = \sum_{m=0}^{\infty} \frac{1}{m! \Gamma(m+\alpha+1)} \left( \frac{x}{2} \right)^{2m+\alpha} \\
K_\alpha(x) &= \frac{\pi}{2} \frac{I_{-\alpha}(x) - I_\alpha(x)}{\sin(\alpha\pi)}
\end{align*}
\]

Eq. (32.11)

when \( \alpha \) is not an integer. If \( \alpha \) is an integer, then the limit is used. These are chosen to be a real and positive arguments \( x \). The series expansion for \( I_\alpha(x) \) is like that for \( J_\alpha(x) \), but without the alternating \((-1)^m\) factor. \( I_\alpha \) and \( K_\alpha \) are the two linearly independent solutions to the modified Bessel's equation:

\[
x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} - (x^2 + \alpha^2) y = 0
\]

Eq. (32.12)

Unlike the ordinary Bessel functions, which are oscillating as functions of a real argument, \( I_\alpha(x) \) and \( K_\alpha(x) \) are exponentially growing and decaying functions, respectively. Like the ordinary Bessel function \( J_\alpha(x) \), the function \( I_\alpha(x) \) goes to zero at \( x = 0 \) for \( \alpha > 0 \) and is finite at \( x = 0 \) for \( \alpha = 0 \). Analogously, \( K_\alpha(x) \) diverges at \( x = 0 \) with the singularity being of logarithmic type.

Figure 32.4 Modified Bessel function of (a) the 1st kind, \( I_\alpha(x) \) for integer orders \( \alpha = 0, 1, 2, 3 \) and (b) the 2nd kind, \( K_\alpha(x) \) for integer orders \( \alpha = 0, 1, 2, 3 \)
32.7 Spherical Modified Bessel functions: \( j_n \) and \( y_n \)

When solving the Helmholtz equation in spherical coordinates by separation of variables, the radial equation has the form:

\[
x^2 \frac{d^2 y}{dx^2} + 2x \frac{dy}{dx} + \left[ x^2 - n(n+1) \right] y = 0
\]

Eq. (32.13)

The two linearly independent solutions to this equation are called the spherical Bessel functions \( j_n(x) \) and \( y_n(x) \), and are related to the ordinary Bessel functions \( J_n(x) \) and \( Y_n(x) \) by:

\[
\begin{align*}
  j_n(x) &= \sqrt{\frac{\pi}{2x}} J_{n+\frac{1}{2}}(x) \\
  y_n(x) &= \sqrt{\frac{\pi}{2x}} Y_{n+\frac{1}{2}}(x) = (-1)^{n+1} J_{-n-\frac{1}{2}}(x)
\end{align*}
\]

Eq. (32.14)

where \( y_n \) is also denoted \( n_n \) or \( \eta_n \); some authors call these functions the spherical Neumann functions.

The spherical Bessel functions can also be written as Rayleigh's formulas:

\[
\begin{align*}
  j_n(x) &= (-x)^n \left( \frac{1}{x} \frac{d}{dx} \right)^n \frac{\sin(x)}{x} \\
  y_n(x) &= -(-x)^n \left( \frac{1}{x} \frac{d}{dx} \right)^n \frac{\cos(x)}{x}
\end{align*}
\]

Eq. (32.15)
32.8 Spherical Hankel functions: $h_n^{(1)}$ and $h_n^{(2)}$

There are also spherical analogues of the Hankel functions:

\[
\begin{align*}
    h_n^{(1)}(x) &= j_n(x) + iy_n(x) \\
    h_n^{(2)}(x) &= j_n(x) - iy_n(x)
\end{align*}
\]

Eq. (32.16)

In fact, there are simple closed-form expressions for the Bessel functions of half-integer order in terms of the standard trigonometric functions, and therefore for the spherical Bessel functions. For non-negative integers $n$ and $h_n^{(2)}$ is the complex-conjugate of this for real-valued $x$.

\[
\begin{align*}
    h_n^{(1)}(x) &= (-i)^{n+1} e^{ix} \sum_{m=0}^{n} \frac{i^m}{m!(2x)^m} \frac{(n+m)!}{(n-m)!} \\
    h_n^{(2)}(x) &= \left(h_n^{(1)}(x)\right)^* \\
\end{align*}
\]

Eq. (32.17)

The spherical Hankel functions appear in problems involving spherical wave propagation, for example in the multipole expansion of the electromagnetic field.

Figure 32.5 \hspace{1cm} The real and imaginary parts $h_n^{(1)}$ in the complex plane

Figure 32.6 \hspace{1cm} The real and imaginary parts $h_n^{(2)}$ in the complex plane
33 Maxwell's Equations

Maxwell's equations are a set of partial differential equations that, together with the Lorentz force law, form the foundation of classical electromagnetism, quantum field theory, classical optics, and electric circuits. They underpin all electric, optical and radio technologies, including power generation, electric motors, wireless communication, cameras, televisions, computers etc.

*Maxwell's equations describe how electric and magnetic fields are generated by charges, currents, and changes of each other.* One important consequence of the equations is that they demonstrate how fluctuating electric and magnetic fields propagate at the speed of light. Known as electromagnetic radiation, these waves may occur at various wavelengths to produce a spectrum from radio waves to γ-rays. The equations are named after the physicist and mathematician James Clerk Maxwell, who between 1861 and 1862 published an early form of the equations, and first proposed that light is an electromagnetic phenomenon.

The equations have two major variants. *The microscopic Maxwell equations have universal applicability but are unwieldy for common calculations.* They relate the electric and magnetic fields to total charge and total current, including the complicated charges and currents in materials at the atomic scale. *The "macroscopic" Maxwell equations define two new auxiliary fields that describe the large-scale behavior of matter without having to consider atomic scale details.* However, their use requires experimentally determining parameters for a phenomenological description of the electromagnetic response of materials.
33.1 Notation Conventions

The operators, symbols and notation conventions are unified in the following sections.

33.1.1 Operator Notations

Symbols in bold represent vector quantities, and symbols in italics represent scalar quantities, unless otherwise indicated.

The equations introduce the electric field, \( \mathbf{E} \), a vector field, and the magnetic field, \( \mathbf{B} \), a pseudovector field, each generally having a time and location dependence. The sources are

- the electric charge density (charge per unit volume), \( \rho \), and
- the electric current density (current per unit area), \( \mathbf{J} \).

The universal constants appearing in the equations are

- the permittivity of free space, \( \varepsilon_0 \), and
- the permeability of free space, \( \mu_0 \).

Differential equations

In the differential equations,

- the nabla symbol, \( \nabla \), denotes the three-dimensional gradient operator,
- the \( \nabla \cdot \) symbol denotes the divergence operator,
- the \( \nabla \times \) symbol denotes the curl operator.

Integral equations

In the integral equations,

- \( \Omega \) is any fixed volume with closed boundary surface \( \partial \Omega \)
- \( \Sigma \) is any fixed surface with closed boundary curve \( \partial \Sigma \)
Here a fixed volume or surface means that it does not change over time. The equations are correct, complete and a little easier to interpret with time-independent surfaces. However, since the surface is time-independent, we can bring the differentiation under the integral sign in Faraday's law:

\[
\frac{d}{dt} \int_S \mathbf{B} \cdot d\mathbf{S} = \int_S \frac{d\mathbf{B}}{dt} \cdot d\mathbf{S}
\]

Eq. (33.1)

The Maxwell's equations can be formulated with possibly time dependent surfaces and volumes by substituting the lefthand side with the righthand side in the integral equation version of the Maxwell equations.

- \[\oint_{\partial \Omega} \] is a closed surface integral over the surface \( \partial \Omega \)
- \[\iiint_{\Omega} \] is a volume integral over the volume \( \Omega \),
- \[\iint_{\Sigma} \] is a surface integral over the surface \( \Sigma \),
- \[\oint_{\partial \Sigma} \] is a line integral around a closed boundary curve \( \partial \Sigma \)

The volume integral over \( \Omega \) of the total charge density \( \rho \), is the total electric charge \( Q \) contained in \( \Omega \)

\[
Q = \iiint_{\Omega} \rho dV
\]

Eq. (33.2)

where \( dV \) is the volume element.

The net electric current \( I \) is the surface integral of the electric current density \( \mathbf{J} \) passing through a fixed surface \( \Sigma \)

\[
I = \iiint_{\Sigma} \mathbf{J} \cdot d\mathbf{S}
\]

Eq. (33.3)

where \( d\mathbf{S} \) denotes the vector element of surface area, \( S \), normal to surface, \( \Sigma \). (Vector area is also denoted by \( A \) rather than \( S \), but this conflicts with the magnetic potential, a separate vector field).

\[
\oint_{\partial \Sigma} \mathbf{B} \cdot d\mathbf{l} = \mu_0 \iiint_{\Sigma} \mathbf{J} \cdot d\mathbf{S} + \varepsilon_0 \frac{d}{dt} \iiint_{\Sigma} \mathbf{E} \cdot d\mathbf{S} \quad \Leftrightarrow \quad \nabla \times \mathbf{B} = \mu_0 \left( \mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right)
\]

Eq. (33.4)
33.1.2 Electrical Notations

Electrical parameters are defined as

\[
\begin{align*}
\mathbf{D} &= \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 (1 + \chi) \mathbf{E} = \varepsilon \mathbf{E} \\
\mathbf{P} &= \varepsilon_0 \chi \mathbf{E} \\
\mathbf{F} &= q \mathbf{E}
\end{align*}
\]

where

1) \( \mathbf{E} \) electric field \((N/C) (V/m) (kg \cdot m / s^3 / A)\)

An electric field is a vector field that associates to each point in space the Coulomb force that would be experienced per unit of electric charge, by an infinitesimal test charge at that point. Electric fields are created by electric charges and can be induced by time-varying magnetic fields. The electric field combines with the magnetic field to form the electromagnetic field.

2) \( \mathbf{D} \) electric displacement field, electric flux density \((C/m^2)\)

It accounts for the effects of free and bound charge within materials while its sources are the free charges only.

3) \( \mathbf{P} \) electric polarization, polarization density \((C/m^2)\)

In classical electromagnetism, polarization density (or electric polarization, or simply polarization) is the vector field that expresses the density of permanent or induced electric dipole moments in a dielectric material.

4) \( \varepsilon_0 \) electric permittivity \((F/m) (A^2 s^4 / kg / m^3)\)

The permittivity relates to a material's ability to resist an electric field.

5) \( \chi \) electric susceptibility \((\text{dimensionless})\)

The greater the electric susceptibility, the greater the ability of a material to polarize in response to the field, and thereby reduce the total electric field inside the material (and store energy).
33.1.3 Magnetic Notations

Magnetic parameters are defined as

\[
\begin{align*}
\mathbf{H} &= \frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \quad \rightarrow \quad \mathbf{B} = \mu_0 (\mathbf{H} + \mathbf{M}) = \mu_0 (1 + \chi) \mathbf{H} = \mu \mathbf{H} \\
\mathbf{M} &= \chi \mathbf{H} \\
\mathbf{F} &= q (\mathbf{E} + \mathbf{v} \times \mathbf{B})
\end{align*}
\]

Eq. (33.6)

where

1) \( \mathbf{H} \) magnetic field strength, magnetic field intensity \( (A/m) \)

A magnetic field is the magnetic effect of electric currents and magnetic materials.

2) \( \mathbf{B} \) magnetic field, magnetic flux density \( (T) \) \( (Wb/m^2) \) \( (N \cdot s/C/m) \) \( (N/m/A) \)

It is most commonly defined in terms of the Lorentz force it exerts on moving electric charges.

3) \( \mathbf{M} \) magnetization \( (A/m) \)

The magnetization vector field represents how strongly a region of material is magnetized.

4) \( \mu_0 \) magnetic perpeability \( (H/m) \) \( (N/m^2) \)

In electromagnetism, permeability is the measure of the ability of a material to support the formation of a magnetic field within itself.

5) \( \chi \) magnetic susceptibility \( (dimensionless) \)

The susceptibility indicates whether a material is attracted into or repelled out of a magnetic field, which in turn has implications for practical applications. Quantitative measures of the magnetic susceptibility also provide insights into the structure of materials, providing insight into bonding and energy levels.
33.2 Microscopic Formulation

In the electric and magnetic field formulation there are four equations. The two inhomogeneous equations describe how the fields vary in space due to sources. Gauss's law describes how electric fields emanate from electric charges. Gauss's law for magnetism describes magnetic fields as closed field lines not due to magnetic monopoles. The two homogeneous equations describe how the fields "circulate" around their respective sources. Ampère's law with Maxwell's addition describes how the magnetic field "circulates" around electric currents and time varying electric fields, while Faraday's law describes how the electric field "circulates" around time varying magnetic fields.

A separate law of nature, the Lorentz force law, describes how the electric and magnetic field act on charged particles and currents. A version of this law was included in the original equations by Maxwell but, by convention, is no longer included.

The precise formulation of Maxwell's equations depends on the precise definition of the quantities involved. Conventions differ with the unit systems, because various definitions and dimensions are changed by absorbing dimensional factors like the speed of light $c$. This makes constants come out differently. The most common form is based on conventions used when quantities measured using SI units, but other commonly used conventions are used with other units including Gaussian units based on the cgs system [156] Lorentz–Heaviside units (used mainly in particle physics), and Planck units (used in theoretical physics).

The vector calculus formulation below has become standard. It is mathematically much more convenient than Maxwell's original 20 equations and is due to Oliver Heaviside [155], [156]. The differential and integral equations formulations are mathematically equivalent and are both useful. The integral formulation relates fields within a region of space to fields on the boundary and can often be used to simplify and directly calculate fields from symmetric distributions of charges and currents. On the other hand, the differential equations are purely local and are a more natural starting point for calculating the fields in more complicated (less symmetric) situations, for example using finite element analysis [152], [153]. For formulations using tensor calculus or differential forms, see alternative formulations. For relativistically invariant formulations, see relativistic formulations.
33.2.1 Formulation in SI Unit Convention

**Gauss's law**

The electric flux leaving a volume is proportional to the charge inside.

\[
\oint_{\partial V} \mathbf{E} \cdot d\mathbf{S} = \frac{1}{\varepsilon_0} \iiint_{V} \rho dV \quad \Leftrightarrow \quad \nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}
\]

Eq. (33.7)

**Gauss's law for magnetism**

There are no magnetic monopoles; the total magnetic flux through a closed surface is zero.

\[
\oint_{\partial \Sigma} \mathbf{B} \cdot d\mathbf{S} = 0 \quad \Leftrightarrow \quad \nabla \cdot \mathbf{B} = 0
\]

Eq. (33.8)

**Maxwell–Faraday equation (Faraday's law of induction)**

The voltage induced in a closed circuit is proportional to the rate of change of the magnetic flux it encloses.

\[
\oint_{c_2} \mathbf{E} \cdot dl = -\frac{d}{dt} \iint_{c_2} \mathbf{B} \cdot d\mathbf{S} \quad \Leftrightarrow \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}
\]

Eq. (33.9)

**Ampère's circuital law (with Maxwell's addition)**

The magnetic field induced around a closed loop is proportional to the electric current plus displacement current (rate of change of electric field) it encloses.

\[
\oint_{c_2} \mathbf{B} \cdot dl = \mu_0 \iint_{c_2} \mathbf{J} \cdot d\mathbf{S} + \mu_0 \varepsilon_0 \frac{d}{dt} \iint_{c_2} \mathbf{E} \cdot d\mathbf{S} \quad \Leftrightarrow \quad \nabla \times \mathbf{B} = \mu_0 \left( \mathbf{J} + \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t} \right)
\]

Eq. (33.10)

33.2.2 Formulation in Gaussian Unit Convention

Gaussian units are a popular system of units that are part of the centimetre–gram–second system of units (cgs). When using Gaussian units it is conventional to use a slightly different definition of electric field \( \mathbf{E}_{\text{cgs}} = c^{-1} \mathbf{E}_{\text{SI}} \). This implies that the modified electric and magnetic field have the same units (in the SI convention this is not the case making dimensional analysis of the equations different: e.g. for an electromagnetic wave in vacuum \( | \mathbf{E}_{\text{SI}} | = c | \mathbf{B}_{\text{SI}} | \)).
The Gaussian system uses a unit of charge defined in such a way that the permittivity of the vacuum \( \varepsilon_0 = 1/4\pi c \) hence \( \mu_0 = 4\pi / c \). These units are sometimes preferred over SI units in the context of special relativity [152] in which the components of the electromagnetic tensor, the Lorentz covariant object describing the electromagnetic field, have the same unit without constant factors. Using these different conventions, the Maxwell equations [151] become:

**Gauss's law**

The electric flux leaving a volume is proportional to the charge inside.

\[
\iiint_{\mathcal{V}} \mathbf{E} \cdot d\mathbf{S} = 4\pi \iiint_{\mathcal{V}} \rho dV \quad \iff \quad \nabla \cdot \mathbf{E} = 4\pi \rho
\]

**Gauss's law for magnetism**

There are no magnetic monopoles; the total magnetic flux through a closed surface is zero.

\[
\iiint_{\mathcal{V}} \mathbf{B} \cdot d\mathbf{S} = 0 \quad \iff \quad \nabla \cdot \mathbf{B} = 0
\]

**Maxwell–Faraday equation (Faraday's law of induction)**

The voltage induced in a closed circuit is proportional to the rate of change of the magnetic flux it encloses.

\[
\oint_{\mathcal{C}} \mathbf{E} \cdot dl = -\frac{1}{c} \frac{d}{dt} \iint_{\mathcal{S}} \mathbf{B} \cdot d\mathbf{S} \quad \iff \quad \nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}
\]

**Ampère's circuital law (with Maxwell's addition)**

The magnetic field induced around a closed loop is proportional to the electric current plus displacement current (rate of change of electric field) it encloses.

\[
\oint_{\mathcal{C}} \mathbf{B} \cdot dl = \frac{1}{c} \left( 4\pi \iint_{\mathcal{S}} \mathbf{J} \cdot d\mathbf{S} + \frac{d}{dt} \iint_{\mathcal{S}} \mathbf{E} \cdot d\mathbf{S} \right) \quad \iff \quad \nabla \times \mathbf{B} = \frac{1}{c} \left( 4\pi \mathbf{J} + \frac{\mathbf{E}}{\partial t} \right)
\]
33.2.3 Relationship between Differential and Integral Formulations

The equivalence of the differential and integral formulations are a consequence of the Gauss divergence theorem and the Kelvin–Stokes theorem.

Flux and divergence

The "sources of the fields" (i.e. their divergence) can be determined from the surface integrals of the fields through the closed surface $\partial \Omega$. E.g. the electric flux is

$$\iint_{\partial \Omega} \mathbf{E} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{E} \, dV \quad \text{Eq. (33.15)}$$

where the last equality uses the Gauss divergence theorem. Using the integral version of Gauss's equation, we can rewrite this to

$$\iiint_{\Omega} \left( \nabla \cdot \mathbf{E} - \frac{\rho}{\varepsilon_0} \right) dV = 0 \quad \text{Eq. (33.16)}$$

Since $\Omega$ can be chosen arbitrarily, e.g. as an arbitrary small ball with arbitrary center, this implies that the integrand must be zero, which is the differential equations formulation of Gauss equation up to a trivial rearrangement. Gauss's law for magnetism in differential equations form follows likewise from the integral form by rewriting the magnetic flux

$$\iint_{\partial \Omega} \mathbf{B} \cdot d\mathbf{S} = \iiint_{\Omega} \nabla \cdot \mathbf{B} \, dV = 0 \quad \text{Eq. (33.17)}$$

Here, $\mathbf{F}$ could be the $\mathbf{E}$ field with source electric charges, but not the $\mathbf{B}$ field which has no magnetic charges as shown. The outward unit normal is $\mathbf{n}$.

Figure 33.1 Volume $\Omega$ and its closed boundary $\partial \Omega$, containing (respectively enclosing) a source (+) and sink (−) of a vector field $\mathbf{F}$. 

308
**Circulation and curl**

The "circulation of the fields" (i.e. their curls) can be determined from the line integrals of the fields around the closed curve $\partial \Sigma$, e.g. for the magnetic field via Kelvin – Stokes theorem so that

$$\oint_{\partial \Sigma} \mathbf{B} \cdot dl = \oiint_{\Sigma} (\nabla \times \mathbf{B}) \cdot d\mathbf{S}$$  \hspace{1cm} \text{Eq. (33.18)}

Using the modified Ampere law in integral form and the writing the time derivative of the flux as the surface integral of the partial time derivative of $\mathbf{E}$ we conclude that

$$\oiint_{\Sigma} (\nabla \times \mathbf{B} - \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}) \cdot d\mathbf{S} = 0$$ \hspace{1cm} \text{Eq. (33.19)}

Since $\Sigma$ can be chosen arbitrarily, e.g. as an arbitrary small, arbitrary oriented, and arbitrary centered disk, we conclude that the integrand must be zero. This is Ampere's modified law in differential equations form up to a trivial rearrangement.

![Figure 33.2](image)

**Figure 33.2** Surface $\Sigma$ with closed boundary $\partial \Sigma$ where $\mathbf{F}$ could be the $\mathbf{E}$ or $\mathbf{B}$ fields.

Again, $\mathbf{n}$ is the unit normal. (The curl of a vector field doesn't literally look like the "circulations", this is a heuristic depiction).

Likewise, the Faraday law in differential equations form follows from rewriting the integral form using the Kelvin – Stokes theorem. The line integrals and curls are analogous to quantities in classical fluid dynamics: the circulation of a fluid is the line integral of the fluid's flow velocity field around a closed loop, and the vorticity of the fluid is the curl of the velocity field.
33.2.4 Vacuum Equations, Electromagnetic Waves and Speed of Light

In a region no charges \( \rho = 0 \) and no currents \( \mathbf{J} = 0 \), such as in a vacuum, Maxwell's equations reduce to:

\[
\nabla \cdot \mathbf{E} = 0 \quad \text{and} \quad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad \text{Eq. (33.20)}
\]

\[
\nabla \cdot \mathbf{B} = 0 \quad \text{and} \quad \nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t} \quad \text{Eq. (33.21)}
\]

Using the properties of curl identity \( \nabla \times (\nabla \times \mathbf{X}) = \nabla (\nabla \cdot \mathbf{X}) - \nabla^2 \mathbf{X} \) on wave equations

\[
\frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} = 0 \quad \text{Eq. (33.22)}
\]

\[
\frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2} - \nabla^2 \mathbf{B} = 0 \quad \text{Eq. (33.23)}
\]

and identifying \( c = 1 / \sqrt{\mu_0 \varepsilon_0} = 2.99792458 \times 10^8 \text{ m/s} \) as the speed of light in free space. In materials with relative permittivity, \( \varepsilon_r \), and relative permeability, \( \mu_r \), the phase velocity of light becomes \( \nu_p = 1 / \sqrt{\mu_0 \mu_r \varepsilon_0 \varepsilon_r} \) which is usually less than \( c \).

In addition, \( \mathbf{E} \) and \( \mathbf{B} \) are mutually perpendicular to each other and the direction of wave propagation and are in phase with each other. A sinusoidal plane wave is one special solution of these equations. Maxwell's equations explain how these waves can physically propagate through space.

The changing magnetic field creates a changing electric field through Faraday's law. In turn, that electric field creates a changing magnetic field through Maxwell's addition to Ampère's law. This perpetual cycle allows these waves, now known as electromagnetic radiation, to move through space at velocity \( c \).
33.3 Macroscopic Formulation

The microscopic variant of Maxwell's equation is the version given above. It expresses the electric $E$ field and the magnetic $B$ field in terms of the total charge and total current present, including the charges and currents at the atomic level. The "microscopic" form is sometimes called the "general" form of Maxwell's equations. The macroscopic variant of Maxwell's equation is equally general, however, with the difference being one of bookkeeping.

*The "microscopic" variant is sometimes called "Maxwell's equations in a vacuum".* This refers to the fact that the material medium is not built into the structure of the equation; it does not mean that space is empty of charge or current. They are also known as the "Maxwell-Lorentz equations". Lorentz tried to use these equations to predict the macroscopic properties of bulk matter from the physical behavior of its microscopic constituents [150]. "Maxwell's macroscopic equations", also known as Maxwell's equations in matter, are more similar to those that Maxwell introduced himself.

**Gauss's law**

$$\iiint\limits_{\Omega} D \cdot dS = \iiint\limits_{\Omega} \rho_f dV \quad \Leftrightarrow \quad \nabla \cdot D = \rho_f$$  
\hspace{1cm} \text{Eq. (33.24)}

**Gauss's law for magnetism**

$$\iiint\limits_{\Omega} B \cdot dS = 0 \quad \Leftrightarrow \quad \nabla \cdot B = 0$$  
\hspace{1cm} \text{Eq. (33.25)}

**Maxwell–Faraday equation (Faraday's law of induction)**

$$\oint\limits_{\Gamma} E \cdot dl = -\frac{d}{dt} \iint\limits_{\Sigma} B \cdot dS \quad \Leftrightarrow \quad \nabla \times E = -\frac{\partial B}{\partial t}$$  
\hspace{1cm} \text{Eq. (33.26)}

**Ampère's circuital law (with Maxwell's addition)**

$$\oint\limits_{\Gamma} H \cdot dl = \iint\limits_{\Sigma} J_f \cdot dS + \frac{d}{dt} \iint\limits_{\Sigma} D \cdot dS \quad \Leftrightarrow \quad \nabla \times H = J_f + \frac{\partial D}{\partial t}$$  
\hspace{1cm} \text{Eq. (33.27)}

Unlike the "microscopic" equations, the "macroscopic" equations separate out the bound charge $Q_b$ and bound current $I_b$ to obtain equations that depend only on the free charges $Q_f$ and currents $I_f$. This factorization can be made by splitting the total electric charge and current as follows:
Correspondingly, the total current density $\mathbf{J}$ splits into free $\mathbf{J}_f$ and bound $\mathbf{J}_b$ components, and similarly the total charge density $\rho$ splits into free $\rho_f$ and bound $\rho_b$ parts.

The cost of this factorization is that additional fields, the displacement field $\mathbf{D}$ and the magnetizing field $\mathbf{H}$, are defined and need to be determined. Phenomenological constituent equations relate the additional fields to the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$, often through a simple linear relation.

For a detailed description of the differences between the microscopic (total charge and current including material contributes or in air/vacuum) and macroscopic (free charge and current; practical to use on materials) variants of Maxwell's equations, see below.

### 33.3.1 Bound Charge and Current

When an electric field is applied to a dielectric material its molecules respond by forming microscopic electric dipoles – their atomic nuclei move a tiny distance in the direction of the field, while their electrons move a tiny distance in the opposite direction. This produces a macroscopic bound charge in the material even though all of the charges involved are bound to individual molecules.

For example, if every molecule responds the same, similar to that shown in the figure, these tiny movements of charge combine to produce a layer of positive bound charge on one side of the material and a layer of negative charge on the other side. The bound charge is most conveniently described in terms of the polarization $\mathbf{P}$ of the material, its dipole moment per unit volume. If $\mathbf{P}$ is uniform, a macroscopic separation of charge is produced only at the surfaces where $\mathbf{P}$ enters and leaves the material. For non-uniform $\mathbf{P}$, a charge is also produced in the bulk [149].

How an assembly of microscopic current loops add together to produce a macroscopically circulating current loop (right). Inside the boundaries, the individual contributions tend to cancel, but at the boundaries no cancelation occurs. Somewhat similarly, in all materials the constituent atoms exhibit magnetic moments that are intrinsically linked to the angular momentum of the components of the atoms, most notably their electrons. The connection to angular momentum suggests the picture of an assembly of microscopic current loops. Outside the material, an assembly of such microscopic current loops is not different from a
macroscopic current circulating around the material's surface, despite the fact that no individual charge is traveling a large distance. These bound currents can be described using the magnetization $M$ [149]

![Figure 33.3](image)

Figure 33.3 A schematic view of how an assembly of microscopic dipoles produces opposite surface charges as shown at top and bottom (left).

The very complicated and granular bound charges and bound currents, therefore, can be represented on the macroscopic scale in terms of $P$ and $M$ which average these charges and currents on a sufficiently large scale so as not to see the granularity of individual atoms, but also sufficiently small that they vary with location in the material. As such, Maxwell's macroscopic equations ignore many details on a fine scale that can be unimportant to understanding matters on a gross scale by calculating fields that are averaged over some suitable volume.

### 33.3.2 Auxiliary Fields, Polarization and Magnetization

The definitions (not constitutive relations) of the auxiliary fields are:

$$D(r,t) = \varepsilon_0 E(r,t) + P(r,t) \quad \text{Eq. (33.30)}$$

$$H(r,t) = \frac{1}{\mu_0} B(r,t) - M(r,t) \quad \text{Eq. (33.31)}$$

where $P$ is the polarization field and $M$ is the magnetization field which are defined in terms of microscopic bound charges and bound currents respectively. The macroscopic bound charge density $\rho_b$ and bound current density $J_b$ in terms of polarization $P$ and magnetization $M$ are then defined as

$$\rho_b = -\nabla \cdot P \quad \text{Eq. (33.32)}$$
If we define the total, bound, and free charge and current density by

$$ \rho = \rho_b + \rho_f $$

Eq. (33.34)

$$ J = J_b + J_f $$

Eq. (33.35)

and use the defining relations above to eliminate $D$, and $H$, the "macroscopic" Maxwell's equations reproduce the "microscopic" equations.

### 33.3.3 Constitutive Relations and Electromagnetism

In order to apply 'Maxwell's macroscopic equations', it is necessary to specify the relations between displacement field $D$ and the electric field $E$, as well as the magnetizing field $H$ and the magnetic field $B$. Equivalently, we have to specify the dependence of the polarization $P$ (hence the bound charge) and the magnetisation $M$ (hence the bound current) on the applied electric and magnetic field. The equations specifying this response are called constitutive relations.

For real-world materials, the constitutive relations are rarely simple, except approximately, and usually determined by experiment. See the main article on constitutive relations for a fuller description.

$$ D = \varepsilon_0 E + P = \varepsilon_0 (1 + \chi_e)E = \varepsilon E $$

Eq. (33.36)

$$ B = \mu_0 (H + M) = \mu_0 (1 + \chi_m)H = \mu H $$

Eq. (33.37)

where $\chi_e$ is electric susceptibility and $\chi_m$ is magnetic susceptibility.

For materials without polarization and magnetization, the constitutive relations are (by definition)

$$ D = \varepsilon_0 E \quad \text{and} \quad H = \frac{1}{\mu_0} B $$

Eq. (33.38)

where $\varepsilon_0$ is the permittivity of free space and $\mu_0$ the permeability of free space. Since there is no bound charge, the total and the free charge and current are equal.
An alternative viewpoint on the microscopic equations is that they are the macroscopic equations together with the statement that vacuum behaves like a perfect linear "material" without additional polarization and magnetization. More generally, for linear materials the constitutive relations are

\[ \mathbf{D} = \varepsilon \mathbf{E} \quad \text{and} \quad \mathbf{H} = \frac{1}{\mu} \mathbf{B} \]  

Eq. (33.39)

where \( \varepsilon \) is the permittivity and \( \mu \) is the permeability of the material.

For the displacement field \( \mathbf{D} \) the linear approximation is usually excellent because for all but the most extreme electric fields or temperatures obtainable in the laboratory (high power pulsed lasers) the interatomic electric fields of materials of the order of \( 10^{11} V / m \) are much higher than the external field. For the magnetizing field \( \mathbf{H} \), however, the linear approximation can break down in common materials like iron leading to phenomena like hysteresis. Even the linear case can have various complications, however.

For homogeneous materials, \( \varepsilon \) and \( \mu \) are constant throughout the material, while for inhomogeneous materials they depend on location within the material (and perhaps time) [148]. For isotropic materials, \( \varepsilon \) and \( \mu \) are scalars, while for anisotropic materials (e.g. due to crystal structure) they are tensors [147].

Materials are generally dispersive, so \( \varepsilon \) and \( \mu \) depend on the frequency of any incident EM waves. Even more generally, in the case of non-linear materials (see for example nonlinear optics), \( \mathbf{D} \) and \( \mathbf{P} \) are not necessarily proportional to \( \mathbf{E} \), similarly \( \mathbf{H} \) or \( \mathbf{M} \) is not necessarily proportional to \( \mathbf{B} \). In general \( \mathbf{D} \) and \( \mathbf{H} \) depend on both \( \mathbf{E} \) and \( \mathbf{B} \), on location and time, and possibly other physical quantities.

In applications one also has to describe how the free currents and charge density behave in terms of \( \mathbf{E} \) and \( \mathbf{B} \) possibly coupled to other physical quantities like pressure, and the mass, number density, and velocity of charge – carrying particles, e.g., the original equations given by Maxwell (see History of Maxwell's equations) included Ohms law in the form

\[ \mathbf{J} = \sigma \mathbf{E} \]  

Eq. (33.40)

33.4 Overdetermination of Maxwell's Equations

Maxwell's equations seem overdetermined, in that they involve six unknowns (the three components of \( \mathbf{E} \) and \( \mathbf{B} \)) but eight equations (one for each of the two Gauss's laws, three vector components each for
Faraday's and Ampere's laws). (The currents and charges are not unknowns, being freely specifiable subject to charge conservation.)

This is related to a certain limited kind of redundancy in Maxwell's equations: It can be proven that any system satisfying Faraday's law and Ampere's law automatically also satisfies the two Gauss's laws, as long as the system's initial condition does [143], [146]. This explanation was first introduced by Julius Adams Stratton [145] in 1941. Although it is possible to simply ignore the two Gauss's laws in a numerical algorithm (apart from the initial conditions), the imperfect precision of the calculations can lead to ever-increasing violations of those laws. By introducing dummy variables characterizing these violations, the four equations become not overdetermined after all. The resulting formulation can lead to more accurate algorithms [144] that take all four laws into account.

33.5 Limitations of the Maxwell Equations as A Theory of Electromagnetism

While Maxwell's equations (along with the rest of classical electromagnetism) are extraordinarily successful at explaining and predicting a variety of phenomena, they are not exact, but approximations. In some special situations, they can be noticeably inaccurate. Examples include extremely strong fields (see Euler–Heisenberg Lagrangian) and extremely short distances (see vacuum polarization). Moreover, various phenomena occur in the world even though Maxwell's equations predict them to be impossible, such as "nonclassical light" and quantum entanglement of electromagnetic fields (see quantum optics).

Finally, any phenomenon involving individual photons, such as the photoelectric effect, Planck's law, the Duane–Hunt law, single-photon light detectors, etc., would be difficult or impossible to explain if Maxwell's equations were exactly true, as Maxwell's equations do not involve photons. For the most accurate predictions in all situations, Maxwell's equations have been superseded by quantum electrodynamics.